2021 ICDXA

International Conference on
Digital Transformation and Applications

IEM
The Institution of Engineers, Malaysia

IEM Journal Special Edition

25th - 26th October 2021

Tunku Abdul Rahman University of Management and Technology

Organised by:

TARUMT
TUNKU ABDUL RAHMAN UNIVERSITY OF MANAGEMENT AND TECHNOLOGY

DIGITAL TRANSFORMATION

CONTENTS

International Conference on Digital Transformation and Applications

IEM Journal Special Edition

THE INSTITUTION OF ENGINEERS, MALAYSIA
Bangunan Ingenieur, Lots 60 & 62, Jalan 52/4,
P.O. Box 223 (Jalan Sultan),
46720 Petaling Jaya, Selangor Darul Ehsan.

Tel: 03-7968 4001/4002
Fax: 03-7957 7678
E-mail: sec@iem.org.my
Homepage: http://www.myiem.org.my

REVIEWER

Assoc. Prof. Dr Abbas Mahmoudabadi
Assoc. Prof. Dr Abdelrahman Zaher Al-Attili
Assoc. Prof. Dr Khurram Ejaz
Assoc. Prof. Dr Lokman Bin Abdullah
Assoc. Prof. Ir. Dr David Chuah Joon Huang
Assoc. Prof. Ir. Dr Low Cheng Yee
Assoc. Prof. Ts. Dr Lee Wah Pheng
Assoc. Prof. Ts. Dr Tew Yiqi
Dr Chan Choon Kit
Dr Chang Kai Ming
Dr Chiew Tsung Heng
Dr Chin Wan Yoke
Dr Gloria Jennis Tan
Dr Goh Ching Pang
Dr Hu Chunxiao
Dr Ioan-Alexandru Barbu
Dr Koh Yit Yan
Dr Lim Yee Mei
Dr Lu Ping
Dr Norhaida Binti Mohd Suaib
Dr Prameen Kalikavunkal
Dr Riccardo Reale
Dr Roeland H.G. Mingels
Dr Shanmuga Vivekananda Nadarajan
Dr Sumit Kalsi
Dr Tan Chi Wee
Dr Wong Siaw Lang
Dr Wong Thein Lai
Dr Yacine Izza
Dr Yeo Kwok Shien
Dr Yu Yong Poh
Ir. Dr Teoh Bak Aun
Prof. Dr Atty Renato S. Pacaldo
Prof. Dr Lim Tong Ming
Prof. Dr Zamberi Bin Jamludin

2021
ICDXA

EVALUATION OF VIRTUAL CLASSROOM WITH ARTIFICIAL INTELLIGENCE COMPONENTS

(Date received: 11.01.2022/Date accepted: 18.04.2022)

Kah Yee Lim[1], Hau Joan[1], Yiqi Tew[1]

[1] Faculty of Computing and Information Technology,
Tunku Abdul Rahman University College, 53300 Kuala Lumpur, Malaysia

**Corresponding author: yiqi@tarc.edu.my*

ABSTRACT

The recent advancement of information technology allows educators and students to interact with Artificial Intelligence (AI) through smart classroom channels. This channel is one of the latest technology-enhanced learning (TEL) that provides a learning environment with educators and students interaction during the learning process. Currently, smart classrooms are believed to change current dull teaching methods and enhance the students' learning experience. Hence, this paper shows a comprehensive investigation of applying AI components to an intelligent classroom system (a.k.a virtual classroom system) that provides hand gestures and face detection through e-learning classrooms. Machine Learning libraries are implemented and compared on three machines with varying hardware specifications and capabilities. As a result of this study, Tensorflow Handpose provides more accuracy than MediaPipe Hands, although it requires higher computational capabilities. Face-api.js outperforms TensorFlow and MediaPipe when it comes to executing face detection functions. In addition to the study, the presented face and hand APIs can be adopted in a real time implementation for smart classroom systems.

Keywords: Smart Classroom, Google Meet, Face Detection, Hand Gesture Detection, Object Recognition

1.0 INTRODUCTION

Classroom plays an imperative role in the modern technological world, as it is an important growing environment for one's better future. Through the rapid development of AI and Internet technology, virtual classrooms have been utilized for modern education in order to provide better teaching and learning services [1]. A virtual classroom differs from a traditional classroom in that it takes place in real-time and is synchronized. There are several software used to conduct virtual classrooms, such as Zoom, Google Meet, Microsoft Teams, etc. Although online education often entails viewing pre-recorded asynchronous content, virtual classroom environments entail real-time interaction between lecturers and students [2]. Through the investigation by other researchers [3], virtual classrooms that realize AI have become a reality and can assist interactive education. Virtual classrooms have the benefit of collective intelligence; students can share what they find relevant and interesting to the particular concepts taught in the classroom. Again, participation in the classroom ultimately depends on the students, and what AI can ensure is to improve the chances of that happening. AI opens many creative doors for students and teachers alike. Students' work can be unconventional; demonstrating their abilities and knowledge beyond the prescribed books, that in turn makes them more confident in their work. Then, teachers can figure out each student's tendencies from a fairly young age [4]. In addition, facial biometrics contribute to competitive authentication methods and advances while ensuring the reliability and validity of e-learning systems. To ensure the authenticity of users, the use of facial biometrics is recommended. This will provide an effective authentication method for learners and reduce the probability of cheating and other user authentication anomalies [5]. In this paper, we are going to study the efficiency of machine learning libraries for face detection and hand gesture detection in order to have proper guidance in future development of virtual classrooms.

2.0 BACKGROUND STUDIES

Several domains are studied and examined based on the components and feasibility of deployment to facilitate the intelligent components of virtual classrooms. These include face detection and hand gesture detection.

2.1 Face Detection

Face detection and face recognition are frequently misunderstood by most of the users. Facial detection identifies face segments or areas from a picture, whereas face recognition identifies an individual's face based on personal information. Face detection and identification are advanced in today's culture, but they will encounter certain challenges throughout the way [6]. Table 1 shows a list of the issues.

In each image, the face is detected and cropped out for further processing in Singh's work [7]. Any colored image will convert to grayscale for image pre-processing. Also, the face detected will then be aligned based on the eye's position and

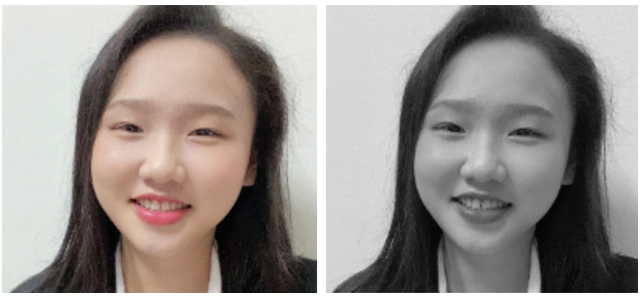
the scale of the image. Several publications by Akshara J. *et al*., Arun K. *et al*. and Chintalapati, S. *et al*. advocated using histogram equalisation to facial images and preprocessing the images by scaling [8, 9, 10].

Table 1: Difficulties of Face Detection

Difficulties	Explanation
Background	Changes in the background and surrounding of the person in the image will influence the face detection accuracy.
Light Level	Various lighting environments reduce the ability to detect facial features.
Pose	The different angles of the captured facial images distort the face recognition process.
Expression	Changes in expressions cause changes in spatial relationships and changes in the shape of facial features
Occlusion	If there is a part of the face that is not observable, it will affect the performance and face recognition due to the not enough information provided.
Rotation, scaling and translation	Transformation of the image may distort the original information of the image.

In order to improve the system performance, pre-processing of input images is required [6]. It is important for enhancing the accuracy of facial recognition. One of the required preparatory stages for processing the image's size is scaling. Due to the reduced number of pixels, scaling of images can increase the processing speed by reducing the system computation. The image's size and pixels contain its unique spatial information. The spatial information is important since it provides a measurement of the image's least identifiable detail. As a result, spatial data must be treated with care to avoid picture distortion and tessellation effects. For normalization and standardization purposes, the dimensions of all images should be the same. The length and width of the image are preferred to be the same size based on the proposed Principal Component Analysis (PCA).

In the pre-processing stage, photographs in color mode are commonly converted to grayscale mode, as shown in Figure 1. A grayscale image is commonly referred to as a black and white image, but the name emphasizes that such an image will also include many shades of gray. Grayscale images are considered to be less sensitive to lighting conditions and to calculate faster. A colour image is a 24-bit image with pixels ranging from 0 to 16777216, whereas a grayscale image is an image with 8-bit and pixels ranging from 0 to 255 [6]. As a result, colour photographs demand more storage space and processing power than grayscale ones [11]. If the colour picture is not required for the computation, it is referred to as noise. Furthermore, preprocessing is required to improve the image's contrast. Histogram equalisation is one way of pre-processing to increase the image's contrast [12]. It may decrease the effect of uneven lighting while providing a consistent intensity distribution on the horizontal axis of intensity.



(A) Coloured Image *(B) Grayscale Image*

Figure 1: Image convert from (A) to (B)

2.2 Face Recognition

Facial recognition identifies or verifies an individual from either image, video frame or real-time. It is commonly used as access control in security systems and can be compared to other biometric techniques such as fingerprint or eye iris recognition systems [13]. The face recognition will capture the face from a group of faces, and then it will identify the details of the face. After that, it will match the faces available in the current storage location, and it will find the exact face that corresponds with the current face in the database. The process of face recognition is shown in Table 2.

Table 2: Process of Face Recognition

Process	Explanation
Face Detection	Face detection is a process of detecting the face that is located in a frame or image by finding the landmarks of the face such as eyes and nose.
Face Alignment	Face alignment plays an important role during the image pre-processing. The face and eye regions are automatically detected and faces are aligned according to translation, scaling and rotation. The face alignment is essential in image pre-processing as it is able to make the face detected more readable for the system to track the features.
Features Extraction	The facial recognition will be producing 128 dimensions embeddings once the pre-processed face is ready. The face that passes into the system, the system will check the face and generate 128 numbered measurements from the image in order to let the computer read the image as it cannot actually observe the complete visual appearance of the image. The number generated is used by the computer for comparison.
Face Recognition or Classification	The system will compare the measurements that are obtained from the input image to what is already in the database. After that, the score for each and every match will be generated, if the score is more than a particular threshold, then it is considered as a match face.

2.3 Hand Gesture Detection

Hand gesture can be described as one of the most natural and intuitive ways of communication between humans and machines, especially in the Human Computer Interaction (HCI) field, because it closely mimics the way of interaction between humans [14]. In order to detect hand gestures, these processes must be passed through, that is, input images or frames through the sensor, execute the Application Programming Interface (API) for image processing, and finally display the returned results [15]. In these processes, efficient API has played a very important role in Hand Gesture Detection. Until now, a lot of Hand Gesture Detection APIs have been released by others, such as Tensorflow Handpose and MediaPipe Hands. These APIs have different architectures to process the input that result in different accuracy and efficiency of Hand Gesture Detection.

3.0 PROPOSED METHOD

We use Google Meet as the main platform for the virtual classroom. Plugins are required for our prototype development using Google Meet through Google Chrome extension. Figure 2 shows the results of using hand gesture detection and face detection in Google Meet Platform where the hand and face landmarks will be drawn while the feature is detected.

Figure 2: Snapshot of hand gesture and face detection in Google Meet

Our experimental work utilizes the following hardware and software as shown in Table 3 for realizing the virtual classroom prototypes. For instance, a higher hardware specifications machine, i.e., machine C, as shown in Table 3 is utilized to evaluate the prototype performance and the software used in the tested proposed solution. In addition, a 0.922 megapixel 1080P high-definition webcam is used in this work.

Table 3: Hardware and software specification for proposed work

Machine	Hardware Specification	Software Specification
A	Intel Core i5-8300H, 2.30GHz 16GB RAM, GTX 1050	Python 3.6 with OpenCV2, Tensorflow, MediaPipe, Visual Studio Code with (HTML, CSS, JSON, JS) Google Firebase Google Meet
B	Intel Core i7-7700HQ, 2.8GHz 24GB RAM, GTX 1050	
C	Intel Core i9-9900KF, 3.60GHz 128GB RAM, RTX 2080	

4.0 RESULT AND DISCUSSION

We collected the results on libraries' efficiency used for face detection and hand gesture detection. For hand gesture detection, results of each Tensorflow and MediaPipe models in detecting hand landmarks are recorded. In order to ensure the consistency of the generated results from the same API, we use a series of recorded videos with the same hand gesture movement as a baseline video. We have implemented a frame per second (FPS) counter in the code itself instead of using Google Chrome's default FPS meter to achieve a more reliable FPS. In addition, we use the confidence provided by API and counts to calculate the accuracy of the model's recognition of Hand Landmark in the recorded video.

The model load time of Tensorflow Handpose and MediaPipe Hands are collected in machine A, as shown in Figure 3. The model is loaded for 10 times and its average value is calculated. It is analysed that Tensorflow Handpose (TFJS), i.e., in blue line, requires more time to load the model in the Google Meet compared to MediaPipe Hands. MediaPipe Hands shows faster performance, 333.5 times faster than TFJS, because it requires very less Graphic Processing Unit (GPU) computing power to execute. However, higher hardware specifications, i.e., machine C, can reduce the time required to load Tensorflow Handpose, as shown in Figure 4.

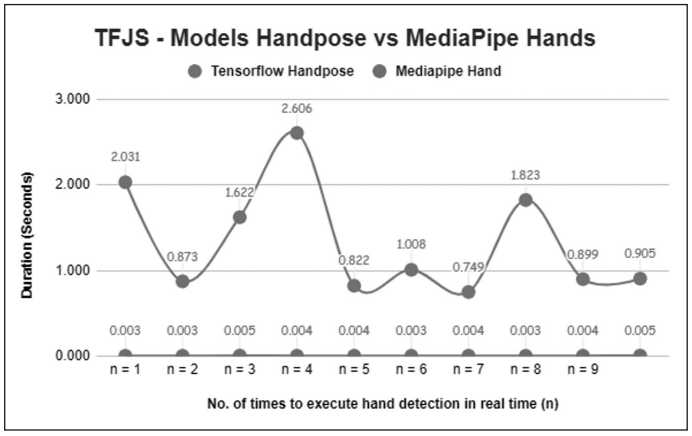


Figure 3: Results of TFJS Handpose vs MediaPipe Hands in Machine A

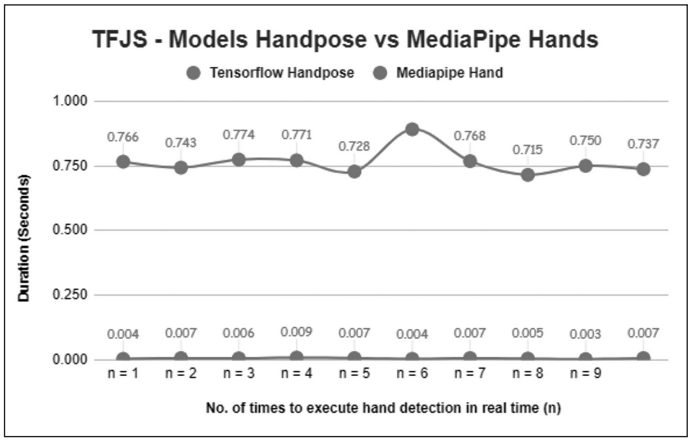


Figure 4: Results of TFJS Handpose vs MediaPipe Hands in Machine C

Table 4 shows the backend library, FPS, confidence level and accuracy of the model used in detecting the hand of the user. Under the low and high hardware specification requirements, i.e., machine A and C, both of them showed that the performance of MediaPipe Hands has a slightly higher FPS than Tensorflow Handpose due to less computational resources required for execution. Nevertheless, due to GPU support, the average accuracy of Tensorflow Handpose is higher than MediaPipe Hands.

The results of time taken for the face detection in three different libraries that include face-api.js, Tensorflow and MediaPipe are collected and discussed at the section below. For the comparison of face detection between the three different libraries, an image video is used for gathering the results of the time taken of face detection for each library in both machine B and C.

In Figure 5, it is analysed that MediaPipe requires more time to execute the face detection compared to others. The performance for the libraries to execute the face detection can be improved by using a machine with higher specification. The same step is carried out for gathering the execution time of face detection in machine B where each of the libraries will run for 10 times and an average time taken for the face detection is calculated. By using machine B, the performance of the face-api.js is 1.78 times and 2.14 times better than TensorFlow and MediaPipe respectively. Due to the lower hardware specification of machine B, the time taken among each of the libraries provides a bigger gap compared to the same libraries running in machine C. A performance analysis of executing different libraries in machine C is illustrated in Figure 6.

Figure 6 illustrates the time taken for face detection among different types of library used in machine C. In order to calculate the average time taken for each of the libraries used, a face detection program with each of the libraries is executed 10 times. Based on the figure, MediaPipe provides the highest time taken for face detection execution that represents that MediaPipe has the worst performance among these libraries. In addition, the sequence of performance from high to low of the libraries for face detection is face-api.js, TensorFlow and MediaPipe. Face-api.js shows the least execution time for face detection where it is 1.89 times and 1.76 times faster than the MediaPipe and TensorFlow respectively. Besides, the performance for different libraries are affected by the hardware limitations where the performance for the libraries in machine C is better than in machine B.

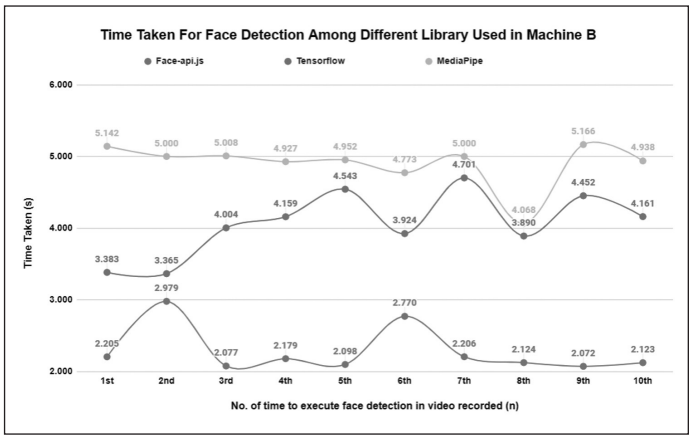


Figure 5: Results of Face-api.js, Tensorflow and MediaPipe Face Detection in Machine B

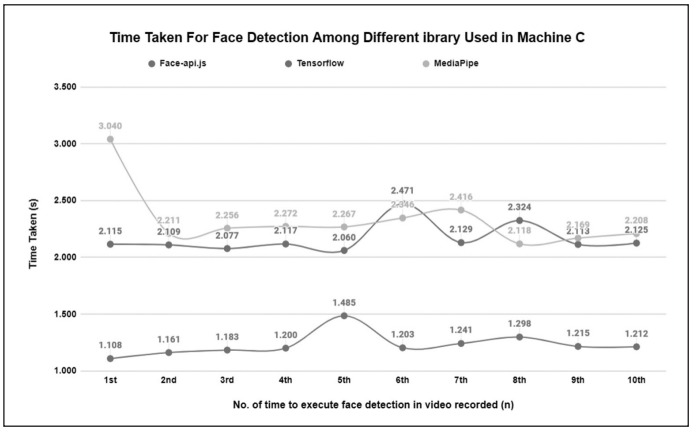


Figure 6: Results of Face-api.js, Tensorflow and MediaPipe Face Detection in Machine C

5.0 CONCLUSION

In a nutshell, through the analysis of Tensorflow Handpose and MediaPipe Hands, in terms of accuracy, Tensorflow Handpose (~99.4%) is higher than MediaPipe Hands (~95.4%). However, in terms of FPS, MediaPipe Hands is more stable in both low and high hardware specification requirements (i.e., machine A and machine C) than Tensorflow Handpose. Through the research on hand gesture detection, in order to have high efficiency on

Table 4: Comparison of Tensorflow Handpose and MediaPipe Hands

Model		**TensorFlow - Handpose**						**MideaPipe - Hands**				
Backend		Web Graphics Library (WebGL)						Web Assembly (Wasm)				
Machines		A			C			A			C	
FPS		~ 23 FPS			~55 FPS			~ 50 FPS			~ 60 FPS	
Total Confidence from API	590	512	364	1733	1711	1719	989	1156	998	1904	1883	1891
Total Count Allocated	595	516	370	1744	1782	1732	1041	1209	1046	1995	1973	1985
Accuracy	99.22	99.23	98.43	99.39	99.41	99.30	94.99	95.64	95.41	95.43	95.44	95.27
Average Accuracy		98.96			99.37			95.35			95.38	

low-specification hardware, it is recommended to use MediaPipe Hands. On the contrary, Tensorflow Handpose is more suitable for high-specification hardware.

Moreover, the face-api.js provides the best performance in executing the face detection function compared to TensorFlow and MediaPipe, therefore the library is used for implementing the face detection in a virtual classroom system. Although there is a difference of average time execution for face detection among each of the libraries in different machines, each of the libraries in different machines shows similar trends where the face-api.js provides the highest performance compared to others. According to the research, face-api.js is recommended for high efficiency on low specification hardware, whereas MediaPipe is more suitable for high specification hardware.

The performance of the libraries may be affected based on the programming languages, therefore a comparison for the performance of the libraries in different programming languages such as python versus javascript can be carried out in the future. Furthermore, in addition to detecting face and hand landmarks, other AI components such as face emotion, liveness detection can be examined and included in virtual classroom systems. Besides, there are enormous face detection library resources that are able to implement the AI work. Nevertheless, each resource may come with different compatibility with recent hardware and updated version software tools and requires further analysis and investigation in order to improve the performance of the virtual classroom system in the future.

6.0 ACKNOWLEDGMENTS

With our sincerest greetings, we would like to express our gratitude to all those who provided great help, moral assistance and cooperation to our research team. We are grateful to the Faculty of Computing and Information Technology (FOCS) of Tunku Abdul Rahman University College (TAR UC) for providing the facilities at the Centre of Computational Intelligence [16] and Centre of Excellence Big Data Analytics and AI to make the project successfully implemented with preliminary results, as shown in this paper. ■

REFERENCES

[1] Yu Fei, L, Saleh, S, Jiahui, H, Mohamad, S & Abdullah, S, 2020, 'Review of the Application of Artificial Intelligence in Education', International Journal of Innovation, Creativity and Change. www.ijicc.net, vol. 12, no. 8, p. 548 - 555, viewed 20 September 2021, <https://www.ijicc.net/images/vol12/iss8/12850_Yufei_2020_E_R.pdf>.

[2] Rapanta, C, Botturi, L, Goodyear, P, Guàrdia, L & Koole, M 2020, 'Online University Teaching During and After the Covid-19 Crisis: Refocusing Teacher Presence and Learning Activity', Postdigital Science and Education, vol. 2, no. 3, pp. 923–945, viewed 20 September 2021, <https://link.springer.com/article/10.1007/s42438-020-00155-y>.

[3] Olszewska, J. I. (2021), The virtual classroom: A new cyber physical system, in '2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)', pp. 187-192.

[4] AIT Staff Writer 2021, Role of Artificial Intelligence in Online Education, AiThority, viewed 20 September 2021, <https://aithority.com/technology/education-and-research/role-of-artificial-intelligence-in-online-education/>.

[5] J. Valera, J. Valera and Y. Gelogo, 2015, "A Review on Facial Recognition for Online Learning Authentication," 2015 8th International Conference on Bio-Science and Bio-Technology (BSBT), pp. 16-19, doi: 10.1109/BSBT.2015.15.

[6] Howard, C. (2018), 'Face recognition based automated student system', Universiti Tunku Abdul Rahman.

[7] Singh, S., Rastogi, R. & Sharma, P. S. (2015), Automatic lecture attendance system using face recognition, in 'MATRIX Academic International Online Journal of Engineering and Technology', Vol. III, pp. 36–40.

[8] Akshara J., Akshay J. T. L., Krishna Y. (2017), Automated attendance system using face recognition, in 'International Research Journal of Engineering and Technology (IRJET)', Vol. 4, pp. 1467–1471.

[9] Arun K., Sudesh V. K., Amar. P. Z., Nikhil. D. B. & Chetan. J. B. (2017), Attendance system using face recognition and class monitoring system, in 'International Journal on Recent and Innovation Trends in Computing and Communication', Vol. 5, pp. 273–276.

[10] Chintalapati, S., Raghunadh, M. V. (2013), Attendance system using face recognition and class monitoring system, in 'International Journal on Recent and Innovation Trends in Computing and Communication'.

[11] Kanan, C. and Cottrell, G., 2012. Color-to-Grayscale: Does the Method Matter in Image Recognition?. PLoS ONE, vol. 7, no. 1, pp. 29740.

[12] Pratiksha M. Patel, 2016, Contrast Enhancement of Images and videos using Histogram Equalization. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 4, no. 11.

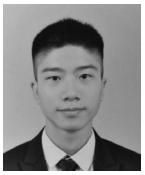
[13] Virgil Petrescu, R. V., 2019, Face Recognition as a Biometric Application. Journal of Mechatronics and Robotics, vol. 3, no. 1, pp. 237–257. <https://doi.org/10.3844/jmrsp.2019.237.257>.

[14] Ren, Z., Meng, J., Yuan, J. & Zhang, Z. (2011), Robust hand gesture recognition with kinect sensor, pp. 759–760.

[15] Zhang, F., Bazarevsky, V., Vakunov, A., Thachenka, A., Sung, G., Chang, C.-L. & Grundmann, M. (2020), 'Mediapipe hands: On-device real time hand tracking'.

[16] Centre of Computational Intelligence (2020), Tunku Abdul Rahman Universiti College, viewed 10 Sept 2021,<https://www.tarc.edu.my/focs/research/cci/>.

PROFILES



KAH YEE LIM is a student of Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kuala Lumpur. He received his Bachelor's Degree from Tunku Abdul Rahman University College in 2022. His areas of interest are web application development, game and graphic design.
Email address: limky-wm18@student.tarc.edu.my

HAU JOAN is a student of Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kuala Lumpur. She received her BS degree from Tunku Abdul Rahman University College in 2022. In 2022, she joined the Faculty of Computing and Information Technology to pursue her MS degree at Tunku Abdul Rahman University College, Kuala Lumpur. Her research interests are video compression and video coding standard.
Email address: joanh-wm18@student.tarc.edu.my

YIQI TEW obtained his B.Eng. (Hons.) electronics degree from Multimedia University, MComSc. from National University of Malaysia, and PhD degree from University of Malaya in 2018, 2011 and 2016 respectively. He is an experienced Associate Professor and skilled engineer in Embedded Systems, Video (HEVC) Processing, Security (Information Hiding). His research involves Multi-view, Depth Sensing and Real-time Video Streaming Mechanisms in Industry 4.0 and Smart Agricultural projects. With his profession, he became the lead researcher in Computational Intelligence with publication of more than 30 papers, appointed as National Expert in IoT and Cloud Computing by the Ministry of Human Resource, Malaysia for WorldSkills events in 2018, 2020 and 2021.
Email address: yiqi@tarc.edu.my

DESIGN OF INTEGRATION FRAMEWORK TOWARDS INDUSTRY 4.0: FROM SHOP FLOOR TO CLOUD

(Date received: 12.01.2022/Date accepted: 22.03.2022)

Lee Yoon Ket[1]*, Chiew Tsung Heng[1], Ong Jia Jan[1], Chang Kai Ming[1], Chan Tai Wei[2], Chan Wah Beow[3], Eyo Geak Loo[4]

[1] Department of Mechanical Engineering, Tunku Abdul Rahman University College, Jalan Genting Kelang, 53300 Setapak, Kuala Lumpur, Malaysia.

[2] Business Development & Data Analysis Department (BDDA), Asia Roofing Industries Sdn. Bhd. (subsidiary of Ajiya Berhad), No.4, Jalan Sungai Pelubung 32/149, Seksyen 32, 40460 Shah Alam, Selangor, Malaysia.

[3] Engineering Department, Asia Roofing Industries Sdn. Bhd. (subsidiary of Ajiya Berhad), No. 4, Jalan Sungai Pelubung 32/149, Seksyen 32, 40460 Shah Alam, Selangor, Malaysia.

[4] Centre for Technology Research, Innovation and Sustainability (C-TRIS), Asia Roofing Industries Sdn. Bhd. (subsidiary of Ajiya Berhad), No. 4, Jalan Sungai Pelubung 32/149, Seksyen 32, 40460 Shah Alam, Selangor, Malaysia.

**Corresponding author: leeyk@tarc.edu.my*

ABSTRACT

The capability of the fourth industrial revolution in fulfilling the ever-changing demand of customers has attracted great attention from the small medium enterprises. Industry 4.0 is one of the ideal solutions to advance the manufacturing to a new level to tackle the high variation in productions of roofing industry. A three-layer Industry 4.0 integration framework, adoptable by a considered roofing company was proposed in this study. A conceptual mechanical designs of automated stacking systems also have been proposed to automate the considered stacking process in the roofing company to minimize the manual labour workforce. The proposed framework realized the complete integration from field layer to the Cloud system using Automation Markup Language and OPC Unified Architecture technologies as the communication backbone. The feasibility of the proposed framework was demonstrated through experiment conducted on an experimental system model where the Google Cloud, the OPC UA-AutomationML server, and a microcontroller-controlled system was used as the enterprise layer, communication layer, and the field layer, respectively. The success of integration was proven by controlling the light intensity and ON/OFF status of a physical LED through the Cloud. For future works, the proposed framework could be implemented on a simulated model of real-time prototype or a real-time manufacturing line to test its robustness.

Keywords: AutomationML, Cloud, Industry 4.0, Integration, OPC UA

1.0 INTRODUCTION

The fourth industrial revolution has becomes the main interest of the world especially the small medium enterprises (SMEs) and multinational corporations. The essence of Industry 4.0 that allows the manufacturers to cope with the ever-changing demand greatly attracts the interest of the roofing industries that require a series of manufacturing processes in their production lines to produce a wide range of roofing with different types of profiles, sizes, materials composition, and number of layers in order to cater various purposes and users, from residential houses to industrial commercial buildings. The presence of high variability in the demand that involved multiple manufacturing processes posed a great challenge to the roofing industries. Consequently, the fourth revolution is desirable to satisfy the ever-changing demands of users by advancing the overall manufacturing structures to the next level. A similar challenge was faced by the roofing industries in Malaysia. Although a national policy on Industry 4.0 (MITI, 2018) was launched by the government to drive the revolution of the manufacturing sectors in Malaysia, majority of the performed works observed in the industries were limited to digitalization of data and big data analytic (Hizam-Hanafiah and Soomro, 2021). In addition, many architectures and models of Industry 4.0 have been introduced (Pethig *et al.*, 2017; Tantik and Anderl, 2017; Xun and Seung, 2019) but the practical application of such models in roofing industries was still inadequate.

A roofing company based in Malaysia was considered and studied in this work. The considered roofing company manufactures various types of roofing products with different lengths in a manufacturing line. The main concern of the

considered roofing company is the transferring and stacking processes of roofing products that relied on an extensive labor workforce due to heavy weight and to avoid deflection, especially for long length roofing products. Moreover, a flipping process is also required to stack the roofing products in an odd and even arrangement. However, this high number of workers is not cost efficient as the quantity of annual orders for such a long length of roofing products is relatively low. Furthermore, the manual labor transfer process also led to low efficiency in terms of time. Hence, an automation in the stacking process including the flipping feature is desired.

This paper proposes a conceptual mechanical design of automated stacking systems to automate the considered stacking process. In addition, this paper also proposes an Industry 4.0 integration framework based on the Reference Architectural Model Industry 4.0 (RAMI 4.0) model that is able to be adopted by the considered roofing company. The proposed framework would allow the complete integration from shop floor to Cloud system in which the automated stacking system would respond in correspondence to to the orders received through the Cloud system.

2.0 CONCEPTUAL DESIGN FOR AUTOMATED STACKING SYSTEM

An on-site primary study was conducted to ensure the feasibility of the designs, by considering: (i) the on-site available space, (ii) the weight and sizes of the roofing products, and (iii) the holding force of the vacuum suction cup.

The main purpose of the primary study was to determine the suitable number of vacuum suction cups in order to hold certain sizes of roofing products with respective weight. The dimension of the overall conceptual designed system was also estimated through the primary study to ensure that the overall designed system would be able to fit into the available spaces in the factory. Based on the study through on-site measurement and review of specification of products, the total maximum weight of the roofing product with a length of 20 m is 115.418 kg. By setting a vacuum level of 0.3 bar, a diameter of 100 mm for suction cup, and the total load of a product with 20 m in length = 2390.88/4 N (assumption; number of lifter modules = 4), the number of suction cups required per module was calculated. The proposed suction cup lifter for a single module comprises ten suction cups and arranged evenly in two columns with five suction cups per column. More than one lifter module could be used depending on the length of the roofing products. The considered safety factor for a module is four to avoid any immediate threat if any malfunction occurred in one of the vacuum cups.

Three conceptual designs were proposed through the extensive research on the market available mechanisms and technologies, e.g. flipping using multi-layer conveyors, slotting discs, L-beam flipper, overhead lifters, and guided rails. The overall conceptual designs of the stacking system were produced using the SOLIDWORKS software and the vacuum suction mechanisms (Jaiswal, 2017) were utilized in the proposed designs. The Quality Function Deployment (QFD) tool was used to determine the product design specifications by considering the safety, size, installation, maintenance, quality, and reliability. The final conceptual design was chosen for the considered roofing company based on the Pugh analysis matrix.

2.1 Final Conceptual Design

Figure 2-1 depicts the final conceptual design for the automated stacking process. A linear gantry system with vacuum suction lifter module was proposed for the transportation of the products. The foam rubber type of suction cup was proposed as it could provide suction grip evenly on uneven surfaces. A series of resting tables was proposed for the flipping process in the zone A. Two L-beams were proposed to perform the flipping mechanism on the resting tables as shown in Figure 2-2. The incoming roofing products could be lifted by the gantry system, transferred and stacked at zone B. The series of roller beds at the zone B was used to transfer out the stacked products as outgoing products. This design required lower cost as no robotics arms were used but still able to provide the ease in transferring out the stacked products.

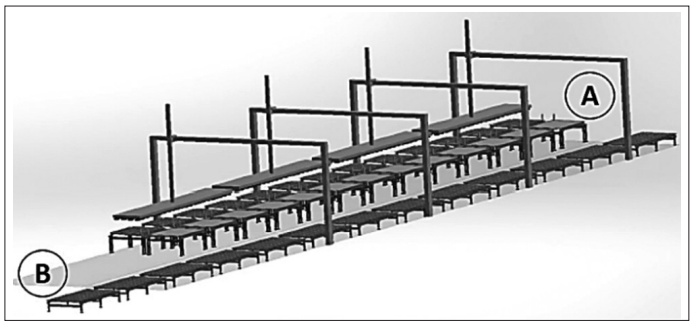


Figure 2-1: Third conceptual design of automated stacking system

Figure 2-2: The flipping process was demonstrated: (a) resting table with two L-beams, (b) product on roller table being flipped up, (c) the flipped up product received by two L-beams of resting table, and (d) the product flipped and located on the resting table

3.0 PROPOSED INTEGRATION FRAMEWORK

The proposed conceptual design of the automated stacking system could be controlled through industrial controllers such as programmable logic controllers (PLCs) or other industrial microcontrollers. An Industry 4.0 integration framework based on the RAMI 4.0 model that enabled the complete integration from shop floor to Cloud system was proposed. In general, the RAMI 4.0 is a reference architecture model for Industry 4.0 that describes the fundamental requirements to comply with the Industry 4.0 systems by linking the system lifecycle, value stream, hierarchy, and functional layers (Xun and Seung, 2019).

In the proposed integration framework, two vital technologies, namely; the Automation Markup Language (AutomationML) and Open Platform Communications Unified Architecture (OPC UA) were utilized as the backbone of the communication and data exchange. The AutomationML (IEC, 2014) is an open and neutral XML-based object oriented data modelling language that realizes the data exchange

between multiple fields including the mechanical engineering, control programming, electrical design, communication and management systems throughout the lifecycle of a production system (Luder *et al*., 2010; Schleipen *et al.,* 2014). It is able to close the data exchange gap of heterogeneous autonomous engineering tools in industry by extending, adapting and merging the existing standardized data formats (Xun *et al.*, 2018). On the other hand, the OPC UA is a client-server mechanism that is able to realize the interconnectivity and interoperability by serving as the communication interface for heterogeneous network fields in industry. The works of Xun *et al.* (2018) and Fuchs *et al.* (2020) explained the details of AutomationML and OPC UA. Technically, the overall proposed integration framework consists of three layers, namely; (i) enterprise layer, (ii) communication layer, and (iii) field layer; complied with the RAMI 4.0 model, as shown in Figure 3-1.

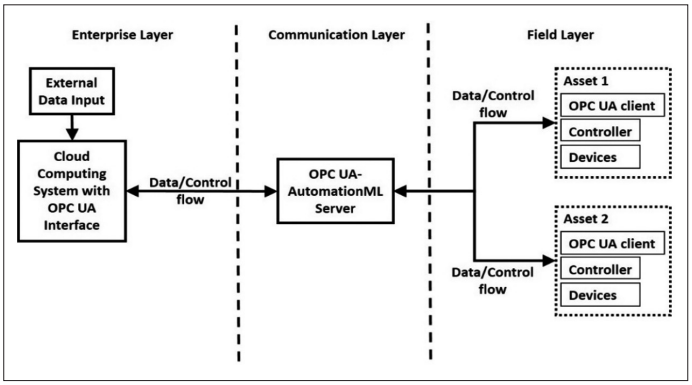


Figure 3-1: The proposed integration framework from field level to Cloud system for roofing industry

3.1 Enterprise Layer

A Cloud platform accessible by all parties would be used in this layer. The management team, engineers or human operators could provide any control commands at this layer through the Cloud platform to alter or adjust the operation or field devices state. External databases or files could be input, saved and backed up into the Cloud platform. These input and updates also would be reflected in the field devices. An OPC UA client is required to be established at this layer to connect with the OPC UA-AutomationML server at the communication layer. In short, all information provided by the OPCUA-AutomationML server at the communication layer could be displayed and acquired in the enterprise layer to allow the decision making of users. Conversely, users also could change the operation states in the OPC UA-AutomationML server.

3.2 Communication Layer

The communication layer is vital in the overall framework as it acts as a bridge to connect the enterprise and field layers. The OPC-UA-AutomationML server acquires and hosts the overall information of the assets in the field layer. It is also able to enable the information flow to any connected external clients such as the enterprise layer and vice-versa. For example, the information of a robotic arm system could be acquired from the server and passed to the Cloud system in the enterprise layer. The order of products could also be acquired from the Enterprise layer to activate suitable number of lifter modules for the respective length of products.

3.3 Field Layer

The filed devices and their respective controllers are considered in the field layer. These field devices would be defined as the assets of a manufacturing line. For example, the robotic arm systems, roller bed systems and flipping systems as well as their respective controllers such as PLCs or microcontrollers of the automated stacking system for the roofing products can be considered as the assets. Respective interface such as OPC UA client for each controller of field devices is required to be established in order to be able to communicate with higher layer.

4.0 FRAMEWORK IMPLEMENTATION AND PRELIMINARY TESTING

Figure 4-1 shows an experimental system model corresponding to the proposed framework shown in Figure 3-1. For the enterprise layer, the Google Cloud platform was used to keep the MySQL Workbench database. The MySQL Workbench is also able to read the external data such as the sales order of roofing products in Comma Separated Values (.CSV) files format and update the database accordingly. For example, MySQL Workbench could read the received orders from the customers in .CSV format, and update the respective database.

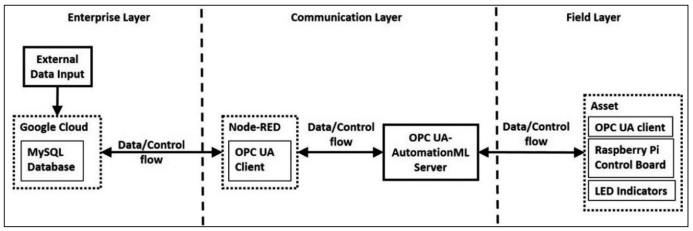


Figure 4-1: The proposed integration framework from field level to Cloud system for the experimental system model

The communication layer consists of two important elements: (i) Node-RED, and (ii) OPC UA-AutomationML server. The Node-RED was used to acquire the order data from the Google Cloud platform and to establish the OPC UA client as shown in Figure 4-2. The creation of the OPC UA client allowed the connection between the Enterprise layer and the OPC UA-AutomationML server. The client was used to send/receive OPC UA service request/response messages to/from the OPC UA-AutomationML server. This server was created based on the developed hierarchy of all related assets from AutomationML Editor, by using AML2OPCUA tool (developed by Fraunhofer IOSB) and acted as the backbone to connect both enterprise and field layers through OPC UA networking.

For the field layer, the LED indicators and its respective controller, the Raspberry Pi control board were used as the field devices in the experimental system model as shown in Figure 4-3. An OPC UA client was created using Python programming in the Raspberry Pi control board to realize the connectivity between the field layer and the OPC UA-AutomationML server. Any data or command change in the Cloud or server could alter and update the status of field devices in the client and vice-versa. A control command on the field devices, such as turning ON/

OFF and changing the light intensity of the LED indicator was input into the Cloud through MySQL workbench (Figure 4-3). The status of changes could be observed through the UaExpert client tool and the physical field device, the LED indicator.

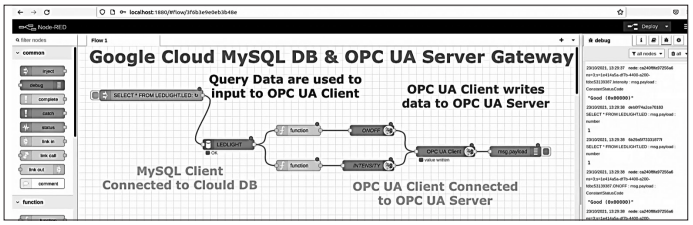


Figure 4-2: The developed Node-RED program used in the experimental system model

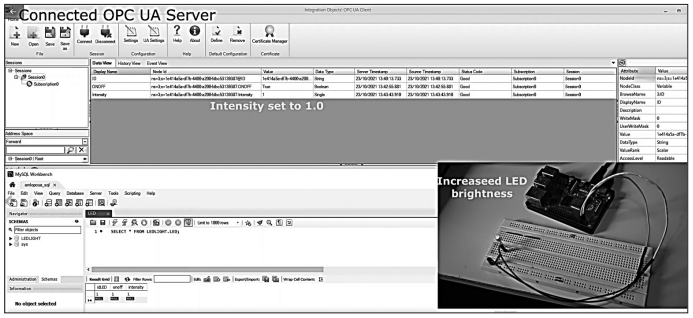


Figure 4-3: The connected physical LED indicator with the OPC UA-AutomationML server to the Google Cloud platform for the experimental system model

Experiment was conducted to test the connectivity of the proposed integration framework by varying the data in the Google Cloud platform through the MySQL Workbench to turn on/off the LED and change the light intensity of the LED. For the ON/OFF status of the LED, two commands were used where the value of 0 indicates the OFF status while the value of 1 indicates the ON status. On the other hand, the range of value between 0.1 and 1 were used to adjust the light intensity of the LED. The value of 0.1 indicates a very low intensity (dim light) while the value of 1 indicates a very high intensity (bright light). For the ease of comparison, only 3 sets of results with values of 0.1, 0.5 and 1 were used in the light intensity experiment.

For the ON/OFF status experiment, the data or command of either 0 or 1 was input into the MySQL Workbench. The new data would update the database in the Google Cloud platform. This updated data would then be written to the OPC UA-AutomationML server through the established OPC UA client in the NodeRED. The updated data or status in the OPC UA-AutomationML server was observed through the UaExpert client tool. The new status in the server would be acquired by the OPC UA client and change the status of the physical LED indicator through the Raspberry Pi control board. The same flow of procedure was applied for the light intensity experiment. The results for both experiment were summarized in Table 4-1 and Table 4-2 respectively.

Based on the results in Table 4-1 and Table 4-2, it was clearly observed that the connectivity of the proposed integration framework from the Cloud to the field devices has been established successfully. These preliminary experiment proven that the data able to flow from the Cloud and update the status of the physical field devices. According to the observation in Table 4-1, the LED indicators in the field layer can be turned ON/OFF by updating the status through the Cloud. A similar observation also has been recorded in the light intensity experiment where the physical LED indicator was able to successfully respond, corresponding to the light intensity set through the Cloud. The success of the preliminary experiment has proven the feasibility of the proposed integration framework. This feasible framework could be used as the fundamental reference and be adapted into the roofing company by adding necessary assets at the shop floor such as the PLC, motors and sensors of automated stacking systems to form the connection to the enterprise layer. Engineers and top management team would be able to monitor and manage the shop floor processes through the considered Cloud system.

Table 4-1: The summary of results for the LED ON/OFF status experiment

Data/ command in the Google Cloud platform	Status observed in OPC UA-AutomationML server through UaExpert client tool	Status of physical LED indicator
0 (OFF)	0	Light off
1 (ON)	1	Light on

Table 4-2: The summary of results for the light intensity of LED indicator experiment

Data/ command in the Google Cloud platform	Status observed in OPC UA-AutomationML server through UaExpert client tool	Status of physical LED indicator
0.1	0.1	Very dim
0.5	0.5	Moderate intensity
1	1	Very bright

5.0 CONCLUSION AND FUTURE WORKS

This study proposed a three-layer integration framework for the industrial process control system of the roofing industry towards the Industry 4.0. In the proposed framework, the OPC UA and AutomationML technologies were utilized as the vital communication and bridging between the enterprise and field layers. Two experiment were conducted on the experimental system model and the obtained results proven the connectivity from the Cloud system at the enterprise layer to the physical LED indicators at the field layer, through to the AutomationML-OPC UA server at the communication layer. The proposed framework also suggested the feasibility in applying any of the conceptual designs for automated stacking systems as the asset of the field layer. For future works, the proposed framework could be applied on a real-time prototype or industrial manufacturing line to test its robustness. Furthermore, it also could be extended to incorporate other elements in the manufacturing management system such as the manufacturing execution system.

6.0 ACKNOWLEDGEMENTS

Authors would like to acknowledge the Centre for Autonomous Systems and Robotics Research (CASRR), Faculty of Engineering and Technology, Tunku Abdul Rahman University College and Asia Roofing Industries Sdn. Bhd. (subsidiary of Ajiya Berhad) for the financial and facilities support. ■

REFERENCES

[1] Fuchs, J., Schmidt, J., Franke, J., Rehman, K., Sauer, M., & Karnouskos, S. (2019). I4.0-compliant integration of assets utilizing the asset administration shell. 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, 1243–1247. 10.1109/ETFA.2019.8869255.

[2] Hizam-Hanafiah, M., & Soomro, M. A. (2021). The situation of technology companies in Industry 4.0 and the open innovation. Journal of Open Innovation: Technology, Market, and Complexity, 7(34), 1–20. https://doi.org/10.3390/joitmc7010034.

[3] IEC (2014). IEC Standard 62714-1:2014. Engineering Data Exchange Format for Use in Industrial Automation Systems Engineering – Automation Markup language – Part 1: Architecture and General Requirements.

[4] Jaiswal, A. K., & Kumar, B. (2017). Vacuum gripper – An important material handling tool. International Journal of Science & technology, 7(1), 1–8.

[5] Luder, A., Hundt, L., & Keibel, A. (2010). Description of manufacturing processes using AutomationML. IEEE 15th Conference on Emerging Technologies and Factory Automation (ETFA), Bilbao, 1–8. 10.1109/ETFA.2010.5641346.

[6] MITI (2018). Industry 4WRD: National policy on Industry 4.0. Perpustakaan Negara Malaysia, Kuala Lumpur, Malaysia.

[7] Pethig, F., Niggemann, O., & Walter, A. (2017). Towards Industrie 4.0 compliant configuration of condition monitoring services. IEEE 15th International Conference on Industrial Informatics (INDIN), Emden, 271–276. 10.1109/INDIN.2017.8104783.

[8] Schleipen, M., Henben, R., Damm, M., Luder, A., Schmidt, N., Sauer, O., & Hoppe, S. (2014). OPC UA and AutomationML – collaboration partners for one common goal: Industry 4.0. 3rd AutomationML User Conference, Blumberg, 1–3.

[9] Tantik, E., & Anderl, R. (2017). Potentials of the asset administration shell of Industrie 4.0 for service-oriented business models. Procedia CIRP, 64, 363–368. 10.1016/j.procir.2017.03.009.

[10] Xun, Y., & Seung, H. H. (2019). Toward Industry 4.0 Components: Insights into and implementation of asset administration shells. IEEE Industrial Electronics Magazine, 13(1), 13–18. 10.1109/MIE.2019.2893397.

[11] Xun. Y, Yuemin, D., & Seung, H. H. (2018). Implementation of a production-control system using integrated Automation ML and OPC UA. 2018 Workshop on Metrology for Industry 4.0 and IoT, Brescia, 1–6. 10.1109/METROI4.2018.8428310.

PROFILES

TS. DR LEE YOON KET is a principal lecturer of the Mechanical Engineering Department in the Faculty of Engineering and Technology, Tunku Abdul Rahman University College (TAR UC). He completed his PhD at Queen University of Belfast, U.K. in 2004. In 2017, Ts. Dr Lee has registered under the Engineering Council U.K. as a member of Institute of Mechanical Engineers with the title of Chartered Engineer. Besides, Ts. Dr Lee also registered with Malaysia Board of Technologists in the Manufacturing and Industrial Technology area with the title of Ts. in 2018. Ts. Dr Lee possesses 7 years of industrial working experiences and 8 years of academic experiences. His research interests mainly focus on application of Industry 4.0 to SMI, Autonomous system and Manufacturing planning and control. Currently, Ts. Dr Lee is leading the Centre of Autonomous and Robotics Research in TAR UC that focuses on applications of Industry 4.0.
Email: leeyk@tarc.edu.my

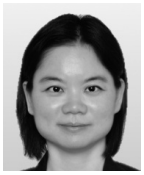
DR CHIEW TSUNG HENG is currently a Programme Leader for Mechatronics Engineering programme and senior lecturer in Department of Mechanical Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University College (TAR UC). His research interest is in the motion control system of machine tools, industrial automation and robotic technologies, and Industry 4.0. He received the PhD, MSC in Manufacturing Engineering, and Bachelor in Manufacturing Engineering (Robotics and Automation) from Universiti Teknikal Malaysia Melaka. Dr Chiew has 4 years of teaching experience and is now actively involved in industrial collaboration research and projects. He is the core member of the Centre of Autonomous and Robotics Research in TAR UC that is focused on the applications of Industry 4.0. Recently, Dr Chiew has registered under the Engineering Council U.K. as a member of Institute of Mechanical Engineers with the title of Chartered Engineer.
Email address: chiewth@tarc.edu.my

DR ONG JIA JAN is currently a Senior Lecturer in the Department of Mechanical Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University College. He completed his PhD in 2016 and MEng (Honours) Mechatronic Engineering in 2009 at University of Nottingham Malaysia. Since 2017, Dr Ong has been registered under the Institute of Engineering and Technology (IET) U.K. as a Member of IET. Dr Ong possesses 4 years of industrial working experience and 2 years of academic experiences. His research interest area is in Industry 4.0, image data compression, encryption, error correction and minimal instruction set computer.
Email address: ongjj@tarc.edu.my

DR CHANG KAI MING is currently a Senior Lecturer in the Department of Mechanical Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University College. His research interest is in robotics, automation and networking for Industry 4.0. He received his PhD in Electronics and Electrical Engineering, MSc in Bionanotechnology and MEng Mechanical Engineering (Mechatronics) from the University of Southampton, U.K. Dr Chang has 2 years of teaching experience and is a core member of the Centre of Autonomous and Robotic Research (CASRR) in TAR UC. Dr Chang is a member of the IMechE, IET and IEEE. Current projects involve automated vehicles and computer vision for remote monitoring and control of manufacturing activities.
Email address: changkm@tarc.edu.my

MR. CHAN TAI WEI is the Head of Business Development and Data Analysis at Asia Roofing Industries Sdn. Bhd. He completed his Master of Engineering in Electrical and Computer Engineering at University of Nottingham in 2014. He had 3 years of experience in corporate finance and advisory, and 5 years in manufacturing of metal building materials. In 2019, Mr. Chan Tai Wei was certified as TRIZ consultant by Malaysia TRIZ Innovation Association. He has also been invited to speak for prominent events such as International Greentech & Eco Products Exhibition & Conference Malaysia 2021, Pesta MyTRIZ 2018 – TRIZ on Corporate Strategies, Asia-Pacific Digital Technology Symposium 2017, etc. In Asia Roofing Industries Sdn. Bhd., Mr. Chan Tai Wei is in charge of strategizing and driving the business performance and sustainability, planning and implementation of Digital Transformation towards Industry 4.0, and advising the management team on new business developments and corporate exercises such as partnerships, joint-ventures, merger and acquisitions.
Email address: chantw@ajiya.com

MR. CHAN WAH BEOW joined Asia Roofing Industries Sdn. Bhd. in 1992. He has 30 years of industrial experience and is currently the Head of Production (Metal Frame division) and Engineering of the company. In 2019, he was certified as Level 2 TRIZ Practitioner by Malaysia TRIZ Innovation Association. He is also involved in the Industry 4.0 - Digitization project in collaboration with TARUC.
Email address: arisgtpro@ajiya.com



TS. EYO GEAK LOO is the Head of Centre for Technology Research, Innovation and Sustainability (C-TRIS) at Asia Roofing Industries Sdn. Bhd. She completed her engineering degree at the University of Malaya in 2000. She has 22 years of industry experience in manufacturing, quality assurance, product development and process improvement. In 2019, she was certified as Level 2 TRIZ Practitioner by Malaysia TRIZ Innovation Association. In 2021, she is registered with Malaysia Board of Technologists with the title Professional Technologist (Ts.). Currently Ts. Eyo leads the company's projects on Industry 4.0 initiatives, including process automation, shop floor digitization and ERP integration. She is also involved in the Industry 4.0 - Digitization project in collaboration with TARUC.
Email address: eyogl@ajiya.com

EXTRACTIVE SUMMARIZATION ON FOOD REVIEWS

(Date received: 12.01.2022/Date accepted: 10.04.2022)

Yuen Kei Khor[1]*, Chi Wee Tan[1], Tong Ming Lim[2]

[1] *Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Malaysia.*

[2] *Centre for Business Incubation and Entrepreneurial Ventures, Tunku Abdul Rahman University College, Malaysia.*

**Corresponding author: khory0k-wm17@student.tarc.edu.my*

ABSTRACT

Text summarization is a technique to summarize the content of a sizeable text but meanwhile, it keeps the key information. Extractive summarization and abstractive summarization are the main techniques for text summarization. TextRank algorithm, an extractive summarization technique is applied to perform automatic text summarization in this study. Furthermore, GloVe pre-trained word embedding model is used to map each word from the reviews to a vector representation. In the end, the PageRank algorithm is applied to rank the sentences based on their sentence ranking scores. The more important and relevant sentences which can be the representatives of a summary will be placed in a higher rank. The objective of our study is to extract the top five reviews with the highest sentence ranking scores which can form a summary to provide a conspectus of a cookies brand in Amazon food reviews. Besides, a detailed description of the implementation is discussed to provide an overview on using TextRank to create a summary. An analysis of the customer perception based on the summary generated is conducted to understand their needs and level of satisfaction. The final summary demonstrates that Amazon customer reviews for certain cookies brand are generally positive.

Keywords: Text Summarization, Extractive Summarization, TextRank

List of notations

n is the number of individual sentences

1.0 INTRODUCTION

With the advancement of technology, social media platforms and websites have evolved into a place for the public to freely share their opinions, experiences and thoughts about products, services, and breaking news. This vast volume of text contains essential information but reading it all and creating a summary is inefficient for humans so text summarization comes in handy.

Maybury (1999) defined text summarization as the process of distilling the most important information from one or more sources to produce an abridged version for one or more users and one or more tasks. To be more specific, text summarization produces a summary from one or multiple plain texts while retaining important information. Abstractive summarization and extractive summarization are the two main techniques of automatic text summarization. Abstractive summarization made use of advanced natural language techniques such as the deep learning approach to create a completely new and shorter text which consists of the key information from original source. On the other hand, extractive summarization is the extraction of a subset of important sentences from the original source.

This study is motivated by the desire to replace human power in the task of summarizing a lengthy text into a few sentences in a short period of time. Aside from time-consuming issues, human knowledge and language ability level also greatly affect the quality of summaries. Humans may occasionally misinterpret the meaning of text documents. We choose to apply the extractive summarization approach in this study is because it always outperforms abstractive summarization. This is due to the reason that abstractive summarization needs to address issues such as natural language generation, semantic representation, and inference which is difficult for sentence extraction (Allahyari *et al.*, 2017). TextRank is applied due to the research result of Mihalcea and Tarau (2004) proved that TextRank is competitive or better in some cases when compared to previously proposed algorithm that using supervised system. Besides, TextRank also adaptable to different languages and domains as training corpus is not required.

The objective of this study is to extract the top five reviews with the highest sentence ranking scores which can summarize the overall reviews of a cookies brand from Amazon fine food reviews. The top sentence represents the highest chances of the topic discussed by customers in overall product reviews of the product. Therefore, we can have an overview of the customers' perceptions toward certain food products based on the summary.

2.0 LITERATURE REVIEW

This section mainly discusses the main approaches of extractive summarization which are widely used in research works. Basically, there are three independent tasks to perform extractive summarization: create an intermediate representation of the document, score sentences based on the representation, and create a summary by selecting the few most important sentences based on their scoring. There are a few common approaches widely used in text summarization that will be further discussed below.

In the early research on extractive summarization, researchers use features from the sentences such as their position in the text, word frequency, or key phrases indicating the importance of the sentences (Erkan & Radey, 2004). Term Frequency-Inverse Document Frequency (TF-IDF) method is used to determine how important a word is to a collection of documents. TF-IDF scores increase when the number of times that a word appears is increased in a document. This method works in the weighted term-frequency and inverse sentence frequency. Sentence frequency refers to the total number of sentences containing a specific term in the document. The sentence vectors will be scored by similarity and the sentences with the greatest similarity scores are chosen as a part of the summary (Saranyamol & Sindhu, 2014). Christian *et al.* (2017) proposed an automatic text summarizer that uses TF-IDF to extract three to five sentences with the highest TF-IDF scores to be the final summary, where the number of sentences is decided by users. When compared to another online automatic summarizer, their proposed text summarizer yields a 67% accuracy.

Machine learning can be applied for text summarization if the dataset or documents consists of a summary for each observation. This is because machine learning required a large amount of labelled data to train the model. Machine learning models will learn the patterns by identifying those relevant features values that are correlated with the labelled data. Feature extraction take an important role to improve the accuracy of the summarization result. Having more training data leads to better accuracy of the model as they can learn more different patterns. As a result, an extractive summary can be produced for each document when new documents are given to the model. Neto *et al.* (2002) present a text summarizer using two well-known algorithms, Naive Bayes and C4.5 decision tree algorithm with a set of features that are classified into two categories: statistics-oriented and linguistic-oriented. The performance of these two algorithms is compared with two baseline methods with two sets of experiments by employing automatically-produced extractive summaries and manually-produced summaries. Results show that Naïve Bayes outperforms all the summarizers. Besides, the deep learning approach is also quite common in automatic text summarization (PadmaPriya, 2014; Day & Chen, 2018; *Patel et al.*, 2018).

Another well-known approach of extractive summarization is a graph-based approach. The graph-based approach consists of two elements which are nodes and edges. Nodes refer to the sentences while edges are the similarity between sentences. If two sentences share certain common words, they are connected with an edge. When a node has a large number of edges connected to it, then it is considered as an important sentence that should be included in the summary. TextRank is a well-known graph-based approach for text summarization and it is inspired by PageRank (Brin & Page, 1998) which is implemented by Google. Simply, TextRank is used to rank sentences while PageRank rank web pages in Google search engine results. Important pages will have a higher PageRank score and rank higher in the search engine results. Actually, PageRank can be used in text summarization to select the most important sentences from the original text document. In the study of Mallick *et al.* (2019), they proposed a modified PageRank algorithm that assumes that the important sentences are linked (similar) to other important sentences in the text document. Li and Zhao (2016) also proposed a TextRank algorithm by exploiting Wikipedia for short keywords extraction. Their findings show TextRank model constructed based on Wikipedia as external knowledge works better than traditional TextRank which uses TF-IDF.

3.0 METHODOLOGY AND FRAMEWORK

This section shows the framework of the TextRank algorithm in Figure 1 and a detailed description for each step is discussed.

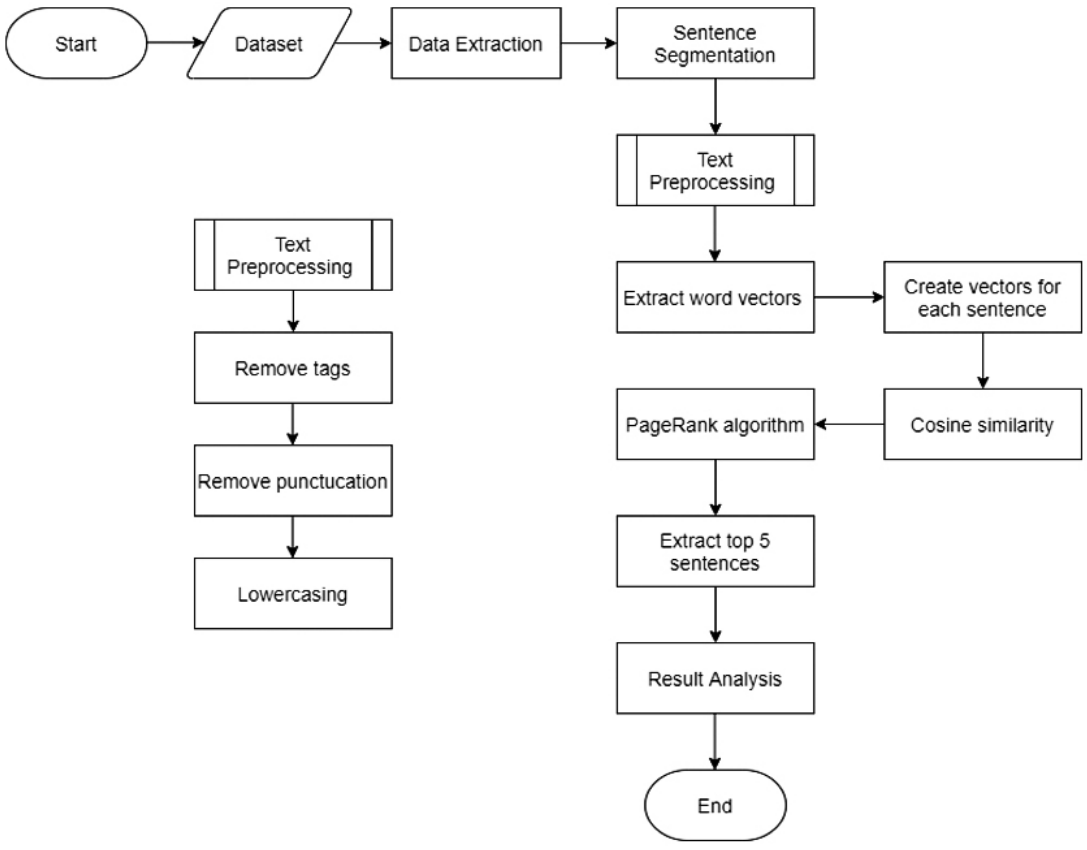


Figure 1: TextRank flowchart of Amazon fine food reviews summarization

3.1 Experiment

This section discusses how to generate a summary of food reviews from Amazon fine food dataset using TextRank algorithm. A total of eight steps that includes dataset, data extraction, sentence segmentation, text preprocessing, extract word vectors, cosine similarity scores, PageRank algorithm and extract top-ranked sentences are explained as follows:

3.1.1 Dataset

The dataset used in our study is Amazon fine food reviews which can be accessed from Kaggle. This dataset contains 568,454 reviews to 74,258 products which are collected from October 1999 to October 2012.

3.1.2 Data Extraction

A summary is generated on the product with the highest number of reviews in the dataset which is a cookie brand but we will not disclose the brand name due to privacy reasons. 910 reviews left after extracting the related reviews of this cookies brand and dropping duplicates.

3.1.3 Sentence Segmentation

zof a few sentences, so it is necessary to segment it into individual sentences for further processing. Segmentation occurred when the sentence ends at the segmentation point such as full stop, question marks, and exclamation marks. Therefore, a total of 910 reviews are segmented to 3661 individual sentences. Figure 2 shows a few sentences from 3661 individual sentences and ranking score is given to each of them to indicate their importance (refer to 3.1.7 PageRank Algorithm). 5 out of 3661 sentences will be selected for inclusion of summary according to the ranking score.

3.1.4 Text Preprocessing

Amazon reviews are usually in the form of unstructured which consists of noises and affect the performance of text summarizing if noise removal is not done perfectly. Noises such as HTML tags, punctuations and stop words are removed and converted all letters to lowercase.

3.1.5 Extract Word Vectors

Techniques are applied to map each word to a real-valued vector which is called word embedding because machines are not able to recognize the semantic and syntactic similarity between words in a text document. Word embedding is typically in the form of a real-valued vector that is used for the representation of words in a vector space. Figure 3 shows an example graph of word embedding, each word represented as a real-valued vector in a vector space. Words that are close to each other in the vector space tend to have associated meanings (McDonald & Ramscar, 2001). Based on the graph, 'cookie' and 'biscuit' are close to each other so they are expected to have a similar meaning. Therefore, Global Vectors (GloVe) (Pennington *et al.*, 2014) which is an unsupervised learning algorithm for obtaining word vector representations are used to convert each word in the sentences to word vectors. It is trained on the global word-to-word co-occurrence statistics by estimating the frequency of words co-occurs with one another in a given corpus. Pre-trained word vectors with 100 dimensional of 400k words computed on 2014 dump of English Wikipedia is used to create vectors for sentences and are available at (https://nlp.stanford.edu/projects/glove/). Each word will have 100 vectors in the 100 dimensional pre-trained word vectors. Next, 100 vectors of each word will be fetched, and calculate the total vectors of each word in the sentence. The final vector is computed by taking the sum of vectors and dividing by the total number of words in a sentence.

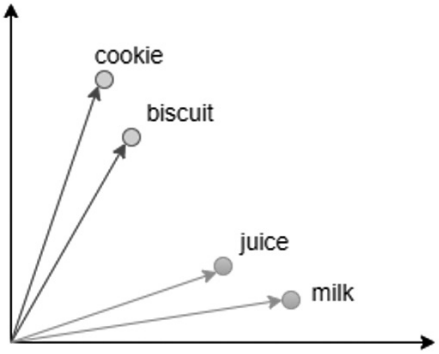


Figure 3: Word Embedding

```
1) I love these cookies!
2) Not only are they healthy but they taste great and are so soft!
3) I will definitely add these to my grocery list!
4) Quaker Soft Baked Oatmeal Cookies with raisins are a delicious treat, great for anytime of day.
5) For example:<br /><br />--at breakfast, I had one with a large banana and a cup of coffee, and felt I'd had a relatively "he
althy" start to the day.<br /><br />--the next day at lunch, following a tuna sandwich, I had one with a glass of milk, and was
satisfied enough to not need a snack before dinner at 6:30.<br /><br />--the following night, after dinner, I had one with the
remainder of my glass of wine.
6) (Delicious!)
7) And again, didn't feel the need to snack later in the evening.<br /><br />Each cookie is individually packaged, and their te
xture is soft and moist, with just the right amount of sweetness.
8) Natural flavors used in the making are Cinnamon and All Spice.
9) These flavorings give the cookies a real old-fashioned, homemade taste.<br /><br />Nutritionally, the cookies have 170 calor
ies each, 1.5g saturated fat, 150 mg sodium, and 12g sugar.
10) They also have 2g of protein, and contain 25g of fiber.<br /><br />While the calorie count may seem a bit high for one cook
ie, they are good sized, and 1 cookie per serving is certainly enough to satisfy.<br /><br />Because of their great taste and t
exture, kids will probably enjoy them also.<br /><br />If you like oatmeal raisin cookies, give these a try!
```

Figure 2: Part of Sentences from the Segmented Individual Sentences

3.1.6 Cosine Similarity Score

Cosine similarity is used to find the similarity between sentences even though TF-IDF is commonly used in text summarization to calculate the relevance and importance of sentences. TF-IDF is too long and sparse because sentences may not share the same words. Despite the fact that no common words appear in two sentences but this does not imply that the sentences have no associated meaning (Han *et al.* 2012). In contrast, cosine similarity measures only focus on common words between sentences and measure their similarity. Therefore, a zero similarity matrix *(n*n)* is created where the size of the matrix is equal to the number of individual sentences. Cosine similarity is used to compute similarity scores between sentences vectors and assigned to the matrix. There will be no relationship between two sentences if the score is 0.

3.1.7 PageRank Algorithm

The similarity matrix is then converted into a graph that has two elements: nodes and edges. Nodes represent the sentence whereas edges reflect the similarity scores between sentences. With the aid of a graph, the PageRank algorithm is used to compute the sentence rankings scores. The scores are used to determine the importance and relevance of sentences in generating a summary. Figure 4 shows the sentence ranking scores for each individual sentences. The first sentence (0.0002837) is less important compared to second sentence (0.0003018) so the ranking of first sentence must lower than second sentence.

```
0: 0.00028375778117883694,
1: 0.00030181592601811114,
2: 0.00028563178137795760,
3: 0.00029746127804224017,
4: 0.00031063658983022780,
5: 0.00022497797519289870,
```

Figure 4: Sentence Ranking Scores of Individual Sentences

3.1.8 Extract Top-Ranked Sentences

Sentences are sorted in descending order based on their sentence ranking scores to generate a summary. The higher the scores, the more relevant the sentences to be extracted for being a part of the summary. In our study, the top five reviews with the highest sentence ranking scores are extracted to form the summary of cookies reviews. This is because five reviews are often the ideal length of a summary, three is too short while ten might be too long for a summary.

4.0 RESULT

Table 1 shows the top five sentences with the highest sentence ranking scores which can be used to form a summary of cookie reviews. Customers who commented on the first and third sentences got a cookie sample from Influenster, product discovery and review platform, and they really liked it because of the softness or freshness. Unlike the second and fifth comments, customers dislike the cookies as they are not fresh, crumbled, or dry. Meanwhile, the third commenter enjoys the taste and softness as well. Based on the summary, we can infer that customers enjoy the oatmeal flavour of this cookies brand and it could be the most popular flavour among customers. In an overall view, customers praised the taste, softness, and freshness of this cookies brand but nevertheless, it is also disliked by customers because the cookies were too dry, not crumble, and not fresh enough.

Table 1: Top 5 highest-ranked sentences

Ranking	Summary
1	I GOT TO TRY THIS QUAKER SOFT BAKED OATMEAL COOKIE THROUGH THE GOOD FOLKS FROM INFLUENSTER AFTER RECEIVING THEIR 2012 MOMVOX BOX, AND I MUST SAY I LOVE IT, FIRST OF ALL OATMEAL COOKIES ARE MY FAVORITE, SO THERE WASNT ANY DISAPPOINTMENT THERE, THE COOKIE RETAIN ITS SOFTNESS/FRESHNESS OFTER BEING OPENED BY ME FOR A WEEK NOW, AND THAT WAS GOOD, PLUS IT TASTE GREAT SO THUMBS UP
2	Maybe it was the baking process? These cookies, although individually packed (so good for school lunches), came out a bit dry and crumbly. Sure, maybe I am just a messy eater but a soft baked cookie just not crumble as much as the cookies I got crumbled. Maybe if you get them at the supermarket they would be less dry. Maybe if is just a general problem with the way they are produced.
3	i received a free sample from Influenster and let me tell you it was so good and soft it crumbles up right in your mouth and its a big cookie my daughter also loved it i would definitely recommend buying it if you like oatmeal and raisins
4	yummy great cookie just like my momma makes this is definitely a second best of course after my mom's cooking love how soft and chewy they are a must buy
5	I love soft baked cookies, but I find that whenever i try to buy ones that are already made, they don't taste fresh.

4.1 Limitations

One of the drawbacks of this experiment is TextRank takes a long time to compute. The process of computing similarity matrix and sentence ranking scores for roughly 3600 sentences takes a few hours to complete. The computation time increases as the number of sentences extracted to perform summarization grows. This is because the increase in similarity matrix size required a longer time to compute the similarity scores between the sentences. Other than that, the summary is deemed lengthy to read even though it is made out of the top five original reviews with the highest-ranking score from the cookies reviews. Sometimes the reviews can be wordy and difficult to

read at a single glance. Researchers may be dissatisfied with the summarising outcome because it is still not a thorough summary of the reviews. In addition, the accuracy of the cookies brand summary should be taken into account. This is owing to the fact that only 100 vectors are used for each word to compute the sentence vector representations.

4.2 Future Work

Computation time can be reduced by expanding the stopwords list from Natural Language Toolkit (NLTK) library. Expanding the stopwords list helps to remove more words from the sentences during text preprocessing which can greatly reduce the number of words to create vectors for each of them. As a result, the computation time needed to calculate the average vectors of each word and the sum of vectors for each sentence is decreasing. We will also study more techniques to summarize each review into a few words and apply our algorithm to obtain a shorter and more relevant summary. In addition, we can fetch more vectors for each word in the sentence to increase the accuracy of summarization. GloVe word embedding consists of pre-trained word vector models with 50, 100, 200 and 300 dimensions for each word. Since 50 dimensions pre-trained model is used to create sentence vector representations in this study, word vectors can be extracted from 200 dimensions or 300 dimensions pre-trained word vector model in our future study. However, there is a trade-off between accuracy and computation time. Increasing the dimensionality of word embedding shall improve the accuracy as it implies the ability to compute more accurate word representation but longer computation time is required. We will explore more in order to take account of the processing speed and accuracy.

5.0 CONCLUSIONS

A summary of a cookies brand's public reviews is created by using the TextRank algorithm to extract the top 5 reviews with the highest sentence ranking score. According to the summary, the majority enjoy this cookie brand because it is fresh, soft, and tastes well but some reviewers may think it is dry, not fresh, and crumble enough. Other than that, the oatmeal cookie could be the most popular product among customers as many compliments have been received by them. The overall summary tends to be positive however there is room for improvement in terms of the moistness and crumble. We can now understand customers' perception of the food product without any human power to read and produce a summary. This study has shown that summarizing reviews required long computation time and unacceptable summary length produced which are needed to be improved in the future.

6.0 ACKNOWLEDGMENTS

The authors would like to express our deepest appreciation to Tunku Abdul Rahman University College for providing financial and resource supports for the implementation of this study. ■

REFERENCES

[1] Allahyari M, Pouriyeh S, Assefi M *et al.* (2017) Text summarization techniques: a brief survey. International Journal of Advanced Computer Science and Applications (IJACSA) 8(10), http://dx.doi.org/10.14569/IJACSA.2017.081052.

[2] Brin S and Page L (1998) The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems 30(1-7): 107-117, http://dx.doi.org/10.3844/jcssp.2014.1.9.

[3] Christian H, Agus MP and Suhartono D (2016) Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). ComTech: Computer, Mathematics and Engineering Applications 7(4): 285-294, http://dx.doi.org/10.21512/comtech.v7i4.3746.

[4] Day MY and Chen CY (2018) Artificial intelligence for automatic text summarization. In 2018 IEEE International Conference on Information Reuse and Integration (IRI), pp. 478-484, http://dx.doi.org/10.1109/IRI.2018.00076.

[5] Erkan G and Radev DR (2004) Lexrank: Graph-based lexical centrality as salience in text summarization. Journal of artificial intelligence research 22(1): 457-479.

[6] Han JW, Kamber M and Pei J (2012) Getting to know your data. In Data mining (Third Edition), pp. 39-82, https://doi.org/10.1016/B978-0-12-381479-1.00002-2.

[7] Li W and Zhao J (2016) TextRank algorithm by exploiting Wikipedia for short text keywords extraction. In 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), pp. 683-686, https://doi.org/10.1109/ICISCE.2016.151.

[8] Maybury M (1999) Advances in automatic text summarization. MIT press.

[9] McDonald S and Ramscar M (2001) Testing the distributioanl hypothesis: The influence of context on judgements of semantic similarity. In Proceedings of the Annual Meeting of the Cognitive Science Society 23(23).

[10] Mihalcea R and Tarau P (2004) Textrank: Bringing order into text. In Proceedings of the 2004 conference on empirical methods in natural language processing, pp. 404-411.

[11] Neto JL, Freitas A and Kaestner CA (2002) Automatic text summarization using a machine learning approach. In Brazilian symposium on artificial intelligence, Berlin, Heidelberg, vol. 2507, pp. 205-215, http://dx.doi.org/10.1007/3-540-36127-8_20.

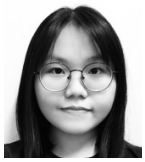
[12] PadmaPriya G (2014) An approach for text summarization using deep learning algorithm. International journal of trends in computer science 10(1): 1-9, http://dx.doi.org/10.3844/jcssp.2014.1.9.

[13] Patel M, Chokshi A, Vyas S and Maurya K (2018) Machine Learning Approach for Automatic Text Summarization Using Neural Networks. International Journal of Advanced Research in Computer and Communication Engineering 7(1), http://dx.doi.org/10.17148/IJARCCE.2018.7133.

[14] Saranyamol CS and Sindhu L (2014) A survey on automatic text summarization. International Journal of Computer Science and Information Technologies, 5(6): 7889-7893.

[15] Pennington J, Socher R and Manning CD (2014) Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543.

PROFILES



KHOR YUEN KEI received her BCS (Hons) in Data Science in the year of 2021 from Tunku Abdul Rahman University College (TAR UC) and is currently pursuing her master's degree in Computer Science at TAR UC. Her research interest focuses on Natural Language Processing (NLP), particularly in code-mixed multiword expressions identification, sentiment and emotion analysis. She is also a data scientist in Work At Cloud Sdn. Bhd. and majorly involved in sentiment analysis tasks and turning data into valuable insights to customers.
Email address: khoryk-wm17@student.tarc.edu.my

DR TAN CHI WEE received BCompSc(Hons) and PhD degrees in year 2013 and 2019 respectively in Universiti Teknologi Malaysia. Currently, he is a Senior Lecturer cum Programme Leader at Tunku Abdul Rahman University College and actively involved in the Centre of Excellence for Big Data and Artificial Intelligent (CoE) and become the research group leader for Audio, Image and Video Analytics Group under Centre for Data Science and Analytics (CDSA). Dr Tan's main research areas are Computer Vision (CV), Image Processing (IP) and Natural Language Processing (NLP) and Artificial Intelligence (AI). He is an enthusiastic researcher experienced in conducting and supporting research into Image Processing. Being a meticulous and analytical researcher with Train-The-Trainer certificate of many years of educational and hands-on experience, he was invited to Université d'Artois (France) under Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) programme for collaborative research between European countries with Southeast Asian countries on motion detection and computer vision and being involved in industry project as professional consultant.
Email address: chiwee@tarc.edu.my

PROFESSOR LIM has about 10 years of industry experiences in the design, development, implementation and maintenance of commercial software from 1989 to 1999 after departing from TARC where he spent his early days with TARC as an IT lecturer from 1987 to 1989 after returning from Mississippi State University USA with a Master of Computer Science degree. He is currently the Director for CBIEV at TAR UC, Professor at FOCS at TAR UC and Head for Big Data Analytics Centre. His research interest involving Natural Language Processing, Sentiment Analysis and Code-Mixed language analysis. In the last 15 years, his work has consistently focused on organizational knowledge sharing and technology acceptance, social media analytics and social influence maximization in Sunway University and Tunku Abdul Rahman University College (TAR UC). Professor Lim has graduated more than 20 master and 2 PhD students while he was with Monash, UTAR and Sunway University.
Email address: limtm@tarc.edu.my

A STUDY ON THE CENTRALITY MEASURES TO DETERMINE SOCIAL MEDIA INFLUENCERS OF FOOD-BEVERAGE PRODUCTS ON TWITTER

(Date received: 12.01.2022/Date accepted: 10.04.2022)

Wai Beng Tan[1], Tong Ming Lim[2]

[1,2] Department of Information and Communication Technology, Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia.

**Corresponding author: tanwb@tarc.edu.my[1], limtm@tarc.edu.my[2]*

ABSTRACT

This research aims to study and identify the Social Media Influencers (SMIs) in the Twitter community in Pizza Hut Industry. In social network analysis (SNA), Eigenvector Centrality (EC) will give the most influential node in a network. A node with the highest eigenvector value among the other nodes is the most influential/important node in a network. Data was collected from Twitter using the Twitter API with the hashtag #pizzahut. It applied the eigenvector centrality to observe the effect of the centrality value for Twitter data. The result shows that there is a significant difference between the three most influential users. This result will be used for future research that will be focused on small and medium enterprise (SME) Twitter data. This research is held a comparison analysis between the four centrality measurements approach: Degree Centrality, Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality. for determining the most influential user with the social network Twitter as its case study.

Keywords: Social Media Influencer, Social Network Theory, Centrality Measures, Food and Beverage (F&B)

1.0 INTRODUCTION

In 2021, the total population of the world is 7.9 billion. The number of Internet users is 4.66 billion and approximately 3.96 billion or 85% of them are active social media users (Dean, 2021). They are actively participating in many online activities on different social media platforms such as Facebook, Twitter, and Instagram.

In these recent years, many people have been influenced by stars, celebrities, and influencers on social media platforms in promoting their online products. Many celebrities who have many followers would start to promote certain products, especially F&B to all their audiences, fans and followers. The influenced audiences, fans and followers have to share and show with their close friends and family members to gain more popularity among other audiences. Celebrities have played a significant role in corporate brands as they manage to promote the brand products and also attracted many people who are interested in purchasing the products. Therefore, the group of celebrities can be called social media influencers (SMI).

A previous study has reported that food products are top interest for Gen Z and millennials (Hanifawati, Dewanti, & Saputri, 2019). Many of Gen Z prefer to spend their money to buy F&B (Cheung, Davis, & Heukaeufer, 2017). Pricing strategy and providing variety are effective approaches for F&B. Gen Z has higher engagement with brands on social media compared to millennials. Either Gen Z or millennials are generally used multichannel to engage with the brands (Hanifawati *et al.*, 2019).

In social network analysis (SNA), a node value influential a network called centrality. Centrality is defined as a value that represents how many connections are from nodes to other nodes (Wasserman & Faust, 1994). There are many methods to define centrality to identify the effect of each node in a social network such as Degree Centrality (DC), Betweenness Centrality (BC), Closeness Centrality (CC), and Eigenvector Centrality (EC). Among these, the eigenvector centrality will give the most influential node in a network. A node with the highest eigenvector value among the other nodes is the most influential/important node in a network.

Social media platforms have become an essential medium of communication among individuals, and they also play a vital role in brand promotion and marketing (Arora, Bansal, Kandpal, Aswani, & Dwivedi, 2019). In the past two decades, social media content has been used by various brands to stay competitive by promoting products and offering offers to maintain market position and reputation among stakeholders (Croft & Brennan, 2006). One of the key drivers of this change (B. F. Liu *et al.*, 2012) is social media influencers (SMIs), whom (Freberg, Graham, McGaughey, & Freberg, 2011) identify as 'a new type of independent third party endorser who shape audience attitudes through blogs, tweets, and the use of other social media' (p. 90). These "social media creators" (B. F. Liu, Jin, Briones, & Kuch, 2012) engage themselves in content creation of particular issues.

SMIs play an important role in social media platforms. They are considered influential as their opinions have an effect on their followers, media coverage, and organizations. They affect their followers by providing issue-relevant opinion leadership that meets followers' information and emotional needs on particular issues (B. F. Liu *et al.*, 2012). In turn, their followers then influence non-followers through "word of mouth communication" (B. F. Liu *et al.*, 2012).

SMIs exhibit credibility and persistence in persuading their audience to notice and support their topics of interest (Goodman, Booth, & Matic, 2011). They also affect media coverage by shaping the media agenda. Journalists can tap on alternative information and newsworthy content generated by SMIs, which can be repackaged and disseminated to news audiences to create issue awareness in public. Furthermore, the ability of SMIs to leverage influence could significantly impact a brand's reputation. It also becomes critically essential for brands to identify the right influencers on the Web through social media to promote their products and services (Huang, Zhang, Li, & Lv, 2013). Brands can directly leverage this to improve and enhance public relations by promoting their offerings for higher engagements (De Vries, Gensler, & Leeflang, 2012).

The power of SMIs lies in their ability to affect media coverage, improve key publics' issue awareness, and persuade their followers to assume a course of action (Freberg *et al.*, 2011). A strong relationship with SMIs can help organizations maximize positive media coverage and ultimately create a stronger public presence.

The recent rise in popularity of social network platforms have prompted modern F&B companies to switch their marketing objectives from the traditional top-down strategies to a peer-to-peer approach. Many of this companies have started to utilized SMI to promote products online. The current study investigated the promotion of F&B products on Instagram tagged with #ad (Reagan, Filice, Santarossa, & Woodruff, 2020). According to previous researches, SMI is significantly important tools in promoting F&B on social media platforms.

The presentation of this paper is divided into four sections – The first section discusses the need for Social Media Influencer in different social media platforms. The second section provides research objective, motivation and contribution. Literature in the same direction is discussed in Section 3, followed by network centrality measures and Social Network Theory (SNT). The research methodology for the study is discussed in Section 4 with consists of two subsections: Twitter, Twitter API, and centralities such as Degree Centrality, Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality. Centrality results and results discussions are detailed in Section 5 and 6 accordingly followed by a concluding remark in Section 7.

2.0 RESEARCH OBJECTIVE, MOTIVATION AND CONTRIBUTION

The elementary functionalities of social media platforms differ from each other. The major social media platforms are Facebook as a relationship network, Instagram as a media sharing network, and Twitter as a social publishing network. Influencers end up posting multiple contents across these platforms while availing these services. Normally, influencers post content on multiple social media platforms based on their popularity. Every influencer has a variable influence on varying social media platforms. Influencers on different social media platforms are measured with a set of weighted attributes by that specific application (Arora *et al.*, 2019).

This paper is intended to find out which entities have the most influence in the dissemination of information on pizza in tweets using the hashtag #pizzahut based on the calculation of Degree centrality, Betweenness centrality, Closeness centrality, and Eigenvector centrality on Twitter.

The research question in this research study is "How would centrality measures determine social media influencers in Twitter?"

The purpose of this study is to investigate whether an influencer has distinctive exposure across social network platforms to contribute to different influence measures on a different social network platform. Based on previous research, information spreading speed among the social media is affected by the users' activity connection which can be represented in centrality values. This research applied degree and eigenvector centrality to observe the effect of centrality value for Twitter data. The results show that there is a significant difference among the three most influential users on Twitter.

3.0 LITERATURE REVIEW

The following sections discuss the importance of Social Network Theory (SNT) and network centrality measures and Two-step flow of Communication.

3.1 Network Centrality Measures

Calculating centrality has been a major focus of social network analysis research for some time (Freeman, 1978). Many references discuss social networks on centrality concepts and calculations (Alain & Michel, 1999; Scott, 2000; Wasserman & Faust, 1994). At least eight centrality measures have been proposed such as degree, betweenness, closeness, eigenvector, power, information, flow, and reach. The most frequently used centrality measures are degree, closeness, betweenness, and eigenvector. The first three were proposed by (Freeman, 1978) and eigenvector was proposed by (Bonacich, 1972). Centrality is important because it indicates who occupies critical positions in the network.

3.2 Social Network Theory (SNT)

Online consumer behaviours and profiles on the social network have begun as a huge source of data and marketers have begun to mine these data to understand consumer behaviours and relationships due to its importance for e-marketing (Dolnicar, 2003).

Understanding the relationships of online consumers helps businesses understand and target their current users well, reach out to potential customers, and to improve communication with them at the right time and place to increase their sales volumes. The consumer relationship also helps to gain a competitive advantage in the international e-marketing field, to control the flow of information in consumer networks, and to make innovations to differentiate themselves from the competitors (Bayer & Servan-Schreiber, 2011).

From the perspective of social network theory (SNT), centrality measures are the most frequently used to find key influential consumers in the network (Valente, Coronges, Lakon, & Costenbader, 2008). The theory has proposed three types of network centrality measures to identify the advantageous position that opinion leaders usually occupy: degree, betweenness, and closeness (Freeman, 1978).

- An online consumer with a high degree of centrality means he or she is highly connected with other online consumers

in the network. Therefore, he receives more information, knowledge, and resources. There are two types of degree centrality: in-degree centrality and out-degree centrality.
    - In-degree centrality of a consumer indicates the popularity of the consumer and his or her accessibility to information.
    - Out-degree centrality shows the control of a consumer over the network and the dependence of the network upon him or her.
- An online consumer who has high centrality of closeness shows that he or she can reach all online consumers on the network faster than anyone else.
    - A consumer with high in-closeness centrality may listen to most consumers through indirect or direct connections in the network.
    - A consumer having high out-closeness centrality sends messages to most consumers in the network through indirect or direct connections.
- An online consumer having high betweenness centrality indicates that he bridges the subgroups in the network and plays the role of gatekeeper.
- An online consumer having high eigenvector centrality connects to many other consumers that are also well connected.

3.3 Two-Step Flow of Communication

The two-step flow of communication hypothesis was first proposed by Lazarsfeld, Berelson, and Gaudet in the book The People's Choice (1944). In their study of voting decisions, they found that personal influence, which was largely derived from people's social contacts and friendship networks, significantly affected voting decisions. The effect was pronounced among people who were less committed to their existing beliefs or who changed their minds during the campaign. The hypothesis is called two-step because the social media platforms initially influence opinion leaders, individuals who are perceived as influential, who in turn influence their social contacts (W. Liu, Sidhu, Beacom, & Valente, 2017). Therefore, central to the two-step flow of the communication process is the concept of opinion leaders, a group of individuals influential in specific domains. Numerous studies have attempted to identify the key characteristics associated with being influential along with three terms (Katz, 1957): who one is, the individual characteristics of opinion leaders, such as personality traits; what one knows, the characteristics of individuals' competence, such as their knowledge or ability to provide information on particular issues; and who knows, the characteristics related to an individual's structural position in a network. In other words, individuals may become opinion leaders not only because they possess certain attributes but also because they occupy the right network positions that enable them to effectively spread information and exert personal influence. Centrality measures such as degree, betweenness, and closeness have been particularly useful for identifying leaders based on their network position (W. Liu *et al.*, 2017).

3.4 Twitter

Twitter is a social network that is widely used by social media users. It plays an important role in the dissemination of information to understand the popularity of a particular brand product. The dissemination of information through Twitter social networks can be done quickly and can be spread in a very short time through the posts of Twitter users themselves. The information provided by these users will be visible to other users and may be reposted by that user via retweet. Many researchers use Twitter in their research related to social network analysis(Priyanta & Nyoman Prayana Trisna, 2019).

3.5 Social Media Influencer Marketing for F&B Products

Social media influencer marketing for F&B products has become a widespread and successful marketing tactic (Byrne, Kearney, & MacEvilly, 2017) that have got a lot of attention in recent years. Previous study investigated the types of F&B products that influencers promote on YouTube (Coates, Hardman, Halford, Christiansen, & Boyland, 2019) and Instagram (Qutteina, Hallez, Mennes, De Backer, & Smits, 2019) and how influencer endorsement affects brand attitudes and purchase intention (Evans, Phua, Lim, & Jun, 2017). Research has also investigated advertising methods employed by brands on Instagram (Klassen *et al.*, 2018) and how different advertising methods influence audience engagement with advertisements (Adegbola, Gearhart, & Skarda-Mitchell, 2018).

Despite with the rapid development of influencer marketing and a widely distribution of social media, few research has studied influencer marketing in the market, so little is known about how to determine SMI on Twitter community. Thus, there is a gap in the literature on identifying SMI on Twitter in terms of F&B products.

4.0 RESEARCH METHODOLOGY

In this paper, the data used are tweets from Twitter using the hashtag #pizzahut, the data obtained is represented in a graph and processed and analysed by Centrality Measurement using Gephi, which can determine which entities will influence the most dissemination of information provided.

In this research, data is obtained with the help of Twitter API and the data retrieved is in the period from December 1, 2020, to December 10, 2020. When creating the hashtag network, it contains 23 users and 22 relationships between users.

4.1 Twitter and Twitter API

Twitter is a free social networking tool that is widely used and allows people to share information and newsfeeds with people who have the same views and thoughts in real time. Twitter API (application programming interface) is a program or application provided by Twitter to make it easier for other developers to access the information on the Twitter network. Many theories are supporting the calculation of centrality measures used for the search of the most influential entities in the graph of the dissemination of #pizzahut information on the social network Twitter that can be found in (Freeman, 1978).

4.2 Centrality

The idea of centrality as applied to human communication was introduced by Bavelas (Bavelas, 1948). This study will be used the calculation of four kinds of centrality, such as degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality.

4.2.1 Degree Centrality

Degree centrality is used to search for the entities that have the most influence on the dissemination of information on Twitter by looking at the number of direct relationships an account has with another account. The higher the degree centrality value, meaning the more relationship an account has with another.

4.2.2 Closeness Centrality

Closeness centrality is used to search for the most influential entities by looking at how close an account is to another based on the shortest distance obtained.

4.2.3 Betweenness Centrality

Betweenness centrality is used to search for the most influential entities in the dissemination of information based on the extent to which they are required as a link in the dissemination of information on Twitter social networks.

4.2.4 Eigenvector Centrality

Eigenvector centrality is used to search for the most influential entities by identifying the influence of those entities across the network, not just their influence on directly connected nodes.

4.3 Research Framework

The research framework of this study was developed to integrate the SNT and Two-step flow of communication by showing that by showing that the audiences search for what they need. They are likely to search for information from opinion leaders such as F&B Social Media Influencers before making informed purchase decisions. Two steps of communication occur as described below:

Step 1 of Communication: Brand communicates to opinion leaders

The particular F&B brand chooses SMI as opinion leaders to forward the branded messages to the target audiences because the brand wants to be part of their social networks rather than pushing advertising to them.

Step 2 of Communication: Opinion leaders communicate to consumers

As opinion leaders, SMI create the product review content to express their personal identity as F&B expert and then publish their product reviews through Twitter as a communication channel to reach their target audiences.

5.0 RESULTS

The tweet data posted on in the period from December 1, 2020, to December 10, 2020 was collected for this study. From the data, it is determined whether the tweet is the result of a retweet or not. If the tweet was a result of a retweet, then we can find who is the original writer of the tweet. These data can be represented in a simple graph where the nodes represent the Twitter account. If an account retweets from another account, then the two entities will be linked by a side called edge. There are 23 nodes with 22 edges. #pizzahut is being selected because Pizza Hut is a very well-know F&B brand with 1.6 million followers.

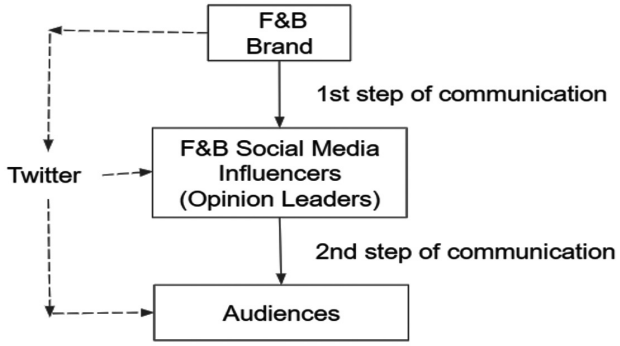


Figure 1: Research Framework

Using Twitter data with hashtag #pizzahut and the help of the Twitter API, a graph is shown in Figure 1.

The basic data roles for network analysis object are Source and Target. In Figure 2, nodes with green represent Target meanwhile nodes with purple represent Source. The Source specifies a data item that contains all of the node values for the plot. The Target specifies a data item that creates the links between nodes.

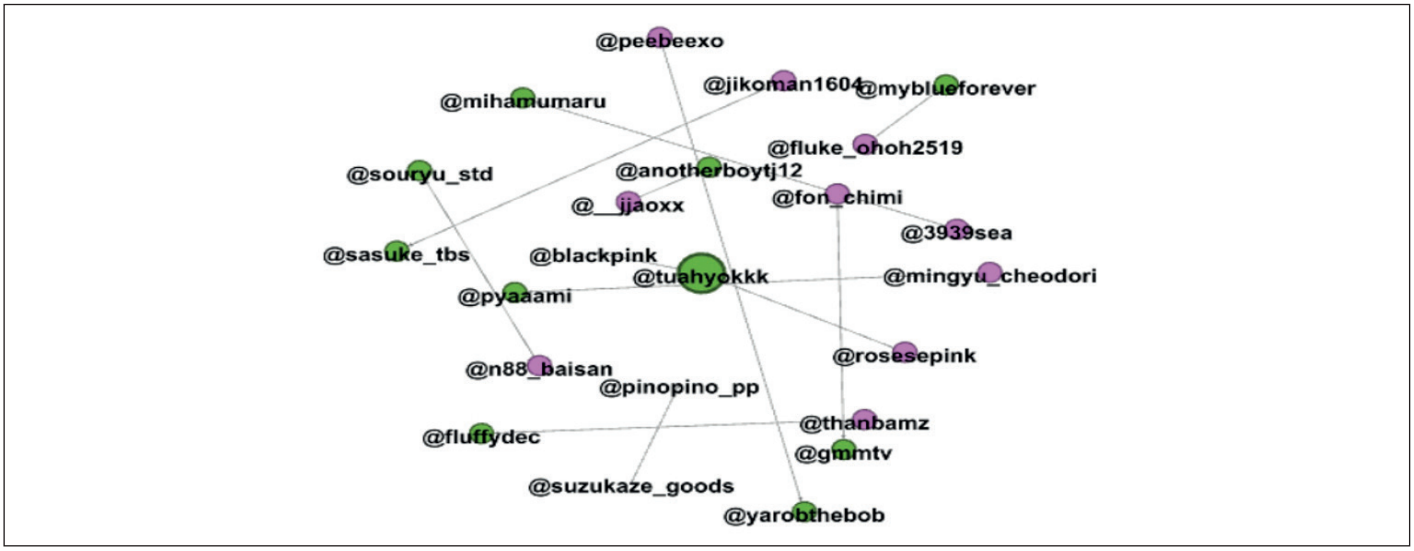


Figure 2: Representation of the tweet data graph with hashtag #pizzahut

5.1 Centrality Measures from Tweet Data with Hashtag #pizzahut

From the data obtained, a tweet data graph is generated. The centrality measure value is calculated for each account from the graph with the purpose of finding the most influential entities. In Table 1, there are four basic centrality measures: Degree Centrality (DC), Closeness Centrality (CC), Betweenness Centrality (BC), and Eigenvector Centrality (EC). The top three entities with the highest value for each centrality are as follow:

Table 1: Top three ranked Id based on centrality measures

Rank.	Id	EC	Id	BC	Id	CC	Id	DC
1	@blackpink	1	@tuahyokkk	1.0	@tuahyokkk	1.0	@tuahyokkk	3
2	@tuahyokkk	0.099886	@blackpink	0.0	@fon_chimi	1.0	@fluffydec	2
3	@fluffydec	0.099886	@fluffydec	0.0	@3939sea	1.0	@fon_chimi	2

In Table 1, it shows the centrality calculation for each Id, @blackpink account always gets the highest eigenvector centrality (EC) value which maximum value is 1, which means that this account is the account that has the most connection with other entities. Besides, it has the closest relationship with other entities, becomes the contact of an account with another account, and has the most interaction with other important entities in the graph.

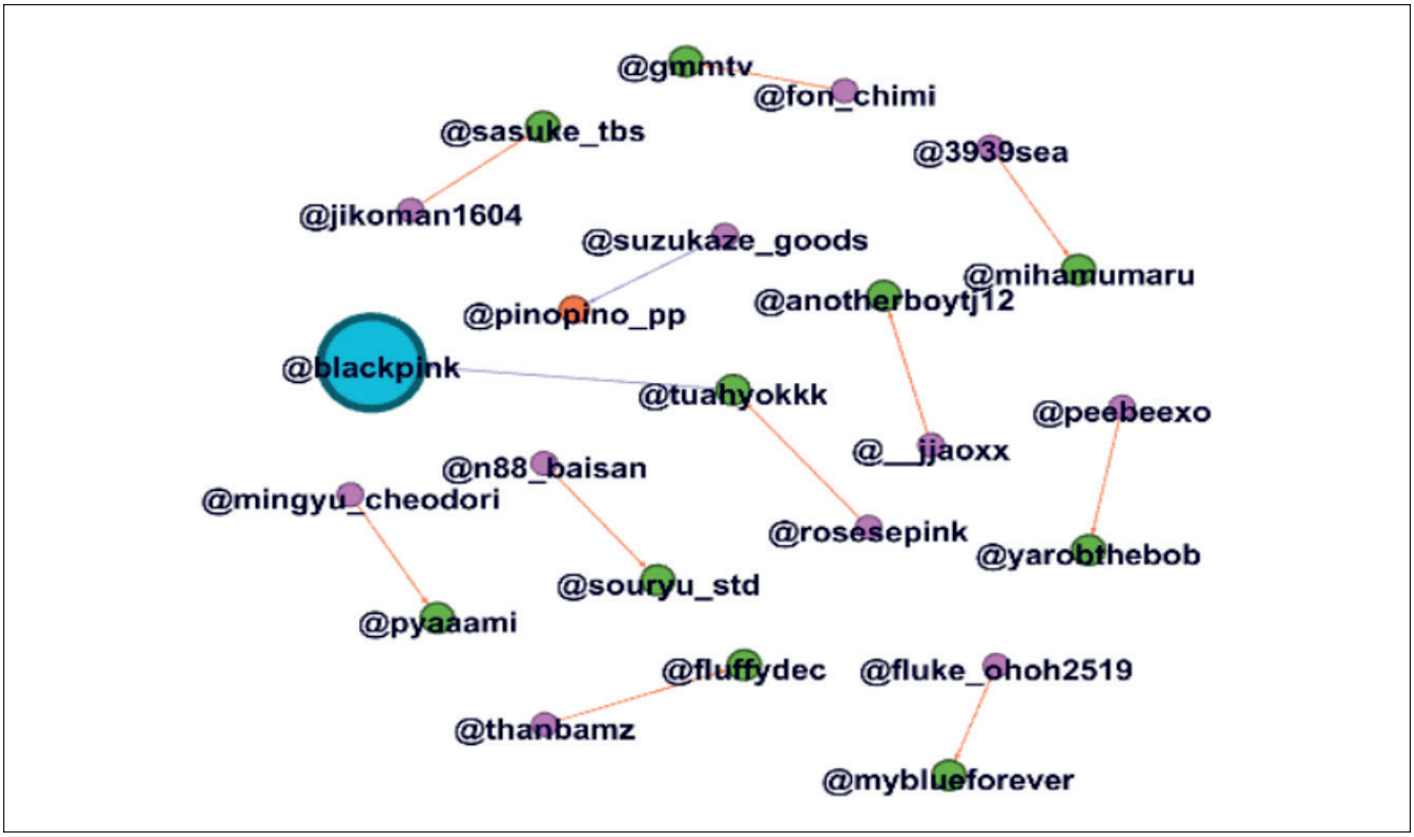


Figure 3: Graph representing the value of eigenvector centrality

From the graph in Figure 3, the node of @blackpink has a bigger size than the other nodes. This means the @ blackpink account has higher centrality measures and also is a more influential account based on the degree centrality measures, closeness centrality measures, betweenness centrality measures, and eigenvector centrality measures. Besides, blue node represents @blackpink account obtained maximum value which is 1 and it has highest EC value among other nodes. It has bigger circle size comparing with other nodes because it has the highest EC value. Orange node represents @pinopino_pp account with EC value 0.049943, purple nodes with 0 value for EC and green nodes with EC value 0.099886. Every edge is carrying same weight with 1.0 value.

6.0 DISCUSSION

The objective of this study is to perform centrality measure analysis on the #pizzahut user network described in earlier section. Figure 3 shows a Gephi graph contains 23 nodes with different EC value. Initially the user network data is loaded into the Gephi tool. Then on, the calculation of centrality measures is carried out. Further using the functions of Gephi tool the centrality measurements are calculated to rank the accounts in the last step of the process.

Visualization of #pizzahut user network based on EC measure by Gephi tool is shown in Figure 3. In this figure, @ blackpink with a bigger size of the node than the other nodes in the network. This bigger size of the node is the result of higher EC value as compared to all other accounts in the network.

Table 1 shows the top three accounts ranked on Degree, Closeness, Betweenness and Eigenvector centrality measures. Eigenvector centrality is a measure of the importance of a node in a network. Here, an account is considered important if he/she is connected to other important accounts. In the analysis, an account with a small number of influential contacts may outrank with a large number of mediocre contacts.

7.0 CONCLUSION

In this paper centrality measure analysis carried out on #pizzahut user network was deliberated. Analysis results assisted in identifying invisible patterns in the user network, for example, relationship between accounts shown by visualization and top-ranking accounts.

Analysing user account information on a larger database of F&B network will assist in identifying groups of people who interact closely together. Focusing future research work on categorization and ranking of accounts based on their preference will assist other people to identify main influencers of their interested preferences. This will aid in strengthening and improving interaction with marketers.

Centrality's calculation in the study was used to study for an account that was most influential in the dissemination of information from tweets that used #pizzahut hashtags on Twitter user social networks based on four centrality measurements: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality.

In this research, we only implemented and analysed centrality measurements, but not analysing the effect of interaction follow, mention and reply. This research still has limitations in measuring the performance of the most influential user rank. Future research will be conducted on an experiment to improve customer engagement by implementing SNA for Twitter SMIs. ■

REFERENCES

[1] Adegbola, O., Gearhart, S., & Skarda-Mitchell, J. (2018). Using Instagram to Engage with (Potential) Consumers: A Study of Forbes Most Valuable Brands' Use of Instagram. The Journal of Social Media in Society Fall, 7(2), 232–251.

[2] Alain, D., & Michel, F. (1999). Introducing Social Network. Sage: Thousand Oaks, CA.

[3] Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index- insights from facebook, Twitter and Instagram. Journal of Retailing and Consumer Services, 49(March), 86–101. https://doi.org/10.1016/j.jretconser.2019.03.012

[4] Bavelas, A. (1948). A Mathematical Model for Group Structures. 7(3), 16–30.

[5] Bayer, J., & Servan-Schreiber, E. (2011). Gaining competitive advantage through the analysis of customers' social networks. Journal of Direct, Data and Digital Marketing Practice, 13(2), 106–118. https://doi.org/10.1057/dddmp.2011.26

[6] Bonacich, E. (1972). A theory of ethnic antagonism: the split labor market. American Sociological Review, 37(5), 547–559. https://doi.org/10.2307/2093450

[7] Byrne, E., Kearney, J., & MacEvilly, C. (2017). The Role of Influencer Marketing and Social Influencers in Public Health. Proceedings of the Nutrition Society, 76(OCE3), 2022. https://doi.org/10.1017/s0029665117001768

[8] Cheung, J., Davis, D., & Heukaeufer, E. (2017). Gen Z brand relationships. IBM Corporation, 1–24. Retrieved from https://nrf.com/research/gen-z-brand-relationships

[9] Coates, A. E., Hardman, C. A., Halford, J. C. G., Christiansen, P., & Boyland, E. J. (2019). Food and Beverage Cues Featured in YouTube Videos of Social Media Influencers Popular With Children: An Exploratory Study. Frontiers in Psychology, 10(September), 1–14. https://doi.org/10.3389/fpsyg.2019.02142

[10] Croft, R., & Brennan, R. (2006). The Use Of Social Media In B2B Marketing and Branding : An Exploratory Study. Journal of Customer Behavior, 1–20.

[11] De Vries, L., Gensler, S., & Leeflang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. Journal of Interactive Marketing, 26(2), 83–91. https://doi.org/10.1016/j.intmar.2012.01.003

[12] Dean, B. (2021). Social Network Usage & Growth Statistics: How Many People Use Social Media in 2021? Retrieved from Backlinko website: https://backlinko.com/social-media-users#social-media-usage-stats

[13] Dolnicar, S. (2003). Established Methodological Weaknesses and Some Recommendations for Improvement.

[14] Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent. Journal of Interactive Advertising, 17(2), 138–149. https://doi.org/10.1080/15252019.2017.1366885

[15] Freberg, K., Graham, K., McGaughey, K., & Freberg, L. A. (2011). Who are the social media influencers? A study of public perceptions of personality. Public Relations Review, 37(1), 90–92. https://doi.org/10.1016/j.pubrev.2010.11.001

[16] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

[17] Goodman, M. B., Booth, N., & Matic, J. A. (2011). Mapping and leveraging influencers in social media to shape corporate brand perceptions. Corporate Communications: An International Journal, 16(3), 184–191. https://doi.org/10.1108/13563281111156853

[18] Hanifawati, T., Dewanti, V. W., & Saputri, G. D. (2019). The Role Of Social Media Influencer On Brand Switching Of Millenial and Gen Z: A Study of Food-Beverage Products. Journal of Applied Management (JAM), 17(4), 625–638. https://doi.org/10.21776/ub.jam.2019.017.04.07

[19] Huang, J., Zhang, J., Li, Y., & Lv, Z. (2013). Business value of enterprise micro-blogs: Empirical study from Weibo.com in Sina. Proceedings - Pacific Asia Conference on Information Systems, PACIS 2013.

[20] Katz, E. (1957). The two-step flow of communication: An up-to-date report on an hypothesis. Public Opinion Quarterly, 21(1), 61–78. https://doi.org/10.1086/266687

[21] Klassen, K. M., Borleis, E. S., Brennan, L., Reid, M., McCaffrey, T. A., & Lim, M. S. C. (2018). What people "like": Analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. Journal of Medical Internet Research, 20(6), 1–9. https://doi.org/10.2196/10227

[22] Liu, B. F., Jin, Y., Briones, R., & Kuch, B. (2012). Managing turbulence in the blogosphere: Evaluating the blog-mediated crisis communication model with the American Red Cross. Journal of Public Relations Research, 24(4), 353–370. https://doi.org/10.1080/1062726X.2012.689901

[23] Liu, W., Sidhu, A., Beacom, A. M., & Valente, T. W. (2017). Social Network Theory. The International Encyclopedia of Media Effects, (September), 1–12. https://doi.org/10.1002/9781118783764.wbieme0092

[24] Priyanta, S., & Nyoman Prayana Trisna, I. (2019). Social network analysis of twitter to identify issuer of topic using PageRank. International Journal of Advanced Computer Science and Applications, 10(1), 107–111. https://doi.org/10.14569/IJACSA.2019.0100113

[25] Qutteina, Y., Hallez, L., Mennes, N., De Backer, C., & Smits, T. (2019). What Do Adolescents See on Social Media? A Diary Study of Food Marketing Images on Social Media. Frontiers in Psychology, 10(November), 1–12. https://doi.org/10.3389/fpsyg.2019.02637

[26] Reagan, R., Filice, S., Santarossa, S., & Woodruff, S. J. (2020). #ad on Instagram: Investigating the Promotion of Food and Beverage Products. The Journal of Social Media in Society, 9(2), 1–28.

[27] Scott, J. (2000). Social Network Analysis: A handbook. Newbury Park, CA: Sage.

[28] Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How Correlated Are Network Centrality Measures? Connections (Toronto, Ont.), 28(1), 16–26. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20505784%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2875682

[29] Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press.

PROFILES

TS. TAN WAI BENG is Lecturer of Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kuala Lumpur. She received Bachelor's Degree from Campbell University, United States and Master's Degree from Universiti Utara Malaysia (UUM). She obtained Professional Technologist in the area of Information and Computing Technology (IT) in 2020.
Email address: tanwb@tarc.edu.my

PROFESSOR LIM has about 10 years of industry experiences in the design, development, implementation and maintenance of commercial software from 1989 to 1999 after departing from TARC where he spent his early days with TARC as an IT lecturer from 1987 to 1989 after returning from Mississippi State University USA with a Master of Computer Science degree. He is currently the Director for CBIEV at TAR UC, Professor at FOCS at TAR UC and Head for Big Data Analytics Centre. His research interest involving Natural Language Processing, Sentiment Analysis and Code-Mixed language analysis. In the last 15 years, his work has consistently focused on organizational knowledge sharing and technology acceptance, social media analytics and social influence maximization in Sunway University and Tunku Abdul Rahman University College (TAR UC). Professor Lim has graduated more than 20 master and 2 PhD students while he was with Monash, UTAR and Sunway University.
Email address: limtm@tarc.edu.my

CAPABILITIES AND SERVICES OFFERED BY MALAYSIAN LEARNING CENTRES IN ASSISTING MANUFACTURING ENTERPRISES IN THEIR DIGITAL TRANSFORMATION JOURNEY

(Date received: 12.01.2022/Date accepted: 01.04.2022)

Yong C H[1]*, Lee W P[1], Lazarus C[1], Tan K W[1], Lim Y M[2]

[1] *Tunku Abdul Rahman University College, Jalan Genting Kelang, Setapak, 53300 Kuala Lumpur, Malaysia.*

[2] *GMCM Sdn. Bhd., Kawasan Perindustrian Bukit Serdang, Seri Kembangan, 43300 Selangor, Malaysia.*

**Corresponding author: yongch-wr20@student.tarc.edu.my*

ABSTRACT

Industry 4.0 encompasses a plethora of notions and technologies aimed at assisting manufacturing entities, particularly small and medium-sized firms, in gaining a competitive edge. Hence, an effective implementation of technology in the domains of asset integration, digitization, and automation is dependent on meeting plenty of demanding requirements. The role of learning centres in assisting these firms is critical to ensuring the country's overall digital transformation is a success. An exploratory quantitative survey was carried out in Malaysia to ascertain the learning centres' strengths and gaps in aiding the country in spearheading the Industry 4.0 (I4.0) agenda. The survey's primary result is that learning centres must further equip themselves to act as a catalyst in assisting the manufacturing industry in realising the I4.0 potential, albeit there is also a need to strengthen commercialization between learning centres and manufacturing enterprises. Additionally, the survey discovered that learning centres should strengthen their collaboration and affiliation with the country's primary industries in order to accelerate the adoption of I4.0 in Malaysia. Another significant outcome of the survey is the call for the establishment of a one-stop-centre for I4.0 services such as funding and policy inquiries, as well as the promotion of learning centres' services, which are critical to assisting manufacturing enterprises in their digital transformation journey.

Keywords: Manufacturing value chain, digital transformation, Industry 4.0

1.0 INTRODUCTION

The First Industrial Revolution was driven by steam and waterpower, which enabled the mechanisation of manufacturing processes, whereas the Second Industrial Revolution was spurred by electric power and mass manufacturing techniques. The Third Industrial Revolution was ushered in by information technology. This has resulted in the subsequent phase of evolution, dubbed Industry 4.0 (I4.0) (Neugebauer *et al*., 2016).

The European Commission estimates that 99.8% of enterprises in the European Union are classified as Small and Medium-sized Enterprises (SME) (Safar, *et al*., 2018). According to Malaysia's National Policy on Industry 4.0 (2018), the manufacturing industry is a significant economic sector, accounting for around 22% of GDP between 2014 and 2018. Malaysia's manufacturing sector is predicted to increase at a rate of roughly 5.1% under the 11th Malaysia Plan, with 98.5% of SMEs accounting for 42% of total employment.

Recognizing the importance of manufacturing and small and medium-sized enterprises (SMEs), Malaysia, like other countries, has developed a National Policy on Industry 4.0, commonly known as Industry 4WRD. The purpose of this Industry 4WRD strategy is to create a more organised national agenda through measures that will hasten Malaysia's transformation into a smart and modern manufacturing system in the age of Industry 4.0.

Likewise, Germany has established its High-Tech Strategy 2020 Plan, while the United States of America has established the Advanced Manufacturing Partnership (AMP) and National Advanced Manufacturing Strategic Plan. The European Commission also sponsored the Factories of the Future Program (2008-2020) with the same objective, while the United Kingdom has its own Industrial 2050 Strategy (Zhang, *et al*., 2019).

China developed the Made in China 2025 policy, whereas the Republic of Korea developed the Strategy for Innovation in Manufacturing Industry 3.0, and Taiwan developed Productivity 4.0. (Kuo, *et al*., 2019). On the other hand, Japan devised the Industrial 4.1J and Revitalization Strategy, which included the establishment of the Science and Technology Industry Alliance to lead the I4.0. Singapore developed the Future of Manufacturing (FoM) through the Advanced Remanufacturing and Technology Centre (ARTC) and the Singapore Institute of Manufacturing Technology (SIMTech).

In Malaysia, the Industry 4WRD policy encourages the participation of I4.0 service providers and their engagement with manufacturing entities to assist in the implementation of relevant processes, technologies, and skill development. This plan is contingent upon the existence of I4.0 learning centres in

Malaysia capable of providing such services. This is consistent with Moeuf *et al*. (2020) assertion that research need to also consider strategic opportunities and operational factors in addition to I4.0 technology.

As a result, Collaborative Research in Engineering, Science, and Technology (CREST), a consortium of academia and industry, conducted an exploratory survey. The purpose of this survey is to ascertain the present capacity of I4.0 learning centres to provide consultation and training services to manufacturing entities in order to function as a catalyst for these entities' success in implementing I4.0. The analysis of the results is explored in greater detail in the remaining sections of this article, and the article's main research contribution is summarised in the conclusion segment.

2.0 METHODOLOGY

This exploratory quantitative survey was conducted from March 1 to May 31, 2021, with the population frame consisting of universities, research centres, and state-sponsored skill development centres in Malaysia; and the sampling frame consisting of technology centres that offer I4.0 services such as training, implementation, and consulting. These learning centers are associated with CREST, undertaking I4.0 initiatives in Malaysia. To reach out to Malaysia's widely spread learning facilities, a quantitative survey was chosen over a qualitative one. The sample size is sixteen recognised learning centres participating in this exploratory survey. Each state with learning centres replied to the survey, with the exception of Johor. The response rate was 75%.

The questionnaire consisted of six components and was administered via an online survey platform. The survey questions were distributed via an online link to the learning centres for completion. The sections include information about the centres' demographics, the technology and services they provide, the vertical and horizontal technologies they adopt, their affiliations or partnerships with other entities, their existing industrial engagement, and a section on the challenges and proposals for I4.0 from the centres' perspective.

3.0 RESULTS ANALYSIS

3.1 Demographic and Distribution of Technology Learning Centres

One of the analyses conducted using the 16 identified I4.0 learning centres is to look into the distribution of the centres in relation to the concentration of manufacturing enterprises in Malaysia. According to the Malaysian Department of Statistics' Economic Census 2016 Manufacturing Sector, manufacturing establishments are concentrated in Selangor (20.4%) and Kuala Lumpur (10.7%). In comparison to the location of the learning centres, the distribution is shown in Table 1.

It was discovered that 53.3% of learning centres cater to the top three locations with the biggest distribution of industrial establishments, namely Selangor, Kuala Lumpur, and Johor, which account for 47.5% of the population. This indicates that learning centres are available in the top three locations with manufacturing companies. However, seven states in Malaysia, accounting for 46.7% of the total, lack learning centres in their respective states. This could be perceived as a shortage of important industrial and training facilities.

Table 1: Distribution of manufacturing establishments in Malaysia and the availability of learning centers

	Location: State (and Federal Territory)	Percentage of manufacturing entities (%)	Availability of learning centers
Klang Valley			
1	Selangor	20.4	2
2	Kuala Lumpur	10.7	4
Northern Corridor Economic Region (NCEC)			
3	Perak	8.9	0
4	Pulau Pinang	8.5	3
5	Kedah	6.7	1
6	Perlis	0.8	0
East Coast Economic Region (ECER)			
7	Terengganu	4.1	0
8	Pahang	3.6	2
9	Kelantan	3.8	0
Sarawak Corridor of Renewable Energy (SCORE)			
10	Sarawak	5.2	1
Sabah Development Corridor (SDC)			
11	Sabah	3.7	1
Others			
12	Johor	16.4	2
13	Negeri Sembilan	3.9	0
14	Melaka	3.1	0
15	WP Labuan	0.2	0

3.2 Capabilities and Services Offered by Learning Centres

The services rendered by the I4.0 technology learning centres is depicted in the Table 2.

To drive I4.0 in Malaysia, training, solution workshops, and the availability of solution demonstrations and hands-on labs are all essential. Malaysia's labour productivity has improved by 3-4% in recent years, according to the Malaysia Standard Classification of Occupations (MASCO) 2013, as mentioned by Economic Census 2016, although the country's worldwide position and use of high-skilled labour has remained stationary. In 2016, Malaysia's labour productivity was rated 44th, the same as it had been since 2009. The proportion of high-skilled workers in the workforce has decreased from 19% in 2010 to 18% in 2017. As a result, the availability of I4.0 learning centres' training and skill enhancement workshops is considered as a significant driver in empowering Malaysia's human resources to adopt I4.0. This aligns with the Industry 4WRD goal of raising the number of high-skilled people in manufacturing from 18% to 35% by 2025.

Table 2: Services provided by learning centres

	Services Rendered	Description of services	Percentage %
1	Training	I4.0 awareness, conceptual, best practices including tools and solution training from learning centres	100
2	Demo machines or software	Availability of prototype or facility within the learning centres as an aid to facilitate learning	67
3	Solution workshop	Hands-on lab on tools, software and hardware/ machines related to I4.0 technology enablers	58
4	Consultancy	Provision of best practices and implementation of I4.0 technologies for the manufacturing entities	58
5	System integration	Services by learning centres to assist with the assets integration in the manufacturing value chain in the hierarchical and vertical axis of architectural model	50
6	Deployment project support	Provision of project management services for manufacturing entities in their I4.0 related projects	33
7	Funding Facilitation	Cooperation and joint-project arrangement with sponsorship between learning centres and manufacturing entities to realise specific I4.0 initiatives adoption	8

Table 3: Technologies offered by learning centres

	Technologies at learning centres	Availability percentage (%)
1	Internet of Things (IoT)	92
2	System integration	92
3	Autonomous robots	67
4	Simulation	75
5	Cloud computing	67
6	Artificial Intelligences (AI)	58
7	Big data analytics	58
8	Augmented reality	33
9	Cybersecurity	33
10	Additive manufacturing	33
11	Advance material	8

One of the pillars of Industry 4WRD is funding and outcome-based incentives. According to the survey, technology learning institutions only provide finance facilitation in 8% of cases. This finding is in line with the findings of the study, which classify learning centre responsibilities as accelerators (8%), technology funding facilitators (8%), and commercialization facilitators (17%). Learning centres identified themselves as training provider (83%), technical and skill development centre (92%).

This suggests that learning centres in Malaysia are not pursuing commercialization initiatives with manufacturing companies, preferring to focus on training and skill development. While funding for the I4.0 journey can come from a variety of sources at the macro or national level, the survey shows that manufacturing entities can benefit from partnership with learning centres. In addition, there is a need to improve commercialization between learning institutes and industrial companies.

Equally important is the availability of technology that can be used and supported by I4.0 learning centres. Table 3 shows the capabilities of the learning centres based on the various technologies used, as well as the experience and information that can be imparted through know-how trainings.

Adoption of these enabling technologies adds a substantial new dimension to the production landscape, resulting in a noticeable boost in industrial productivity. According to the exploratory study, learning centres are prepared to act as a catalyst in assisting the manufacturing industry in realising the I4.0 potential. Some of the available capabilities cited by the survey respondents include Drone Center providing development and commercialization of drone technologies, Autonomous Mobile Robot (AMR) Center which focuses on mechanical design, fabrication, PCB design, and application development, 3D Printing Center for rapid prototyping, XR Center with augmented reality programs, and other Center of Excellence (CoE) labs equipped with the latest systems for applied engineering to further develop I4.0 future ready workers.

These high-tech services supplied by I4.0 learning centres are critical for Malaysia's I4.0 implementation. According to the World Economic Forum's 2018 Readiness for the Future of Production Report, Malaysia's competitive position is jeopardised by increasing competition from countries such as Indonesia, the Philippines and Vietnam, as well as established global manufacturing leaders such as Germany, China, Japan, and the Republic of Korea.

3.3 Affiliation and Collaboration with Industries and Countries

Additionally, the accomplishment of I4.0 learning centres is contingent upon their affiliation with industrial partners engaged in genuine manufacturing. Close partnership to better connect market needs with the assessment of the centres' training and technological capabilities would be critical to ensuring that the explicit manufacturing sectors that require I4.0 catalysts receive adequate backing. Table 4 indicates the many forms of industrial involvement that the I4.0 learning centres engage in.

Electrical and electronics, machinery and equipment, chemical, medical device, and aerospace are the five primary industries mentioned in the Industry 4WRD policy for I4.0. According to the survey's findings, learning centres should

increase their partnership with industry associates, particularly in the rubber, aerospace, chemical and medical device sectors, to ensure that industrial requirements are considered more fully in their R&D and skillset competency training to avoid programme gaps. Simultaneously, the majority of learning centres (75%) are directly affiliated with manufacturers in the electrical, electronic, and mechanical sub-sector. The survey's respondents also indicated that they expect to collaborate with industry partners more in the future, both locally and internationally, with the latter including associates from Europe and China, notably Germany. This is in line with the World Economic Forum's (WEF) Readiness for Future of Production Report (2018), which places countries like Germany, China, Italy, Poland, and France in the 'Leader' quadrant of the report, where learning centres' collaboration with industry partners from these countries would boost closer knowledge sharing that could be employed for the benefit of local manufacturing.

Table 4: Types of industry engaged by learning centres

	Manufacturing sub-sectors	Affiliation and engagement by learning centres (%)	Identified as focus sectors of I4.0 in 4WRD
1	Education	83	Not identified
2	Electronics, Electrical and Mechanical	75	Primary
3	R&D	75	Not identified
4	Agriculture	67	Not identified
5	Food and beverages	58	Secondary
6	IT and digital business	58	Not identified
7	Public sector	50	Not identified
8	Medical device	43	Primary
9	Telecommunication	42	Not identified
10	Construction	42	Not identified
11	Aerospace	33	Primary
12	Rubber	33	Primary
13	Hospitality	33	Not identified
14	Professional services	29	Secondary
15	Chemical / petro chemical	25	Primary

Learning centres should, nevertheless, explore forming partnerships with industry partners from the Republic of Korea and Japan as well, as these nations are also ranked in the report's 'Leader' quadrant.

3.4 Vertical and Horizontal Integration of Learning Centres

A segment of the questionnaire was devoted to determining the digitalization maturity of the learning centres' systems. The survey questions were developed using the Industrie 4.0 Reference Architecture Model (RAMI 4.0) as a guide. The German Electrical and Electronic Manufacturers' Association (ZVEI) developed this model to support Industrial 4.0 aspirations. This three-dimensional framework establishes a shared knowledge of standards among community stakeholders (Birtel, *et al*., 2019).

The 3-dimensions of RAMI 4.0 consist of:

a) Hierarchy level – where product, field devices, control devices, station, work units, enterprise to connected world are interfaced through standards defined by ISO/IEC62264 International Standard: Enterprise-control System Integration and ISO/IEC61512, where it is made up of four concerted standard documentations
b) Lifecycle and value stream – where product development to post product production activities standard is defined by ISO/IEC62890 Industrial-process Measurement, Control and Automation
c) Architectural layer – where assets are represented in the physical world with a digital twin and illustrate how asset integration up to the functional and business layers benefits manufacturing entities in a value chain. The Asset Administration Shell (AAS) provides a guiding principle on assets' digital representation (Grangel-Gonzalez, *et al*. 2016).

There are ten steps to perform vertical and horizontal integration based on RAMI 4.0 through the technology enablers or the I4.0 technology pillars, in Industry 4.0 deployment. As a result, it is critical to comprehend how the technological pillars of learning centres map to the vertical and horizontal integration technology solutions, as shown in Tables 5(a) and 5(b). Every stage of the vertical and horizontal integration is assisted with the required technology enablers. By mapping the technology services provisioned by the learning centers, it is possible to deduce the capabilities of the learning centers in assisting manufacturing companies in their I4.0 quests.

Learning centres will be able to support the execution of steps one to five in vertical integration. At the highest levels, the learning centres offer very rudimentary support. Learning centres' help for horizontal integration is limited due to its technical pillars' specialisation, which is mostly focused on vertical integration.

Figure 1 shows the number of centres required to support vertical integration (VI) and horizontal integration (HI). According to RAMI 4.0, VI and HI must take ten steps to achieve Industry 4.0 (Tables 5(a) and 5(b)). From the survey, there are 7, 9, and 10 learning centres supporting the execution of vertical integration phases 1 to 3, with 6 learning centres focusing on networking and communication, integration and interoperability, and database management. In addition, in vertical integration, 5 and 8 centres offer visualisation dashboards with analytics, and in horizontal integration, digital enterprise, business analytics, and AI, respectively. The result corresponds to the services given by the previously mentioned centres.

Figures 2(a) and 2(b) demonstrate the breakdown of the number of centres that support vertical and horizontal integration based on investment corridors. The Klang Valley and the NCER are two investment corridors that can enable vertical and horizontal integrations from steps 1 through 5 and 9 in general. Nonetheless, from steps 6 to 8, there is a general lack of support for vertical and horizontal integration. It's also worth noting that the I4.0 deployment in East Malaysia, such as Sabah and Sarawak, indicating insufficiency to support manufacturing organisations within the investment corridors.

Table 5(a): Technology pillars of I4.0 learning centres map to vertical integration table

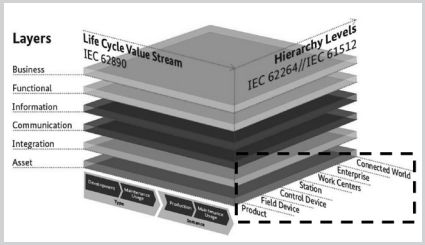
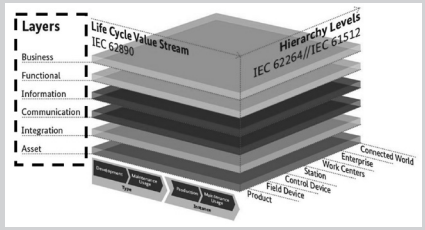
HI Steps	Reference Architecture: RAMI 4.0  Vertical Integration (VI) Technology Solutions	Advanced Materials	Additive Manufacturing	Simulation	Internet of Things	Autonomous Robots	Augmented Reality	System Integration	Big Data Analytics	Artificial Intelligences	loud Computing	Cybersecurity
1	Product, tracking and traceability	●	●	●	●							
2	Field devices, smart devices and robot				●	●	●					
3	Control devices (PLC, IPC and others)				●	●	●					
4	SCADA and data acquisition solution				●							
5	Control devices and SCADA integration							●				
6	Manufacturing Operations Management											
7	SCADA and MES integration							●				
8	Enterprise Resource Planning											
9	MES and ERP integration							●				
10	Visualization dashboard with analytics								●	●	●	●

Table 5(b): Technology pillars of I4.0 learning centres map to horizontal integration table

HI Steps	Reference Architecture: RAMI 4.0  Horizontal Integration (HI) Technology Solutions	Industry 4.0 Technology Enablers										
		Advanced Materials	Additive Manufacturing	Simulation	Internet of Things	Autonomous Robots	Augmented Reality	System Integration	Big Data Analytics	Artificial Intelligences	loud Computing	Cybersecurity
1	Asset identification and definition				●							
2	Asset digitalization and integration				●							
3	Networking and communication							●				
4	Integration and interoperability							●				
5	Database management										●	
6	Hybrid data management										●	
7	Manufacturing chain management							●				
8	Data protection and access control											●
9	Data and information brokering							●			●	
10	Digital enterprise, business analytics, AI										●	●

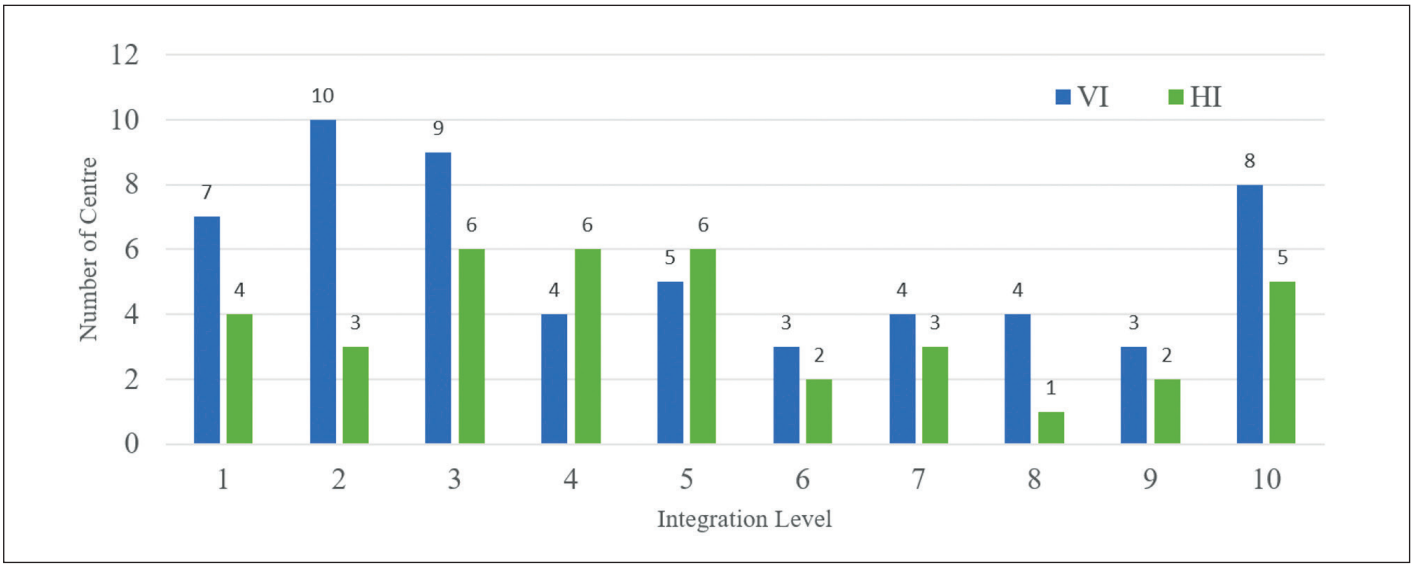


Figure 1: Support of learning centres in the vertical and horizontal integration

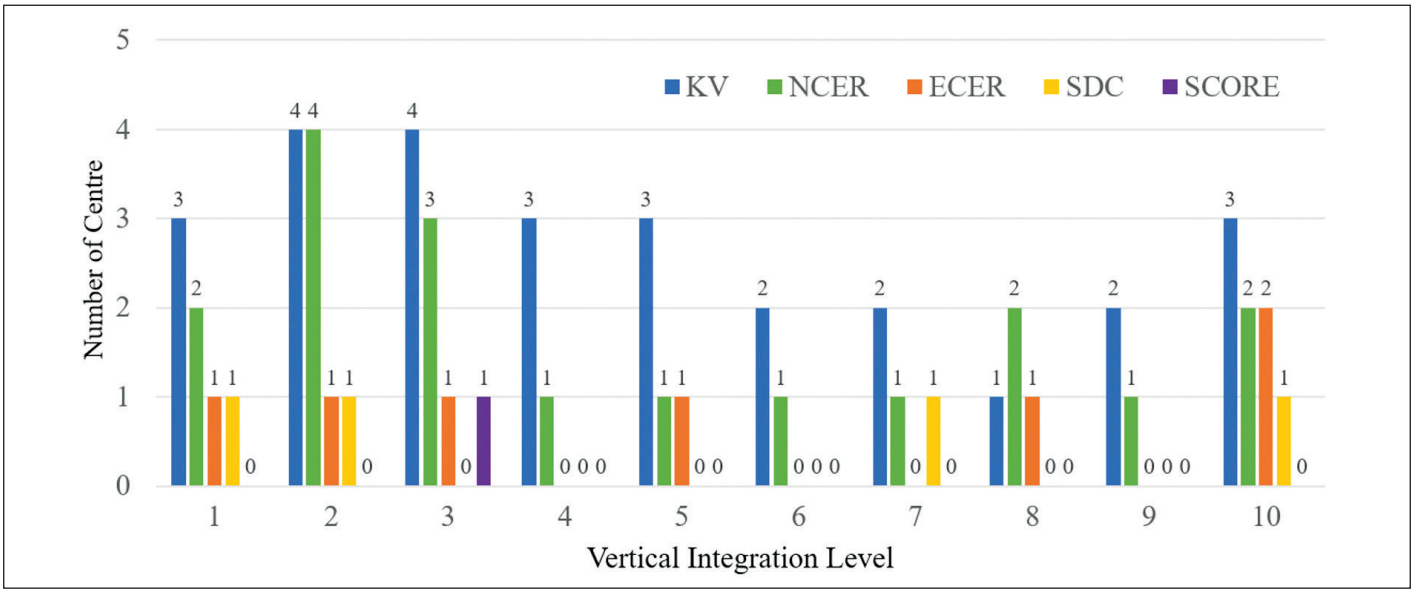


Figure 2(a): Learning centres to support the vertical integration based on investment corridors

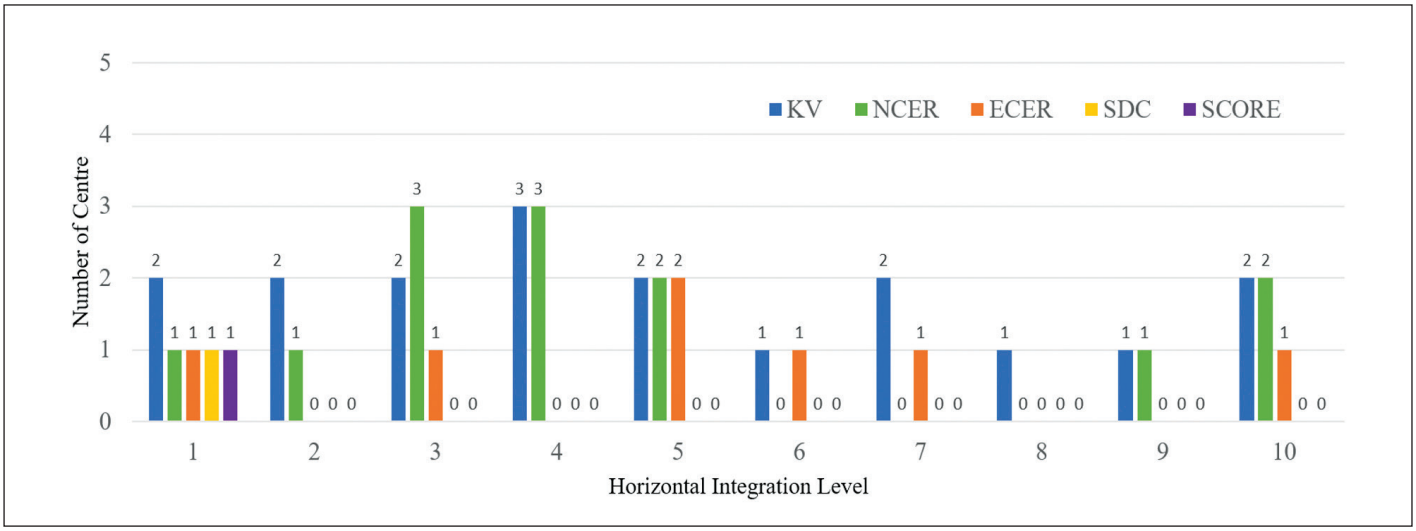


Figure 2(b): Learning centres to support the horizontal integration based on investment corridors

3.4 Challenges and Opportunities

The common issue mentioned in the survey on the topic of obstacles faced by manufacturing companies as assessed by learning centres is funding, which is divided into two categories. The first is learning centres and service providers' funding. For example, the Ministry of International Trade and Industries (MITI) and the Ministry of Science, Technology and Innovation (MOSTI) should put more resources into advancing learning centres' facilities with cutting-edge platforms to show how technology can help manufacturing businesses adopt I4.0 roadmaps faster.

Another type of support is funding for manufacturing companies to enable them to implement I4.0-related technologies. Industry 4.0 deployment results in horizontal, vertical, and end-to-end integration for manufacturing entities (Wang, *et al*., 2016). A significant initial financial and time commitment is necessary to create and execute an architecture that is tailored to the needs of such enterprises (Singer, 2015). To implement Industry 4.0, significant capital expenditures are required, and funding must be raised (Rojko, 2017).

Other form of financial assistance could be tax incentives or tax exemptions on the adoption of technology and even on the services obtained by the manufacturing enterprises. Under the funding programme, the 4WRD Policy specifies two strategies. The much-anticipated action plans stated in the said strategy, according to survey respondents, should be hastened, such as the formation of a government-led expansion fund for I4.0.

One more issue raised in the survey is that manufacturing companies lack the know-how to get started with I4.0 adoption and application. There are few local success stories to promote I4.0 since manufacturing companies believe such initiatives will have a low return on investment (ROI). These notions are in line with the findings of Sony's (2020), which found that cybersecurity concerns, trade union concerns, the initial high cost of implementation, workforce readiness and the negative impact of data sharing between enterprises in the value chain would all pose challenges to a successful I4.0 adoption. Furjan *et al*. (2020) add to this by stating that digital transformation faces a high chance of failure if the business processes and environment are neglected.

The necessity for a coordinated effort among the various parties in the country to establish a one-stop-centre for I4.0 services such as funding and policy inquiries, and the promotion of learning centre services is among the suggestions made by the survey respondents. The provision of a readiness assessment for manufacturing companies, followed by the referral of appropriate learning centres to provide guidance, training and implementation services, is one of the anticipated services from this one-stop-centre.

The function of the one-stop-centre in promoting increased awareness of learning centres' capabilities and services was also mentioned by respondents in the survey as a way to address the problem of learning centres' lack of visibility in the country. Another crucial success factor (CSF) in I4.0 in the country is the adoption of a single marketing platform, which includes the use of social media to boost learning centres' visibility.

Respondents also recommended that learning centres continue to put up actual assembly lines to demonstrate I4.0 competencies, as well as a call for learning centres to undertake vertical and horizontal integration among themselves to further demonstrate their competency. Other learning centres around the country may be able to acquire data for training purposes by connecting to the digital twin, which is a virtual version of the real-time physical production line in both centres.

4.0 CONCLUSION

The survey's conclusions are provided in this article, with the primary finding pointing to the learning centres' inability to deliver the essential services and training to manufacturing companies in Malaysia to promote I4.0 development. A defined vertical and horizontal execution strategy is necessary to support I4.0 learning centres in planning and enriching their training and services based on the technology pillars in which they specialized in. Additional learning centres should be developed in other states throughout the country, as seven locations in Malaysia, accounting for 46.7% of the country, now lack the provision of learning centres.

Additionally, there is a need to strengthen commercialization between technological learning centres and manufacturing firms. To further accelerate the adoption of I4.0 in Malaysia, learning institutes should strengthen their collaboration and affiliation with the country's primary industries, as well as with other countries. Another potential comprises the need for a coordinated effort among the country's various stakeholders to establish a one-stop-centre for I4.0 services such as funding and policy inquiries as well as the promotion of learning centre services. ■

REFERENCES

[1] Birtel, M., David, A., et al. (2019) FutureFit: A Strategy for Getting a Production Asset to an Industry 4.0 Component – a Human-Centred Approach, 29th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2019), 24-28

[2] Flatt, H., Schriegel, S., Jasperneite, J., Trsek, H., & Adamczyk, H. (2016) Analysis of the cyber-security of industry 4.0 technologies based on RAMI 4.0 and identification of requirements, Emerging Technologies and Factory Automation (ETFA), 2016 IEEE 21st International Conference On, 1-4

[3] Furjan, M., Tomicic-Pupek, K. & Pihir, I. (2020) Understanding Digital Transformation Initiatives: Case Studies Analysis, Business Systems Research, 11 (1)

[4] Grangel-González, L. Halilaj, S. Auer, S. Lohmann, C. Lange and D. Collarana, (2016). An RDF-based approach for implementing industry 4.0 components with Administration Shells. 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), 1-8

[5] Ministry of International Trade and Industry (2016) National Policy on Industry 4.0, Kuala Lumpur, Ministry of International Trade and Industry

[6] Malaysia Department of Statistics (2016) Economic Census 2016 Manufacturing Sector, Putrajaya.

[7] Moeuf, A., Lamouri, S., Pellerin, R., Tamayo-Giraldo, S, Tobon-Valencia, E. and Eburdy, R. (2020) Identification of Critical Success Factors, Risks and Opportunities of Industry 4.0 in SMEs, International Journal of Production Research, 58 (5), 1384-1400.

[8] Neugebauer, R., Hippmann, S., Leis, M. and Landherr, M., (2016) Industrie 4.0 - From the Perspective of Applied Research. Procedia CIRP, 2–7.

[9] Rojko, A. (2017). Industry 4.0 concept: Background and overview. International Journal of Interactive Mobile Technologies (IJIM), 11(5), 77–90.

[10] Safar, L., J. Sopko, S. Bednar, and R. Poklemba. (2018) Concept of SMEBusiness Model for Industry 4.0 Environment. TEM Journal 7 (3), 626.

[11] Singer, P. (2015). Are you ready for Industry 4.0? Deloitte Review, 1(22), 1–136

[12] Sony, M. (2020) Pros and Cons of Implementing Industry 4.0 for the Organizations: a Review and Synthesis of Evidence, Production & Manufacturing Research, 8 (1)

[13] Wang, S., Wan, J., Li, D., & Zhang, C. (2016) Implementing smart factory of Industrie 4.0: An outlook, International Journal of Distributed Sensor Networks, 12(1)

[14] World Economic Forum (2018) Readiness for the Future of Production Report 2018, Geneva.

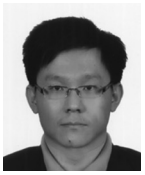
[15] Zhang, X., Ming, X. & Qu, Y. (2019) Top-level scenario planning and overall framework of smart manufacturing implementation system (SMIS) for enterprise, The International Journal of Advanced Manufacturing Technology, 104, 3835-3848

PROFILES

YONG CHAN HUANG is a PhD candidate at Tunku Abdul Rahman University College. He received his BSc in Information Systems and Management (Hons) from University of London and MSc in Information Technology Management (Distinction) from University of Sunderland. His working experience includes being IT Director for global IT hub and VP of IT driving IT strategic initiatives. His research interest includes areas in strategic IT management, change management and Industry 4.0.

ASSOC. PROF. TS. DR LEE WAH PHENG is the Associate Professor in Tunku Abdul Rahman University College. He worked in the manufacturing industry for 10 years and more than 20 years of business and education industry experiences. Dr Lee is a pioneer and consultant in Industry 4.0. He works with a team of researchers and industry partners to develop a holistic digital solution suitable for the small and medium enterprises.

DR CHRISTOPHER LAZARUS received his Bachelor in Computing from the University of Portsmouth and his PhD in Computer Science from Essex University (UK) in 2011. His thesis work includes the use of multi-objective optimisation in genetic programming to evolve robot controllers. He is a senior lecturer teaching information technology and computer science courses at Tunku Abdul Rahman University College, Malaysia. His research interest includes areas in Artificial Intelligence, Machine Learning, Data Science, Robotics and Industry 4.0.



DR TAN KONG WOUN is a PhD holder majoring in Technology Management (Manufacturing Flexibility a.k.a Agile Manufacturing). He received his Bachelor in Technology Management in 2007 and PhD in Technology Management in 2017 from Universiti Utara Malaysia. His teaching interests are in the areas of industrial engineering, basic electronics, construction management, CAD/CAM, production management, research methodology, computer networking, operations management and statistical data analysis.

DR LIM YEE MEI received her PhD in Artificial Intelligence from De Montfort University, United Kingdom. During her tenure with TAR UC. She was appointed as Associate Dean of Department of Computer Science and Mathematics in 2017, Associate Dean of Department of ICT in 2018, the Lead of Research Centre for ICT Innovations and Creativity, and the Project Lead of TAR UC Smart Campus IoT Applications. She is currently the Director of OMIS Consulting Sdn. Bhd., a consultancy company that offers Industry 4.0 education and consultancy services for SMEs.

ONLINE GAMING ADDICTION FACTORS AMONG TAR UC STUDENTS IN KL

(Date received: 12.01.2022/Date accepted: 25.04.2022)

F Z Yaw[1], K M Cheok[1], K Z Ng[1], J X Teo[1], T T Ting[2], Siew Mooi Lim[1]*

[1] Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia.

[2] Faculty of Information Technology, INTI International University, Negeri Sembilan, Malaysia.

**Corresponding author: : siewmooi@tarc.edu.my*

ABSTRACT

Online gaming is a popular digital entertainment that people worldwide, including university students, are well-received. Consequently, addiction to online gaming is worrying, and this issue has received significant attention. We collected and analysed primary data to reveal the factors relating to online gaming addiction. We performed a bivariate correlation test to examine the relationship between online games addictions with the characteristics of depression, loneliness, motivation for escapism and motivation for achievement on the 118 responses that we collected from online questionnaires. We also calculated Cohen's effect size, f^2, for each path. The results show that those identified factors positively correlate with online game addiction with a large effect size.

Keywords: Online game addiction, depression, loneliness, motivation for achievement, motivation for escapism

1.0 OVERVIEW

Online gaming addiction (OGA) has been explored in various aspects, including social psychology and psychiatry (Hsu *et al.*, 2009; Xu *et al.*, 2012; Kim *et al.*, 2008). For example, pathological gaming was conducted on 3,034 elementary students in Singapore, resulting in dysfunctional family, friends, and school relationships. It was also related to depression, social phobias, anxiety, and lower grades in 9% of the study participants (Gentile, 2011). Based on the 9% estimate, over 11 million Fortnite players may exhibit a harmful gaming pathology (Fortnite, 2018). Moreover, given the prevalence of online gaming, with an estimated 2.2 billion active gamers worldwide, the problem is alarming (McDonald, 2017). World Health Organisation (WHO) is monitoring this situation seriously and has classified 'gaming' under the category of "Disorders due to addictive behaviours" (WHO, 2021).

1.1 Depression and OGA

Increased levels of depression are related to different forms of addiction (Griffiths *et al.*, Stavropoulos *et al.*, 20162016). Whereas loneliness and depression were proven related to symptoms of pathological gaming in a mutually upholding cycle (Krossbakken *et al.*, 2018). The researcher Taechoyotin (2020) has discussed that the person might feel stressed, depressed, or anxious by the problems in the real world and may choose to use the game world (where they feel safe and secure) to escape these feelings. Burleigh (2018) had identified that depressed adolescents were significantly more likely to be addicted to online games when they experienced stronger Game Avatar Relationships.

1.2 Loneliness and OGA

Loneliness is not only related to social isolation but people can be lonely even when other people surround them. Based on a cross-sectional study conducted by Kim *et al.* (2009) a reciprocal relation between pathological gaming and loneliness among adolescents cognitive-behavioural model of PIU. The study showed that lonely individual or did not have good social ability may develop strong compulsive Internet use behaviours. Jeong *et al.* (2015) have discussed that loneliness is positively related to game addiction. Access to online games is an easy way to release tension because online games are a channel close at hand. Furthermore, Chen and Leung (2016) have discussed that loneliness was significantly linked to mobile game addiction.

1.3 Motivation for Escapism and OGA

Yee (2006) debated that motivation for escapism is one of the four components of game immersion. Escapism refers to a person's attempt to avoid thinking about or to run away from real-life problems by engaging in an online experience (Yee, 2006). The research published by Bányai *et al.* (2019) constructed a questionnaire that collected 4284 results from e-sport and recreational gamers.

The results stated that the escapism motive appeared to be the common predictor of problematic gaming among e-sport and recreational gamers. Another research analysed 27 studies, with only 7 studies with negative outcomes, 9 studies with positive outcomes and 11 studies having an escapism relationship with both mixed outcomes in a given independent study (Hussain *et al.*, 2021). It was found that in western countries, escapism via video games held a stronger association with negative outcomes

while in non-Western countries, escapism via video games is more likely to lead to positive outcomes. Another study, published by Šporčić and Glavak-Tkalić (2018) had gathered 509 young adults via questionnaire with the hierarchical regression analyses suggested that escapism is a significant predictor of problematic online gaming.

1.4 Motivation for Achievement and OGA

Yee (2006) suggested that in gaming, the sense of achievement originates from three components: advancement, mechanics, and competition. Advancement is the players' interest in gaining power and accumulating in-game wealth. Mechanics refers to players' interests in analysing the underlying rules, levelling up characters, and optimising character performance. Finally, competition refers to the desire to challenge and compete with other players.

Specific psychological characteristics drive OGA. Yee (2006) collected online survey data from 30,000 users of Massively Multi-User Online Role-Playing Games (MMORPGs) over three years to explore users' demographics, motivations, and derived experiences. His study reveals that male players were significantly more likely to be motivated by the achievement and manipulation factors (Yee, 2006).

Following this framework, Chang *et al.* (2018) examined the mediational effects of multiple gaming motives, from online game involvement to problematic Internet use. They discovered that advancement motives have a positive relationship with online game involvement. T'ng and Pau (2021) assessed 1175 Malaysia MOBA gamers to study the avatar in the relationship between motivations of gaming and OGA. The findings revealed that motivation of achievement, motivation of immersion, and identification of avatar positively predict OGA. Besides that, Khan and Muqtadir's (2016) research indicated that problematic gamers had stronger motivation for socialisation, achievement, and immersion than non-problematic gamers.

1.5 Recap

Online games are the major contributor to the video games market. This type of digital entertainment has an estimated revenue of over USD90 billion by 2020 (McDonald, 2017). Because of COVID-19 pandemic lockdowns, it is estimated that the online gaming industry will continue to grow exponentially. Popular online games like Fortnite have 125 million players (Fortnite, 2018).

Research shows that online games provide a form of escapism from the world's reality. Therefore, it continues to gain popularity among the youth and young adults (André *et al.*, 2018). However, on the downside, online games cause many cases of online gaming addiction (OGA).

In a study for an online game called Massively Multiplayer Online Roleplaying Game, the author, Yee (2006) showed that motivation for achievement and motivation for escapism has caused people to be engaged in the games. In another study, loneliness and depression have a mutual enforcing loop with online gaming addiction (Kim *et al.*, 2009; Burleigh, 2018).

In short, this paper aims to prove that psychological factors (depression and loneliness) and gaming motivations (escapism and achievement) are positively associated and are predictors of online gaming addiction.

2.0 RESEARCH METHODOLOGY

This section justified the chosen factors of depression, loneliness, motivation for escapism, and motivation for achievement as the possible causes of gaming addiction in TAR UC students.

Past studies suggest that MMO players create an avatar in which they often imbue part of their identity and idealised identity (Bessiere *et al.*, 2007). This may prompt them to project their idealised selves onto their avatars to regulate related depressive emotions (Bessiere *et al.*, 2007). Therefore, we hypothesise that depression is related to online gaming addiction. Furthermore, psychopathologies, including Attention-Deficit/Hyperactivity Disorder (ADHD) and depression, were the most significant factors of online gaming addiction in individuals.

People who suffer from psychological problems (e.g. loneliness) may use online or video games to satisfy their needs to escape from negative moods. Consequently, emotionally susceptible individuals may be deeply immersed in virtual life. Thus, we hypothesise that loneliness is related to online game addiction.

A handful of research projects have suggested that escapism motivation increases the extent of online game playing (Yoo, Sanders and Cerveny, 2018). We suggest that higher levels of engagement and more time spent on the game can afford players the opportunity to be more familiar with the game world and to acquire a sense of belonging and closeness, which, in turn, can lead to online game addiction.

A study by Chang, Grace M.Y.Hsieh & Sunny S.J.Lin (2018) shows that the desire for advancement encourages players to stay in the game. Sepandar Sepehr & Milena Head (2018) also suggest that the perception of video game competitiveness is a strong predictor of gameplay satisfaction.

Within a gaming environment, a player with increased competence is likely to seek more power, higher-performing characters, and rare items to outperform others, which help generate feelings of capability. Therefore, we propose that the motivation for achievement keeps gamers engaged in the gaming environment, which, in turn, facilitates online game addiction. With this in view, the followings are the hypotheses of this study:

H1: Depression is positively related to TAR UC students' online gaming addiction.

H2: Loneliness is positively related to TAR UC students' online gaming addiction.

H3: Motivation for escapism is positively related to TAR UC students' online gaming addiction.

H4: Motivation for achievement is positively related to TAR UC students' online gaming addiction.

Our research instrument, online questionnaire, was distributed to TAR UC students in Google Forms from the 1st of August, 2021 until the 19th of September, 2021 (Figure 1). We applied a simple random sampling technique in this research where the respondents were chosen randomly through the indiscriminate distribution of the questionnaire. We filtered the respondents who have no experience with online games and play games for less than an hour each week.

For determining the OGA of the individual, we use the Lemmens *et al.* (2009) OGA scale for determining the OGA of the individual. We adapted Yee's gaming motivation items to measure achievement motivation and escapism motivation. The questions were crafted with the UCLA loneliness scale (Version 3) for loneliness. As for depression, we referenced the

Center for Epidemiologic Studies Depression Scale (CES-D) for the questions.

With a thorough overview and background study of the research topic, we prepared our questionnaire consisting of the sections below with their respective questions. Then, before actual data collection from the intended respondents, pre-test and pilot test were carried out by distributing the questionnaires to some peers to ensure the smoothness of the research process.

Game Addiction

1. I think about playing games all day long.
2. I often find I have to increase my playing time to get the desired enjoyment.
3. Me or others unsuccessfully tried to reduce my game use.
4. I feel anxious when I am unable to play games frequently.
5. I often have arguments with others (e.g. family, friends) over the time spent on playing games.
6. I often neglect other important activities (e.g., school, work, sports) to play games.

Depression

1. I lack the motivation to do simple things such as cleaning and showering.
2. I do not have hope for the future.
3. I have no goals, or have given up on them.
4. I feel worthless, and guilty when people care for me.
5. I often feel lost and confused.

Loneliness

1. I often feel that I am not close to anyone.
2. I often feel my interests and ideas are not shared by those around me.
3. I often feel I am isolated from others.
4. I often feel left out.

Escapism

1. I enjoy being immersed in a game world.
2. I often play so I can avoid thinking about some of my real-life problems or worries.
3. I often play to relax from the day's work.
4. It is important for me that the game allows me to escape from the real world.

Motivation for Achievement

1. It is important for me to level up my character as fast as possible.
2. It is important for me to acquire rare items that most players will never have.
3. It is important for me to become powerful in games I play.
4. It is important for me to accumulate resources, items or money.

Figure 1: Questionnaire Items

3.0 RESULTS AND DISCUSSION

The result of 0.94 of Cronbach's alpha test indicated a robust internal consistency of the questions. Table 1 shows the respondents' demographic data.

Table 1: Questionnaire respondents' demographic statistics

	Frequency	Percent
Gender		
Female	35	29.7%
Male	83	70.3%
Programme		
Accounting	2	1.7%
Advertising	1	0.8%
Computer Science	1	0.8%
Corporate Administration	1	0.8%
Data Science	6	5.1%
Engineering	1	0.8%
Enterprise Information Systems	10	8.5%
FCCI	1	0.8%
Graphic Design	1	0.8%
Information Security	9	7.6%

Interactive Software Technology	6	5.1%
Internet Technology	15	12.7%
Logistics and Supply Chain Management	1	0.8%
Marketing	2	1.7%
Mass Communication	1	0.8%
Mechatronic Engineering	1	0.8%
Multimedia Design	1	0.8%
Software Engineering	3	2.5%
Software Systems Engineering	33	28.0%
Software Systems Development	18	15.3%
Architecture	1	0.8%
Finance and Investment	1	0.8%
International Business	1	0.8%
Mechatronic Engineering	1	0.8%
	118	100%

The results of Pearson Correlation and Effect Size (Cohen's f^2), as presented in Table 2 shows the significant positive relationship between Depression and Online Game Addiction (r = 0.531, sig = 0.000) with large effect ($f^2 = 0.39$). Thus, H1 is accepted. The result is consistent with the findings of Burleigh (2018), which demonstrated that depression is statistically significant related to OGA.

Table 2: Pearson Correlation and Effect Size

	Pearson Correlation	**Sig**	**Mean**	**Std Dev**	**Cohen's f^2**
	Online Game Addiction				
Depression	0.531	.000	2.42	1.16	0.39
Loneliness	0.531	.000	2.79	1.13	0.39
Motivation for Escapism	0.571	.000	3.60	0.97	0.48
Motivation for Achievement	0.662	.000	3.22	1.11	0.78
Dependant Variable: Online Game Addiction (OGA)					
$f^2 \geq .02$ = small effect; $f^2 \geq .15$ = medium effect; $f^2 \geq .35$ = large effect (Cohen, 1988)					

Pearson Correlation for Loneliness and OGA is positive, r = 0.531 with effect size of $f^2 = 0.39$ (large effect). This concludes there is a high and positive correlation between Loneliness and OGA and is practically significant in the real world (effects size Cohen's f^2 is large (Pritha, 2021)). Therefore, H2 is accepted in which there is a correlation between Loneliness and OGA.

The social aspect of online games provides an easy way for social interaction for lonely people. Other than that, online games provide anonymity and are less socially demanding. This could explain loneliness leading to addiction. The result is consistent with the findings of Jeong *et al.* (2015), which revealed that loneliness, aggression and depression predict OGA.

Pearson Correlation for Motivation for Escapism and OGA is also positive, r = 0.571 with large effect size, $f^2 = 0.48$. Thus, H3 is accepted in which motivation for escapism is related to online gaming addiction, which is significant in the real practical world. The relationship of Motivation for Escapism is higher compared to Depression and Loneliness. This could be attributed to trending online games now that feature immersive worldbuilding. The result is consistent with the findings of (Hussain *et al.*, 2021), which states that motivation for escapism is correlated with OGA.

Based on Table 2, the Pearson Correlation for Motivation for Achievement and OGA is r = 0.662, a high correlation between Motivation for Achievement and OGA. Thus, H4 is accepted with the largest effect size, $f^2 = 0.78$. The acceptance of H4 suggests that our initial assumptions were correct, as people who view achievement as important may be attracted to online gaming. The result is consistent with the findings of T'ng *et al.* (2021), which revealed that motivation of achievement positively predicts OGA.

Table 3: Pearson Correlation between factors

		DP	**LON**	**ME**	**MA**
DP	Pearson Correlation	-	0.642	0.362	0.443
	Sig. (2-tailed)		0.000	0.000	0.000
LON	Pearson Correlation	0.642	-	0.443	0.321
	Sig. (2-tailed)	0.000		0.000	0.000
ME	Pearson Correlation	0.362	0.443	-	0.516
	Sig. (2-tailed)	0.000	0.000		0.000
MA	Pearson Correlation	0.443	0.321	0.516	-
	Sig. (2-tailed)	0.000	0.000	0.000	
DP – Depression; LON – Loneliness; ME – Motivation for Escapism; MA – Motivation for Achievement					

Table 3 shows the correlation results among the chosen. There is a significant positive relationship between the four factors, especially between Depression and Loneliness (r=0.642, sig=0.00). It is interesting to find out further the correlation between factors that could affect students addiction to online gaming in future research. For example, lonely students would probably have depression, which could cause the students to immerse in the virtual world of gaming.

4.0 CONCLUSIONS AND FUTURE RECOMMENDATIONS

Based on the data collected from 118 respondents through the online questionnaire, we have proven that depression, loneliness, motivation for escapism and achievement positively correlate with OGA and the effect sizes are large. Their respective Pearson Correlation results (r) are 0.531, 0.531, 0.571 and 0.662. Whereas their Effect Size results (Cohen's f^2) are 0.39, 0.39, 0.48 and 0.78, respectively.

Our results show that motivations for achievement and escapism are closely related to OGA. Although not as strong as the previous two factors, loneliness and depression are still associated with OGA. This shows that all four factors are predictors of the tendency of OGA in a person. This suggests that all four factors and OGA form a mutually reinforcing loop in which an individual unsatisfied with real-life is becoming increasingly addicted to online games.

For the sake of the young generation, further studies can include more factors of OGA and preventive measures for OGA. Moreover, the research outcomes can be a reference for other higher institutions to learn more about their students. ■

REFERENCES

[1] André JP, Jan KC, Dominika BO, Florian E and Leane A (2018) Online gamers, lived experiences, and sense of belonging: students at the University of the Free State, Bloemfontein. Qualitative Sociology Review 14(4): 122–137, https://doi.org/10.18778/ 1733-8077.14.4.08.

[2] Bányai F, Griffiths MD, Demetrovics Z and Király O (2019) The mediating effect of motivations between psychiatric distress and gaming disorder among Esport gamers and recreational gamers. Comprehensive Psychiatry 94, https://doi.org/10.1016/j.comppsych. 2019.152117.

[3] Bessière K, Seay AF and Kiesler S (2007) The ideal elf: Identity exploration in world of warcraft. Cyberpsychology and Behavior 10(4), https://doi.org/10.1089/cpb.2007.9994.

[4] Burleigh TL, Stavropoulos V, Liew LWL, Adams BLM and Griffiths MD (2018) Depression, Internet gaming disorder, and the moderating effect of the gamer-avatar relationship: an exploratory longitudinal study. International Journal of Mental Health and Addiction 16(1): 102–124, https://doi.org/10.1007/s11469-017-9806-3.

[5] Chang SM, Hsieh GMY and Lin SSJ (2018) The mediation effects of gaming motives between game involvement and problematic Internet use: Escapism, advancement and socialising. Computers and Education 122, https://doi.org/10.1016/j.compedu.2018.03.007.

[6] Chen C and Leung L (2016) Are you addicted to Candy Crush Saga? An exploratory study linking psychological factors to mobile social game addiction. Telematics and Informatics 33(4), https://doi.org/10.1016/j.tele.2015.11.005.

[7] Cohen J (1988) Statistical power analysis for the social sciences. Hillsdale, New Jersey, Lawrence Erlbaum Associates.

[8] Fortnite (2018) Announcing 2018-2019 Fortnite competitive season. Epic Games' Fortnite. USA.

[9] Gentile, DA, Choo H, Liau A, Sim T, Li D, Fung D and Khoo A (2011) Pathological video game use among youths: A two-year longitudinal study. Pediatrics 127(2), https://doi.org/ 10.1542/peds.2010-1353.

[10] Griffiths MD, Kuss DJ, Billieux J and Pontes HM (2016) The evolution of Internet addiction: A global perspective. Addictive Behaviors 53, https://doi.org/10.1016/j.addbeh.2015. 11.001.

[11] Hsu SH, Wen MH, and Wu MC (2009) Exploring user experiences as predictors of MMORPG addiction. Computers and Education 53(3), https://doi.org/10.1016/j.compedu.2009.05. 016.

[12] Hussain U, Jabarkhail S, Cunningham GB and Madsen JA (2021) The dual nature of escapism in video gaming: A meta-analytic approach. Computers in Human Behavior Reports 3, https://doi.org/10.1016/j.chbr.2021.100081.

[13] Jeong EJ, Kim DJ and Lee DM (2015) Game addiction from psychosocial health perspective. ACM International Conference Proceeding Series, https://doi.org/10.1145/2781562. 2781587.

[14] Khan A and Muqtadir R (2016) Motives of problematic and non problematic online gaming among adolescents and young adults. Pakistan Journal of Psychological Research 31(1).

[15] Kim EJ, Namkoong K, Ku T and Kim SJ (2008) The relationship between online game addiction and aggression, self-control and narcissistic personality traits. European Psychiatry 23(3), https://doi.org/10.1016/j.eurpsy.2007.10.010.

[16] Kim J, Larose R and Peng W (2009) Loneliness as the cause and the effect of problematic internet use: the relationship between internet use and psychological well-being. Cyberpsychology and Behavior 12(4), https://doi.org/10.1089/cpb.2008.0327.

[17] Krossbakken E, Pallesen S, Mentzoni RA, King DL, Molde H, Finserås TR and Torsheim T (2018) A cross-lagged study of developmental trajectories of video game engagement, addiction, and mental health. Frontiers in Psychology 9(NOV), https://doi.org/10.3389/ fpsyg.2018.02239.

[18] Lemmens JS, Valkenburg PM, Peter J (2009) Development and validation of a game addiction scale for adolescents. Media Psychology 12(1): 77-95.

[19] McDonald E (2017) Report: Insights into the $108.9Bn global games market. Newzoo. San Francisco, CA, USA.

[20] Pritha B (2021) Effect size in statistics. Scribbr. Amsterdam.

[21] Sepehr S and Head M (2018) Understanding the role of competition in video gameplay satisfaction. Information and Management 55(4), https://doi.org/10.1016/j.im.2017. 09.007.

[22] Šporčić B and Glavak-Tkalić R (2018) The relationship between online gaming motivation, self-concept clarity and tendency toward problematic gaming. Cyberpsychology 12(1), https://doi.org/10.5817/CP2018-1-4.

[23] Stavropoulos V, Gentile D and Motti-Stefanidi F (2016) A multilevel longitudinal study of adolescent Internet addiction: the role of obsessive–compulsive symptoms and classroom openness to experience. European Journal of Developmental

Psychology 13(1): 99–114, https://doi.org/10.1080/17405629.2015.1066670.

[24] T'ng ST and Pau K (2021) Identification of avatar mediates the associations between motivations of gaming and Internet gaming disorder among the Malaysian youth. International Journal of Mental Health and Addiction 19(4): 1346–1361, https://doi.org/10.1007/s11469-020-00229-9.

[25] Taechoyotin P, Tongrod P, Thaweerungruangkul T, Towattananon N, Teekapakvisit P, Aksornpusitpong C, Sathapornpunya W, Hempatawee N, Rangsin, R, Mungthin M and Piyaraj P (2020) Prevalence and associated factors of internet gaming disorder among secondary school students in rural community, Thailand: A cross-sectional study. BMC Research Notes 13(1), https://doi.org/10.1186/s13104-019-4862-3.

[26] WHO (2021) ICD-11 for mortality and morbidity statistics – 6C51 Gaming disorder.

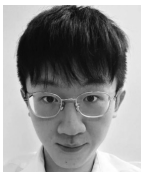
[27] Xu Z, Turel O and Yuan Y (2012) Online game addiction among adolescents: Motivation and prevention factors. European Journal of Information Systems 21(3), https://doi.org/ 10.1057/ejis.2011.56.

[28] Yee N (2006) The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. Presence: Teleoperators and Virtual Environments 15(3), https://doi.org/10.1162/pres.15.3.309.

[29] Yoo CW, Sanders GL and Cerveny RP (2018) Exploring the influence of flow and psychological ownership on security education, training and awareness effectiveness and security compliance. Decision Support Systems 108, https://doi.org/10.1016/j.dss. 2018.02.009.

PROFILES

YAW FOONG ZENG is a final year BSc (Software Engineering) student at Tunku Abdul Rahman University of Management and Technology. He is adept in different fields of computer science, with notable knowledge in software engineering, artificial intelligence and data science. In addition, with a deep interest in interactive software such as video games, he likes to discover how people interact with the software and how it affects them.
Email address: yawfz-wm19@student.tarc.edu.my



CHEOK KAH MING is a final year BSc (Software Engineering) student at Tunku Abdul Rahman University of Management and Technology. He excels in various computer science specialities, particularly software engineering, artificial intelligence, blockchain, web apps, and data science. In addition, he enjoys teamwork projects and has a strong interest in team sports such as basketball and badminton. Moreover, he is also a big fan of interactive software such as video games.
Email address: cheokkm-wm19@student.tarc.edu.my

NG KAI ZHUN is a final year BSc (Software Engineering) student at Tunku Abdul Rahman University of Management and Technology. He has special knowledge in software engineering, data science and blockchain. He is interested in video games and badminton.
Email address: kzng-wm19@student.tarc.edu.my

TEO JIAN XIANG is a final year BSc (Software Engineering) student at Tunku Abdul Rahman University of Management and Technology. He is an upcoming software engineer with expertise in web and mobile applications. He enjoys problem-solving, and video games which lead him to investigate the topic of video games that cause mental distress and how to solve it.
Email address: teojx-wm19@student.tarc.edu.my

DR TING TIN TIN received her BSc and PhD in Computer Sciences from University of Science, Malaysia. She joined Gemalto as telecommunication software engineer in Singapore before joined academic industry after her PhD graduation. She has more than 12 years of lecturing, supervising projects, and research. Her research interests including big data analytics, information systems engineering, educational data mining, psycho-academic research, and software engineering. Dr Ting received her professional certification in project management from PMI and data analytics from SAS. Dr Ting is currently attached to INTI International University, responsible for research and postgraduate supervision. At the same time, Dr Ting serves as a freelance lecturer at Monash University, Tunku Abdul Rahman University of Management and Technology, and Methodist College Kuala Lumpur.
Email address: tintin.ting@newinti.edu.my

DR LIM SIEW MOOI (SANDY) obtained her undergraduate and postgraduate degrees from Universiti Teknologi Malaysia and Universiti Putra Malaysia, respectively. Currently, she is a Senior Lecturer cum Program Leader for BComp Sc in Data Science at Tunku Abdul Rahman University of Management and Technology. Dr Lim leads the students and is involved in collaborative research and industrial consultancy projects in data science, machine learning, natural language processing and predictive modelling.
Email address: siewmooi@tarc.edu.my

FLOWER RECOGNITION MODEL BASED ON DEEP NEURAL NETWORK WITH VGG19

(Date received: 13.01.2022/Date accepted: 03.04.2022)

Zi Yuan Ong[1], Kah Kien Chye[1], Huay Wen Kang[1], Chi Wee Tan[1]*

[1] Faculty of Computing And Information Technology, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300, Wilayah Persekutuan Kuala Lumpur, Malaysia.

**Corresponding author: chiwee@tarc.edu.my*

ABSTRACT

Computer vision is one way to streamline processes like robotic process automation and digital asset management. It has come a long way in terms of its capabilities and what it can provide and do for different industries. Applications provided by computer vision include object detection and image detection. This field of technology is still relatively young and faces many challenges however. Challenges faced in this field include the lack of comprehensively annotated images to use for training the optimal algorithms, and lack of accuracy for application to real-life images which differ from the training dataset. To tackle these issues, this paper is aiming to adjust pre-trained machine learning models, which are ResNet50 and VGG19 respectively, while also training and tuning a new SqueezeNet inspired model to create a flower recognition model that is able to process and remember large amounts of flower species data. From the research carried out, VGG19 was discovered to have the best performance on both the 5 Categories and Flower-102 dataset, with an accuracy of 88 percent and 84 percent respectively.

Keywords: VGG19, Transfer Learning, Deep Learning, Flower Recognition, Neural Network

1.0 INTRODUCTION

The world has approximately 369,000 named flowering plant species (Liu *et al.*, 2016). In general, experienced plant taxonomists are able to recognize plants based on the flower characteristics such as sepals, petals, stamens, and carpels. However, distinguishing these flowers is difficult for most people. Furthermore, similar flower species can lead to confusion. This is where object recognition comes in, as it is able to understand and analyse images effortlessly and instantaneously (Ong *et al.*, 2021). Therefore, the main objective of this project is to create a flower recognition model that is able to accurately recognise the class of flower in an image. This model is able to analyse the input image and determine the specific type of flower in the image. The end-goal is to train a computer to do what comes naturally to humans, which is to understand what is included in an image and gain insight from it. This project is interested in both modifying existing pre-trained machine learning models and training a flower recognition model based upon SqueezeNet. It is critical to both pursue novel and innovative methods, and also to conduct thorough research on existing methods in order to gain new insights and make new discoveries.

2.0 LITERATURE REVIEW

Lv *et al.* (2021) proposed a model to do flower classification with saliency detection and VGG-16 deep neural network model that is trained on the Oxford Flower-102 data set. An optimization algorithm of stochastic gradient descent was done to reduce computing consumption and training time to improve the model. In order to reduce model overfitting, a dropout method was used by randomly removing training information. To deal with the issue of insufficient image data and also to reduce model training time, transfer learning methods were used and an accuracy of 91.9 percent was achieved, which demonstrates better results than other conventional methods for image classification tasks and proves the possibility of flower identification using this model.

Cibuk *et al.* (2019) used pre-trained DCNN models, AlexNet and VGG-16, for feature extraction and concatenated features from both models to construct efficient feature sets. For the feature selection algorithm, the minimum Redundancy Maximum Relevance (mRMR) model was implement. The extracted features were then used to classify the flower species using a support vector machine (SVM) classifier with a Radial Basis Function (RBF) kernel. Their experimental results showed that they were able to achieve an accuracy performance of 96.39 percent and 95.70 percent for the Flower-17 and Flower-102 datasets, respectively.

Feng, Wang, Zha and Cao (2019) proposed the approach of using transfer learning and Adam deep learning optimization algorithm to fix the defects of current mainstream CNN, which are deep depths, long parameters, long training time and slow convergence. A modified and supplemented VGG-16 model was used, and the transfer learning method and Adam optimization algorithm are used to accelerate the network convergence. They used partial sets of images of the Flower-102 dataset combined with the Flower-17 dataset to form 30 sets of images, which are then randomly divided with Stratified Shuffle Split. With this they were able to obtain a 98.99 percent accuracy on their test set, while maintaining fast convergence.

Liu *et al.* (2016) proposed a flower classification approach which uses a convolutional neural network to extract features. They have also obtained the luminance map which is created

by converting RGB pixels to YUV, and the brightness of the colour is extracted from the Y component, which allows better performance as flowers have high brightness. They compute a bottom-up saliency map using a regional contrast-based salient object detection algorithm, which simultaneously evaluates global contrast difference and spatial weighted coherence scores. The algorithm is simple, efficient, and naturally multi-scale, and it generates full-resolution, high-quality saliency maps, which improves performance. They achieved an accuracy of 76.54 percent in their dataset and 84.02 percent in the Oxford Flower-102 dataset.

2.1 Existing Method

SqueezeNet is a novel convolutional neural network notable for having 112 times fewer parameters than another CNN, Alexnet, while also being able to maintain an accuracy top-5 performance comparable to that of AlexNet (Iandola *et al.*, 2016). Because it is such a small mode, SqueezeNet is more suitable for on-chip implementations on FPGAs (Iandola *et al.*, 2016). The SqueezeNet model has been studied in various use cases, and the results have been promising. Sayed, Soliman, and Hassanien (2021) used a SqueezeNet model optimised with a bald eagle search (BES) optimization to find the best hyperparameter to predict melanoma skin cancer on ISIC 2020 and ISIC 2019. The proposed melanoma skin cancer prediction model obtained an overall accuracy of 98.37 percent, specificity of 96.47 percent, sensitivity of 100 percent, f-score of 98.40 percent, and area under the curve of 99 percent. The results showed the robustness and efficiency of the proposed model compared with VGG-19, GoogleNet, and ResNet50. Therefore, a SqueezeNet-inspired model was chosen to be explored.

ResNet, which is an abbreviation for Residual Networks, enables engineers to train hundreds or even thousands of layers while still achieving impressive results (He *et al.*, 2016). The model managed to win the ImageNet challenge in 2015. It has been discovered that increasing training error in deep neural networks is caused by the network's initialization, optimization function, or one of the most well-known problems, the vanishing gradient problem (He *et al.*, 2016). It is an issue that occurs during the training of artificial neural networks with gradient-based learning and backpropagation. Gradients are known and can be used to update the weights in a network during backpropagation. However, the gradient can become increasingly small at times, thereby functionally preventing the weights from changing values. Because the same values are propagated over and over again, the network stops training, resulting in no useful work being done. To solve such problems, residual neural networks are used. ResNet alleviates this vanishing gradient problem through employing skip connections, which works by adding the output of an earlier layer to a later layer (He *et al.*, 2016).

VGG is a novel object-recognition model with support for up to 19 layers (Simoyan and Zisserman, 2015). It is pre-trained with ImageNet datasets and is still able to outperform with other unseen datasets, which makes it one of the most used image recognition architectures. The VGGNet has several variants, like the VGG-16 and VGG-19 variants, which differ only in the total number of layers in the neural network. Several studies have been conducted using the VGG-19 model, with impressive results. Victor Ikechukwu *et al.* (2021) conducted research using ResNet-50, ChexNet, VGG-19, and their own proposed Iyke-Net models to identify pneumonia from chest x-ray images, where VGG-19 achieved a high accuracy of 93.5 percent, coming in close second after their proposed Iyke-Net which is 93.6 percent accurate.

3.0 METHODOLOGY

This section explains and elaborates on the datasets used and the design of the models.

3.1 Datasets

The first dataset is the Kaggle flower recognition dataset, which includes 4242 images from Flickr, Google Images, and Yandex Images (Mamaev, 2021). Daisy, tulip, rose, sunflower, and dandelion images are divided into five categories. Each class has about 800 images, with each image measuring about 320x240 pixels. The photos are not reduced to a single size, but rather come in a variety of sizes. It is henceforth referred to as the 5-category dataset.

The Oxford Flower-102 dataset, which consists of 102 flower categories, is also employed (Nilsback & Zisserman, 2008). This dataset is even more specific than the 5-category dataset in that the category is based on flower species specified in their scientific name, instead of a general flower category like “daisy”. The images depict flowers that are common in the United Kingdom. Each class contains between 40 and 258 images, with varying scales, poses, and lighting. The difficulty of classification is exacerbated by large variations within the same category and several very similar categories. There are 8189 images in the dataset. This dataset is henceforth referred to as the 102-category dataset.

3.2 Model Architecture

The first model developed is based upon the original SqueezeNet model (Iandola *et al.*, 2016). It utilises the Fire module architecture as suggested by the original developers, consisting of a squeeze convolution layer of only 1x1 filters, which feeds into an expand layer that has a mix of 1x1 and 3x3 convolution filters. The abundant use of 1x1 filters greatly reduces the parameters by 9 times compared to using 3x3 filters. The parameters in the layer are further reduced by using the squeeze layers, which reduces the number of input channels to 3x3 filters. Therefore, SqueezeNet is capable of achieving a size more than 50 times smaller than the AlexNet model while still achieving a reasonable accuracy. This particular implementation is a stripped down version of SqueezeNet with fewer layers which consists of one input into a conv2d layer, followed by batch normalisation, the first fire module, the first MaxPooling2D layer, the second fire module, the second MaxPooling2D layer, the third fire module, the first GlobalAveragePooling2D layer and the final Dense layer with softmax activation to obtain the categories to be predicted.

With comparisons to different ResNet variants like ResNet18, ResNet34, ResNet101 and ResNet152, ResNet50 is chosen for transfer learning because of its lower requirement of computational power and encouraging accuracy (He *et al.*, 2016). Firstly, the pre-trained ResNet50 model (resnet50 weights tf dim ordering tf kernels.h5) is downloaded from Github (Fchollet, 2016) and which uses the weights already pre-trained from the imagenet datasets. The first layer of ResNet50 is frozen and made untrainable to avoid taking a long time during the training

process by having lesser trainable parameters. Extra dense layers shall also be added to the final layers of the pre-trained network, which allows learning the combinations of the previously learned features which are useful in recognising new objects in the new dataset. Therefore, additional layers were implemented after the output of the ResNet50 model. A total number of 7 layers were added in this project including the flatten layer, batch normalization layers, two customised layers, a Rectified Linear Unit (ReLU) activation function layer and the softmax layer. The flatten layer converts data into a 1-dimensional array for input to the next layer, and batch normalisation is a layer that allows each layer of the network to learn more independently. Finally, a softmax layer is included as the output layer in order to predict fixed types of flowers whereby 5 or 102 classes of classification are produced, depending on whether it is for the 5-category or 102-category dataset. The 5 classes are created in order to predict the 5 types of flowers for the 5-category dataset, which are daisy, tulip, rose, sunflower and dandelion. So, after applying this layer, a transfer learning model that can classify the input images into various types of flowers based on predictions from the pre-trained ResNet50 model is developed.

VGG-19 is composed of 16 convolutional layers, 5 pooling layers, 3 fully-connected layers and a final layer of softmax function (Simonyan & Zisserman, 2015). The matrix was shaped (224, 224, 3) as the fixed input size of 224x224 RGB image is passed into this network. Small kernel size of 3x3 with stride of 1 pixel is used to cover every part of the image rather than using a large kernel size, whereas a 2x2 pixel window with a stride of 2 pixels is used to perform max pooling. The multiple layers of small kernels are able to effectively cover the images without the use of large kernels such as 11x11 kernel in AlexNet and 7x7 kernel in ZFNet. Therefore, the number of parameters and the overfitting problem is mitigated. All hidden layers are equipped with ReLU which uses tanh or sigmoid functions to introduce non-linearity for better classification compared to the previous models. In this project, the pre-trained weights are used by setting the parameter weights to the one trained with imagenet. The first 19 layers are frozen to prevent the weights from being modified and similarly to the aforementioned ResNet50 model, additional layers are added after the pre-trained model. A max-pool layer is added to downsample the input features. A flattening layer is then added before a dense layer with a softmax function since the dense layer accepts 2D input.

3.3 Logical Flow

Data Loading: The flowers dataset consists of images of flowers with different class labels and stored in respective directories. For the 5-category dataset, there are 4242 images and 5 class labels. For the 102-category dataset, there are 8189 images and 102 class labels. This stage loads the data from their directories and concatenates it into a single dataframe. As a result, the flower images end up with an image dimension of 244 x 244, which minimises image dimensions while maintaining image readability with efficient computational complexity and accommodating the input shape for the pre-trained models.

Data Understanding: The flowers dataset contains examples of labelled flower images. Each example includes a JPEG flower image as well as the class label. The exploration of the image data helps to authenticate the class distribution of each type of flower in the dataset and ensures a balanced dataset, in order to an imbalanced dataset that can lead to poor predictive power. Data visualization is needed to sample and study the input data to ensure image readability by randomly viewing the 10 images in the dataframe via 2D representation with the selected new image dimensions.

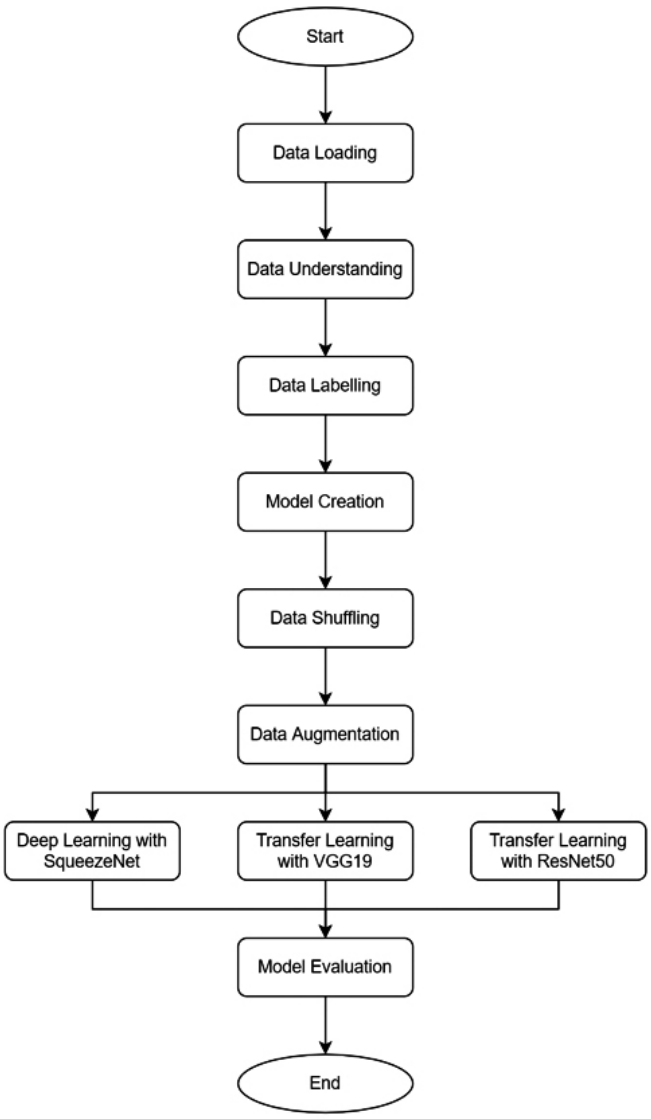


Figure 1: Logical Flowchart

Data Labelling: This stage is used to transform categorical data (textual data) into numerical values for the prediction functions so that the deep learning predictive models can understand the input data. Label encoding technique is performed to convert categorical values to numbers in this step.

Model Creation: The same optimizer and loss function is used to compile all models. Cross-entropy is the loss function chosen to evaluate a set of weights in multi-class classification problems for flower recognition. Furthermore, the Adam optimizer with the learning rate of 0.01 is used to search through different weights for the network. Finally, because this is a classification problem, classification accuracy is collected and reported, which will be defined via the metrics argument.

Data Shuffling: This stage is used to redistribute the training and testing data samples in the dataset to ensure that each data sample produces an "independent" change on the model, without being influenced by the points that came before it. Data shuffling

is required since the images were added sequentially from subfolders into the dataframe during data loading Otherwise, the model can only learn what is "daisy" from the first 800 images, which does not optimise the model's parameters. The seed is set to 100 and is applied to both images and labels to ensure that each image matches the correct label.

Data Augmentation: This stage is used to increase the amount of data by adding slightly changed copies of already existing data or newly created artificial data from existing data. This refers to randomly changing the images in ways that shouldn't impact their interpretation, such as horizontal flipping, zooming, and rotating. Through data augmentation, this stage regularises the images and prevents overfitting problems when training a machine learning model. This technique is used to overcome the problem of overfitting by creating more data and making the model generalise well on the unseen data.

Model Training: The model is using the predefined train-test split for the Oxford Flower-102 dataset, with 1020 training samples and 6149 testing samples. The 5-category dataset is split into train-test sets using a ratio of 80:20. Different batch sizes and epochs are adjusted in different models in order to achieve the optimal result. This stage generates tensor image data in batches, which will be looped over for both training and testing. The neural network in each model takes in inputs, which are then processed in hidden layers using weights that are adjusted during training. Then the model spits out a prediction. The weights are adjusted to find patterns in order to make better predictions.

Model Evaluation: In this process, validation accuracy measures were calculated after the model had gone through all the data. The network had been fully trained when these scores were calculated. The final model selection is ultimately based on validation accuracy.

4.0 RESULTS AND DISCUSSION

This section shows the results obtained for each model, on both datasets. The metric used is the validation accuracy, because metrics such as training accuracy do not accurately reflect real-world performance. The validation accuracy is calculated by taking the True Prediction/Total Number of Predictions using the validation dataset. The models were validated with N = 1020 samples for the Oxford Flower-102 dataset and N= 848 samples for the 5-category dataset.

Table 1: Validation accuracy for each model for the two datasets

	Validation Accuracy (5 Category)	Validation Accuracy (102 Category)	Status
SqueezeNet Inspired Model	77%	67%	Rejected
ResNet50	67%	42%	Rejected
VGG19	88%	84%	Accepted

According to our findings, the partially pre-trained VGG-19 model performed the best on both datasets, achieving a high validation accuracy of 88 percent on the 5 Categories dataset and 84 percent on the Flower-102 dataset. This could be due to the VGG-19's simple structures and hidden layers with ReLU functions, which can better introduce non-linearity for better classification than other models. The model's generalisation improves as the number of features produced decreases.

The SqueezeNet Inspired model had the second-highest validation accuracy of 77 percent and 67 percent on the 5 Categories and Flower-102 dataset respectively. This might be due to the fact that while this SqueezeNet Inspired model is not quite as complex and has less layers and performance than both the other models, this model is not pre-trained. Instead, the whole model was trained solely relying on the two datasets individually for each scenario, and therefore was able to perform better than the partially pre-trained ResNet50. This is a lightweight model, even when compared to the already lightweight original SqueezeNet implementation, and therefore it should be expected performance was sacrificed for it's small size.

Finally, the ResNet50 model performs the worst, with 67 percent and 42 percent on the 5 Categories and Flower-102 datasets, respectively. This could be because the architecture of the ResNet50 model is overly complicated for this task, resulting in poor generalisation. The optimal learning rate, batch size, and identification of the best freezing layer all play important roles in the model's performance.

5.0 CONCLUSION

The objectives to create a flower recognition model that is able to accurately recognise the class of flower in an image have been met because a model based on VGG19 has been developed that can perform flower classification tasks with 88 percent accuracy with 5 flower categories and 84 percent accuracy with 102 flower species. However, there are some limitations in this research, such as hardware limitations. This project's algorithms and models are all computationally intensive due to the complexity of the models and the size of the dataset. This project is also more concerned with which models perform better on the two flower datasets, rather than why the models perform the way they do. Other pre-trained models such as AlexNet, VGG16 and so on, which were not used in this project, are another potential avenue for investigation. More comprehensive studies on other models can be conducted in the future.

6.0 ACKNOWLEDGMENTS

Authors thank the Faculty of Computing and Information Technology, Tunku Abdul Rahman University College for financial support and resources to carry out this project. ■

REFERENCES

[1] Cıbuk, M., Budak, U., Guo, Y., Cevdet Ince, M. and Sengur, A., (2019) 'Efficient deep features selections and classification for flower species recognition'. Measurement, 137, pp.7-13.

[2] Fchollet, (2016). Release VGG16, VGG19, and ResNet50 · fchollet/deep-learning-models., GitHub. Available at: https://

github.com/fchollet/deep-learning-models/releases/tag/v0.1 (Accessed: 10 September 2021)

[3] Feng, J., Wang, Z., Zha, M. and Cao, X. (2019) 'Flower Recognition Based on Transfer Learning and Adam Deep Learning Optimization Algorithm'. Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence - RICAI 2019.

[4] He, K., Zhang, X., Ren, S., Sun, J. (2015) 'Deep Residual Learning for Image Recognition' 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778

[5] Iandola, F.N., Moskewicz, M.W., Ashraf, K., Han, S., Dally, W.J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. ArXiv, abs/1602.07360.

[6] Liu, Y., Tang, F., Zhou, D., Meng, Y. and Dong, W. (2016) 'Flower classification via convolutional neural network'. 2016 IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA), Qingdao, China, pp. 110-116.

[7] Lv, R., Li, Z., Zuo, J., Liu, J. (2021) 'Flower Classification and Recognition Based on Significance Test and Transfer Learning'. 2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021. Institute of Electrical and Electronics Engineers Inc., pp. 649–652.

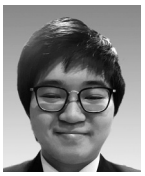
[8] Mamaev, A. (2021). Flower Recognition, Kaggle [Online]. Available at: https://www.kaggle.com/alxmamaev/flowers-recognition (Accessed: 19 April 2021)

[9] Nilsback, M., Zisserman, A. (2008). 102 Category Flower Dataset, Visual Geometry Group [Online]. Available at: https://www.robots.ox.ac.uk/~vgg/data/flowers/102/ (Accessed: 19 April 2021)

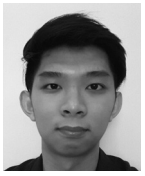
[10] Simonyan, K., Zisserman, A. (2015) 'Very deep convolutional networks for large-scale image recognition', 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, International Conference on Learning Representations, ICLR, San Diego, California, United States of America.

[11] Ong, ZY., Chye, KK., Kang, HW., Tan, CW. (2021) 'A Flower Recognition System using Deep Neural Network Coupled with Visual Geometry Group 19 Architecture', International Conference On Digital Transformation And Applications 2021, ICDXA 2021, pp. 121-128 Kuala Lumpur, Malaysia.

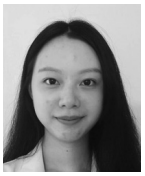
PROFILES



ONG ZI YUAN received his Bachelor's degree of Computer Science (Honours) in Data Science from Tunku Abdul Rahman University College (TAR UC) in 2022. He is a fresh graduate who is passionate about Artificial Intelligence (MI), Machine Learning (ML) and Data Science. His research interests include artificial intelligence, natural language processing, image processing, data analytics as well as computer vision.
Email address: ongzy-wp17@student.tarc.edu.my



CHYE KAH KIEN graduated from Tunku Abdul Rahman University College (TAR UC) in 2022 with a BCompSc (Hons) (Computer Science in Data Science). He is a recent graduate who is passionate about applying data science and machine learning skillsets to SAP systems, an enterprise software to manage business operations and customer relations in which 77% of the world's transaction revenue touches an SAP system. His research interests include natural language processing, image processing, data mining, process mining, as well as advanced analytics.
Email address: chyekk-wp17@student.tarc.edu.my



KANG HUAY WEN received BCompSc (Hons) Computer Science in Data Science from Tunku Abdul Rahman University College (TAR UC) in 2022. She is passionate about Artificial Intelligence (AI) and Data Science and aspires to be a professional data scientist. Her research interests involved Data Analytics, Code Mixed Sentiment Analysis and Computer Vision.
Email address: kanghw-wp17@student.tarc.edu.my

DR TAN CHI WEE received BCompSc(Hons) and PhD degrees in year 2013 and 2019 respectively in Universiti Teknologi Malaysia. Currently, he is a Senior Lecturer cum Programme Leader at Tunku Abdul Rahman University College and actively involved in the Centre of Excellence for Big Data and Artificial Intelligent (CoE) and become the research group leader for Audio, Image and Video Analytics Group under Centre for Data Science and Analytics (CDSA). Dr Tan's main research areas are Computer Vision (CV), Image Processing (IP) and Natural Language Processing (NLP) and Artificial Intelligence (AI). He is an enthusiastic researcher experienced in conducting and supporting research into Image Processing. Being a meticulous and analytical researcher with Train-The-Trainer certificate of many years of educational and hands-on experience, he was invited to Université d'Artois (France) under Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) programme for collaborative research between European countries with Southeast Asian countries on motion detection and computer vision and being involved in industry project as professional consultant.
Email address: chiwee@tarc.edu.my

SENTIMENT ANALYSIS ON MALAYSIAN AIRLINES WITH BERT

(Date received: 13.01.2022/Date accepted: 03.04.2022)

Huay Wen Kang[1], Kah Kien Chye[1], Zi Yuan Ong[1], Chi Wee Tan[1]*

[1]Faculty of Computing And Information Technology, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300, Wilayah Persekutuan Kuala Lumpur, Malaysia.

**Corresponding author: chiwee@tarc.edu.my*

ABSTRACT

Sentiment analysis has been a popular research area in Natural Language Processing (NLP), where sentiments expressed through text data including positive, negative and neutral sentiments are analyzed and predicted. It is often performed to evaluate customer satisfaction and understand customer needs for businesses. In the airline industry, millions of people today use social networking sites such Twitter, Skytrax, TripAdvisor and more to express their emotions, opinions, reviews and share information about the aircraft service. This creates a treasure trove of information for the airline company, showcasing different points of views about the airline's brand online and providing insightful information. Hence, this paper experiments with six different sentiment analysis models in order to determine and develop the best model to be used. The model with the best performance was then used to determine the social status, company reputation, and brand image of Malaysian airline companies. In conclusion, the BERT model was found to have the best performance out of the six models tested, scoring an accuracy of 86 percent.

Keywords: Supervised Learning, Ensemble Learning, Deep Learning, Transfer Learning, Airline Sentiment

1.0 INTRODUCTION

Social media is a huge source of opinionated text data that comes in various forms such as Facebook posts, tweets, comments, replies, product reviews and others. The Internet era has altered how people express their thoughts and opinions. It is now primarily done through blog posts, online forums, product review websites, social media, and so on. It is extremely difficult to manually track multiple social media platforms filled with this data for large airlines, hence this is where natural language processing comes in useful. As a result, the goal of this project is to conduct research and analysis on Malaysian airline companies' social status, company reputation, and brand image, specifically Malaysia Airlines, AirAsia, and Malindo. In order to develop a sentiment analysis model about airlines, various methods and techniques were explored including supervised learning, ensemble learning, deep learning, and transfer learning in this research work.

2.0 LITERATURE REVIEW

Sousa *et al*. (2019) has predicted the following movements of the Dow Jones Industrial (DJI) Index in the task of stock market analysis to evaluate BERT. 582 financial news were crawled from various news websites such as CNBC, Forbes and New York Times using Selenium tool to serve as the dataset and it was manually annotated with sentiment classes such as positive, negative and neutral. WordPiece is used to tokenize and transform each document into a token sequence. Then, the parameter for attention head was set to BERT BASE as it is smaller and is able to cater to their limited computational power. A 10-fold cross validation was done before training the models using labeled data. Based on the results, BERT clearly outperformed other methods compared to Support Vector Machines, Naive Bayes and Convolutional Neural Network with the highest F1-score of 72.5 percent. This analysis was used as an indicator of falling and rising of the economy in the day.

Zhang, Wang and Liu (2018) give a detailed overview on deep learning and also a comprehensive survey of the current applications of deep learning in sentiment analysis. Through their analysis, they have concluded that many deep learning techniques have shown state-of-the-art results for sentiment analysis tasks. This paper touches upon the concept of Neural Networks, Deep Learning, Word embedding, Autoencoder and denoising autoencoder, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory Network (LSTM), Attention Mechanism with RNN, Memory Network (MemNN) and Recursive Neural Network (RecNN). This paper also surveys the various kinds of sentiment analysis tasks, like document-level sentiment classification, sentence-level sentiment classification, aspect-level sentiment classification, aspect extraction and categorization, sentiment composition, opinion holder extraction, temporal opinion mining, sentiment analysis with word embedding, sarcasm analysis, emotion analysis, the use of multimodal data for sentiment analysis, and resource-poor language and multilingual sentiment analysis.

Agarwal, Xie, Rambow and Passonneau (2011) proposed a method of analysis sentiment for Twitter data. They have prepared an emoticon dictionary by labeling 170 emoticons listed on Wikipedia with their emotional state, such as "Extremely-positive", "Extremely-negative", "Positive", "Negative" and "Neutral". An acronym dictionary was created for 5,184

acronyms. They consider all words found inside WordNet as English words, while the rest are stop words or non-English words. A combination of the Dictionary of Affect in Language (DAL) and WordNet to get the polarity of words. Several models were investigated, including a Unigram model, a senti-features model, a tree kernel model, a unigram + senti-features model and tree kerna + senti-features model. In the feature-based model, feature analysis was done and revealed that the most important features are those that combine prior polarity of words and their parts-of-speech tags. The unigram model, senti-features, tree kernel, unigram + senti-features and kernel + senti-feature models managed an average accuracy of 56.58 percent, 56.31 percent, 60.60 percent, 60.50 percent and 60.83 percent respectively.

2.1 Existing Method

Naive Bayes is a classifier family based on Bayes' popular probability theorem that is well suited for creating simple but powerful models, particularly in the area of textual classification (Lewis, 1998). Compared to many models, Naive Bayes do not need much data to train and tune the required parameters (Abbas *et al.*, 2019). Therefore, it is less complicated to implement, and is more reliable in comparison to more complex and slower algorithms. The Naive Bayes Classifier has multiple variations, which are Multinomial Naive Bayes (MNB), Bernoulli Naive Bayes (BNB) and Gaussian Naive Bayes (GNB). A Naive Bayes classifier is essentially a probabilistic machine learning model for classification that employs the Bayes theorem, which describes the probability of an event based on prior knowledge of the conditions that may be associated with the event. Multinomial Naive Bayes was selected instead of the other Naive Bayes types because MNB is an improved version of Multivariate Bernoulli Naive Bayes model (BNB) and considers word frequency and information and thus obtains better accuracy.

Support Vector Machine (SVM) (Cristianini & Shawe-Taylor, 2000) is a learning technique that excels at sentiment analysis because it can significantly reduce the need for labelled training instances in both the standard inductive and transductive settings (Phienthrakul, 2009). SVM creates that hyperplane by transforming data using mathematical functions known as "Kernels." Types of Kernels are linear, sigmoid, RBF, non-linear, polynomial, etc. The kernel function used in SVM determines its performance. As a result, selecting the appropriate kernel will improve classification efficiency. Previously, research on sentiment classification using SVM and multiple kernel functions yielded promising results with high accuracy. However, in order to adapt to new changes and be more flexible in today's world, SVM has evolved into various versions. Other researchers had proposed various extensions of SVM such as support vector classifier, transductive support-vector machines, multiclass SVM, structured SVM, bayesian SVM, and others during their research. Hence, it is possible that other combination methods may improve sentiment classification efficiency compared to conventional SVM. The support vector classifier is then used in this research based on this idea.

The Random Forest algorithm (Breiman, 2001) consists of a large number of individual decision trees that operate in an ensemble manner, resulting in a forest of trees. Rather than using best split among all variables to split each node, Random Forest randomly selects a subset of predictors and narrows down the best among them. The tree grows with randomly selected features and is not pruned, resulting in the Random Forest algorithm having great accuracy (Breiman and Cutler, 2004). Random Forest algorithm requires each individual tree to have low correlation with each other to perform well. Random Forest has been proven to be a suitable classifier to perform sentiment analysis in multiple research projects. Based on a sentiment analysis research on Malaysian mobile digital payment applications done by Balakrishnan *et al.* (2020), Random Forest achieved the highest accuracy of 75.62 percent and f1-score of 71.99 percent amongst other algorithms including Support Vector Machine, Naive Bayes and Decision Tree. Likewise, Hedge and Padma (2017) also managed to obtain an accuracy of 72 percent when using Random Forest to perform sentiment analysis for mobile product reviews in Kannada.

Compared to a single model, the ensemble method improves predictive performance by combining the results of several models. This eliminates the possibility of overfitting while improving overall performance (Yu *et al.*, 2010). Furthermore, ensemble methods are frequently used to solve the class imbalance problem using multi-objective optimization algorithms (Yang *et al.*, 2020) and to minimise the number of features better than existing ensemble algorithms such as AdaBoost and Gradient Boosting. In Wan and Gao's (2016) experiment with an ensemble sentiment classification system for airline service analysis, the ensemble classifier using a majority vote method with five classifiers achieved the highest accuracy of 84.2 percent when compared to other single classifiers.

The Recurrent Neural Network (RNN) (Marhon *et al.*, 2013) family of models, particularly the LSTM networks, has been demonstrated to be the most effective sequence model used in practical applications (Goodfellow, Bengio and Courville, 2016). To learn from both forward and backward time dependencies, Bi-directional LSTMs are used. Each unit in a Bi-derectional LSTM is divided into two units that share the same input and are connected to the same output. One unit is used for the forward time sequence, while the other is used for the reverse time sequence. As a result, Bi-directional LSTM is useful for learning from long-spanning time-series data, and it produces better results without increasing training time.

In 2018, the Bidirectional Encoder Representations from Transformers (BERT) (Devlin, Chang, Lee, & Toutanova, 2019) model was introduced to quickly and effectively create a high-quality model with minimal effort and training time using the PyTorch interface, regardless of the specific NLP task, and produce state-of-the-art results. Recently, a sentiment analysis using the BERT model on the impact of coronavirus on social life achieved 94 percent validation accuracy on the collected data sets (Singh *et al.*, 2021). In short, BERT is one of the most powerful NLP models available today, requiring only a small amount of data while achieving cutting-edge results with minimal task-specific adjustments for a wide range of NLP tasks such as named entity recognition, language inference, semantic similarity, question answering, and classification like sentiment analysis.

3.0 METHODOLOGY

This section talks about the datasets used and the design of the models.

3.1 Datasets

The training and testing dataset are sentiment datasets about US airlines (US Airways, United, Virgin America, Delta, Southwest, and American) from Twitter which can be obtained from Kaggle. There is a total of 9,178 negative tweets, 3,099 neutral tweets, and 2,363 positive tweets. For model training, it is divided into a train and test set with an 80:20 ratio. The models trained with this dataset are expected to be generalizable and capable of handling all airline reviews, regardless of company or country, as long as they are all written in English. The following are the 13 columns from the mentioned dataset:

1. tweet_id: Unique ID of the tweet
2. airline_sentiment: Sentiment of the tweet
3. airline_sentiment_confidence: Confidence level of the sentiment of the tweet
4. negativereason: Negative reason if sentiment is negative
5. negativereason_confidence: Confidence level of the negative reason if sentiment is negative
6. airline: Name of the airline
7. name: Username of the tweet author
8. retweet_count: Number of retweets
9. text: Text data of the tweet
10. tweet_coords: Location coordinates of the person who posted the tweet
11. tweet_created: date and time of the tweet posted
12. tweet_location: Location of the person who posted the tweet
13. user_timezone: Timezone of the person who tweeted

Table 1: Sample Data for Malaysian Airlines

	Username	Review	Airline
0	Ritesh Aggarwal	Booked flight from Delhi to Sydney for my pare...	mas
1	Tara-lets T	I was looking for affordable airfare for multi...	mas
2	w KING	I booked a flight from Delhi to KL on the 2 Ju...	mas
3	paul	Booked a flight with these thieves from Jakart...	mas
4	Mickey Pearson	With the E-Voucher Malaysia Airlines is forcin...	mas

3.2 Text Preprocessing

Data preprocessing was done on the training data before being used to train the models for traditional machine learning methods. The steps are as follows:

1. Emoticons and emojis are converted to text that expresses their textual meaning using the emoticon library (Shah & Rohilla, 2018) as emoticons and emojis carry sentiment.
2. All words in the message are converted into lowercase in order to normalize the text. To reduce noise and ease further process, symbols, punctuations, hyperlink, extra whitespace, new lines along with digits are removed.
3. Stopwords are removed and lemmatization is done to return the words back to their root word since the word would be disrupted by an irrelevant inflection like a simple plural or present tense inflection. On the other hand, BERT does not require text preprocessing such as stopword removal, lemmatization and others because the model has a specific way of dealing with out-of-vocabulary words using its own fixed vocabulary and the BERT tokenizer for text formatting.
4. Feature selection is done to transform the text into unigram word vectors using CountVectorizer.
5. Following that, resampling techniques are used to address the data imbalance between positive, neutral, and negative sentiment. This is due to the fact that class imbalance may have an impact on the accuracy and performance of the models. As a result, random undersampling is used to randomly reduce the majority class to the desired ratio against the minority class, and it is combined with SMOTE oversampling (Chawla *et al.*, 2002), which is oversampling by creating "synthetic" examples of the minority class. Both undersampling and SMOTE use the default sampling strategy, which is "not minority" for undersampling and resamples all classes except the minority, and "not majority" for SMOTE, which oversamples all classes except the majority. Both will bring the unequal classes into balance. The training dataset will have N = 21,867 samples from the initial 11,712 samples at the end of the process.

3.3 Model Applications

Naive Bayes (NB) is based on the Bayes theorem, and can be used for classification challenges. Multinomial Naive Bayes (MNB) is one probabilistic learning method variant of NB which is commonly used in Natural Language Processing (NLP). MNB performs calculation of probabilities based on probabilities of causal factors and is useful to model feature vectors where each value represents the number of occurrences of a term, where in the case of NLP a text can be considered as a particular instance of a dictionary and the relative frequency of all terms in the text provide enough information for inferring a belonging class. To avoid the possibility of getting zero probability, additive laplace smoothing parameter is used as one of the hyperparameters, whether to learn class prior probabilities or not and prior probabilities of the classes. Lower alpha values are used to avoid getting a likelihood of around 0.5, which will not be helpful for the results. The optimized set of parameters for the model after hyperparameter tuning: `MultinomialNB(alpha=0.5)`.

A simple, linear Support Vector Classifier (SVC) is proposed to classify the airline data into three different classes. SVC is a different implementation of the SVM algorithm which is implemented in terms of liblinear rather than libsvm (Asif *et al.*, 2020). It is just a thin wrapper around libsvm but has more options for penalties and loss functions and has higher scalability to larger samples. It supports both dense and sparse inputs and also handles multi classes. The classifier will consider each unique word present in the sentence, along with multiple word expressions which are suitable for text classification. A linear hyperplane is used to identify the text involved in the airline dataset and separate them into three categories by implementing this mechanism. Common features such as gamma are set at 0.5, regularisation parameter (c) of 100, and the use of probability estimates in conjunction with a balanced mode after parameter tuning. The tuned parameters for this module are as follows: `SVC (C=100, class_weight='balanced', gamma=0.5, kernel='linear', probability=True)`.

The hyperparameters of the Random Forest algorithm include the number of decision trees in the forest, number of features considered by each tree when splitting a node, number of

levels in each decision tree, min number of data points allowed in a leaf nod, min number of data points placed in a node before the node is split, and number of trees in the forest. The parameters of the random forest are the variables and thresholds used to split each node learned during training. The line below shows the optimized parameters for the model after hyperparameter tuning: `RandomForestClassifier(max_depth=100, max_features='sqrt', min_samples_leaf=2, min_samples_split=10, n_estimators=800).`

For the voting classifier, the previous three algorithms (Multinomial Naive Bayes, Random Forest and LinearSVC) with their respective optimized parameters set are passed into the classifier and used soft voting, which predicted the class with the highest summed probability from models.

Bidirectional LSTMs are an extension of traditional LSTMs in which two LSTMs are trained instead of one on the input sequence, with the second LSTM trained on the reversed copy of the input sequence, providing additional context to the network and resulting in faster and even more complete learning on the problem. The data was preprocessed before being fed into the Bidirectional LSTM. To summarise, the sentiments were converted into a binary class matrix before tokenizing the text, converting it to an encoded form, and finally padding all of the encoded text to the same length. The proposed Bidirectional LSTM model is simple, with only 7 layers, including one embedding input layer and one embedding layer, followed by one Convolutional layer, one max pooling layer, one bidirectional LSTM layer, one dropout layer, and finally one dense layer to generate 3 classification classes.

BertForSequenceClassification model with an added single linear layer on top for classification that serves for sentence classifier purposes is used in this project. After the airline data sets are fed in, the entire pre-trained BERT model and the additional untrained classification layer is trained on the specific task, that is the multi-class classification. With the exception of the batch size, learning rate, and number of training epochs, most model hyperparameters remain at their default values for fine-tuning. The author discovered a set of feasible values that work well across all NLP tasks, so the batch size is set to 32, with a 2e-5 learning rate (Adam) from the recommended range values and a maximum epoch value of 4. In addition, the model will be trained using the PyTorch framework. The model will go through the standard PyTorch training cycle and loop four times, iterating through the mini-batches, performing a feedforward pass for each batch, computing the loss, performing backpropagation for each batch, and finally updating the gradients. In short, the power of the pre-trained transformers (BERT) model and the PyTorch framework will be leveraged by this fine-tuned model.

4.0 RESULTS AND DISCUSSION

The following table shows the results of all models. Unigram features were used for Multinomial Naive Bayes, Linear Support Vector Classifier, Random Forest and Voting Classifier. The testing dataset has N = 2,928 samples, split from the US Airline dataset.

Among the traditional machine learning techniques (Multinomial Naive Bayes, Linear Support Vector Classifier and Random Forest), Multinomial Naive Bayes achieved the highest results. Hence it is further used to test using bigram and trigram features.

Table 2: Results for all models

	Accuracy	Precision	Recall	F1-Score	Status
Multinomial Naive Bayes	70.4%	77.7%	78.2%	77.9%	Rejected
Linear Support Vector Classifier	66.7%	74.0%	73.8%	73.9%	Rejected
Random Forest	66.5%	76.3%	77.3%	76.5%	Rejected
Voting Classifier	68.4%	79.2%	80.2%	78.7%	Rejected
Bidirectional LSTM	77.0%	77.0%	77.0%	77.0%	Rejected
BERT	86.0%	86.0%	86.0%	86.0%	Accepted

Table 3: Results for Multinomial Naive Bayes using different n-gram orders

	Accuracy	Precision	Recall	F1-Score
Unigram	70.4%	77.7%	78.2%	77.9%
Bigram	53.4%	64.9%	66.9%	65.6%
Trigram	41.4%	61.3%	65.8%	59.1%

According to the experimental results, the BERT model outperformed the other six models with an accuracy of 86.0 percent. It demonstrates that in airline sentiment analysis, transfer learning approaches outperform traditional machine learning algorithms. This may be due to that it is trained on available plain text corpora that are larger than reviews only, as it already encode a lot of information of generic English text as it was trained from Wikipedia and book corpus (in total 3,300 million words) (Devlin *et al.*, 2019), therefore allowing better performances. Experiments show that BERT outperforms unsupervised text classification, such as the standard preprocessing process, which includes decapitalization, punctuation removal, stopword removal, and emoji conversion to text (Kang *et al.*, 2021).

The Bidirectional LSTM was the second-best performer, with an accuracy of 77.0 percent. This result outperforms all traditional machine learning methods tested, proving the hypothesis that deep learning models outperform traditional machine learning algorithms. This could be because the Bidirectional LSTM considers both backward and forward time dependencies.

Among the various n-gram models tested for Multinomial Naive Bayes, unigram features produce the best results (70.4 percent). With each increase in n-gram, the accuracy decreases even further. This could be because as n-gram length increases, so does the frequency of any given n-gram, and it may not generalise well to a different data set, resulting in lower accuracy.

5.0 CONCLUSION

In conclusion, this paper investigated various sentiment analysis models and discovered that Multinomial Naive Bayes is the best machine learning model out of the four chosen, with an accuracy of 70.4 percent. It was also demonstrated that the Unigram model of the Multinomial Naive Bayes outperforms the Bigram

and Trigram models. On the other hand, deep learning methods outperform traditional machine learning methods, as evidenced by the results, with both the Bidirectional LSTM (77 percent accuracy) and BERT model (86 percent accuracy) outperforming all traditional machine learning methods. Finally, we discovered that the BERT model outperforms all six models tested.

A high-accuracy sentiment analysis model has been successfully developed and applied to a new airline reviews dataset. According to the crawled data, the majority of the 1371 reviews were classified as negative (N=1118, 81.55 percent), 226 positive (16.48 percent), and 27 neutral (1.97 percent). This suggests that the majority of Malaysian Airlines' customers are dissatisfied with the airline's services. Companies can get better directions and make better business decisions based on sentiment analysis insights, such as providing better training for their flight crews if negative sentiments toward their services are discovered.

However, the crawled data used in this project is relatively small; such a small and sparse dataset may not be truly representative of the larger population, and it also limits several types of data analysis, such as trend analysis, due to data scarcity. As a result, this project can be improved by crawling more data for the dataset and reducing the sparsity of the data, making it more representative of the actual population and allowing for more complex analysis.

6.0 ACKNOWLEDGEMENTS

Authors thank the Faculty of Computing and Information Technology, Tunku Abdul Rahman University College for financial support and resources to carry out this project. ■

REFERENCES

[1] Abbas, M., Memon, K., Jamali, A., Memon, S. and Ahmed, A. (2019). 'Multinomial Naive Bayes Classification Model for Sentiment Analysis'. International Journal of Computer Science and Network Security, 19(3), pp.62-67.

[2] Agarwal, A., Xie, B., Rambow, O. and Passonneau, R. (2011). 'Sentiment Analysis of Twitter Data'. Proceedings of the Workshop on Languages in Social Media.

[3] Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H. and Shah, J. (2020). 'Sentiment analysis of extremism in social media from textual information'. Telematics and Informatics, 48, p.101345.

[4] Balakrishnan, V., Selvanayagam, P.K., Yin, L.P. (2020). 'Sentiment and Emotion Analyses for Malaysian Mobile Digital Payment Applications'. ACM International Conference Proceeding Series. Association for Computing Machinery, pp. 67–71.

[5] Breiman, L. (2001). Machine Learning, 45(1), 5-32. https://doi.org/10.1023/a:1010933404324

[6] Breiman, L., Cutler, A. (2004). 'RFtools for Predicting and Understanding Data'. Interface Workshop 1–62.

[7] Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal Of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

[8] Cristianini, N., & Shawe-Taylor, J. (2000). 'An Introduction to Support Vector Machines and Other Kernel-based Learning Methods'. https://doi.org/10.1017/cbo9780511801389

[9] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv, abs/1810.04805.

[10] Goodfellow, I., Bengio, Y. and Courville, A. (2016). 'Deep Learning'. MIT Press.

[11] Hegde, Yashaswini & Padma, S.K. (2017). Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada. 777-782. IEEE 7th International Advance Computing Conference (IACC) 2017, Hyderabad, India. IEEE, 10.1109/IACC.2017.0160.

[12] Kang, HW., Chye, KK., Ong, ZY., Tan, CW. (2021). 'The science of emotion: Malaysian airline sentiment analysis using BERT Approach', International Conference On Digital Transformation And Applications 2021, ICDXA 2021, Kuala Lumpur, Malaysia.

[13] Lewis, D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. Machine Learning: ECML-98, 4-15. https://doi.org/10.1007/bfb0026666

[14] Marhon, S., Cameron, C., & Kremer, S. (2013). Recurrent Neural Networks. Intelligent Systems Reference Library, 29-65. https://doi.org/10.1007/978-3-642-36657-4_2

[15] Neel Shah & Shubham Rohilla (2018). Description of the emot library, v2.2. Github [Online]. Available at: https://github.com/NeelShah18/emot (Accessed: 19 April 2021)

[16] Phienthrakul, T., Kijsirikul, B., Takamura, H. and Okumura, M. (2009). 'Sentiment Classification with Support Vector Machines and Multiple Kernel Functions'. Neural Information Processing, pp.583–592.

[17] Singh, M., Jakhar, A.K. and Pandey, S. (2021). 'Sentiment analysis on the impact of coronavirus in social life using the BERT model'. Social Network Analysis and Mining, 11(1).

[18] Sousa, M.G., Sakiyama, K., Rodrigues, L.S., Moraes, P.H., Fernandes, E., Matsubara, E.T. (2019). 'BERT for stock market sentiment analysis', International Conference on Tools with Artificial Intelligence, ICTAI. IEEE Computer Society, pp.1597–1601.

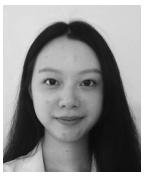
[19] Wan, Y., Gao, Q. (2016). 'An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis'. 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015. Institute of Electrical and Electronics Engineers Inc., pp. 1318–1325.

[20] Yang, K., Yu, Z., Wen, X., Cao, W., Chen, C, Wong, H., You, J. (2020). 'Hybrid Classifier Ensemble for Imbalanced Data'. IEEE Transactions on Neural Networks and Learning Systems 31, 1387–1400.

[21] Yu, H., Lo, H., Hsieh, H., Lou, J., McKenzie, T., Chou, J., Chung, P., *et al.* (2010). 'Feature engineering and classifier ensemble for KDD cup 2010'. JMLR: Workshop and Conference Proceedings 1–12.

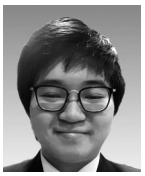
[22] Zhang, L., Wang, S. and Liu, B. (2018). 'Deep learning for sentiment analysis: A survey'. WIREs Data Mining and Knowledge Discovery, 8(4).

PROFILES



HUAY WEN KANG received BCompSc (Hons) Computer Science in Data Science from Tunku Abdul Rahman University College (TAR UC) in 2022. She is passionate about Artificial Intelligence (AI) and Data Science and aspires to be a professional data scientist. Her research interests involved Data Analytics, Code Mixed Sentiment Analysis and Computer Vision.
Email address: kanghw-wp17@student.tarc.edu.my

KAH KIEN CHYE graduated from Tunku Abdul Rahman University College (TAR UC) in 2022 with a BCompSc (Hons) (Computer Science in Data Science). He is a recent graduate who is passionate about applying data science and machine learning skillsets to SAP systems, an enterprise software to manage business operations and customer relations in which 77% of the world's transaction revenue touches an SAP system. His research interests include natural language processing, image processing, data mining, process mining, as well as advanced analytics.
Email address: chyekk-wp17@student.tarc.edu.my



ZI YUAN ONG received his Bachelor's degree of Computer Science (Honours) in Data Science from Tunku Abdul Rahman University College (TAR UC) in 2022. He is a fresh graduate who is passionate about Artificial Intelligence (MI), Machine Learning (ML) and Data Science. His research interests include artificial intelligence, natural language processing, image processing, data analytics as well as computer vision.
Email address: ongzy-wp17@student.tarc.edu.my

DR CHI WEE TAN received BCompSc(Hons) and PhD degrees in year 2013 and 2019 respectively in Universiti Teknologi Malaysia. Currently, he is a Senior Lecturer cum Programme Leader at Tunku Abdul Rahman University College and actively involved in the Centre of Excellence for Big Data and Artificial Intelligent (CoE) and become the research group leader for Audio, Image and Video Analytics Group under Centre for Data Science and Analytics (CDSA). Dr Tan's main research areas are Computer Vision (CV), Image Processing (IP) and Natural Language Processing (NLP) and Artificial Intelligence (AI). He is an enthusiastic researcher experienced in conducting and supporting research into Image Processing. Being a meticulous and analytical researcher with Train-The-Trainer certificate of many years of educational and hands-on experience, he was invited to Université d'Artois (France) under Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) programme for collaborative research between European countries with Southeast Asian countries on motion detection and computer vision and being involved in industry project as professional consultant.
Email address: chiwee@tarc.edu.my

MALAYSIAN ENGLISH LARGE VOCABULARY CONTINUOUS SPEECH RECOGNIZER: AN IMPROVEMENT USING ACOUSTIC MODEL ADAPTATION

(Date received: 13.01.2022/Date accepted: 09.05.2022)

Kah Chung Yoong[1], Kai Sze Hong[1]

[1] Department of Electrical & Electronic Engineering, Faculty of Engineering & Technology, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia.

**Corresponding author: yoongkc-wg17@student.tarc.edu.my, hongks@tarc.edu.my*

ABSTRACT

This research project aims to develop Malaysian English Continuous Speech Recognition system by adapting US English acoustic model with Malaysian English speech corpus using Maximum a posteriori reasoning (MAP) and Maximum Likelihood Linear Regression (MLLR). During feature extraction stage, the Mel-Frequency Cepstral Coefficients (MFCC) technique was used. The Hidden Markov Model was used as the back end pattern comparison technique. For the purpose of implementation, the CMU Sphinx toolkit, which includes Pocketsphinx and Sphinxtrain as well as an acoustic model, was used to develop a speech recognition system for Malaysian English. Malaysian English speech samples were recorded and transcribed to produce the training database required for acoustic model adaptation. The adaptation speech corpus were collected from a number of speakers. The outcome of this research could increase the application of Malaysian English speech recognition in Malaysia due to accent problem. As a result, speech recognition systems that have gone through the MAP adaptation had the best performance. Its average word error rate achieved was 32.84%. Average word recognition rate was 72.52% and average sentence error rate was 78.89%.

Keywords: Speech Recognition, Acoustic Model, MAP, MLLR, Pocketsphinx

1.0 INTRODUCTION

Speech is a form of language-based individual vocalization. The audio within each linguistic vocabulary is formed by pronunciation arrangements of vowel and consonant tones. While using several terms in their linguistic context as vocabulary in a linguistic dictionary in accordance with the grammatical restrictions that regulate the role of lexical parts of speech (Houghton Mifflin, 2015). Speakers may use pronunciation, inflection, volume of voice, rhythm, and several other non-representational or vocabulary elements of vocalization to express meaning in several specific deliberate speech acts, such as telling, announcing, questioning, convincing, and guiding. Speakers unwittingly express many facets of their social status in their speeches, including gender, aged, point of birth, cognitive abilities, psychological condition, physio-psychic condition, background, or knowledge, etc. Several components of speech are investigated by researchers, including speech processing and voice detection (Houghton Mifflin, 2015). Voice duplication, voice defects, and the failure to translate hearing speaking terms into the vocalizations required to reproduce them are all examples of these. This is important for kid's vocabulary development and mind development in various fields. Sociology, theoretical physics, information science, sociology, software engineering, forensic linguistics, ophthalmology, and sound systems are all fields that research speech. Speech is related to linguistic knowledge, which can distinguish from spoken language in terms of pronunciation, grammar, and phonology, a disorder known as diglossia (Houghton Mifflin, 2015; NIDCD, 2021).

The conversion of individual voice signals into vocabulary or commands is known as speech recognition. Speech recognition is dependent on the sound of a person's voice. That is a subdivision of information processing and a significant research path in speech signal processing. Software engineering, machine learning, digital signal processing, information processing, sound systems, linguistics, and cognitive science all play a role in speech recognition study. It's an interdisciplinary, all-encompassing field of study. One of the research objectives are to improve the speech recognition accuracies. Besides, speech recognition systems are designed to work in either a constrained environment or an opened environment. Different applications have different requirements of specific type of speech recognizer. Based on research objectives and restrictions, numerous research areas have arisen. These fields could be separated into isolated words, connected words, and continuous speech recognition systems, depending upon the needs of the presenter's style of communicating. The above fields could be classified into speech recognition systems for individuals and non - specific persons depending on the level of reliance on the person speaking (Science Direct, 2021). These could be categorized into small vocabulary, medium vocabulary, big vocabulary, and infinite vocabulary speech recognition systems based also on scale of their pronunciation. The concept of the voice recognition model is predicated on information processing, according to the speech

recognition model. That aim of speech recognition is to use phonology and textual data to convert a received signal feature vector pattern into a series of text. A full speech recognition system involves information extraction technique, the acoustic template, a language model, and a search method, as according to configuration of a speech recognition system. A multifaceted information processing system is exactly what a speech recognition system is. Researchers use different recognition methodologies for specific speech recognition systems. However, the simple concepts are the same. Function abstraction is applied to the obtained speech recognition. The template database system collects and processes the speech information received by the system. The voice information retrieval module recognizes speech samples based on the template database, and then calculates the segmentation results (Science Direct, 2021).

The overall organization of this paper is described here. Firstly, section 2 describes the literature review of Malay LVCSR system. Next, section 3 elaborates the research methodology. Section 4 shows the results and discussions. Finally, section 5 concludes the whole research work.

2.0 LITERATURE REVIEW

Speech recognition systems allow a computer to carry out human-spoken commands, perform automatic translation, and generate print-ready dictation. The microphone receives the sound signal, which is then transformed into a digital signal by the system hardware. The analysis tool of a speech recognition device uses the produced digital signal as input to extract and recognise the differentiated phoneme. Several of the lowest units of sound is the phoneme, which distinguishes the sound of one word from that of another. Nevertheless, most of the voice of the phrases are similar. As a result, the software must depend on context to distinguish the right punctuation among all these similar sounding words (Abhang *et al.*, 2016).

2.1 Process of Speech Recognition

Those two most popular strategies to speech recognition could be classified as "template matching" and "feature analysis." Template matching is the most efficient method with the highest precision, but it often would have the most drawbacks. That individual should first say a word or statement into a mic, just like in any technique to speech recognition. An "analogue-to-digital (A/D) converter" transforms the electric signals from the receiver into digital form, which is then loaded into memory. This approach is somewhat similar to using a button to enter commands. That software includes an input framework and provides a simple constraint declaration to try and adapt the model to the real input (HITL, 2021). The overview of process of speech recognition general framework is showed in Figure 1:

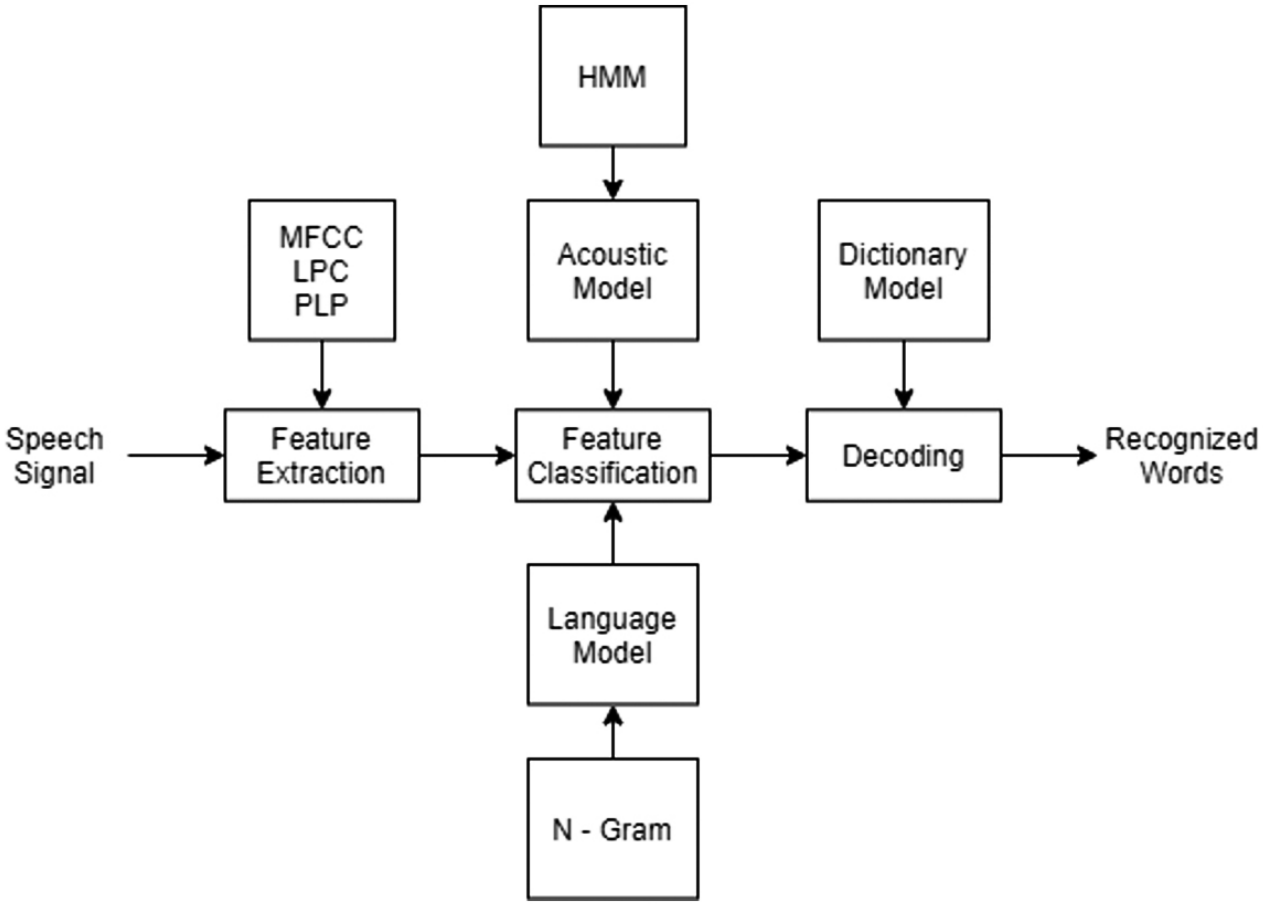


Figure 1: Process of Speech Recognition

Figure 1 depicts the general framework of the speech recognition process. Speech signal features are extracted using feature extraction techniques like MFCC, LPC and PLP. Next, features that have been extracted will be classified using acoustic modelling and language modelling. Finally, during the recognition stage, the features are decoded with the aid of a dictionary model to produce the recognized words.

2.2 Feature Extraction

The key component of a speech recognition program is feature extraction. These are regarded as the program's beating heart. The main objective is to extract features through the source voice signal which will aid the device in recognising the user. That intensity of the source signals is compressed by feature extraction without affecting the voice signal's strength. Furthermore, it attempts to minimise the loss of data kept among the terms during this point. That aids throughout the consistency comparison of its acoustic model's distributional assumption.

Mel frequency cepstral coefficients (MFCC) were first proposed for recognizing idiomatic phrases in consistently voiced statements, but not for determining the user. The MFCC algorithm is really a simulation of the living thing listening process that serves to theoretically enforce the ear's working theory, assuming also that living person ear is really an effective voice recognition system. That MFCC models are based on a known difference between the essential bandwidths of the living person ear and frequency filtering distributed sequentially at lower frequency. That pronunciation critical characteristics of the voice signal were preserved by doing this logarithmically at higher frequency. Sounds of various frequencies are generally used in voice signal, with every voice having its own frequency. The Mel measurement is used to calculate arbitrary pitch. Approximately 1000 Hz, that Mel-frequency level contains linear frequency spaced, and over 1000 Hz, these have linear interpolation frequency width. The 1000 meals was its pitch of even a 1 kHz voice at 40 dB over the perceived detectable level, which is applied as a level of comparison (Alim, 2018; Vergin, R. *et al.,* 1999; X. Zhou, D. Garcia-Romero, R. Duraiswami *et al.*, 2011; Muda, L *et al.*, 2021).

That MFCC algorithm depends upon signal dissolution using a wiener filter. That MFCC generates a discrete cosine transform (DCT) of even a particular logarithm of its simple terms power also on Mel frequency range. Under safety purposes, the MFCC is being used to classify travel arrangements, contact information, and speech recognition. With improved robustness, several improvements to the simple MFCC methodology are being suggested. Besides instance, while implementing the DCT, raise the log-mel-amplitudes to a suitable capacity. That alone lessens the detrimental effects of low-energy materials (Alim, 2018; Vergin *et al.*, 1999; Zhou *et al.*, 2011; Muda *et al.*, 2021).

The Mel Frequency Cepstrum parameters of MFCC are constructed from a distorted frequency range focusing on individual perceptual experience. Its first stage in MFCC processing is windowing the voice signal which divides these into layers. Although the high frequency formants method would have a lower amplitude than the low frequency formants, the high frequencies are stressed in order to achieve a comparable amplification for both formants. The energy spectrum within each frame is determined using the Fast Fourier Transform (FFT) upon windowing. Following that, Mel-scale filter system work is performed on the energy spectrum. In order to determine MFCC parameters, the DCT is added to the voice signal once the energy spectrum being converted to log field (Alim, 2018; Vergin *et al.*, 1999; Zhou *et al.*, 2011; Muda *et al.*, 2021).

$$mel(f) = 2595\, x\, log_{10}\,(1 + \frac{f}{700}) \qquad \textit{Eq. 1}$$

Where Mel (f) represents the mel scale of frequency and f represents the frequency (Hz).

The below formula is used to measure the MFCCs.

$$C_n = \sum_{n=1}^{k} (\log S_k) \cos\,[n\left(k - \frac{1}{2}\right)\frac{\pi}{k}] \qquad \textit{Eq. 2}$$

The k defines the number of Mel Cepstrum parameters S the filter bank production, and C the last MFCC approximation.

Figure 2 shows the block diagram including its MFCC module. That low frequency area could be essentially denoted by MFCC, while the high frequency field can indeed be efficiently denoted by MFCC. This one can approximate and define vocal folds resonances using formants throughout the low frequency region. That is widely acknowledged as a front-end technique for popular Speech Recognition implementations. These have minimised noise disruption uncertainty, provides minute's process instability, and is simple to develop. Often, whenever the useful in the evaluation are balanced and coherent expression, it is a good demonstration for voices. It could also derive data through processed signals with a peak frequency of 5 kHz. This covers the bulk of the power found in human-made voices (Alim, 2018; Vergin *et al.*, 1999; Zhou *et al.*, 2011; Muda *et al.*, 2021).

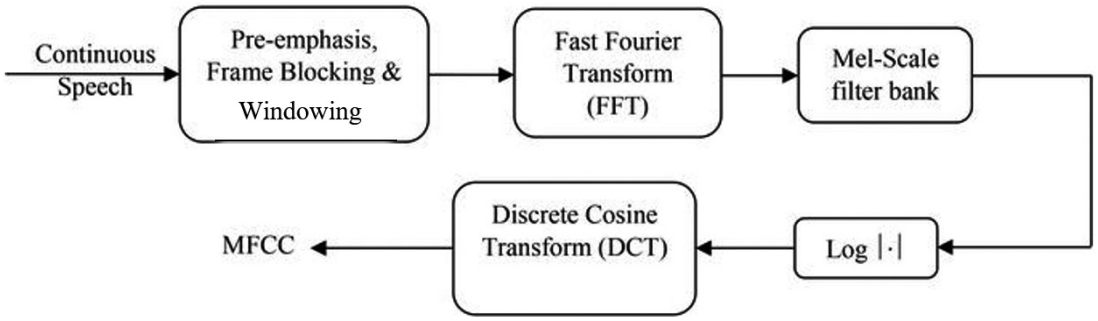


Figure 2: Block Diagram of MFCC

In several pattern recognition difficulties concerning living person speech, cepstral coefficients are shown to be correct. They are commonly utilized in numerous speech recognition and presenter detection. Those certain formants should be over 1 kHz, and indeed the wide filter width throughout the high frequency region does not effectively account for them. Throughout the presence of ambient noise, MFCC functions may not always be reliable, and they are not very well equipped for generalisation (Alim, 2018; Vergin *et al.*, 1999; Zhou *et al.*, 2011; Muda *et al.*, 2021).

2.3 Feature Classification

Within various conditions, a classification model is being used to determine the right voice speaking style. The strong pattern classification scheme is being trained through appropriately labelled illustrations in speech recognition.

The Hidden Markov Model (HMM) is a mathematical method that might be used to explain how measurable phenomena evolve over time (Rabiner and Juang, 1986). The observable element influencing the occurrence is referred to as a 'state,' whereas the observable phenomenon is referred to as a 'symbol'. The unidentifiable procedure of unidentified conditions and the transparent methodology of measurable signs are the two stochastic systems that make up an HMM. That invisible conditions construct a Markov chain, and perhaps the detected sign's probability distribution will be determined by the corresponding system's probability distribution. Mostly as result, a doubly-embedded random variable is another name about an HMM (Rabiner and Juang, 1986).

It is sufficient to define findings within those 2 phases, one transparent and the other hidden. Although several real-world problems include categorising unstructured data into a set of classification or membership functions. That matters to them as well. Remember the issue of voice recognition, where this HMMs had already long been shown. Forecasting the spoken phrase from such a registered voice signal will be the aim of speech recognition. According to interpretations, the voice recognition system attempts to locate the series of consonant conditions that resulted in the real spoken voice. Although real transcription can vary greatly, the basic phonetic symbols cannot be detected explicitly and must be expected (Yoon, 2009; Cuiling, 2016; Xue, 2018).

There are also other research work related to Hidden Markov Models by other researchers (Aymen *et al.*, 2011; Gales, 2009; Abushariah *et al.*, 2010).

2.4 N - Gram

The N-gram is a texting-gram structure in which the "N" objects from a provided series or series are continuously arranged. Text mining, communication theory, text probability, and data compression are only a few of the areas where it is used. When assigning with varying probabilities, the N-gram is useful. That is because it assists in deciding which N-grams have the greatest chance of chunking together to create a single entity (Gadag and Sagar, 2016; Takahashi and Morimoto, 2012).

The N-gram analysis shows the appearance of an interpretive structural on the appearance of N-1 preceding terms. Unigram (N), bigram (N=2), trigram (N=3), and so on are all examples of N-grams. Bigram model N=2 predicts a term occurring based on the previous single word (N-1) and bigram model N=3 forecasts a term phenomenon based on the previous two terms (N-2) (N-2) (Ito and Kohda, 1996; Hatami, Akbari and Nasersharif, 2013).

2.5 Maximum Likelihood Linear Regression (MLLR)

Maximum Likelihood Linear Regression (MLLR) would be the first adaptation algorithm inspired. Through analysing that acoustic model's mean and variance, MLLR can predict and determine that optimum probability distribution and characteristics. Maximum A-Posteriori (MAP) has been the second adaptation approach applied. MAP performs nearly the similar purpose as MLLR, with the exception that it considers the prior throughout the prediction. The mixture weight and the transition matrices are also observed by MAP in addition to the acoustic model's variance and means. MAP updates the characteristics in the acoustic model, contrasting MLLR, which just generates a matrix that could be passed to the system upon decoding. Furthermore, while MAP appears to become a preferable adaptation approach, that alone necessitated a large amount of data for adaptation in attaining the desired prediction performance. Although MLLR proved possible to provide a noticeable enhancement in recognition rate with small dataset, two adaption methods were used in this study, one simultaneously and one individually, to examine respective impact on speech recognition application efficiency. Research works related to MLLR have been done by some researchers (Lestari and Irfani, 2015; Oh *et al.*, 2007).

2.6 Novelty of this Research Work

There were no work previously done in the area of adapting Malaysian English to English LVCSR. The reason is most speech recognition research works are focusing on other languages. Thus this research project helps to study the possible accuracies when Malaysian English is adapted to English LVCSR.

3.0 METHODOLOGY

3.1 Speech Recognition Process

Figure 3 depicts the flow chart for the overall system process. This overall system process is separated into 2 stages including as adapting stage and decoding stage. For adapting stage, 50 individual phrases of utterances by five separate speakers (Anonymous FYP students) may also being used to adjust the Malaysian English language acoustic model utilising two adaptation strategies, MLLR and MAP. At first, all the speech sample will be collected and convert them into Mel Frequency Cepstral Coefficients (MFCC) format by undergoes the feature extraction. This feature extraction process will extract the main speech information and critical characteristic that is required for the system. So, all the MFCC files with extracted feature are produced. These MFCC files, list of transcripts of the adaption speech samples, list of speech sample filenames, and acoustic model were used in the adaptation.

Prior beginning the decoding stage, those speech samples utilized to analyse the results were first feature extracted by MFCC. This speech recognition system then uses the created MFCC files, including comprise all extracted feature, customised acoustic model, MLLR matrices, language model, and dictionary model, to conduct speech recognition. After the decoding stage,

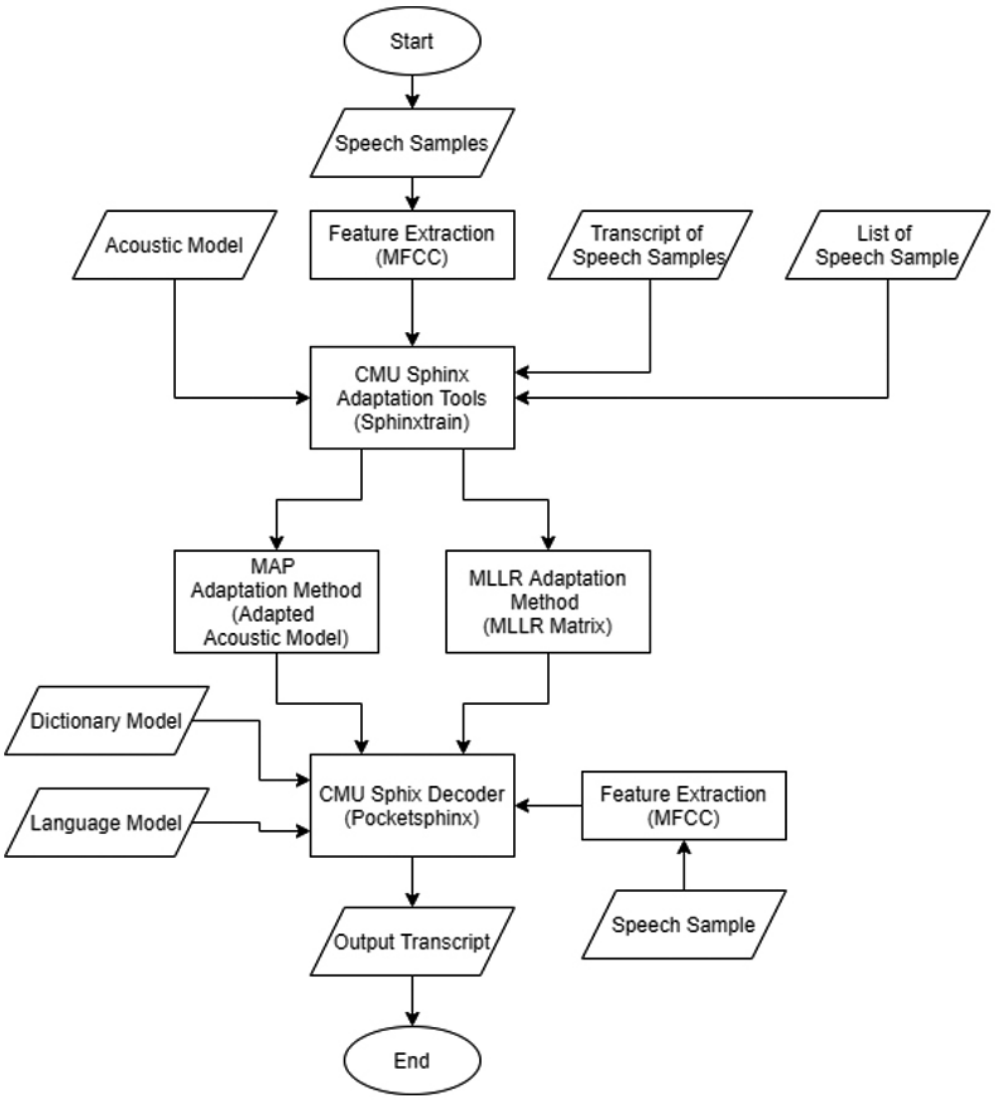


Figure 3: Flow Chart for the overall system process

all the result will be analysed by Sphinxtrain modules with Perl to measure the Word Error Rate (WER), Word Recognition Rate (WRR) and Sentence Error Rate (SER). Regarding comparison purposes, the system's functionality is evaluated under various conditions, such as without adaptation, with MLLR adaptation only, with MAP adaptation solely, and with combined MAP and MLLR adaptation.

3.2 Model Adaptation

Adaptation was employed throughout this research to increase the program's speech recognition performance. 5 separate people's voice samples and one internet-based conversation containing 50 different sentences were used in this research project. For adaption, many of these voice samples were obtained from typical conversations with five different people, as well as a presentation for one online speech. That work was completed using Sphinxtrain and Sphinxbase. Maximum Likelihood Linear Regression (MLLR) and Maximum A-Posteriori (MAP) were really the two adaption strategies used.

Typically implemented to the means through one or perhaps more Gaussian Mixture Models, this MLLR method determines a linear transformation. Pertaining to such models, this also maximises the probability of a sample of information. On the other hand, MAP is a sort of enhancement in which the functional form represents the likelihood of an outcome based (or symbol, or sequence) on prior information. That would be the likelihood with that occurrence, consistent with past prediction with that event's possibility (prior probability distribution) and one or perhaps more occurrences. Furthermore, MLLR adaptation is convenient and ideal for limited sample adaption, but MAP adaptation needs additional sample to achieve the desired precision. These MLLR matrix and customised acoustic models are again incorporated and modified correspondingly only with initial acoustic model.

3.3 Software Required

This research had utilized GoldWave (Goldwave, 2021) as the main recording tool to capture voices from Malaysian English speakers. Besides, CMU Sphinx (CMU Sphinx, 2021) had been used to build speech recognition software decoders. This is a cutting-edge large-vocabulary speech recognition framework introduced at CMU. Continuous Hidden Markov Models (HMMs) with optimised Gaussian Mixture Model (GMM) computations are used to decode conversation. Microsoft Visual Studio (Microsoft, 2021) provides the integrated programming environment (IDE) developed by Microsoft for various sorts of programming production, which including software applications. Implementation programs, compilers, and several other capabilities are also included to make overall program construction procedure easier. In this project, the PyCharm IDE (PyCharm, 2021) is used to construct the Graphical User Interface (GUI) for this project.

3.4 Method/ Tools/ Technologies Involved

3.4.1 Acoustic Model

The quantitative interpretations of different sounds that make up a phrase are contained in the acoustic model. Furthermore, these statistical interpretations of voice are referred to as phonemes. Phonemes were created by training Hidden Markov Models (HMM) on a huge array of voice databases, and each phoneme will have its own hmm value. Throughout decoding, the decoder will compare the hmm value of certain phonemes to the adjusted distinct sound to evaluate whether phrase has been spoken.

Figure 4 depicts the acoustic model raw file used to describe the phonemes and their respective tied states. The acoustic model is important to calculate the probabilities of tri-phone sequences.

```
0.3
46 n_base
137053 n_tri
548396 n_state_map
5138 n_tied_state
138 n_tied_ci_state
46 n_tied_tmat
#
# Columns definitions
#base lft  rt p attrib tmat      ... state id's ...
+BREATH+   -   - - filler    0      0      1      2 N
+COUGH+    -   - - filler    1      3      4      5 N
+NOISE+    -   - - filler    2      6      7      8 N
+SMACK+    -   - - filler    3      9     10     11 N
 +UH+    -   - - filler    4     12     13     14 N
 +UM+    -   - - filler    5     15     16     17 N
   AA    -   - -    n/a    6     18     19     20 N
   AE    -   - -    n/a    7     21     22     23 N
   AH    -   - -    n/a    8     24     25     26 N
   AO    -   - -    n/a    9     27     28     29 N
   AW    -   - -    n/a   10     30     31     32 N
   AY    -   - -    n/a   11     33     34     35 N
    B    -   - -    n/a   12     36     37     38 N
   CH    -   - -    n/a   13     39     40     41 N
    D    -   - -    n/a   14     42     43     44 N
   DH    -   - -    n/a   15     45     46     47 N
   EH    -   - -    n/a   16     48     49     50 N
   ER    -   - -    n/a   17     51     52     53 N
   EY    -   - -    n/a   18     54     55     56 N
    F    -   - -    n/a   19     57     58     59 N
    G    -   - -    n/a   20     60     61     62 N
   HH    -   - -    n/a   21     63     64     65 N
   IH    -   - -    n/a   22     66     67     68 N
   IY    -   - -    n/a   23     69     70     71 N
   JH    -   - -    n/a   24     72     73     74 N
    K    -   - -    n/a   25     75     76     77 N
    L    -   - -    n/a   26     78     79     80 N
    M    -   - -    n/a   27     81     82     83 N
    N    -   - -    n/a   28     84     85     86 N
```

Figure 4: Acoustic Model that describes the features of the sounds

```
\data\
ngram 1=72547
ngram 2=9704821
ngram 3=12264838

\1-grams:
-6.283094       'bout   -0.1851698
-4.577734       'cause  -0.1588756
-3.713352       'em     -0.5801811
-6.23701        'n      -0.07844383
-6.160365       's      -0.08116825
-6.098946       'til    -0.1397758
-1.126058       </s>
-99     <s>     -1.518327
-1.651049       a       -0.9826273
-5.534918       a's     -0.1901456
-6.368959       a.      -0.1376058
-8.792997       a.'s
-7.153554       aachen  -0.1377168
-8.040561       aamodt  -0.02926606
-6.90229        aardvark        -0.1475043
-5.206353       aaron   -0.4841892
-6.668598       aaron's -0.1548517
-7.313623       aarons  -0.1346213
-7.10872        aaronson        -0.1588136
-7.634939       aaronson's      -0.02629348
-5.570579       ab      -0.3932968
-6.608242       ababa   -0.2087529
-7.720179       abacha  -0.1117107
-5.975394       aback   -0.6033721
-8.157355       abaco   -0.1315198
```

Figure 5: N-gram Modeling which predicts the sequence of words

3.4.2 Language Model

The N-gram language model was included in this research. Three-gram, unigram, bigram, and trigram are among the N-grams included in this research. Various probabilities were allocated to the phrases throughout this N-gram. That sequence probabilities of a phrase appearing in a conversation are represented by the probabilities allocated to all of it. In other words, using all these probabilities, the voice recognition software can correctly determine which term will occur next and differentiate separating two words with quite comparable pronunciations.

Figure 5 depicts the language model used in this speech research project. The first column shows the log probabilities of the n-gram while the last column shows the back-off probabilities.

3.4.3 Dictionary Model

The dictionary would be a document which includes all or most of the vocabulary words which can be identified mostly by software, within both vocabulary and phonology transcription. None with the voiced phrases would be identified without the dictionary.

3.5 PyCharm Community Edition Graphical User Interface Design

This graphical user interface for the speech recognition system has been devised and provided with the opportunity for the average consumer to be used. For the Malaysian English Continuous Speech Recognition System, PyCharm Community Edition has been applied to construct and establish the graphical user interface.

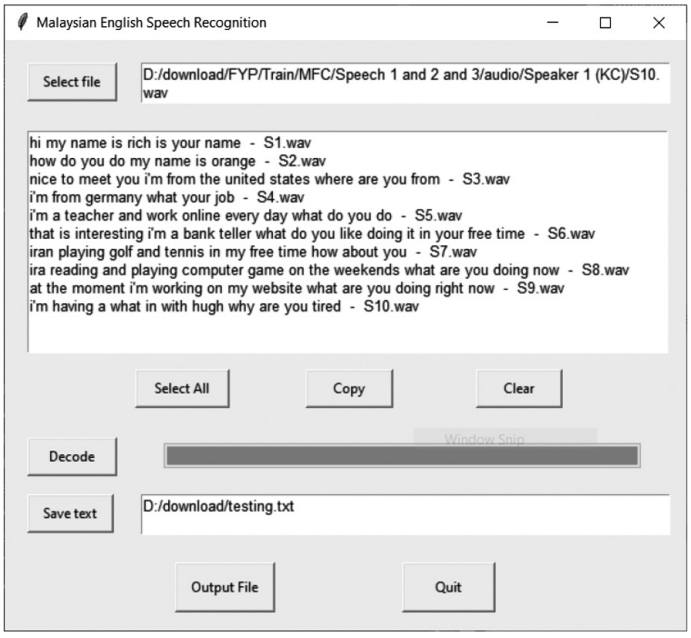


Figure 6: Graphical user interface for Malaysian English Speech Recognition System

The graphical user interface platform designs, as depicted in Figure 6, were created using Window Form App (.Net Framework). The developed graphical user interface GUI is shown in Figure 6. The buttons have been written with their specific roles and implemented to ensure that they functioned as intended.

4.0 RESULT AND DISCUSSION

Overall effectiveness of speech recognition is compared with a popular measurement defined as Word Error Rate (WER) to determine the functionality including its produced speech recognition system. WER is calculated through using words of speech, as stated with in equation following.

$$Word\ Error\ Rate\ (WER) = \frac{Substitutions\ (S) + Insertions\ (I) + Deletions\ (D)}{Number\ of\ Word\ in\ the\ reference\ (N)}$$

Eq. 3

Word Recognition Rate (WER) is calculated by dividing the overall amount of words throughout the comparison transcript through the amount of matching terms. Sentences Error Rate (SER) is calculated by dividing the quantity of incorrect sentences throughout the comparison transcript through the entire quantity of sentences throughout the transcript.

This would calculate the amount of words which are substituted, deleted from its transcription, and unsaid words which are already inserted to evaluate the effectiveness about an automatic speech recognition system. Word Recognition Rate (WRR) and Sentences Error Rate (SER) are two metrics that can be used to evaluate the effectiveness about an ASR system. Several metrics should be applied to evaluate how many words inside the comparison translation fit and how many phrases are erroneous.

The model adaptation will be trained and tested by 1 online presenter and 5 speakers. These speakers will help to adapt the acoustics models and testing the adapted models. For all adapted transcription will used 50 sentences to adapt acoustics model and 30 sentences to test the adapted model.

4.1 Speech Recognition System Performance Test by using Adaptation Transcription

For the MAP average value, the WER is 15.57%, the WRR is 86.41% and the SER is 64.97%. By comparison between the MAP and MLLR, the WER of MAP is lesser than the MLLR with 31.39%, the WRR of MAP is higher than MLLR with 22.29% and the SER of MAP is lower than MLLR with 32.47%. This comparison showed that WER and SER of MAP have much smaller compare with MLLR and the WRR have very large improvement compare with MLLR. The MAP is better than MLLR in improve accuracy around 20% and reduce mistake for the speech recognition system around 30%.

For both adaptation with MAP and MLLR average value, the WER is 17.94%, the WRR is 88.84% and the SER is 70.83%. By comparison between both adaptation and MAP, the WER of both adaptation is more than the MAP with 2.37%, the WRR of both adaptation is higher than MAP with 2.44% and the SER of both adaptation is higher than MAP with 5.86%. This comparison showed that both adaptations have improved the accuracy of the speech recognition system, but it also increases the error rate of the speech recognition system by comparing with MAP. The increase of SER is much higher compare WRR by comparing both adaptation and MAP.

In conclusion, the MAP adaptation method is best choice for the speech recognition system, and it is highest improvement in accuracy with lowest error rate for the speech recognition system. Even through, both adaptations method has the highest WRR but the WER and SER also increased. The increase of the error rate is much higher compare with recognition rate for both adaptation method. Table 1 tabulates the word error rate, word recognition rate and sentence error rate when adaptation transcription was used.

Table 1: Error Rate and Recognition Rate with Adaptation Transcription

	WER	WRR	SER
Original	55.10%	58.25%	97.44%
MLLR	46.96%	64.12%	97.44%
MAP	15.57%	86.41%	64.97%
MLLR and MAP	17.94%	88.84%	70.83%

4.2 Speech Recognition System Performance Test by using Test Transcription

For the MAP average value, the WER is 32.84%, the WRR is 72.52% and the SER is 78.89%. By comparison between the MAP and MLLR, the WER of MAP is lesser than the MLLR with 18.55%, the Word Recognition Rate (WRR) of MAP is higher than MLLR with 12.79% and the SER of MAP is lower than MLLR with 18.33%. This comparison showed that WER and SER of MAP have much smaller compare with MLLR and the Word Recognition Rate (WRR) have very large improvement compare with MLLR. The MAP is better than MLLR in improve accuracy around 12% and reduce mistake for the speech recognition system around 18%.

For both adaptation with MAP and MLLR average value, the WER is 48.01%, the Word Recognition Rate (WRR) is 61.97% and the SER is 81.67%. By comparison between both adaptation and MAP, the WER of both adaptation is more than the MAP with 15.17%, the Word Recognition Rate (WRR) of both adaptation is lower than MAP with 10.55% and the SER of both adaptation is higher than MAP with 2.78%. This comparison showed that both adaptations have reduced the accuracy of the speech recognition system and it also further increases the error rate of the speech recognition system by comparing with MAP.

In conclusion, the MAP adaptation method is also best choice for the speech recognition system with test transcription, and it have highest improvement in accuracy with lowest error rate for the speech recognition system. The accuracy of both adaptation method is decreased compare with MAP by using test transcription and this case is not similar compare with adaptation transcription. By using adaptation transcription, both adaptation method is showed the improvement of the accuracy, but it also increases the error rate. That adaptation method showed the accuracy decrease and further improve the WER by using test transcription. Table 2 tabulates the word error rate, word recognition rate and sentence error rate when test transcription was used.

One of the possible reason for MAP to outperform MLLR using both adaptation and test transcription is due to the small amount of adaptation data. Thus, in future research, it is advisable to record more adaptation data in different settings so that the performances of MLLR could be better.

Table 2: Error Rate and Recognition Rate with Test Transcription

	WER	WRR	SER
Original	58.55%	54.36%	97.22%
MLLR	51.39%	59.73%	97.22%
MAP	32.84%	72.52%	78.89%
MLLR and MAP	48.01%	61.97%	81.67%

5.0 CONCLUSION

In conclusion, adaptation and test scripts were used to evaluate the effectiveness of the constructed speech recognition system. For testing this system by using adaptation transcription, the MAP adaptation method has showed the high accuracy with lowest error rate. The MAP has the lowest WER and lowest SER with 15.57% and 64.97%. The MAP also has high Word Recognition Rate (WRR) with 86.41%. But the Word Recognition Rate (WRR) of MAP adaptation method is lower than both adaptation method and the different of them is 2.44%. Both adaptation method also will increase the WER and SER compare with MAP adaptation method. The improvement accuracy of both adaptation method is lower compare with increment of error rate. The MLLR adaptation method is showed the worst improvement in accuracy and error rate compare with MAP and both adaptation method. So, the MAP is best choice for adaptation transcription.

For testing transcription, the MAP adaptation method is showed the highest accuracy and lowest error rate for speech recognition system. The WER of MAP adaptation method is 32.84%, the Word Recognition Rate (WRR) is 72.52% and the SER is 78.89%. Moreover, both adaptation method has second higher accuracy and second lower error rate. Both adaptation with test transcription have showed the decrease accuracy and increase error rate compare with adaptation transcription. The MLLR adaptation method is showed the worst improvement in accuracy and error rate compare with MAP and both adaptation method.

Finally, the MAP adaptation method is suitable for adaptation and test transcription. The MAP adaptation method is best adaptation method, and it will be implemented into speech recognition system. This Malaysian English speech recognition system's weakness is that it may function poorly if somehow the voice stream includes a lot of distortion and noise. Whereas if surroundings are noisy, this system may not achieve sufficient performance.

This research work found that MAP adaptation approach outperformed MLLR adaptation approach in all settings. It may be due to the limited adaptation data recorded. In future, more adaptation data should be recorded in different environment so that the results could be improved. ■

REFERENCES

[1] Abhang P, Gawali B and Mehrotra S (2016) Introduction to EEG- and speech-based emotion recognition. Academic Press.

[2] Abushariah AAM, Gunawan TS, Khalifa OO and Abushariah MAM (2010) English digits speech recognition system based on Hidden Markov Models. International Conference on Computer and Communication Engineering (ICCCE'10), Kuala Lumpur, Malaysia, 2010, pp. 1-5, doi: 10.1109/ICCCE.2010.5556819.

[3] Alim S (2018) Some Commonly Used Speech Feature Extraction Algorithms, From Natural to Artificial Intelligence - Algorithms and Applications. Ricardo Lopez-Ruiz, IntechOpen.

[4] Aymen M, Abdelaziz A, Halim S and Maaref H (2011) Hidden Markov Models for automatic speech recognition. 2011 International Conference on Communications, Computing and Control Applications (CCCA), Hammamet, Tunisia, 2011, pp. 1-6.

[5] CMU Sphinx (2021) CMU Sphinx website. Available at: https://cmusphinx.github.io/ (Accessed: 11 September 2021).

[6] Cuiling L (2016) English Speech Recognition Method Based on Hidden Markov Model. 2016 International Conference on Smart Grid and Electrical Automation (ICSGEA), Zhangjiajie, China, 2016, pp. 94-97.

[7] Gadag A and Sagar DBM (2016) N-gram based paraphrase generator from large text document. 2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2016, pp. 91-94.

[8] Gales MJF (2009) Acoustic modelling for speech recognition: Hidden Markov models and beyond?. 2009 IEEE Workshop on Automatic Speech Recognition & Understanding, Moreno, Italy, 2009, pp. 44.

[9] Goldwave (2021) Goldwave website. Available at: https://www.goldwave.com/ (Accessed: 11 September 2021).

[10] Hatami A, Akbari A and Nasersharif B (2013) N-gram adaptation using Dirichlet class language model based on part-of-speech for speech recognition. 2013 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, Iran, 2013, pp. 1-5, doi: 10.1109/IranianCEE.2013.6599642.

[11] HITL (2021) Voice Recognition. Available at: <http://www.hitl.washington.edu/projects/knowledge_base/virtual-worlds/EVE/I.D.2.d.VoiceRecognition.html> (Accessed 19 March 2021).

[12] Houghton Mifflin (2015) The American Heritage Dictionary of the English Language. Boston.

[13] Ito A and Kohda M (1996) Language modeling by string pattern N-gram for Japanese speech recognition. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, Philadelphia, PA, USA, 1996, pp. 490-493 vol.1, doi: 10.1109 /ICSLP.1996.607161. ISBN: 978-0-12-804490-2.

[14] Lestari DP and Irfani A (2015) Acoustic and language models adaptation for Indonesian spontaneous speech recognition. 2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2015, pp. 1-5.

[15] Microsoft (2021) Microsoft Studio website. Available at: https://visualstudio.microsoft.com/ (Accessed: 11 September 2021).

[16] Muda L, Begam M and Elamvazuthi I (2021) Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. Available at: <https://arxiv.org/abs/1003.4083> (Accessed 20 March 2021).

[17] NIDCD (2021) What Is Voice? What Is Speech? What Is Language? Available at: <https://www.nidcd.nih.gov/health/what-is-voice-speech-language> (Accessed 17 March 2021).

[18] Oh Y, Yoon J and Kim H (2007) Acoustic model adaptation based on pronunciation variability analysis for non-native speech recognition. Speech Communication, 49(1), pp.59-70.

[19] Rabiner L and Juang B (1986) An introduction to hidden Markov model. IEEE ASSP Magazine, Vol. 3 Issue 1, pp. 4-6.

[20] PyCharm (2021) PyCharm website. Available at: https://www.jetbrains.com/pycharm/ (Accessed: 11 September 2021).

[21] Science Direct (2021) Speech Recognition - an overview ScienceDirect Topics. Available at: <https://www.sciencedirect.com/topics/engineering/speech-recognition> (Accessed 17 March 2021).

[22] Takahashi S and Morimoto T (2012) N-gram Language Model Based on Multi-Word Expressions in Web Documents for Speech Recognition and Closed-Captioning. 2012 International Conference on Asian Language Processing, Hanoi, Vietnam, 2012, pp. 225-228, doi: 10.1109/IALP.2012.55.

[23] Vergin R, O'Shaughnessy D and Farhat A (1999) Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition. IEEE Transactions on Speech and Audio Processing, 7(5), pp.525-532.

[24] Xue C (2018) A Novel English Speech Recognition Approach Based on Hidden Markov Model. 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Hunan, China, 2018, pp. 1-4.

[25] Yoon B (2009) Hidden Markov Models and their Applications in Biological Sequence Analysis. Current Genomics, 10(6), pp.402-415.

[26] Yoong, KC and Hong KS (2021) Development Of Malaysian English Large Vocabulary Continuous Speech Recognizer Using Acoustic Model Adaptation. International Conference on Digital Transformation and Applications (ICDXA), 25-26 October 2021, pp. 36-48.

[27] Zhou X, Garcia-Romero D, Duraiswami R, Espy-Wilson C and Shamma S (2011) Linear versus mel frequency cepstral coefficients for speaker recognition. IEEE Workshop on Automatic Speech Recognition & Understanding, Waikoloa, HI, USA, 2011, pp. 559-564.

PROFILES

YOONG KAH CHUNG is a bachelor degree graduate of Faculty of Engineering and Technology, Tunku Abdul Rahman University College, Kuala Lumpur. His specialization and area of interest are speech recognition and electronic technology.
Email address: yoongkc-wg17@student.tarc.edu.my

DR HONG KAI SZE is a senior lecturer of Faculty of Engineering and Technology, Tunku Abdul Rahman University College, Kuala Lumpur. He received his PhD from Universiti Sains Malaysia. His specialization and area of interest are speech recognition and artificial intelligence.
Email address: hongks@tarc.edu.my

A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR SENTIMENT ANALYSIS OF GAME REVIEWS

(Date received: 13.01.2022/Date accepted: 01.04.2022)

Jie Ying Tan[1], Andy Sai Kit Chow[1], Chi Wee Tan[1]*

[1] Faculty of Computing And Information Technology, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300, Wilayah Persekutuan Kuala Lumpur, Malaysia.

**Corresponding author: chiwee@tarc.edu.my*

ABSTRACT

Sentiment analysis, also known as opinion mining, is the process of analysing a body of text to determine the sentiment expressed by it. In this study, Natural Language Processing techniques and Machine Learning algorithms have been applied to create multiple sentiment analysis models customized for the gaming domain to determine the sentiment of game reviews. The dataset was collected from Steam and Metacritic through the use of web API and web scraping. This was followed by text preprocessing, data labelling, feature extraction and finally model training. In the training phase, the effects of oversampling and hyperparameter tuning on the performance of the models have been evaluated. Through comparison between Support Vector Classifier (SVC), Multi-layer Perceptron Classifier (MLP), Extreme Gradient Boosting Classifier (XGB), Logistic Regression (LR) and Multinomial Naïve Bayes (MNB), it was evident that SVC had the most superior performance.

Keywords: Sentiment Analysis, Natural Language Processing, Machine Learning, Support Vector Machine, Game Reviews

1.0 INTRODUCTION

The video game industry is a multi-billion-dollar industry that has become increasingly competitive due to the advancement of technology that has spurred the accessibility of video games and popularized it. Therefore, in order for video game companies to keep up and stay ahead in the market, it is crucial for them to understand the needs and wants of their users. Sentiment analysis helps game developers to uncover the opinions of users towards their games so that they can design and develop their games according to the users' expectations. In view of the above, the specific objectives of this study are:

1. To train multiple machine learning models to classify sentiment of game reviews and compare their performance.
2. To investigate whether oversampling and hyperparameter tuning improve the models' performance.

2.0 LITERATURE REVIEW

This section describes the supervised machine learning algorithms implemented in this study.

2.1 Machine Learning Algorithms

LR is a machine learning algorithm that is used to solve classification issues based on the concept of probability. There are a few assumptions that must be met for LR which include the dependent variable must be dichotomous and the linear relationship between the dependent and independent variable does not exist (Prabhat & Khullar, 2017). An example of real-world application of LR is in the medical field where it can predict the mortality of injured patients (Boyd *et al.*, 1987).

The Support Vector Machine (SVM) is a statistical classification method that determines a hyperplane in an N-dimensional space, where N is the number of characteristics, that categorises data points. It was thought to be the most effective text categorization approach (Xia *et al.*, 2011). It is a non-probabilistic binary linear classifier that can linearly separate classes by a considerable margin, making it one of the most powerful classifiers because of its capacity to handle infinite dimensional feature vectors (Al Amrani *et al.*, 2018). SVC is developed based on SVM and has various applications which include numerical pattern recognition, face detection, text categorization and protein fold recognition (Lau & Wu, 2003).

The Naive Bayes (NB) algorithm is a classification strategy based on the Bayes' Theorem and the assumption of predictor independence. It is mostly used to classify documents at the document level. The main concept behind the technique is that it can calculate the probability of categories given a test document using the combined probabilities of words and categories. The decision-making time for NB classifiers is computationally short and learning can be started without a large amount of data (Ashari *et al.*, 2013). There are a few variations of the NB classifier, namely Multinomial Naive Bayes (MNB), Bernoulli Naive Bayes (BNB) and Gaussian Naive Bayes (GNB).

The XGB is a variation of the Gradient Boosting Machine. The unrivalled scalability of XGB in all circumstances, which consumes considerably fewer resources than previous systems, is its main selling feature. On a single machine, the system is ten times quicker than existing popular methods, and it is scalable to billions of samples in distributed or memory-limited environments. The scalability of XGB is influenced by a number of factors, including system and algorithmic improvements.

For example, a novel tree learning technique is used to handle sparse data, and a theoretically justified weighted quantile sketch procedure is used to handle instance weights. These distinguishing characteristics have made XGB a well-known system for machine learning and data mining problems (Chen & Guestrin, 2016).

MLP is a feed-forward artificial neural network composed of perceptrons which are organised hierarchically in numerous linked layers, each of which is made up of three types of layers: input, output, and hidden. The input signal is transferred to the input layer which will then be passed to the output layer where prediction and classification are performed while the hidden layer performs computational processing in the network to generate network outputs. The goal of MLP network training is to find the optimum collection of connection weights and biases to reduce prediction error (Alboaneen *et al.*, 2017).

2.2 Related Work

Several studies on machine learning-based sentiment analysis have been carried out in the past. Chakraborty *et al.* (2018) performed sentiment analysis on game reviews obtained from Amazon and Twitter. The algorithms which include NB, SVM, LR and Stochastic Gradient Descent (SGD) were used to train sentiment analysis models and the models were evaluated in terms of their accuracies. The feature extraction method used was the Bag-of-Words method.

Zuo (2018) performed sentiment analysis on game reviews collected from Steam. The algorithms applied were NB and Decision Tree classifiers. Feature selection using information gain was carried out, followed by feature extraction through Term Frequency-Inverse Document Frequency (TF-IDF) and hyperparameter tuning of the models through grid search.

A study was conducted by Britto and Pacifico (2020) on video game acceptance by performing sentiment analysis on game reviews. The dataset used was game reviews written in Brazilian Portuguese language extracted from Steam. Feature extraction was performed using the Bag-of-Words method. The algorithms implemented were Random Forest classifier, SVM and LR.

Based on the previous studies, there exists several research gaps for sentiment analysis of game reviews using machine learning techniques. Firstly, there is a lack of implementation of XGB and MLP algorithms. Besides that, exploration on sentiment analysis for more professional and complex reviews written by game critics such as the Metacritic reviews was absent from previous studies. Furthermore, the effect of resampling techniques such as oversampling along with hyperparameter tuning of TF-IDF have not been studied before. Therefore, this study was conducted to fill the above-mentioned research gaps such that XGB and MLP models were trained to compare their performance, Metacritic reviews were included in the training data and an oversampling technique and hyperparameter tuning of TF-IDF were experimented to investigate their effects on the models' performance.

3.0 METHODOLOGY

This section presents the system framework, datasets, text preprocessing, data labelling, and feature extraction, handling of imbalanced classes, model applications and hyperparameter tuning.

3.1 System Framework

Figure 1 summarises the steps involved in preparing the data, training and optimizing the sentiment analysis models.

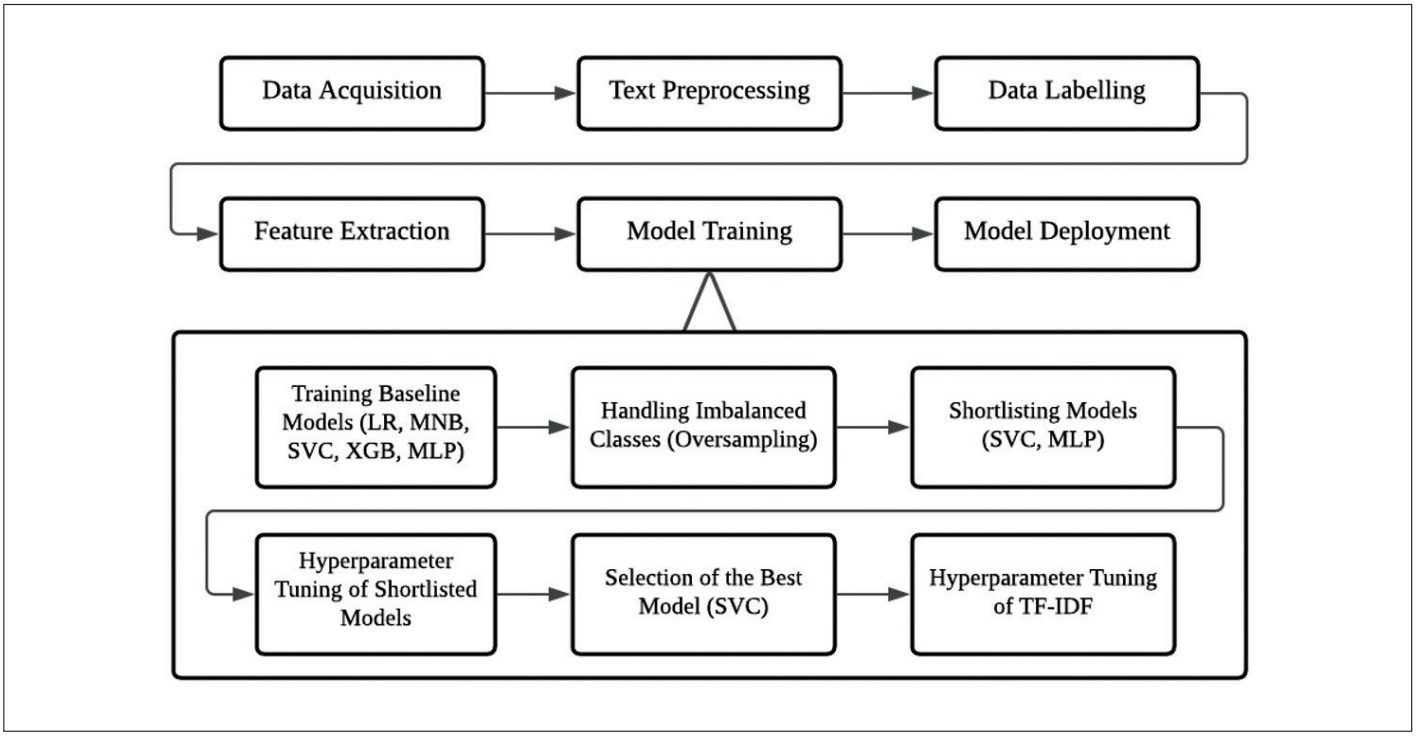


Figure 1: System Framework

3.2 Dataset

The dataset contains 17543 game reviews, 8363 of which were critic reviews on 300 games scraped from Metacritic's website (https://www.metacritic.com/) and 9180 were user reviews on 100 games collected from Steam (https://store.steampowered.com/) through its web API in July 2021. Table 1 shows the sample reviews obtained from Metacritic and Steam.

Table 1: Sample reviews

	Metacritic critic reviews	Steam user reviews
1	Considering how well Valve got the action down pat, I was very unpleasantly surprised that they managed to fumble the storyline so badly. HL2 starts off with such promise and ends with something akin to what you'd find in "[deleted]" which is a travesty of a mockey of a sham, if there ever was one.	This is a game that I've played on and off for several years now. Sometimes I leave because of issues I have with this game, or personal frustration, or some other reason, but I've always come back to my favorite shooty boat game in the end.
2	This jump & run is a must-buy for every fan of the genre. Even beginners should take a look because Limbo shows that you don't need DirectX 11 or a huge story - just two colors and a lot of love.	Very fun and addictive. Me and my daughter played Rise and Generations ultimate on Nintendo switch a bunch, now starting on world. 100% worth playing!!
3	After spending fifty hours with GW2, I have a lot of praise for ArenaNet's work and the way it changes up some of the typical trappings of the MMO. And yet, I find myself thinking less and less about it each day. It's not a declaration against the product, mind you, but simply a fact that this game still is very much an MMO, and your enjoyment will directly relate to how much you enjoy the genre. For many who were hoping for a clean break from MMO design philosophy, Guild Wars 2 will probably come across as a slight disappointment. It pushes the genre slightly forward, however, and could lead to even further development in the future.	Call of Duty: Black Ops III is one of the best zombies experiences I have ever had in my life. The DLC maps really pull the story line together. When Treyarch decided to add another DLC pack to the game (DLC 5/Zombie Chronicles) it truly brought nostalgic experiences to all players that have played the other games in this franchise. The use of Steam Workshop just keeps this games spirit high. If you love zombies this is the game for you. Despite the flaws (campaign) this is a great game.

3.3 Text Preprocessing

The dataset was preprocessed before it was used to train the models. Firstly, HTML tags and hyperlinks were removed. Next, the texts were converted into lowercase and contractions were expanded. Besides that, special characters were removed. This is followed by removal of numbers, single character words, extra whitespaces and stopwords, except for negations such as "no" and "not" because removal of such words would invert the sentiment of the reviews. Then, tokenization and part-of-speech (POS) tagging were performed. The POS tags were passed on to the lemmatizer so that lemmatization can be carried out based on the context of the tokens.

3.4 Data Labelling

The sentiments of the reviews were labelled as positive, negative or neutral by using pretrained sentiment analysis models of three libraries. The models used were NLTK's VADER Sentiment Intensity Analyzer, Textblob's Pattern Analyzer and Flair's TARS Classifier. A majority voting approach was used to determine the final sentiments of the game reviews. There was a total of 10426 positive reviews, 2975 neutral reviews and 2017 negative reviews. The distribution of the labelled reviews is shown in Figure 2.

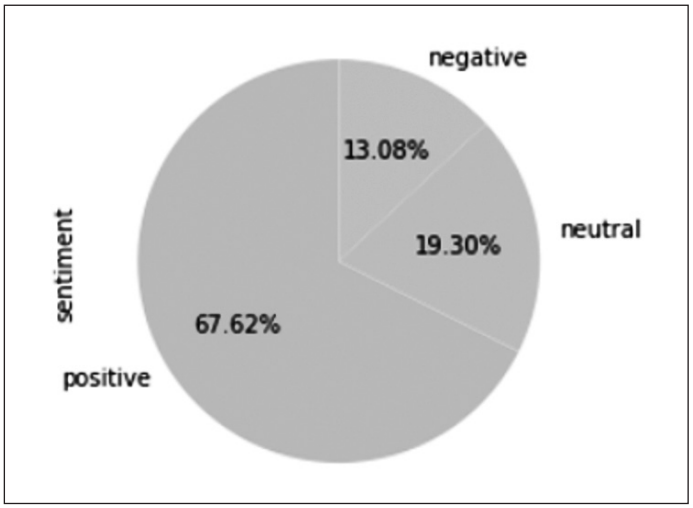


Figure 2: Labelled reviews distribution

3.5 Feature Extraction

The TF-IDF approach has been applied by using Scikit-learn's TfidfVectorizer to perform feature extraction. The "max_features" hyperparameter was set to 2500 while default values were used for other hyperparameters.

3.6 Handling Imbalanced Classes

Oversampling is an approach to deal with data with imbalanced classes by adding more samples to the minority class. Synthetic Minority Oversampling Technique (SMOTE) is one of the oversampling techniques that generates synthetic samples for the minority class. Since the data contains a significantly greater number of positive reviews than neutral and negative reviews which may affect the performance of the models, SMOTE was applied to adjust the distribution of the classes so that all classes have the same number of samples.

3.7 Model Applications

The machine learning algorithms used in this study are as follows:

a. Logistic Regression (LR)
LR is by default used for binary classification but it is extended by the Scikit-learn library to also perform multi-class classification.

b. Multinomial Naïve Bayes (MNB)
MNB is a probabilistic learning method used for classification with discrete features. Scikit-learn's MNB algorithm not only allows the use of integer feature counts, but also fractional counts obtained from TF-IDF.

c. Support Vector Classifier (SVC)
SVC is a classification algorithm that can be used to solve binary and multi-class problems. Scikit-learn's SVC algorithm uses a one-vs-one scheme to support multi-class classification.

d. Extreme Gradient Boosting Classifier (XGB)
XGB is a decision-tree-based ensemble machine learning algorithm that implements gradient boosting. The XGBoost library provides a Scikit-learn wrapper class that allows the XGB algorithm to be used the same way as other Scikit-learn algorithms.

e. Multi-layer Perceptron Classifier (MLP)
MLP is a feedforward Artificial Neural Network (ANN) algorithm that consists of multiple fully connected layers. Scikit-learn's MLP algorithm provides a regularization term that can be used to constraint the size of the weights in the neural network to prevent overfitting.

All the models were trained with their default hyperparameters to obtain their baseline performances except for MLP. Scikit-learn's MLP algorithm has a default architecture that consists of one input layer, one hidden layer with 100 neurons and one output layer, which causes the model to be computationally expensive to train. Therefore, a smaller value for the "hidden_layer_sizes" hyperparameter was set. The MLP model trained comprised 2 hidden layers, with 10 neurons in the first hidden layer and 5 neurons in the second hidden layer. The number of hidden layers and neurons were set arbitrarily as the model only acts as a baseline model before hyperparameter tuning was performed.

3.8 Hyperparameter Tuning

Hyperparameter tuning is the process of determining the combination of hyperparameters which maximizes the model's performance. In order to improve the performance of the models, a Randomized Search Cross Validation with 3 splits was carried out to find the best combination of hyperparameters. In addition, a Grid Search Cross Validation with 3 splits was also performed on TF-IDF to select the best hyperparameters for it to further improve the performance of the models.

4.0 RESULTS AND DISCUSSION

Table 2 and Table 3 show the baseline performance of the models trained on the imbalanced dataset and oversampled dataset obtained through cross validations. Weighted precision, weighted recall and weighted F1-score were used as the metrics as they take into account the number of instances in each class.

Table 2: Baseline performance of all models trained on imbalanced dataset

	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-Score		
				Negative	Neutral	Positive
LR	74.8%	71.7%	74.8%	71.6%	71.1%	72.2%
MNB	68.3%	63.3%	68.3%	57.2%	56.8%	56.9%
SVC	72.9%	69.9%	72.9%	66.6%	67.4%	67.6%
XGB	75.9%	73.5%	75.9%	74.1%	72.6%	73.3%
MLP	69.4%	70.1%	69.4%	70.1%	69.1%	70.0%

Table 3: Baseline performance of all models trained on oversampled dataset

	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-Score			
				Negative	Neutral	Positive	Status
LR	79.3%	79.6%	79.3%	79.1%	79.5%	79.6%	Rejected
MNB	67.4%	67.5%	67.4%	67.2%	66.7%	67.1%	Rejected
SVC	87.7%	88.7%	87.7%	87.4%	87.4%	88.0%	Accepted
XGB	79.7%	80.1%	79.7%	80.0%	79.9%	79.5%	Rejected
MLP	86.8%	87.0%	86.8%	86.5%	86.8%	86.9%	Accepted

Based on the results in Table 2 and Table 3, oversampling has significantly improved the performance of all the models except MNB which was observed to have a drop in accuracy and weighted recall. The improved performance was due to the class distribution being balanced after performing duplication of data to synthesize new data from the minority classes.

MNB performed poorer on the oversampled data and was the worst-performing model most likely due to its assumption that all features are independent which is rarely true in real-world use cases where there are a large number of features.

The most significant improvement of performance was observed in SVC and MLP. These two models worked well with the larger, balanced dataset and they were also the two best-performing models. Therefore, they have been shortlisted for hyperparameter tuning through Randomized Search Cross Validation. The best combinations of the models' hyperparameters and their performances are shown in Table 4 and Table 5 respectively.

Table 4: Best hyperparameter values of SVC and MLP

	Best hyperparameter values
SVC	kernel: rbf, gamma: scale, C: 10
MLP	"solver": "adam", "max_iter": 150, "hidden_layer_sizes": (10,), "alpha": 0.0001, "activation": "relu"

Table 5: Performance of fine-tuned SVC and MLP

	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-Score			
				Negative	Neutral	Positive	Status
SVC	89.7%	90.0%	89.7%	89.2%	89.7%	90.1%	Accepted
MLP	87.0%	87.1%	87.0%	86.7%	87.0%	87.0%	Rejected

Table 5 shows that hyperparameter tuning has improved both models' performance. Hyperparameter tuning is able to improve the models' performance because it determines the best combinations of hyperparameters which produce optimal models that minimize the loss functions.

SVC outperformed MLP in terms of accuracy, weighted precision, weighted recall and weighted F1-score after the hyperparameter tuning. Hence, SVC as the best-performing model among all the models, was selected to test the effect of hyperparameter tuning of TF-IDF on its performance. The tested values of the TF-IDF hyperparameters and the best values determined by Grid Search Cross Validation are shown in Table 6.

Table 6: Tested hyperparameter values and the best values of TF-IDF

Hyperparameter	Tested values	Best value
max_features	2500, 5000, 10000	2500
max_df	0.25, 0.5, 0.75	0.25
ngram_range	(1, 1), (1, 2), (1, 3)	(1, 1)

As can be seen in Table 6, SVC had the best performance under the condition in which the top 2500 terms across the corpus ordered by term frequency were considered, terms that occurred in more than 25% of the documents were ignored and only unigrams were extracted. The performance of SVC trained with the features extracted by the fine-tuned TF-IDF is shown in Figure 3.

	precision	recall	f1-score	support
negative	0.97	0.91	0.94	2044
neutral	0.91	0.87	0.89	2082
positive	0.87	0.96	0.91	2130
accuracy			0.91	6256
macro avg	0.92	0.91	0.91	6256
weighted avg	0.92	0.91	0.91	6256

Figure 3: Classification report of SVC with fine-tuned TF-IDF

Based on the classification report in Figure 3, hyperparameter tuning on TF-IDF has improved SVC's performance. The model has achieved an accuracy of 91%, precision of 92%, recall of 91% and F1-score of 91%.

5.0 CONCLUSION

In conclusion, by training five machine learning models with game reviews acquired from Metacritic and Steam, this study has provided clear evidence that oversampling will lead to an improved performance for most models. In addition to that, better results are also evident after hyperparameter tuning on the models and TF-IDF have been performed.

Support Vector Classifier with an accuracy of 91 percent has emerged as the best-performing model among the five models after comparing the accuracy scores for each model. The way SVC performs classification, which is based on hyperplanes

instead of probabilities, is likely to be the main contribution to its outstanding performance. It is therefore the ideal model to be used for text classification tasks with a large number of features such as sentiment analysis.

Future research into machine learning algorithms for sentiment analysis of game reviews should focus on exploring the sentiment of emojis and emoticons since it is common for users to incorporate them in their reviews. Furthermore, experimenting with ensemble methods are required to gain more insight into building a more robust sentiment analysis model.

6.0 ACKNOWLEDGEMENTS

Authors thank the Faculty of Computing and Information Technology, Tunku Abdul Rahman University College for financial support and resources to carry out this study. ■

REFERENCES

[1] Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. Procedia Computer Science, 127, 511-520. https://doi.org/10.1016/j.procs.2018.01.150.

[2] Alboaneen, D. A., Tianfield, H., & Zhang, Y. (2017). Sentiment analysis via multi-layer perceptron trained by meta-heuristic optimisation. 2017 IEEE International Conference on Big Data (Big Data), 4630-4635. https://doi.org/10.1109/BigData.2017.8258507.

[3] Ashari, A., Paryudi, I., & Tjoa, A. M. (2013). Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. International Journal of Advanced Computer Science and Applications (IJACSA), 4(11). https://doi.org/10.14569/IJACSA.2013.041105.

[4] Boyd, C. R., Tolson, M. A., & Copes, W. S. (1987). Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. The Journal of trauma, 27(4), 370-378. https://doi.org/10.1097/00005373-198704000-00005.

[5] Britto, L. F., & Pacıfico, L. D. (2020) Evaluating Video Game Acceptance in Game Reviews using Sentiment Analysis Techniques. Proceedings of SBGames 2020, 399-402.

[6] Chakraborty, S., Mobin, I., Roy, A., & Khan, M. H. (2018). Rating Generation of Video Games using Sentiment Analysis and Contextual Polarity from Microblog. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 157-161. https://doi.org/10.1109/CTEMS.2018.8769149.

[7] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794. https://doi.org/10.1145/2939672.2939785.

[8] Tan, J. Y., Chow, A. S. K., & Tan, C. W. (2021). Sentiment Analysis on Game Reviews: A Comparative Study of Machine Learning Approaches. International Conference on Digital Transformation and Applications (ICDXA2021), 209-216.

[9] Lau, K. W., & Wu, Q. H. (2003). Online training of support vector classifier. Pattern Recognition, 36(8), 1913-1920. https://doi.org/10.1016/S0031-3203(03)00038-4.

[10] Prabhat, A., & Khullar, V. (2017). Sentiment classification on big data using Naïve Bayes and logistic regression. 2017 International Conference on Computer Communication and Informatics (ICCCI), 1-5. https://doi.org/10.1109/ICCCI.2017.8117734.

[11] Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. Information sciences, 181(6), 1138-1152. https://doi.org/ 10.1016/j.ins.2010.11.023.

[12] Zuo, Z. (2018). Sentiment Analysis of Steam Review Datasets using Naive Bayes and Decision Tree Classifier. Student Publications and Research - Information Sciences. http://hdl.handle.net/2142/100126.

PROFILES

TAN JIE YING received her Bachelor's degree of Computer Science (Honours) in Data Science from Tunku Abdul Rahman University College (TAR UC), Malaysia in 2022. She is passionate about creating value from data using Machine Learning, Natural Language Processing and Computer Vision.
Email address: tanjy-wp17@student.tarc.edu.my

ANDY CHOW SAI KIT is a Bachelor of Computer Science (Honours) in Data Science graduate from Tunku Abdul Rahman University College (TAR UC), Malaysia in 2022. His research interest is in Artificial Intelligence and Machine Learning.
Email address: andycsk-wp17@student.tarc.edu.my

DR TAN CHI WEE received BCompSc(Hons) and PhD degrees in year 2013 and 2019 respectively in Universiti Teknologi Malaysia. Currently, he is a Senior Lecturer cum Programme Leader at Tunku Abdul Rahman University College and actively involved in the Centre of Excellence for Big Data and Artificial Intelligent (CoE) and become the research group leader for Audio, Image and Video Analytics Group under Centre for Data Science and Analytics (CDSA). Dr Tan's main research areas are Computer Vision (CV), Image Processing (IP) and Natural Language Processing (NLP) and Artificial Intelligence (AI). He is an enthusiastic researcher experienced in conducting and supporting research into Image Processing. Being a meticulous and analytical researcher with Train-The-Trainer certificate of many years of educational and hands-on experience, he was invited to Université d'Artois (France) under Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) programme for collaborative research between European countries with Southeast Asian countries on motion detection and computer vision and being involved in industry project as professional consultant.
Email address: chiwee@tarc.edu.my

CAPACITIVE INTERDIGITATED ELECTRODES SENSOR FOR THE FIELD DEVICE TO MEASURE MOISTURE CONTENT IN THE NITRILE GLOVES MANUFACTURING INDUSTRY

(Date received: 13.01.2022/Date accepted: 01.04.2022)

Vishnukumar Rajandran[1,4]*, Lee Wah Pheng[2], Yip Mum Wai[1,3], Lim Joo Eng[1,3], Tan Yoke Meng[4]

[1] Faculty of Engineering and Technology, Tunku Abdul Rahman University College, 53300 Kuala Lumpur, Malaysia.

[2] Centre for Postgraduate Studies and Research, Tunku Abdul Rahman University College, 53300 Kuala Lumpur, Malaysia.

[3] Centre for Systematic Innovation Research, Tunku Abdul Rahman University College, 53300 Kuala Lumpur, Malaysia.

[4] Research and Development Centre, Factory 25, Top Glove Sdn. Bhd., 41050 Klang, Malaysia.

**Corresponding author: vishnukrajandran@gmail.com*

ABSTRACT

This paper investigates the impedance spectroscopy technique in determining moisture content in Nitrile gloves. Interdigitated electrode was designed and fabricated, then evaluated on LCR Impedance meter subjected to frequency range of 100Hz, 120Hz, 1kHz, 10kHz, 20kHz and 100kHz. Samples of Nitrile gloves were compounded and prepared for different moisture content level and regression analysis was performed to evaluate the relationships between capacitance and moisture content of the glove samples. Experimental results indicated that the capacitance value is a strong function of moisture content in gloves and also that the capacitance of moisture content in Nitrile gloves decreased with increasing drying time over the measured frequency range whilst statistical analysis results have confirmed that the 1kHz, 10kHz and 20kHz signal frequencies have highest reliable prediction of the nitrile gloves' moisture content with high R^2 value of 0.96, 0.97 and 0.97, respectively. The ability to determine average moisture content of Nitrile gloves via a non-destructive and online method, utilizing a low-cost instrument, will be of considerable use in the glove industry. This method could also be extended to other types of gloves and rubber products.

Keywords: Interdigitated electrode sensor (IDE), Capacitive sensor, Field Device, Industry 4.0, Moisture content, Nitrile Glove

1.0 INTRODUCTION

Nitrile rubber also named as Acrylonitrile Butadiene rubber or Nitrile Butadiene Rubber (NBR)(Yew *et al.*, 2019), a copolymer of butadiene and acrylonitrile, has many advantages such as low cost, low allergic risk, good chemical resistivity, low gas permeability, excellent penetration resistance, good dexterity and has static dissipation behaviour (Yew *et al.*, 2019). Therefore, gloves are widely used in many applications, for instance, medical, industrial, laboratory, pharmaceutical, food preparation and processing (Critchley & Pemberton, 2020), (Yew *et al.*, 2019). Vulcanisation is a vital process in producing quality gloves. Vulcanisation explains the curing process of the raw (unvulcanised rubber) material mixed or compounded with curatives, antioxidants, and stabilisers to form the dry (vulcanised) rubber at high temperatures between 140°C and 160°C in the curing ovens (Adam *et al.*, 2020) (Yip *et al.*, 2002).

The common manufacturing of nitrile gloves via dipping process consists of the following steps (Yip *et al.*, 2002) (Yew *et al.*, 2019). First, the latex is compounded with a crosslinker, accelerator, initiator, antioxidants, stabilisers, pigments and additives depending on the product requirement and applications by customers. Second, porcelain mould, commonly known as 'formers', are cleaned and dried before being dipped in coagulant nitrate solutions. After drying the coagulant on the former's surface, they dip the formers into the compounded latex at an ambient temperature. Next, the beading of gloves takes place. This process is to ease gripping of the glove during stripping section. Later, the wet latex films are dipped in a hot water bath called leaching process. This process is to remove chemical particles and materials on latex films. Then, the gloves enter the curing Oven at a high temperature of 70°C to 150°C. The curing or drying of a glove is a critical process for the vulcanisation of latex to take place to prevent defects like blisters or porous forming on the latex film (Chambers, 2017). The cured gloves are further dried in leaching ovens to remove excess moisture and chemical particles from the latex films. Figure 1 below shows the common manufacturing process for nitrile gloves. There are post-treatment methods in certain manufacturing process such as chlorination, polymer coating, and hydrogel coatings which can be referred to here (Yew *et al.*, 2019).

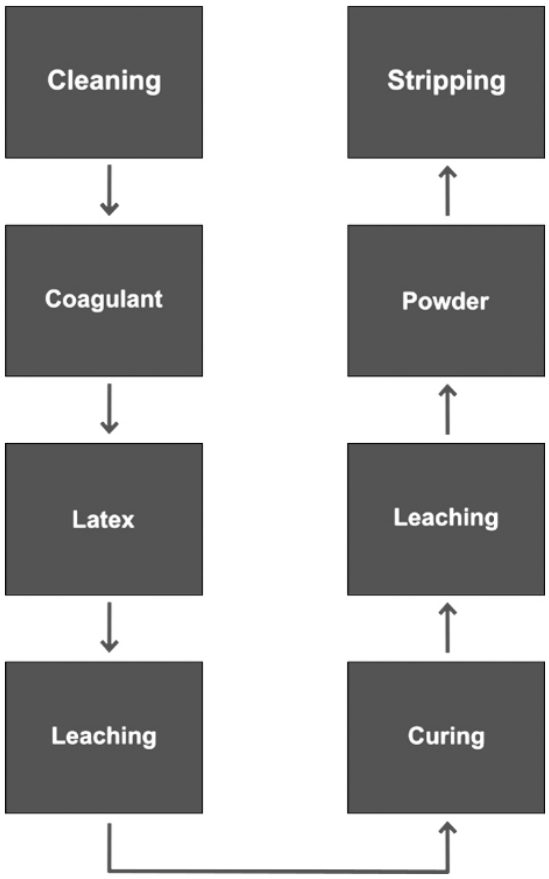


Figure 1: Common Manufacturing process of Nitrile gloves

Moisture content is an essential parameter for quality assessment of latex gloves in the industry; due to its relevance in raw material assessment, mechanical strength and finished product quality as well as in terms of commercial value. Although faster analyser-based techniques are available, the glove industry has traditionally relied on standard inspection method of ASTM D5668 – 19 or familiarly known as the oven drying from monitoring processes up to the release of end product. This is offline, destructive, time-consuming and usually limited to analysing few samples during the process. Furthermore, sampling and preparation can lead to significant analytical errors. Not only in glove making processes but also rubber latex cultivation find moisture content (MC) analysis critical since it greatly affects quality, stability and processing of latex. In the past decades several multiple methods and technologies have been developed to measure MC in rubber latex more rapidly, non-destructively and precisely For instance, infrared, microwave, X-ray, radar, conductive and capacitive.

Near infrared technique uses reflectance and absorbance principles to calculate the moisture content (Puttipipatkajorn & Puttipipatkajorn, 2020; Suchat *et al.*, 2015). The water absorbs certain wavelengths of light in the NIR region. The higher the moisture content, the higher the amount of light absorbed. The filtered beam is directed onto the surface of the sample. A portion of the light is reflected to a detector. Moisture content can be calculated form the amplitude ratio of the reflected wavelengths of the sample beam and the reference beam. Recent studies involving NIR region have shown NIR is a suitable method for quantifying trace amounts of moisture in a rubber sheet due to the strong combination of absorption bands for water at around 1940nm and the first, second and third overtones at 1450nm, 970 and 760nm, respectively. In another method, a calibration models such as Partial Least Square Regression (PLSR), Least Square support and ANN using NIR is constructed to predict the moisture content in a rubber sheet in the wavelength range 900-1700nm. NIR spectroscopy is a fast and non-destructive analytical method. Infrared method can only measure the moisture content of thin films or layers of material, paper, and other thin materials since this non-contact method detects only the material's surface moisture. Therefore, effects of bulk material's particle size, particle shape, particle surface characteristics, and colour may cause high errors in measurement.

Microwave method (Yahaya *et al.*, 2014, 2015) transmits microwaves at a material and then calculates the energy losses emitted from the material and speed variation due to microwave propagation from the material to the moisture content. This technique is based on transmission. An emitter and a receiver are mounted opposite of each other in the process, so they can shine through the material. For this reason, the measuring setup of this method depends on the space between emitter and receiver. Microwave radiation, (frequencies between 1 GHz and 100GHz), has advantages like penetration depth is much larger than that of infrared radiation and permits the sensing of a significant volume of material being transported on a conveyor or in a pipe. Also, water reacts specifically with certain frequencies in the microwave region allowing small amounts of water to be detected. Yahaya and co-workers have conducted studies on the dielectric constant property from the measured reflection coefficient as a function of moisture content at 1 GHz to determine moisture content in rubber latex using Agilent Open Ended Coaxial probe. The lowest mean relative error between actual and predicted moisture contents was 0.02 at 1 GHz when using the Cole-Cole dielectric constant calibration equation. This aforementioned method is an offline, contact method and also, used to measure moisture content of latex in liquid form in the laboratory. Microwave radiation are not convenient to determine the moisture contents of a large number of contact points in gloves. In addition, the equipment of each method is expensive.

X-ray method (Chen *et al.*, 2011) measures the moisture content in material with irradiation of X-ray beams into the material. This method estimates the speed losses of the beam after they pass through the material's water molecules. Synchrotron small-angle X-ray scattering technique was employed to investigate the drying dynamic of latex dispersion. The results obtained were beneficial in understanding the mechanism of latex film formation especially providing clear insights into the effects of temperatures and relative humidity on the evaporation of water after the deformation of latex particles. This method is extremely expensive and not frequently applied in manufacturing process like other methods. Contrary to X-ray method, radar is an advanced technology based on the propagation velocity (PV) of ground penetrating radar (GPR) signals. Radar monitors material's moisture content with the measurement of the travel time through the sample and early time amplitude of the radar signal. The moisture content of testing material with infinite media often cannot be measured by this method. The testing material size effects on accuracy of the measurement. GPR

technology is increasingly being used by researchers to determine moisture content including, but not limited to building material, surface soil and log (Hans *et al.*, 2015; Huisman *et al.*, 2003; Klewe *et al.*, 2021). However, to my knowledge, there have been no literature reports investigating moisture content in gloves or rubber latex using radar technology.

To measure electrical properties of products that are resistance or capacitance, correlating to moisture content is an alternative method widely used in the manufacturing industry. A resistance method (Kueseng *et al.*, 2013; Naphon *et al.*, 2020) uses two electrodes inserted directly into the rubber compound to measure its resistivity. As moisture increases, the rubber compound's electrical resistance decreases. The resistance could vary between several hundred kΩ when wet as opposed to several thousand MΩ and more when dry. In a study conducted to determine the effect of wetting intensity on electrical surface resistance in silicone rubber, it can be observed that in dry conditions the surface resistance of the rubber has a large value of 1085 MΩ. In comparison to wet conditions, the resistance decreases to a low value of 253.99 MΩ. Clearly, the higher the wetting intensity, the lower the surface resistance of rubber.

Capacitive sensing provides better advantages such as measurements can be implemented online, simple, low cost and rapid (Bhuiyan *et al.*, 2015; Döring *et al.*, 2019; Khaled *et al.*, 2015). Online and real-time measurement is important, especially where one must not turn off the machine and collect samples to test the MC. Hence, capacitance can be applicable in the production line for continuous monitoring of MC. Recent studies have established several online capacitive methods of measuring moisture which can have either non-contact or contact measurement depending on its application. One of such is measuring MC of a moving stream of spray-dried gelatin powder (Wang *et al.*, 2017). A non-contact capacitance sensing system encompassing test capacitors, signal processing and data acquisition device was applied in MC determination on spray-dried gelatin with different water content (4 – 44% MC). The copper plate electrodes of test capacitors generate an electric field within the measuring cylinder. The strength of capacitance was related to the permittivity of the material between the electrodes. Each electrode was connected to a capacitance-to-voltage transducer so that the output signal could be collected using a data acquisition device. The application of a 4-electrode system and dual-frequency sensing system has improved the accuracy (R^2 = 0.9) and reliability of the online MC measuring system. However, the measurement or sampling rate needs to exceed the minimum sampling requirement of data acquisition device.

Besides that, the contact capacitive measurement method utilising interdigital electrode (IDE) sensor has gained increasing attention the last decades due its simplicity and low cost. The technique is suitable to evaluate MC in wood, wood chip and wood pellets (Chetpattananondh *et al.*, 2017) because of the high relative permittivity of water ($\varepsilon_r \approx 80$). The variation of the material permittivity due to the change of MC can be measured as a value of capacitance when the test material is placed between two electrodes. In other words, the fringing capacitance measured between the electrode varies with the dielectric constant, which varies with the MC in material. Therefore, measurement of the capacitive values for the material's moisture properly can be operated. The experimental results indicate IDE sensor has good repeatability and linearity. Also, the sensor offers great benefits in being cost effective, easy to use and portable with rapid measurement and non-destructive. To date, there is no literature reports investigating the application of IDE sensor in determining glove's moisture content.

2.0 DETERMINATION OF MOISTURE CONTENT IN NITRILE RUBBERS

Nitrile rubbers are categorized as polar rubber just like other polymers including acrylic rubbers, hydrogenated nitrile rubber and ethylene-acrylate terpolymers. By the term "polar rubber," it is meant that the rubber contains atoms other than hydrogen or carbon such as nitrogen or oxygen as in nitrile rubber, acrylic rubber, or copolymers of acrylic rubber. The nitrile rubbers are combined polymers of acrylonitrile with a conjugated diene having anywhere from 4 to 8 carbon atoms, with butadiene being highly preferred (Patel *et al.*, 1996). Polymers such as nitrile rubbers are widely used as dielectric materials due to their high flexibility, tractable processing as well as good chemical stability and readily tunable properties. Their dielectric constant is lower than non-polymeric materials.

Dielectric constant relates to the permittivity ε, of the material. The permittivity explains the ability of a material to polarize in response to an applied field. In other words, greater the polarization developed by a material in an applied field of given strength, the greater the dielectric constant will be. The mechanism which contributes to the dielectric properties are the interaction of electric field with electronic, atomic and dipole polarization. The relation between permittivity of the dielectric material with polarizability is $\varepsilon_r = 1 + \frac{N_a \alpha}{\varepsilon_0}$,where ε_r is the relative permittivity, ε_0 is the permittivity in vacuum, α is polarizability and N_a is the Avogadro constant.

Polarizability refers to the proportionality constant for the formation of dipole under the influence of electric field. The polarizability depends on applied field frequency, it has a strong frequency dependence, besides the conductivity and permittivity, because it is a complex function (Ahmad, 2012; Kosumphan *et al.*, 2018). The dielectric properties of most materials depend on many factors, including frequency of the applied alternating electric field, chemical composition and structure of the material, and especially permanent dipole moments associated with water and any other molecules making up the material of interest (Wang *et al.*, 2017). Multiple studies to explain the polarizability and the orientational effects of acrylonitrile–butadiene rubber (NBR) was done by observing changes in these electrical properties (Kueseng *et al.*, 2013; Zhao *et al.*, 2015; Zhu and Zhang, 2017).

For example, studies on changes of electrical properties were introduced and some sort of instruments were recommended to be used in the food industry (Mohamad *et al.*, 2015; Sairin *et al.*, 2019). Parallel plate electrodes are one of the generally used probes to sense the moisture content in peanut oil (Butts, 2008). Another example is by measuring capacitance using a pair of copper electrodes in spray dried products for a non-contact measurement (Wang *et al.*, 2017). Apart from this, Son's findings suggest that the electrical resistivity could be used as an effective alternative for estimating the weathering degree of soil (Son *et al.*, 2010).

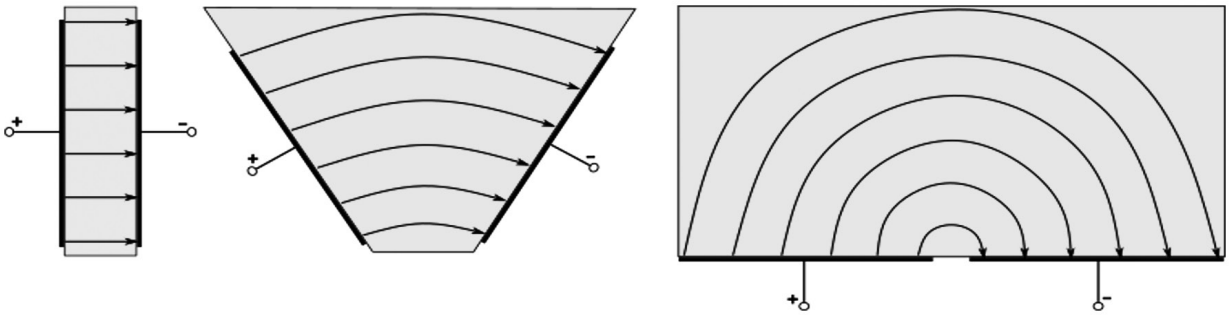


Figure 2: Transition from the parallel-plate capacitor to a planar capacitor

The capacitive interdigitated electrodes (IDE) sensor is a coplanar structure encompassing of multiple interpenetrating comb electrodes. The working principle of the interdigital coplanar capacitive sensor is similar to the two parallel plate capacitors. The parallel plate capacitor is transformed to the interdigital capacitive sensor as shown in Figure 2. When both electrodes are excited by the different voltages to generate fringing electric fields between electrodes, these electric fields then travel from positive electrode to negative electrode while passing through the material in contact with the electrodes. Thus, the material's dielectric properties affect the impedance of electric fields between these electrodes. The sensor behaves as a capacitor in which the capacitive reactance becomes a function of material properties. The fringing capacitance measured between the electrode varies with the dielectric constants, which varies with the moisture contents in material (Afsarimanesh *et al.*, 2019). Therefore, measurement of the capacitive values for the material property measurement can be operated.

As a result of the high sensitivity and simplicity of the sensor, the interdigital capacitive sensor is widely used in different applications such as biosensor for bacterial detection(Varshney & Li, 2009) soil moisture (Markevicius *et al.*, 2012), lard detection (Mohamad *et al.*, 2015), rubberwood (P. Chetpattananondh *et al.*, 2017), concrete moisture (Alam *et al.*, 2010), humidity (Rivadeneyra *et al.*, 2014), and water level measurement (K. Chetpattananondh *et al.*, 2014). Interdigitated electrodes (IDEs) sensor is also effectively being implemented in sensing devices such as, but not limited to, piezoresistive sensors, chemical sensors, environmental monitoring sensors and MEMS biosensors (Ferrari & Prudenziati, 2012). IDE is also used to study oil degradation in determining frying oil quality (Khaled *et al.*, 2015). Bioimpedance measurement utilizing IDE is a well-established method for the detection and characterization of cancerous cells (Alexander *et al.*, 2010). Therefore, IDE sensors could be used in solving complex calibration requirements and improving the accuracy of sensory sensitivity. IDE shape configurations have some advantages such as non-moving parts, ease of fabrication, are flexible in design as well as cost-effectiveness (Bhuiyan *et al.*, 2015; Döring *et al.*, 2019).

Therefore, IDE could be used in solving complex calibration requirements and improving the accuracy of sensory sensitivity. IDE shape configurations have some advantages such as non-moving parts, ease of fabrication, are flexible in design as well as cost effective (Döring *et al.*, 2019; Bhuiyan *et al.*, 2015). The purpose of this research is to develop a new sensor to determine the moisture content of nitrile gloves by measuring the changes occurring in capacitance during the curing process of gloves. To achieve this, a capacitive sensor was designed by integrating IDE platform to assess different moisture content in gloves at varying frequency.

3.0 MATERIALS AND METHODS

3.1 Capacitive Sensor Design

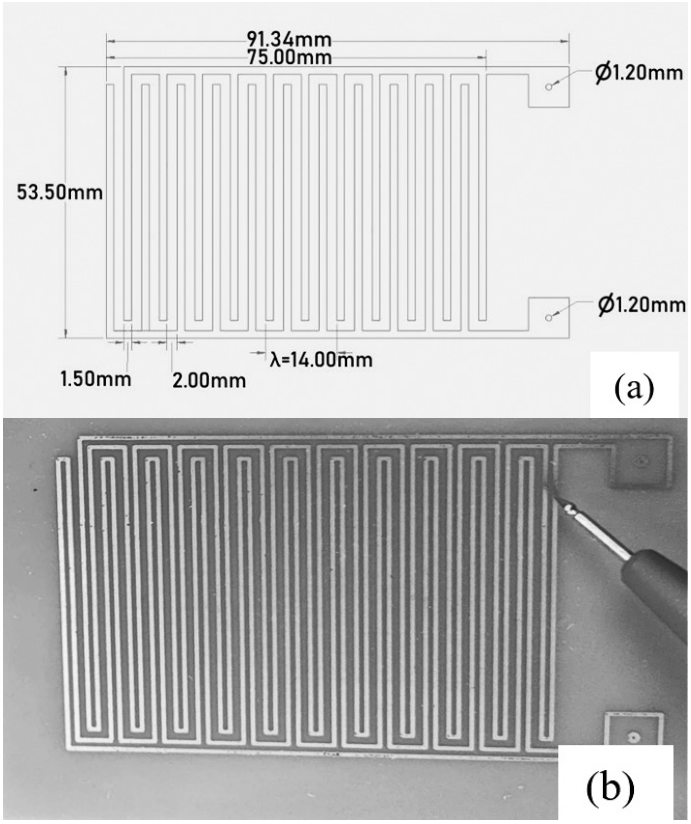


Figure 3: (a) Low-cost IDE sensor drawing in Solidworks (b) IDE sensor with 22 number of electrodes

The capacitive sensor was designed based on the interdigitated electrodes (IDE) as shown in Figure 3. The sensor was drawn using Solidworks software and then fabricated using conventional photolithography and etching process. The aforementioned fabrication process is adapted from previous research conducted by Zoolfakar and his team (Zoolfakar *et al.*, 2010). This IDE sensor is fabricated with copper electrodes and fiberglass as printed circuit board (PCB) substrate. This sensor gives high sensitivity due to the strong effect on signal area in the numbers of electrode pairs which produces uniform electrical field distribution and measurable output signal (Chetpattananondh *et al.*, 2017). The capacitance of the sensor is varied with dielectric constant of material due to change of the moisture content. Neglecting edge effects, the sensor capacitance C can be computed from the capacitance per unit length C_{uc} of a 2D cell formed by an electrode pair yielding equation (1) (Ferrari and Prudenziati, 2012).

$$C = C_{uc}(N-1)L\,, \tag{1}$$

where N and L are the number and length, measured by mm, of the finger electrodes, respectively. The capacitance per unit cell C_{uc} of electrode pair attached with the material is given by equation (2) (Alam *et al.*, 2010).

$$C_{uc} = \varepsilon_0 \frac{\varepsilon_g + \varepsilon_s}{2} \times \frac{K\sqrt{1-(\frac{a}{b})^2}}{K\left(\frac{a}{b}\right)} + \varepsilon_0 \frac{\varepsilon_g h}{a}\,, \tag{2}$$

where ε_0 is the dielectric constant in free space, $\varepsilon_0 = 8.8542 \times 10^{-12} F/m$, ε_g and ε_s are the dielectric constants of moisture content and the substrate, respectively. Also *a, b* and *h* are the finger spacing width, distance (pitch) and thickness, respectively. K[x] is the complete elliptic integral of the first kind given by equation (3) (Abramowitz *et al.*,1965).

$$K[x] = \int_0^{\pi/2} \frac{1}{\sqrt{1-x^2 sin^2\theta}} d\theta \tag{3}$$

Figure 4 shows a unit cell of an interdigitated sensor without the conducting plane. The variables of the sensor are the number of the electrodes N, width of the electrode w, electrode space s, and the length of the electrode L, with dimension of 22, 1.5mm, 2mm and 75mm, respectively. Every other electrode finger is connected electrically together through a common electrode arm. The variables were suggested by the pioneer work on MC determination (Chetpattananondh *et al.*, 2017). The overall capacitance C between the electrode pair is varied because of variation in the electrode pair attached to the dielectric medium of material. Thus, the moisture content measurement in NBR glove can be determined in term of the varied capacitance of the electrode pair attached on the sample.

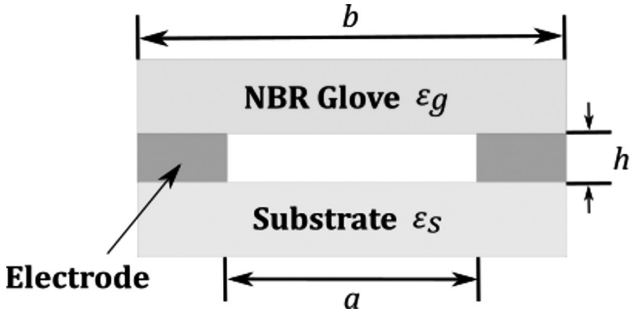


Figure 4: Unit cell of an interdigitated sensor without the conducting plane

3.2 Sample Preparation

Nitrile Butadiene Rubber (NBR) latex is compounded at the chemical lab to prepare a 14% TSC (Total Solid Content) latex. Potassium Hydroxide, pH Adjuster, Accelerator, Metal Oxide Crosslinker, Wetting agent, Surfactant, Opacifier, Antifoam, and water are added during the compounding process. The chemical formulation for preparing 14% TSC Nitrile glove is not disclosed because it is Top Glove's proprietary formula. The compounded sample is stirred at 300rpm for 24 hours for the maturation process to take place.

To prepare sample, a ceramic former mold is used for dipping process throughout this study. The ceramic former is first cleaned at the beginning of the experiment. The ceramic former and coagulant solution are heated to a temperature of 65°C using an immersion heater, this enables the coagulant to coat the former evenly which helps in picking up latex and controlling the thickness of latex film. The ceramic plate is then dipped for 10 seconds in the coagulant solution, which has 8.0 - 9.0% ± 0.5% of Calcium Nitrate. After dipping in the coagulant solution, the former is heated in an oven at 120°C for 5 minutes. Next, the hot former is placed in a desiccator to cool down until temperature drops to 60°C - 65°C. This is then dipped into the latex compound for 8 seconds. Now a wet gel-like film will form over the ceramic former. The ceramic former is withdrawn and is dried in the oven for curing of latex. The curing time of the sample ranges from 1 to 20 minutes, where one sample was removed from the oven every 1 minutes. After curing, all glove samples were kept in petri dish for further analyses.

3.3 Electrical Capacitance and Moisture Content Measurement

The distinction among each glove sample was analyzed by measuring its electrical capacitance and moisture content (MC). The capacitance was measured using the custom built IDE sensor pressed onto the glove sample. The sensor was connected to a LCR meter (4263B, Agilent, Malaysia) with Kelvin clip leads (TH26011AS, Changzhou Tonghui Electronic Co. Ltd, China) as depicted in Figure 5. The LCR meter has a frequency range from 100Hz and 100kHz (Afsarimanesh *et al.,* 2019).

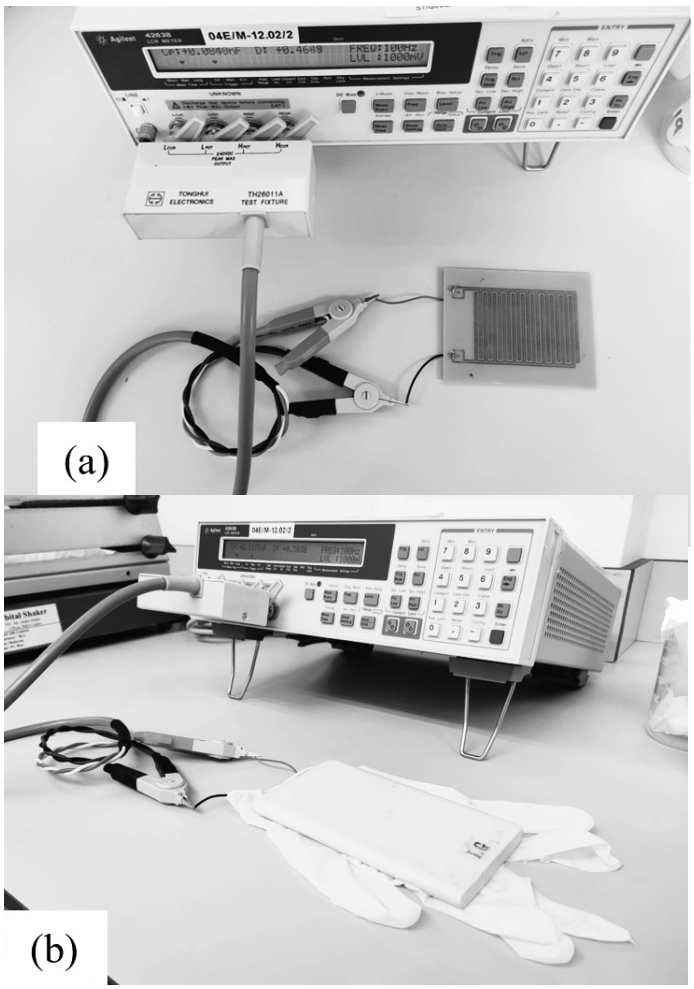


Figure 5: (a) The 4263B Agilent LCR meter connected to interdigitated electrode via TH26011AS Kelvin clip leads (b) Glove is positioned on sensor

Before starting the measurements using the LCR meter, calibration was performed following the standard procedure of the instrument operation manual. Glove samples is pressed on interdigitated electrode with heavy ceramic plate to make sure close contact of sample to sensor. The ceramic plate used is flat and with good surface finish which does not affect the glove's physical properties such as thickness. Then the MC of each dried sample was measured using a Moisture Analyzer (MB120, OHAUS, China). Glove moisture content can be determined by gravimetric method as shown in equation (4).

$$\%MC = \frac{m_w - m_0}{m_0} \times 100 \quad (4)$$

where m_w is mass of wet glove and m_0 is mass of dried glove. Calibration or adjustment of the Moisture Analyzer is not necessary for a correct moisture determination as the measurement is relative. The balance determines the weight of the sample before and after drying and the moisture is calculated on the basis of the ratio between wet and dry weights. After each testing, the IDE sensor and pan in the Moisture Analyzer was cleaned by soft tissues. Each sample is measured thrice for each curing time ranging from 1 to 20 minutes. The experiment is repeated thrice.

3.4 Statistical Analysis

Regression analysis was performed to evaluate the relationship between electrical capacitance with moisture content of the glove samples. The regression equations were evaluated by the coefficient of determination (R^2) and the root mean square error (RMSE) calculated by equation (5) (Chetpattananondh *et al.*, 2017).

$$RMSE = \sqrt{\frac{1}{N_s}\sum_{n-1}^{N}(Y_t - Y_e)^2} \quad (5)$$

where N_s is the number of samples in the dataset, Y_t is the predicted value calculated using the regression equation and Y_e is the measurement obtained through experimental procedures.

4.0 RESULTS AND DISCUSSION

This study is aimed to design a new sensor to measure moisture in Nitrile gloves. The glove's moisture can be quantified with impedance spectroscopy technique using IDE capacitance method. The capacitance measurements were analyzed to evaluate the determination for gloves moisture content. Overall, the capacitance of glove decreased as the heating time increased. For example, as the heating time progressed from 1 to 8 minutes, a rapid drop in capacitance was observed (16.84μF to 62.73pF) at a frequency of 1kHz, along with a decrement of MC values from 50.10% to 2.40% (Figure 6). The capacitance measured by the IDE sensor exhibits good correlation to MC measured using moisture analyzer. Figure 7 shows the regression of capacitance measurements with MC values of gloves at 20kHz during different drying time. This result shows that the electrical capacitance has significant positive correlation with MC. Table 1 shows that the highest correlation between electrical capacitance and MC was computed at 20kHz having R^2 of 0.969 and this was validated using a set validation data and the lowest regression equation RMSE of 2.78 is found at 20kHz. Figure 8 shows the capacitive property of NBR glove with different drying times in a wide range of frequencies. In the high moisture region (between 30 to 60%) lower frequencies exhibited high electrical capacitance and this electrical capacitance sharply decreased as drying time increased. Our findings indicate that Nitrile glove's capacitance is a potential parameter to determine its moisture content. Further large-scale studies are required to calibrate the IDE sensor and accurately predict gloves moisture for online detection method.

Table 1: RMSE of the regression equation and correlation coefficient applied to predict moisture content using electrical capacitance

Frequency	Equation	R^2	RMSE
100 Hz	y = 7.7238x - 14.534	0.944	3.72
120Hz	y = 7.7541x - 14.466	0.945	3.67
1kHz	y = 8.7121x - 14.204	0.960	3.15
10kHz	y = 11.677x - 18.306	0.966	2.91
20kHz	y = 13.449x - 21.105	0.969	2.78
100kHz	y = -0.0022x + 6.9791	0.384	12.30

Consistent with our present findings, Wang *et al.* (2017) also reported that materials with high moisture exhibits greater capacitance due to high dielectric constant of water (ε_r = 80). NBR at 100Hz have a dielectric constant of 10 or more at room temperature (Matsuno *et al.*, 2021). During drying process, water diffuses and dries off from glove surface, causing drastic drop in gloves dielectric constant. Capacitance is related to dielectric constant using the definition $C = \frac{\varepsilon_0 \varepsilon_r A}{d}$ resulting in the rapid drop in capacitance as heating time increases. Hence, this further reinforces the notion that Nitrile glove's capacitance decreases sharply as the moisture content decreased. This finding is also consistent with Butts's where he found capacitance measurement using impedance technique has 87 to 100% predictability having 8 to 21% MC values (Butts, 2008).

It is proven by Zhu *et al.* (2018) and Yang *et al.* (2019), where the permanent dipoles in the NBR attributed to the CN groups orientation polarization was the reason for the large dielectric constant in lower frequencies. Their findings support our study, where at lower frequencies within the same MC level, the electrical capacitance was relatively larger than higher frequencies (Figure 8). The electrical capacitance is proportional to relative dielectric constant as shown in equation 2. The decreased capacitance is attributed to the dipole polarization of CN groups that could not keep up with the increase in frequency. The CN orientation polarization response is slower resulting in more time to reach field of static equilibrium with electronic and atomic polarization. Hence it can be stated that as frequency increases, the electrical capacitance of Nitrile gloves decreases as a result of the lag of CN group orientation polarization in NBR.

Negative electrical capacitance is expected at higher frequencies. The series LC circuit of IDE sensor connection with clip leads of LCR meter behaves such that it measures capacitor at low frequencies and as an inductor at high frequencies. Plonus (2020, p.92) states that at high frequencies

the series circuit is inductive as: Inductive Reactance, X_L > Capacitive Reactance, X_C. When the clip leads are connected to a capacitor at a frequency above its series resonance, the capacitor will appear inductive resulting in a negative value in the LCR meter. Reason is that a capacitor at a frequency above its series resonance is an inductor, hence the voltage leads the current. It is important to note in theory, current leads voltage in a (positive) capacitor whereas in a negative capacitor, voltage leads current. Therefore, if the LCR meter is set up to measure the capacitance in a component, where the voltage leads the current, the meter will read a negative number. (Halpin and Card, 2011; Plonus, 2020).

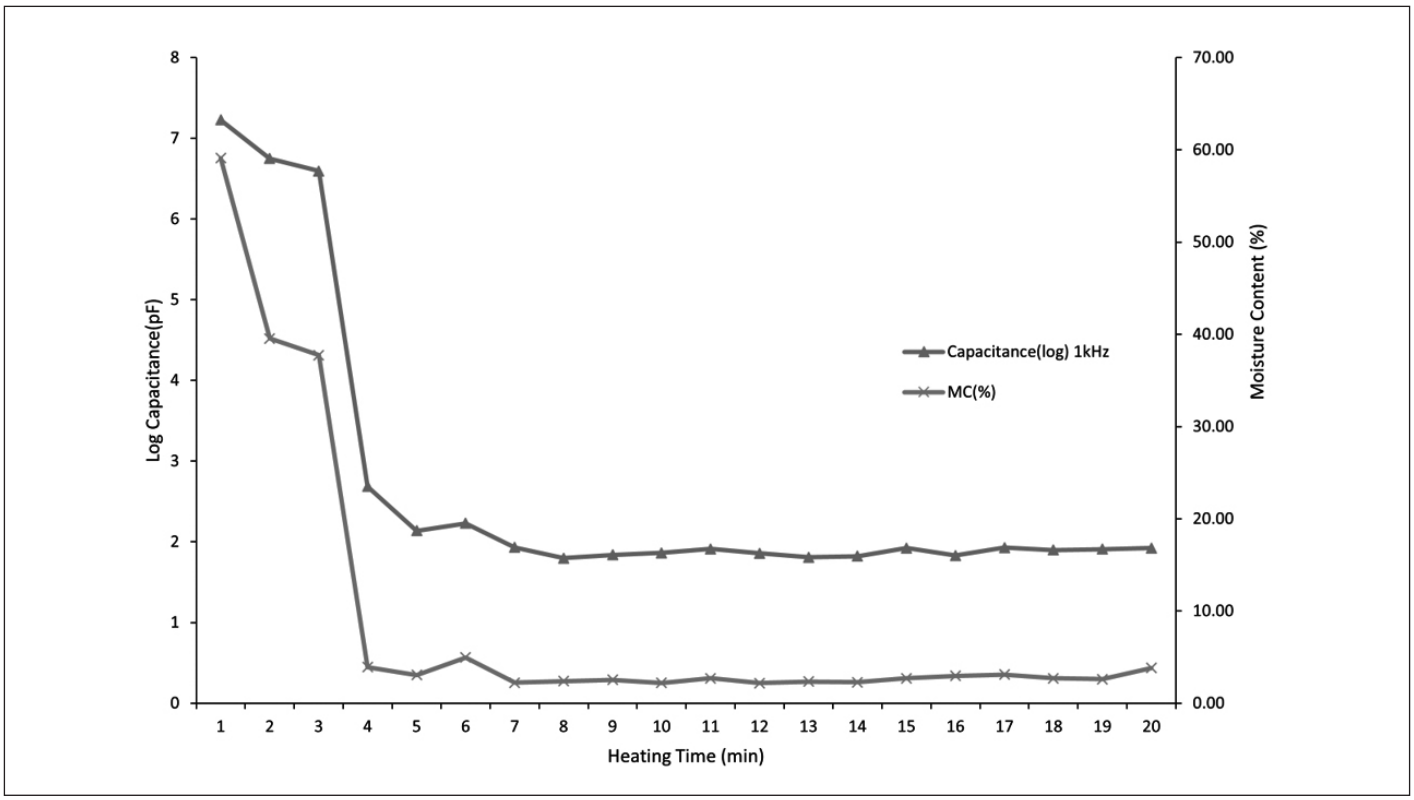


Figure 6: The capacitance and MC measurements at 1 kHz

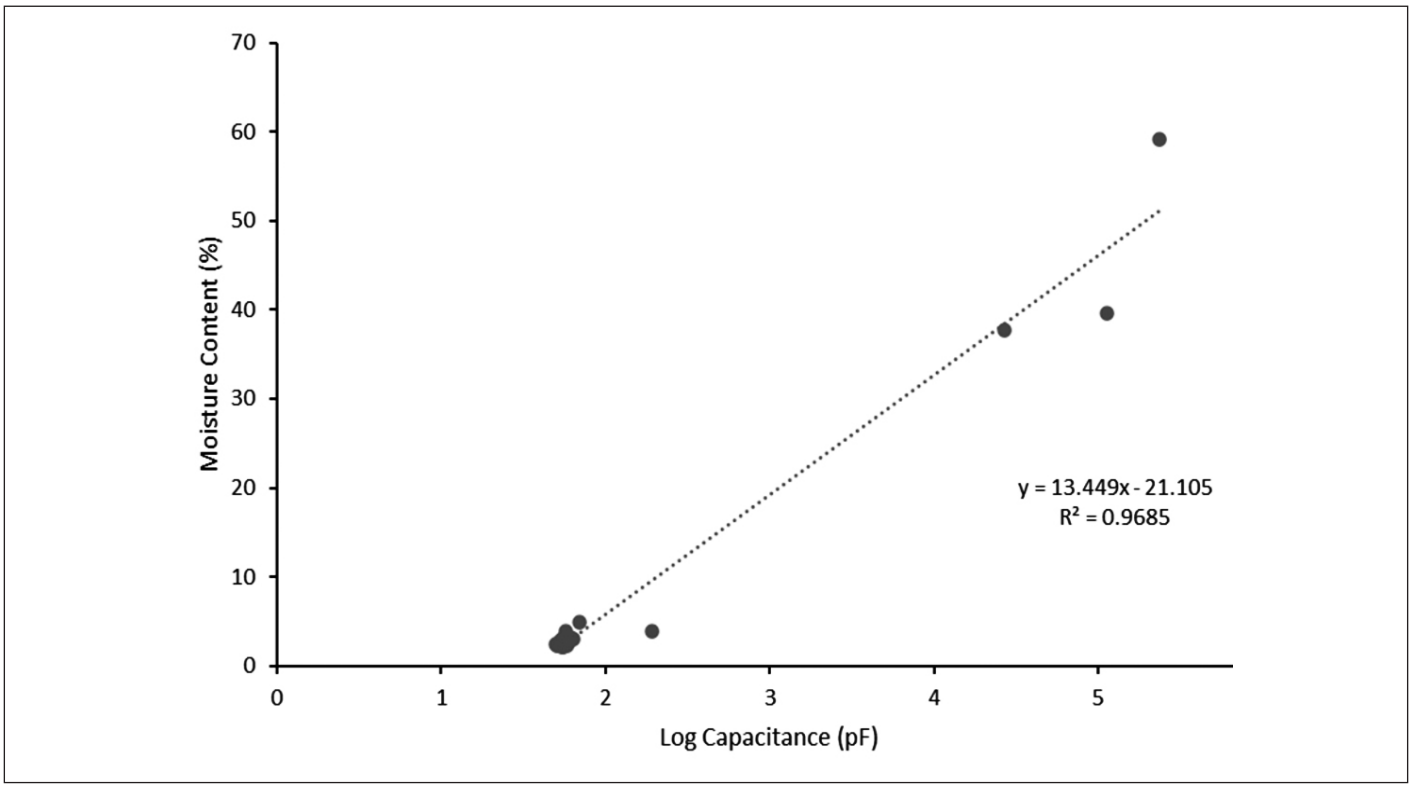


Figure 7: Regression of Capacitance measurements with MC values of gloves at 20 kHz from 1 to 20 minutes during the drying process

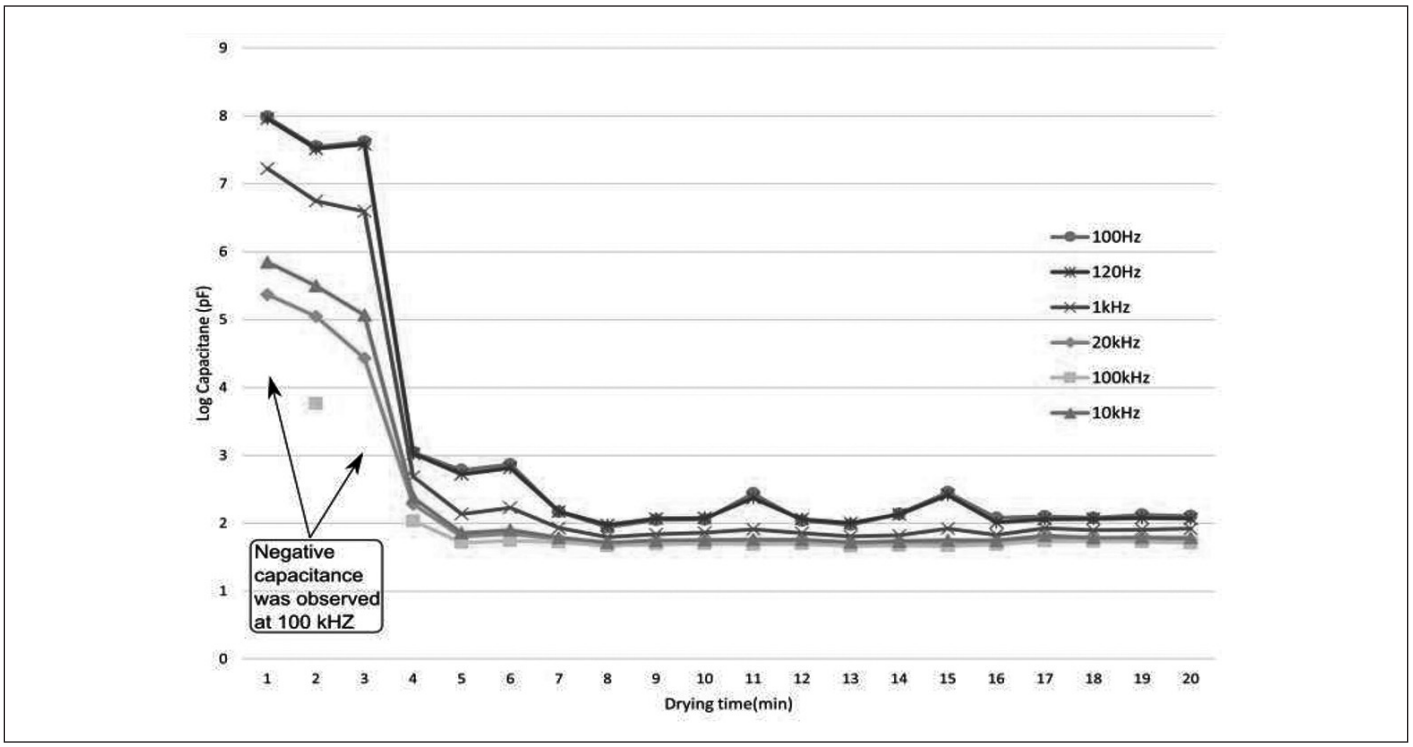


Figure 8: 100Hz, 120Hz, 1kHz, 20kHz, 100kHz, 10kHz frequencies of gloves vs drying times

There are few limitations observed here. Firstly, fluctuations occurring during the experiments in the capacitance might be due to the changes in temperature of the oven. Future study is proposed to stabilize oven temperature to increase accuracy. Next limitation is that the measurement for drying interval of 1 minute is long causing the capacitance dropping drastically. Further study is therefore necessary to determine the accuracy by reducing time interval for assessing the change in capacitance. Another limitation is IDE design which is not the focus in this study. One of the most important factors to be considered in order to analyze the capacitance sensor of the IDE is its electrode geometry. Future studies are required with regards to the output of the capacitance such as selectively designing the dynamic range and penetration depth as well as the ratio of electrode, substrate thickness, shield electrode and placement of coating layer on electrode to study sensor modelling including optimization and performance evaluation. Further experiments should also be conducted to study the effects of glove thickness on the sensor's sensitivity. Results from the experiments have shown that novel interdigital sensing system has the potential to be one of the options to assess the quality of glove products for online monitoring. Findings of this study is significant in automation and IoT device detection. Outcomes from the experiments also provide opportunity for further research in developing a low-cost IDE capacitance sensor with a reliable moisture sensing system for on-line, non-contact measurement of moisture content (MC) of glove products.

5.0 CONCLUSION

This paper proves that the moisture content of Nitrile gloves can be determined with a non- destructive and rapid method using IDE capacitance sensor. Capacitance values of varying moisture in nitrile gloves were characterized with 6 discrete frequencies range from 100Hz, 120Hz, 1kHz, 10kHz, 20kHz, 100kHz and statistical analysis, the coefficient of determination (R^2) and the root mean square error (RMSE) technique, was applied to predict the moisture content in gloves. Experimental results indicate that the capacitance value is a strong function of moisture content measurement. The capacitance of moisture content in nitrile gloves decreased with increasing drying time over the measured frequency range. Statistical analysis results have confirmed that the 1kHz, 10kHz, and 20kHz signal frequencies have highest reliable prediction of the nitrile gloves' moisture content with high R^2 value of 0.96, 0.97 and 0.97, respectively. The findings of this study indicates that with the use of IDE sensor can easily predict the moisture content in Nitrile gloves. Further study is suggested on IDE sensor design with Finite Element Method (FEM) analysis to study sensor modelling, optimization, and performance evaluation to improve accuracy and reliability of MC measuring system. Results from the experiments shows that a low-cost capacitance moisture detection sensing system can be built for an on-line, non-contact measurement of moisture content (MC) of glove products for commercial use by manufacturing industry in their automation process. Future study is proposed to investigate the IDE sensor design configuration such as length, shape and number of electrodes for better sensitivity towards Nitrile Gloves.

6.0 ACKNOWLEDGEMENT

This work was supported and funded by Top Glove Sdn. Bhd., Grant number UC/IC/2019- 0016/2. The authors would like to thank Prof. Dr Tou Teck Yong for his valuable suggestions and all staff in the Top Glove, Factory 25, R&D Centre for providing technical supports and guidance. ■

REFERENCES

[1] Abramowitz, Milton and Stegun, A. I. (1965), 'Handbook of Mathematical Functions.', Dover Publications Inc. pp. 1046.

[2] Afsarimanesh, N., Mukhopadhyay, S. C. and Kruger, M. (2019), 'Planar interdigital sensors and electrochemical impedance spectroscopy', Smart Sensors, Measurement and Instrumentation 30, pp. 33–44.

[3] Ahmad, Z. (2012), Polymer Dielectric Materials, in M. A. Silaghi, ed., 'Dielectric Material', IntechOpen, Rijeka, chapter 1, p. Ch. 1.

[4] Alam, N., Member, S., Bhuiyan, R. H., Member, S., Dougal, R. A., Member, S., Ali, M. and Member, S. (2010), 'Concrete Moisture Content Measurement Using Interdigitated Near- Field Sensors', IEEE Sensors Journal 10(7), pp. 1243–1248.

[5] Alexander, F., Price, D. T. and Bhansali, S. (2010), 'Optimization of interdigitated electrode (IDE) arrays for impedance based evaluation of Hs 578T cancer cells', Journal of Physics: Conference Series 224(1).

[6] Bhuiyan, R. H., Alam, M. N., Caicedo, J. M. and Ali, M. (2015), 'Real-Time Wireless Mois- ture Sensing in Concrete Using Interdigitated Stick-on Sensors', Sensors & Transducers 195(12), pp. 30–38.

[7] Butts, C. V. K. K. Æ. C. L. (2008), 'Design and performance of a capacitor sensor and impedance analyzer for nondestructive moisture content determination', pp. 240–246.

[8] Chetpattananondh, P., Thongpull, K. and Chetpattananondh, K. (2017), 'Interdigital capaci- tance sensing of moisture content in rubber wood', Computers and Electronics in Agriculture 142(November), pp. 545–551.

[9] Chetpattananondh, K., Tapoanoi, T., Phukpattaranont, P., & Jindapetch, N. (2014). A self-calibration water level measurement using an interdigital capacitive sensor. Sensors and Actuators A-physical, 209, 175-182.

[10] Döring, J., Tharmakularajah, L., Happel, J. And Krieger, K. L. (2019), 'A novel approach for road surface wetness detection with planar capacitive sensors', Journal of Sensors and Sensor Systems 8(1), pp. 57–66.

[11] Ferrari, V. and Prudenziati, M. (2012), Printed thick-film capacitive sensors, in M. Prudenziati and J. B. T. P. F. Hormadaly, eds, 'Woodhead Publishing Series in Electronic and Optical Materials', Woodhead Publishing, pp. 193–220.

[12] Fleischmann, D. D., Arbeiter, F., Schaller, R., Holzner, A., Kern, W. and Schlö gl, S. (2018), 'Influence of crosslinker and water on cyclic properties of carboxylated nitrile butadiene rubber (XNBR)', Polymer Testing 67(March), pp. 309–321.

[13] Halpin, S. M. and Card, A. (2011), 'Power Quality', Power Electronics Handbook pp. 1179– 1192.

[14] Khaled, A. Y., Aziz, S. A. and Rokhani, F. Z. (2015), 'Capacitive sensor probe to assess frying oil degradation', Information Processing in Agriculture 2(2), pp. 142–148.

[15] Kosumphan, H., Somwong, S. and Chongcheawchamnan, M. (2018), 'Electrical characteristics of pure and contaminated latex serum', Songklanakarin Journal of Science and Technology 40(2), pp. 329–332.

[16] Kueseng, P., Sae-Oui, P. and Rattanasom, N. (2013), 'Mechanical and electrical properties of natural rubber and nitrile rubber blends filled with multi-wall carbon nanotube: Effect of preparation methods', Polymer Testing 32(4), pp. 731–738.

[17] Ludwig, I. (2008), Drying, Film Formation and Open Time of Aqueous Polymer Dispersions: An Investigation of Different Aspects by Rheometry and Inverse-Micro-Raman Spectroscopy (IMRS).

[18] Ma, Y., Davis, H. T. and Scriven, L. E. (2005), 'Microstructure development in drying latex coatings', 52(April 2003), pp. 46–62.

[19] Markevicius, V., Navikas, D., Valinevicius, A., Andriukaitis, D., & Cepenas, M. (2012). The Soil Moisture Content Determination using Interdigital Sensor. Elektronika Ir Elektrotechnika, 18, 25-28.

[20] Matsuno, R., Takagaki, Y., Ito, T., Yoshikawa, H., Takamatsu, S. and Takahara, A. (2021), 'Relationship between the Relative Dielectric Constant and the Monomer Sequence of Acry- lonitrile in Rubber', 13, p. 54.

[21] Mohamad, F., Sairin, M. A., Nizar, N. N. A., Aziz, S. A., Hashim, D. M., Rokhani, F. Z. and Preparation, A. S. (2015), 'Investigation on Interdigitated Electrode Design for Impedance Spectroscopy Technique Targeting Lard Detection Application', pp. 739–744.

[22] Patel, R., Sabet Abdou-Sabet and Hsien-Chang Wang (1996), 'Thermoplastic Elastomer with Polar and Non-polar Rubber Components'.

[23] Plonus, M. (2020), 'AC circuits', Electronics and Communications for Scientists and Engineers, pp. 79–120.

[24] Rivadeneyra, A., Fernández-Salmerón, J., Banqueri, J., López-Villanueva, J.A., Capitán-Vallvey, L.F., & Palma, A.J. (2014). A novel electrode structure compared with interdigitated electrodes as capacitive sensor. Sensors and Actuators B-chemical, 204, 552-560.

[25] Sairin, M. A., Amira, N., Aziz, S. A., Sucipto, S. and Rokhani, F. Z. (2019), 'Design of portable wireless impedance spectroscopy for sensing lard as adulterant in palm oil', IOP Conference Series: Earth and Environmental Science 230, p. 12021.

[26] Son, Y., Oh, M. and Lee, S. (2010), 'Estimation of soil weathering degree using electrical resistivity', Environmental Earth Sciences 59(6), pp. 1319–1326.

[27] Tirumkudulu, Mahesh S and Russel, W. B. (2005), 'Cracking in Drying Latex Films', Langmuir, 21(11), pp. 4938–4948.

[28] Varshney, M., & Li, Y. (2009). Interdigitated array microelectrodes based impedance biosensors for detection of bacterial cells. Biosensors & bioelectronics, 24 10, 2951-60.

[29] Wang, B., Pan, P., McDonald, T. P. and Wang, Y. (2017), 'Development of a capacitance sensing system for monitoring moisture content of spray dried gelatin powders', Journal of Food Engineering 195, pp. 247–254.

[30] Yahaya, N. Z., Abbas, Z., Norimi, A. M., Yahaya, M. Z., Razak, N. N. A. A. and Mustafa, I. S. (2015), 'A simple rectangular microstrip technique for determination of moisture content in Hevea rubber latex', AIP Conference Proceedings 1674.

[31] Yang, D., Kong, X., Ni, Y., Gao, D., Yang, B., Zhu, Y. and Zhang, L. (2019), 'Novel nitrile- butadiene rubber composites with enhanced thermal conductivity and high dielectric constant', Composites Part A: Applied Science and Manufacturing 124(May), p. 105447.

[32] Yew, G. Y., Tham, T. C., Law, C. L., Chu, D. T., Ogino, C. and Show, P. L. (2019), 'Emerging crosslinking techniques for glove manufacturers with improved nitrile glove properties and reduced allergic risks'.

[33] Zhao, X., Yang, J., Zhao, D., Lu, Y., Wang, W., Zhang, L. and Nishi, T. (2015), 'Natural rubber/nitrile butadiene rubber/hindered phenol composites with high-damping properties', International Journal of Smart and Nano Materials 6(4), pp. 239–250.

[34] Zhu, S. and Zhang, J. (2017), 'Enhanced dielectric constant of acrylonitrile–butadiene rub- ber/barium titanate composites with mechanical reinforcement by nanosilica', Iranian Poly- mer Journal (English Edition) 26(4), pp. 239–251.

[35] Zhu, S., Zhang, W. and Zhang, J. (2018), 'High dielectric acrylonitrile-butadiene rubber with excellent mechanical properties by filling with surface-modified barium/strontium inorganic functional powders', Journal of Materials Science: Materials in Electronics 29(8), pp. 6519– 6529.

[36] Zoolfakar, A. S., Hashim, S. B., Zolkapli, M. and Idros, M. F. (2010), 'Design, fabrication and characterization of conductivity sensor using printed circuit board', 2010 6th International Colloquium on Signal Processing & its Applications (May), pp. 1–6.

PROFILES

VISHNUKUMAR RAJANDRAN, presently working with Top Glove International as Research & Development Engineer, and has completed B.Eng (Mechanical) from University of Malaya in the year 2019, M.EngSc from Tunku Abdul Rahman University College in the year 2022. A patent for Laser Detection system has been filled in the year 2022 as an outcome of his master's research, and he is presently pursuing Ph.D in Advanced Sensor from Nanotechnology and Catalysis Research Centre (NANOCAT), University of Malaya, Malaysia.

ASSOC. PROF. TS. DR LEE WAH PHENG is the Associate Professor in Tunku Abdul Rahman University College. He worked in the manufacturing industry for 10 years and more than 20 years of business and education industry experiences. Dr Lee is a pioneer and consultant in Industry 4.0. He works with a team of researchers and industry partners to develop a holistic digital solution suitable for the small and medium enterprises.

PROF. TS. DR YIP MUM WAI holds a Diploma in Materials Engineering and MSc in Manufacturing Systems Engineering from University of Warwick, United Kingdom. In 2008, he was awarded an Engineering Doctorate in Engineering Business Management (specialised in Knowledge Management) from Business Advanced Technology Centre, Universiti Teknologi Malaysia. Dr Yip is a fellow of IMechE, MATRIZ Practitioner with Level III certified, TRIZ Instructor, certified Knowledge Management Facilitator and Practitioner, Senior Member of International Association of Computer Science and Information (IACSIT), Senior Member of IEDRC, Senior Member of SAISE, Senior Member of SCIEI and also Vice President of Malaysia TRIZ Innovation Association (MyTRIZ).

LIM JOO ENG holds a Bachelor of Chemical Engineering (Honours) from University of Malaya (UM) and Master of Science (MSc) in Manufacturing System Engineering from University Putra Malaysia (UPM). She is a Chartered Engineer of Engineering Council (UK), a registered Graduate Engineer of Board of Engineers Malaysia (Grad. Eng), a member of The Institution of Mechanical Engineers (MIMechE) UK, and a Graduate Member of The Institution of Engineers, Malaysia (Grad.IEM). Currently, she is the Principal Lecturer in Tunku Abdul Rahman University College (TAR UC). Her expertise and field of research are in the area related to project management, operation management and manufacturing system.

DR TAN YOKE MENG has been in working in Top Glove since 2016, focusing on product development in research and development centre. He is passionate on sustainability and currently is leading the new product and sustainable product development team striving to be at the forefront in glove manufacturing industry.

INTERACTIVE DASHBOARD WITH VISUAL SENSING AND FAST REACTIVITY

(Date received: 13.01.2022/Date accepted: 13.04.2022)

Wen Lin Yong[1], Jun-Kit Chaw[2]*, Yiqi Tew[1]

[1] Faculty of Computing and Information Technology, Kuala Lumpur Main Campus, Jalan Genting Kelang, Setapak, 53300 Kuala Lumpur, Malaysia.

[1] Institute of IR4.0, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.

**Corresponding author: chawjk@ukm.edu.my*

ABSTRACT

These days, technology is growing rapidly, and the market has been introduced with lots of fascinating ways to interact with computers. The advancement of deep learning models and hardware technology also enables more applications with fancy features to be built. The importance of hand gesture recognition has increased due to the prevalence of touchless applications. However, developing an efficient recognition system needs to overcome the challenges of hand segmentation, local hand shape representation, global body configuration representation, and a gesture sequence model. This paper proposed an interactive dashboard that could react to hand gestures. This is also an initiative of the Tunku Abdul Rahman University College (TAR UC) Smart Campus project. Deep learning models were investigated in this research and the optimal model was selected for the dashboard. In addition, 20BN Jester Dataset was used for the dashboard development. To set up a more user-friendly dashboard, the data communication stream between the captured input stream and commands among the devices were studied. As to achieve higher responsiveness from the dashboard, evaluation on data communication protocols which were used to pass the input data were included in the study.

Keywords: Computer Vision, Human-Computer Interaction (HCI), Gesture Detection, Real-time systems, Feature Extraction

1.0 INTRODUCTION

Human-Computer Interaction (HCI) is a field of study that focuses on discovering ways of human interacting with computers. Due to the rapid growth in computer technology, touchscreen HCI has evolved in recent years and over the pandemic to interaction with gestures, motion sensors, hand interaction and other touchless interfaces. One of the well-known practices is the visual-based HCI which requires cameras as the input devices instead of computer mouse and keyboard. Although this field has been broadly studied in the computer vision field, the methods proposed in the market show vulnerability in the outdoor environment. A review done by Chakraborty *et al.* suggested that existing classifiers for vision-based gesture identification are unable to handle all types of gesture classification problems at the same time. Each has disadvantages that limit total performance [1].

The development of the interactive dashboard in this research aimed to achieve high performance in responding to user's commands which were performed using gestures movements. Thus, the gesture recognition algorithms were studied and evaluated to achieve real-time capability in returning outcomes for the gesture movements detected. Aside from that, to obtain higher usability for the interactive dashboard, the video streaming protocols were determined for faster transmission of gesture data captured from the input devices.

Every individual in this world is known to act differently in their way. Even though a lot of defined classes and algorithms for human recognition systems have been introduced in the market, there might be chances of miscalculating poses while employing the practices for their use. To fine-tune the best gesture weights, a large gesture dataset maybe required. Besides, gesture recognition performance might be affected while the system is implemented in a real-world environment due to external factors such as visual occlusions [2] and illumination [3].

The user experience while making use of the dashboard might be affected by the sensitivity of the interactive dashboard. One of the factors that might affect the interactive dashboard irritability is the transmission of gesture videos data through the IP cameras and network video recorders. While there is a delay in the video transmission, the system might not achieve responsive controls towards the dashboard. Moreover, the model used to process the gesture would also affect the interactive dashboard reactivity [4].

2.0 RELATED WORKS

2.1 Gesture Recognition with Deep Learning Models

Gesture is a widely used touchless interface that allows humans to engage with technology, and it is the next step in the evolution of motion sensors [5]. Gesture-based touchless interfaces or better real-time hand interaction are possible with devices like Leapmotion, kinect, Azure Kinect. For example, using wired

gloves as input devices is known to perform well in gesture recognition. However, data collection always requires the gloves as extra equipment that caused inconvenience due to wired restrictions [6].

In this paper, RGB cameras were used as the input devices to capture the motions. Hence, only deep learning-based development are discussed here as it could extract discriminative features for classification from RGB images more effectively. Although there are other methods that may perform better in terms of preciseness or pace when detecting the gestures, deep learning approaches are more robust in the real-world environment [7]. We considered two deep learning models to be utilized for detecting gestures: Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) approaches. CNN is recognized to make use of object recognition with the ability to learn features in spatial data. By using the CNN model, gesture recognition can be performed in a fast behaviour with high precision in the results [8]. However, CNN models are incapable of learning temporal data which is crucial in continuous hand gesture recognition. Hence, 3DCNN is proposed to acquire features extraction on 3D data (eg. gesture movements' videos data) which allows detecting dynamic gestures movements. Besides, recurrent neural networks (RNN) which are known for processing sequential data are taken into consideration to incorporate with the CNN model to process longer continuous gesture data [9]. By employing RNN within the CNN model, the gestures could be captured in a continuous manner for the model to predict the next actions done by the users. However, it is challenging for this approach to correctly recognize directional motions like swipe and rotation, as well as motions like push that changes spatial information over time [10]. To maintain information in memory for long periods of time, LSTM is preferred over RNN [11]. Thus, this paper integrated 3DCNN and LTSM to recognize gestures that are commonly used in HCI.

2.2 Data Communication Protocols

Modern communication systems and networks, such as the Internet of Things (IoT) and cellular networks, generate enormous amounts of heterogeneous traffic data [12]. As stated by Gaurav Sinha *et al.*, high bandwidth interaction allows the acceleration of the interaction between human and computer due to its capability to transfer massive data [13]. Either analog-based solutions or digital-based solutions could be used to acquire data for gesture recognition. Analog-based solutions are also known as closed-circuit televisions (CCTV) that have to process the analog signals in a digital video recorder (DVR) for video recording purposes. On the contrary, digital-based solutions use Internet protocol cameras (IP cameras) to transmit digitalized surveillance video to back-end computers for monitoring by using IP-based protocols such as RTP (Real-time Transport Protocol) or HTTP (Hyper-Text Transfer Protocol). For recording, it uses network video recorders (NVR) [14].

Due to the convenience and high resolution of video surveillance data, IP cameras were commonly used for machine learning or deep learning approaches [15], [16]. Together with other data-driven tasks, such as machine learning, it could unlock the potential of big data in many domains in the artificial intelligence era [17]. Furthermore, the goal of improving accuracy and efficiency to better promote the use of data-driven computation has remained unchanged. This paper aims to increase efficiency by utilizing the correct data communication protocol to produce high satisfactory HCI capability to the users. In conjunction with this, the responsiveness of the dashboard could be improved.

3.0 RESEARCH DESIGN AND METHODS

IP cameras were adopted as the input devices to capture the gestures from the users, then the captured data were forwarded to the dashboard for further processing and response. Our scope of work covers the exploration of the video streaming protocols used to allow communication (eg. passing media/files) between devices and the employment of the deep learning model for gesture recognition that was used to allow direct human interaction and controls to the dashboard. This work aims to develop an interactive dashboard with gesture-navigated effects based on artificial intelligence human gestures detection which brings great convenience, flexibility, and efficiency by meeting the benchmarks listed below:

1. **To attain high accuracy results in gesture recognition results.**
   The gesture recognition algorithms and methods applied should be able to achieve high precision results to secure the usability of the interactive dashboard. By acquiring exact gesture recognition, the dashboard will be able to react with proper responses according to what users perform.
2. **To achieve high responsiveness from the interactive dashboard.**
   a. The video streaming protocols which have been employed should be able to transmit the gesture data at a higher speed to reduce the delay of video streaming to the system for gesture recognition processing.
   b. The suitable deep learning models used to perform the gesture recognition should be implemented to accomplish the real-time capability of the interactive dashboard.

Figure 1 illustrates the program flow for the interaction dashboard developed. The aim of this paper is to detect the gestures for direct interaction between humans and the dashboard. Thus, gestures such as swipe left, right, up, and down were learnt by the model to interact with the dashboard. Since the gestures might be performed differently by every individual, the 20BN Jester dataset with a large amount of significant gesture data were utilized in the model learning process.

Figure 2 shows the overall process of video retrieval in our proposed video streaming module. To capture human gesture movements, IP cameras and NVRs were employed to capture the gestures that were performed by each individual. While human gesture recognition was performed in a real-time manner, the gesture data captured by those devices were expected to be transferred between those devices as quickly as possible to bring out a responsive and low latency interaction between the individuals and the dashboard.

A convolutional HIKvision's IP cameras and Network Video Recorders (NVR) were used in this research. As HIKvision's network cameras provide predefined video streaming protocols, specific settings could be configured for video streaming. The protocols provided by the HIKvision products are TCP, UDP,

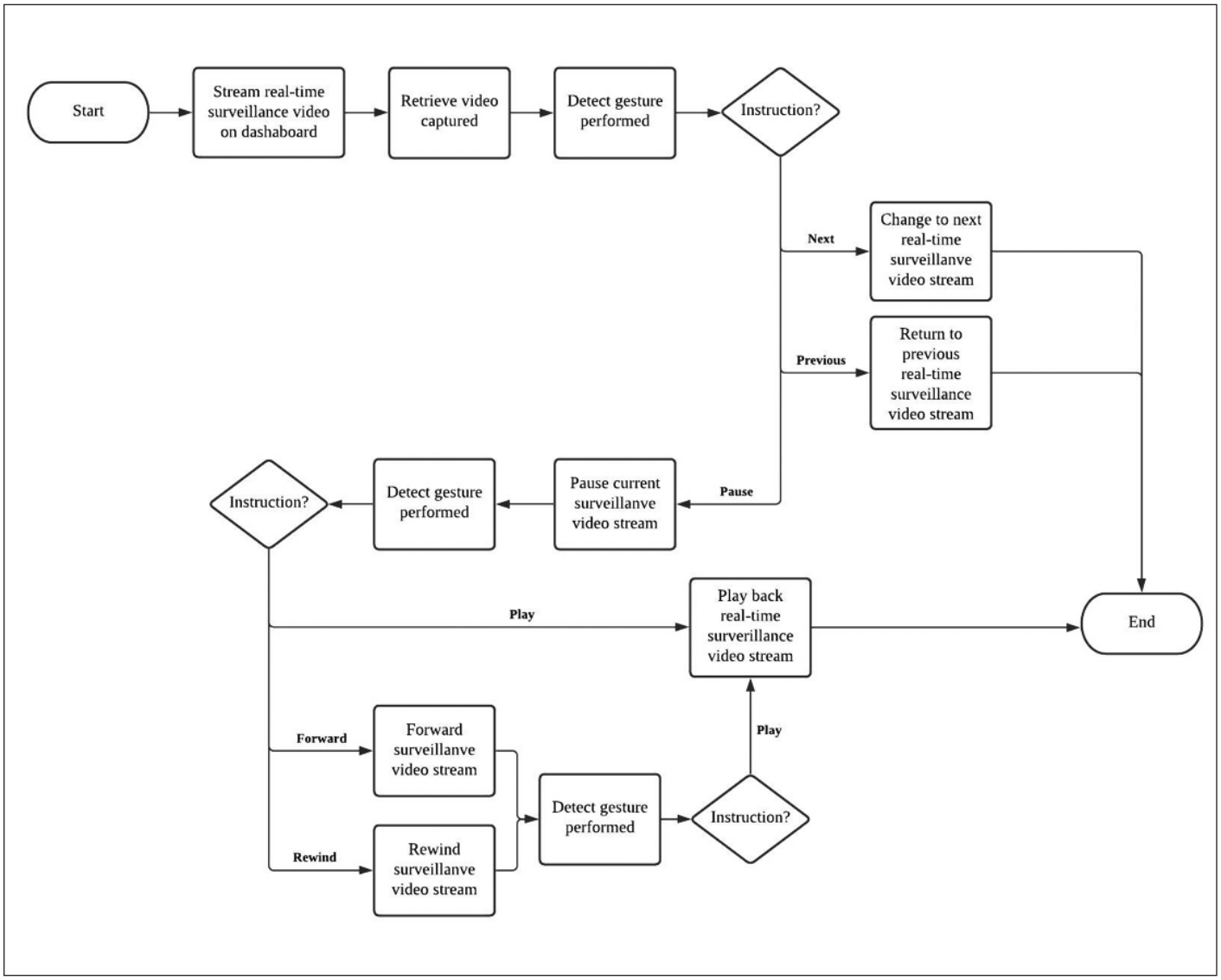


Figure 1: Flowchart of overview interactive dashboard

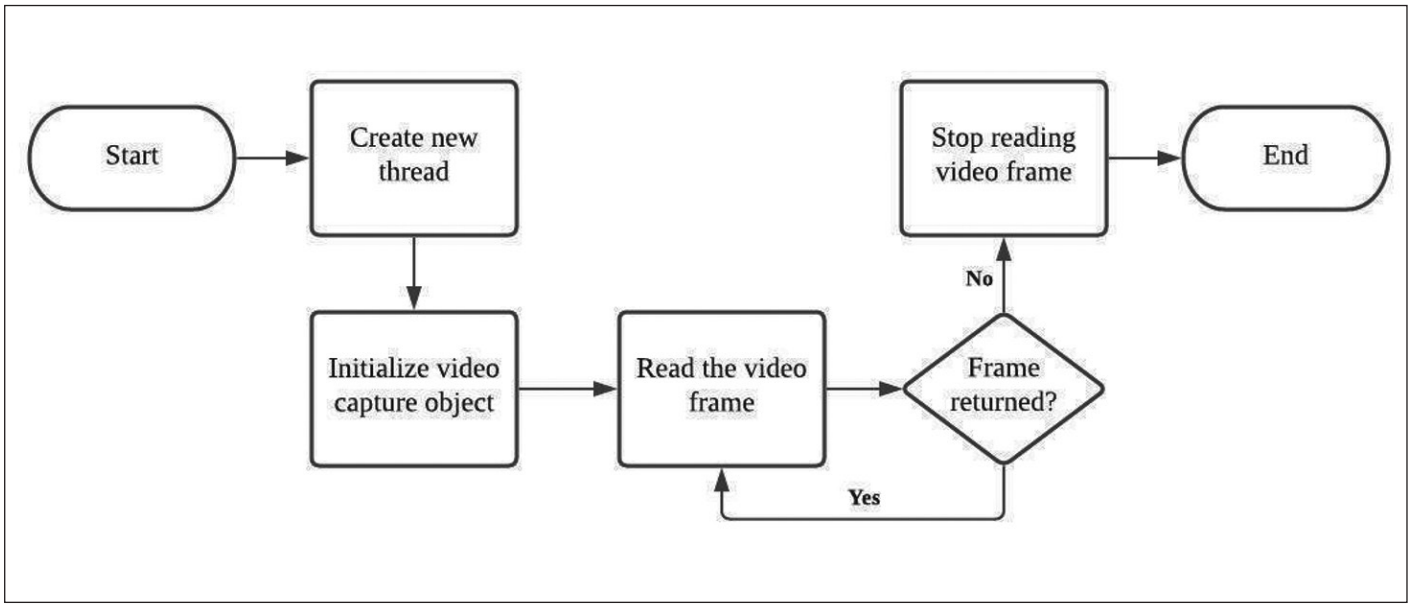


Figure 2: Flowchart of video retrieval in video streaming module

HTTP, and MULTICAST. Every live streaming protocol achieves different goals:

- TCP: Compromising video stream quality and reducing packet loss problems, however, the real-time streaming will have delays.
- UDP: Support real-time video and audio streaming.
- HTTP: Similar to the TCP protocol, yet not required ports specifications
- MULTICAST: Establish multicast group addresses and provide stream acquisition by multiple users simultaneously.

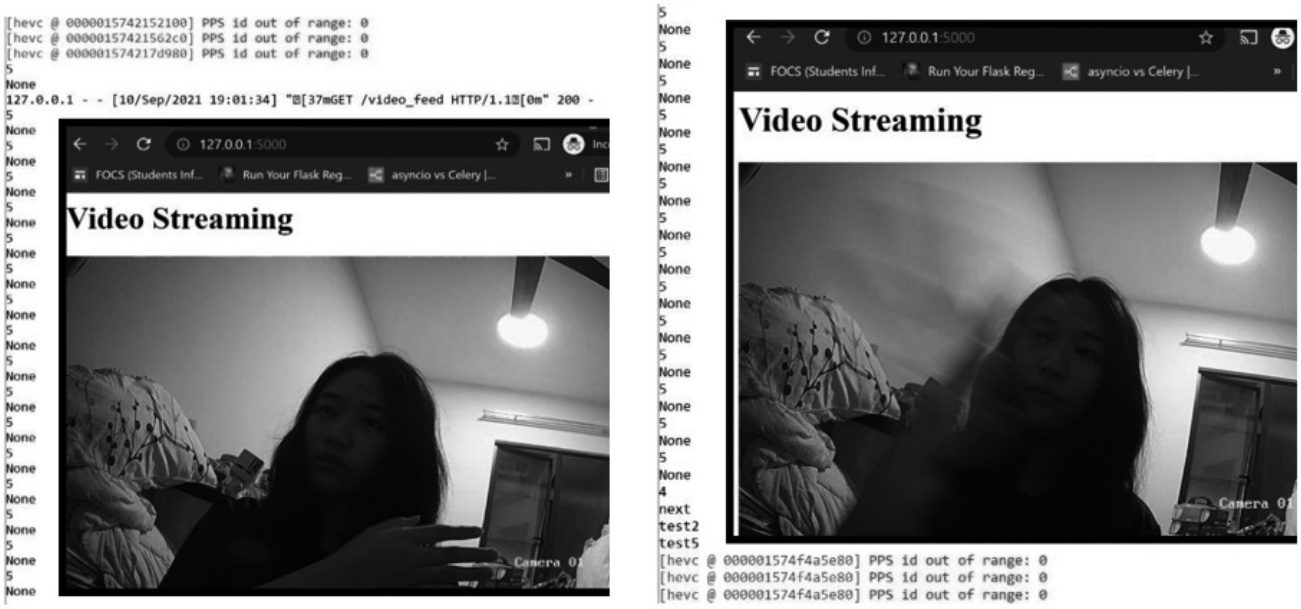
To obtain the video stream at a faster pace from the cameras to the dashboard, the UDP protocol were selected for this work.

For the experimental setup, we utilized the Jester dataset [18] to train the gesture recognition model. This dataset contains an enormous amount of gesture movements video data captured from numerous actors. The gestures which acted in those videos were suitable for HCI, e.g., swipe left, right, down, up and etcetera).

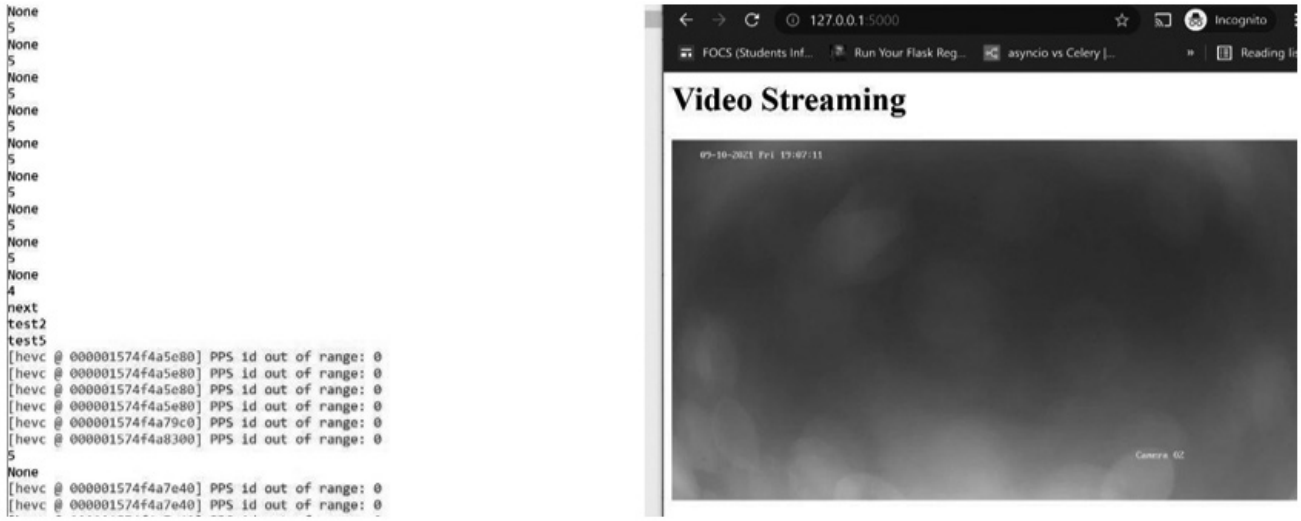
4.0 RESULTS AND DISCUSSION

Before the construction of the interactive dashboard, the gesture recognition approach was tested, and the results (gesture commands detected) were used as input to interact with the dashboard subsequently.

Figure 3 shows the test screens of the gesture recognition implementation in the flask application which was planned to be used for the dashboard set up. The gestures were used to decide which camera view to be shown on the screen. For example, a swipe left gesture was performed and this gesture was expected to change to another video stream, as captured in Figure 3 (left). After a few seconds, the program returned the gesture results which were displayed on the command prompt screen as shown in Figure 3 (right). Another few seconds passed, the gesture result was successfully passed to the controlling section, and action was carried out. Then, the video stream has changed to another in Figure 3 (bottom).



(left) Swipe Left action is performed, (right) Delay response in results



(bottom) Command occurred and the stream has changed

Figure 3: Sample output of testing on gesture recognition capabilities

However, in Figure 3 the whole testing for the program was not running smoothly, there was lagging, and the program took time to return the result. This might be due to the reason that all the functions were run in one individual function, so the previous task has to be completed before proceeding to the next task. Therefore, tasks have been separated into several functions' definitions.

Figure 4 shows the output after the tasks have been separated into individual functions. The screen output in the right shows the results with the detected gesture, returns values 5 and "None" mean that there is no gesture detected, the values 4 and "Next" means that "Swipe Left" gesture detected, the values 3 and "previous" means that "Swipe Right" gesture detected. Yet while running the application, the dashboard keeps on showing a loading screen and is unable to prompt out the dashboard. Meanwhile, gesture recognition is running normally. This might be due to the separated functions, they were not running concurrently. This means that it remains the same that it requires to wait for one task to be done only then proceed to the next task.

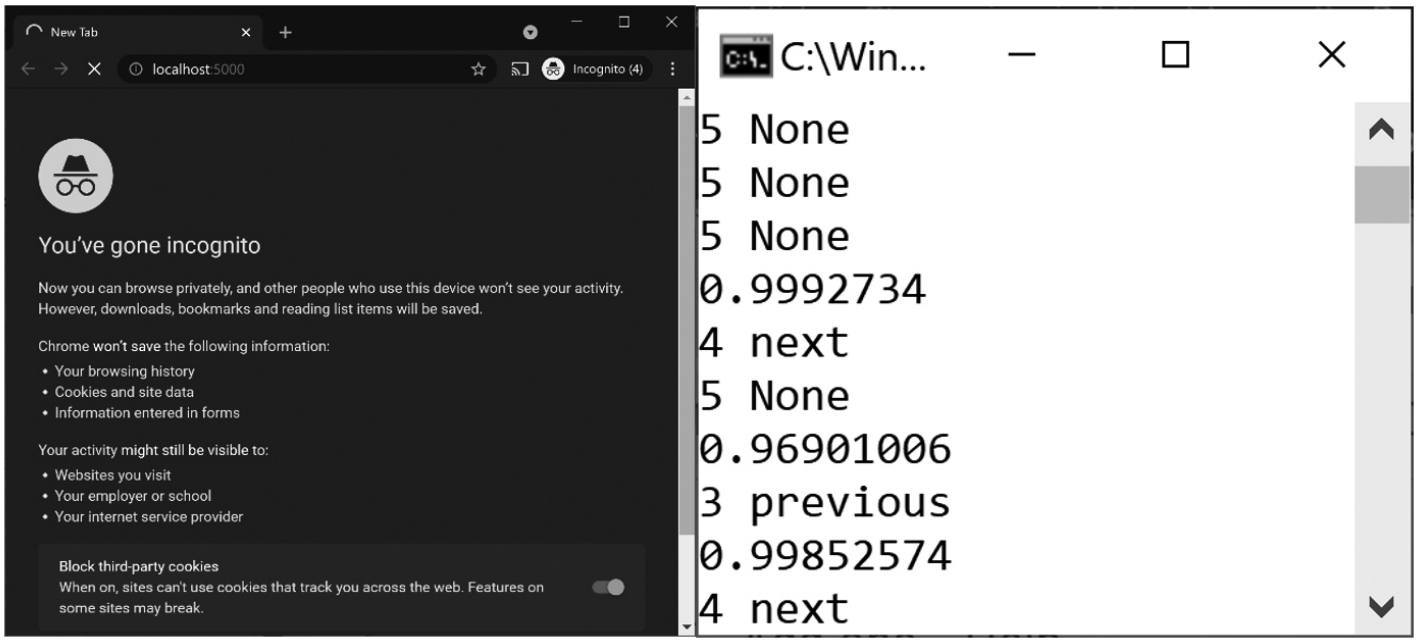


Figure 4: Only run the long-run task (gesture recognition) and dashboard unable to be prompted

To eliminate the blocking of calling other functions that were caused by the long-run task (gesture detection), the functions are expected to run asynchronously. With a focus to conduct the functions concurrently, several task queue management frameworks have been investigated.

Two of the task queues that have been tested to run two tasks concurrently. The two tasks are: (1) Sleep for 3 seconds and print a beer mug; (2) Sleep for 1 second and print a coffee mug. Figure 5 shows the result when two tasks were run concurrently using the Celery task queue. In this testing, the 2 tasks were able to run concurrently and the time elapsed for the program is around 3 seconds in total.

Figure 5: Running 2 tasks concurrently using Celery task queue

Figure 6 shows the result while two tasks were run concurrently using the Redis Queue task queue. In this testing, the tasks failed to run concurrently and the time elapsed for the program is around 4 seconds in total. This might be due to the failure of running them in the background using Redis Queue worker. The Redis Queue worker consists of making use of fork(), which is not available on Windows systems.

Figure 6: Running 2 tasks concurrently using Redis Queue task queue

The testing in Figure 5 and 6 were done in the Windows system, but Redis Queue is more compatible with Unix operating systems due to the usage of fork(). Therefore, the results might not be as precise as they should be. Thus, we recommend to find another suitable approach to replace the celery library to run gesture detection as a background task.

5.0 CONCLUSION

As one of the sub-projects of the TARUC Smart Campus initiative, this research project is aimed to provide artificial intelligence capabilities for the development of the dashboard employed in the i2hub, also known as the integrated innovation

hub, located in the TARC cyber centre which collaborates with works such as Industry 4.0, Agriculture 4.0 and more. The dashboard that will be developed is in the pipeline to cooperate with these initiatives. For example, monitoring the robot motions from the video captured in the Industry 4.0 manufacturing site's cameras and more. This project also envisaged the dashboard with manipulation functionalities to be performed using human gestures detection.

Although there are lots of products with similar functionalities on the market, this research outcome aims to give a better solution for resolving the implementation issues faced in the real-world environment. Especially during this Covid-19 pandemic outbreak, the healthcare industry advocated making use of the interactive dashboard in better remote monitoring of the patients. This reduces unnecessary physical contact with patients or objects to control the dashboard which may be contaminated with viruses. Besides, the interactive dashboard could also be advantageous for the manufacturing industry during this pandemic. By adopting an interactive dashboard for remote monitoring of machines, it could possibly take over the worker's positions and continue the productions during the pandemic [19].

Considering this project as one of the research projects done in TARUC and the sub-project of the TAR UC smart campuses project, this project is aimed to bring new findings and knowledge that are beneficial to related topics. For instance, the video streaming protocols that have been studied in this research may be beneficial for further workings in the TARC Smart Campus's IoT initiative to expand the infrastructure and merge with more systems and other related projects on the campus. For future works, the zero-shot learning capabilities will be explored when dealing with gestures that are not learnt by the model. ■

REFERENCES

[1] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, 'Review of constraints on vision-based gesture recognition for human–computer interaction', IET Comput. Vis., vol. 12, no. 1, pp. 3–15, Feb. 2018, doi: 10.1049/IET-CVI.2017.0052.

[2] S. Liao *et al.*, 'Occlusion gesture recognition based on improved SSD', Concurr. Comput. Pract. Exp., vol. 33, no. 6, p. e6063, Mar. 2021, doi: 10.1002/CPE.6063.

[3] D. Y. Huang, W. C. Hu, and S. H. Chang, 'Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination', Expert Syst. Appl., vol. 38, no. 5, pp. 6031–6042, May 2011, doi: 10.1016/J.ESWA.2010.11.016.

[4] V. Gokul, G. P. Balakrishnan, T. Dubnov, and S. Dubnov, 'Semantic Interaction with Human Motion Using Query-Based Recombinant Video Synthesis', Proc. - 2nd Int. Conf. Multimed. Inf. Process. Retrieval, MIPR 2019, pp. 379–382, Apr. 2019, doi: 10.1109/MIPR.2019.00075.

[5] M. Z. Iqbal and A. G. Campbell, 'From luxury to necessity: Progress of touchless interaction technology', Technol. Soc., vol. 67, p. 101796, Nov. 2021, doi: 10.1016/J.TECHSOC.2021.101796.

[6] X. Han, X. Miao, X. Chen, G. Jiang, and L. Niu, 'Research on finger movement sensing performance of conductive gloves':, https://doi.org/10.1177/1558925019887622, vol. 14, Nov. 2019, doi: 10.1177/1558925019887622.

[7] S. Sharma and S. Singh, 'Vision-based hand gesture recognition using deep learning for the interpretation of sign language', Expert Syst. Appl., vol. 182, p. 115657, Nov. 2021, doi: 10.1016/J.ESWA.2021.115657.

[8] A. Mujahid *et al.*, 'Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model', Appl. Sci. 2021, Vol. 11, Page 4164, vol. 11, no. 9, p. 4164, May 2021, doi: 10.3390/APP11094164.

[9] E. Tsironi, P. Barros, C. Weber, and S. Wermter, 'An analysis of Convolutional Long Short-Term Memory Recurrent Neural Networks for gesture recognition', Neurocomputing, vol. 268, pp. 76–86, Dec. 2017, doi: 10.1016/J.NEUCOM.2016.12.088.

[10] H. R. Lee, J. Park, and Y. J. Suh, 'Improving Classification Accuracy of Hand Gesture Recognition Based on 60 GHz FMCW Radar with Deep Learning Domain Adaptation', Electron. 2020, Vol. 9, Page 2140, vol. 9, no. 12, p. 2140, Dec. 2020, doi: 10.3390/ELECTRONICS9122140.

[11] G. Zhu, L. Zhang, P. Shen, and J. Song, 'Multimodal Gesture Recognition Using 3-D Convolution and Convolutional LSTM', IEEE Access, vol. 5, pp. 4517–4524, 2017, doi: 10.1109/ACCESS.2017.2684186.

[12] M. Abbasi, A. Shahraki, and A. Taherkordi, 'Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey', Comput. Commun., vol. 170, pp. 19–41, Mar. 2021, doi: 10.1016/J.COMCOM.2021.01.021.

[13] G. Sinha, R. Shahi, and M. Shankar, 'Human Computer Interaction', Proc. - 3rd Int. Conf. Emerg. Trends Eng. Technol. ICETET 2010, pp. 1–4, 2010, doi: 10.1109/ICETET.2010.85.

[14] C. F. Lin, H. T. Chiao, R. K. Sheu, Y. S. Chang, and S. M. Yuan, 'A fault-tolerant ONVIF protocol extension for seamless surveillance video stream recording', Comput. Stand. Interfaces, vol. 55, pp. 55–72, Jan. 2018, doi: 10.1016/J.CSI.2017.04.005.

[15] H. C. Kaşkavalci and S. Gören, 'A Deep Learning Based Distributed Smart Surveillance Architecture using Edge and Cloud Computing', Proc. - 2019 Int. Conf. Deep Learn. Mach. Learn. Emerg. Appl. Deep. 2019, pp. 1–6, Aug. 2019, doi: 10.1109/DEEP-ML.2019.00009.

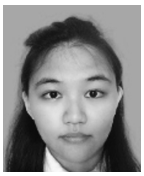
[16] M. Bugeja, A. Dingli, M. Attard, and D. Seychell, 'Comparison of Vehicle Detection Techniques applied to IP Camera Video Feeds for use in Intelligent Transport Systems', Transp. Res. Procedia, vol. 45, pp. 971–978, Jan. 2020, doi: 10.1016/J.TRPRO.2020.02.069.

[17] Y. Wu *et al.*, 'Efficient maliciously secure two-party mixed-protocol framework for data-driven computation tasks', Comput. Stand. Interfaces, vol. 80, p. 103571, Mar. 2022, doi: 10.1016/J.CSI.2021.103571.

[18] J. Materzynska, G. Berger, I. Bax, and R. Memisevic, 'The jester dataset: A large-scale video dataset of human gestures', Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019, pp. 2874–2882, Oct. 2019, doi: 10.1109/ICCVW.2019.00349.

[19] T. Chen and C. W. Lin, 'Smart and automation technologies for ensuring the long-term operation of a factory amid the COVID-19 pandemic: an evolving fuzzy assessment approach', Int. J. Adv. Manuf. Technol., vol. 111, no. 11–12, pp. 3545–3558, Dec. 2020, doi: 10.1007/S00170-020-06097-W/FIGURES/8.

PROFILES



WEN LIN YONG is pursuing her postgraduate studies in Computer Science at Tunku Abdul Rahman University College, Kuala Lumpur. At the same time, she was engaged as a research assistant which contributed to the TARUC Smart Campus initiative. She has completed her undergraduate studies with Diploma in Business Information Systems and bachelor's degree (Honours) in Enterprise Information Systems.
Email address: yongwl-wa16@student.tarc.edu.my

JUN-KIT CHAW received the BEng (Computer) and PhD (Electrical Engineering) from Universiti Teknologi Malaysia (UTM), currently serving as lecturer in the Institute of IR4.0, Universiti Kebangsaan Malaysia (UKM, Bangi), Malaysia. Previously, he served as Senior Lecturer in the Faculty of Computing and Information Technology at Tunku Abdul Rahman University College (TAR UC). He is a SAS certified predictive modeller and machine learning specialist as well as NVIDIA Deep Learning Institute certified for demonstrating competence in deep learning for computer vision. His primary research interests include computer vision and deep learning. His research projects include optimization of manufacturing processes, predictive maintenance and medical data analytics.
Email address: chawjk@ukm.edu.my

YIQI TEW obtained his B.Eng. (Hons.) electronics degree from Multimedia University, MComSc. from National University of Malaysia, and PhD degree from University of Malaya in 2018, 2011 and 2016 respectively. He is an experienced Associate Professor and skilled engineer in Embedded Systems, Video (HEVC) Processing, Security (Information Hiding). His research involves Multi-view, Depth Sensing and Real-time Video Streaming Mechanisms in Industry 4.0 and Smart Agricultural projects. With his profession, he became the lead researcher in Computational Intelligence with publication of more than 30 papers, appointed as National Expert in IoT and Cloud Computing by the Ministry of Human Resource, Malaysia for WorldSkills events in 2018, 2020 and 2021.
Email address: yiqi@tarc.edu.my

AN AGROLINK ECONOMIC SYSTEM THROUGH AN INFORMATION TOOLBOX APPROACH

(Date received: 13.01.2022/Date accepted: 09.05.2022)

Voo Nyuk Mee[1]*, Low Ying Chiang[1], Lim Yee Mei[2], Tan Hui Yin[1], Lee Wah Pheng[1]

[1] Tunku Abdul Rahman University College, Kuala Lumpur.

[2] GMCM Sdn. Bhd.

**Corresponding author: voonm@tarc.edu.my*

ABSTRACT

Malaysia agricultural sector has been severely affected by the pandemic, as a result, the Agri-food supply chain has been disrupted since early 2020.The agri-food supply chain needs to be strengthened through a tighten collaboration between supply chain stakeholders in the sector. Data can be collected through technological solutions; market insights can be achieved through analysis and forecasting. The paper focuses on introducing an Agrolink economic system through an information toolbox approach. It is an aggregator of the information resources connecting all elements of the agri-food supply chain in Malaysia, which accessible by farmers, producers, service providers of the supply chain, agri-food associations, government agencies, and educational and research institutions. The information toolbox is deployed using Apache web server, MySQL database, and Laravel web framework. Information sharing through supply chain stake holders' connection. The information presented using data cataloging, information indexing, interconnection of information resources, and ranking in the information toolbox. Several events have been organized to guide how users could surf, search, find, and connect with what they hoped to achieve, assistance to create and discover the values of their companies and how they could identify their assets into the right places of the Agrolink system.

Keywords: Agricultural economic system, Information toolbox, Food supply chain, Agrolink

1.0 INTRODUCTION

Since March 2020, Malaysia agricultural sector has been affected by Movement Control Order (Prime Minister Office of Malaysia, 2020). Since then, local news has been reporting on these issues such as fresh vegetables being discarded due to supply chain issues, labor shortages leading to reduced productivity, inter-state movement restrictions disrupting transportation and deliveries, closure of non-essential business activities affecting market access, etc. The pandemic situation has severely impacted the supply of farmers to consumer demand activities. The occurrence of all these problems seems to point to an important issue, namely the lack of connectivity among stakeholders in the agricultural supply chain, or to be precise, the absence of a connected supply chain system in Malaysian agriculture.

The rise of the technology and the pandemic have driven a surge in digital transformation, and all sectors of the economy, including agriculture, have been forced to embrace the transformation and move into a new era of digital supply chain systems. A competitive, strongly connected supply chain enables the flow of information and the flow of economic trade. A competitive agricultural supply chain economic system can be achieved by leveraging technology and information, as well as the insights gained through supply chain connectivity. Malaysia's agricultural supply chain system can be strengthened to facilitate agricultural trade flows and empower agricultural business.

There is an urgent need for all stakeholders in the agricultural supply chain to implement technology solutions. Modern agricultural equipment and machinery is needed to replace many manual workers. And internet of things devices can be embedded in farms and production activities to manage, monitor, and collect data. E-commerce marketplaces needed to facilitate business activities, especially allowing small farmers to sell their products and reach customers. Market information can be retrieved through artificial intelligence analysis and forecasting, etc. The data created from technology solutions, automation, E-commerce marketplaces, and market insights can be shared among supply chain stakeholders to create value. Therefore, Malaysia's agri-food supply chain needs to be strengthened through close collaboration between supply chain stakeholders in the sector.

This paper presents an Agrolink system to connect agri-food supply chain stakeholders through an information sharing platform. The information toolbox was created to provide information and search for information on target crops for the agri-food industry in Malaysia. The Information Toolbox is an aggregator of information resources that connects all elements of the agri-food supply chain in Malaysia. It can be used by all farmers, service providers in the supply chain, associations, government agencies and educational institutions. This paper discusses the methodology and framework used in developing the information toolbox. The key findings of supply chain system, data catalog, indexing, interlinking information resources, page ranking, and information toolbox from the literature review below have been applied to the research methodology discussed in the paper.

2.0 LITERATURE REVIEW

According to Malaysia's Digital Economy Blueprint, agriculture is one of the preferred sectors for developing the digital economy. Government initiatives include promoting the adoption of smart agriculture by creating a centralized open data platform among industry players and creating more local digital platforms to enable "farm-to-table" digital marketplaces (Prime Minister's Department, 2021). Several books and academic articles have been reviewed to identify appropriate approaches and frameworks used to develop agroecosystems through information toolbox.

In The Global Supply Chain Ecosystems, the author describes the supply chain as the artery of today's globalized economy, enabling the flow of international trade and empowering global commerce. A complex international network of suppliers, stakeholders, partners, regulators, and customers are involved to ensure the efficient and effective flow of products, services, information, and funds around the world (Millar, 2015).

Through a study of a data catalog project, researchers developed a system that allows users to annotate their data products with structured metadata, providing a discoverable, browsable index of data for data consumers. A data catalog system can provide a roadmap of useful data obtained from experiments or simulations, making it easier for researchers to find and access important data, and understand the meaning of the data and how to obtain the data (Stillerman, Frediana, Greenwald, & Manduchi, 2016).

In the research paper titled 'Discovering, Indexing and Interlinking Information Resources', the authors discuss the crawling and analyzing web resources to populate a crawler database. The processes being implemented in the International System for Agricultural Science and Technology. It applied semantic enrichment to crawled web resources and used this semantic knowledge to enhance the web portal (Celli, Keizer, Jaques, Konstantopoulos, & Vudragović, 2015).

In a book titled 'Indexing It All', the authors describe the importance of indexing and the practical application of indexing in Google PageRank. As with Google PageRank, with its relevancy-based link-analysis algorithm, citation indexing and analysis is introduced into the corpus of what is searchable by the user, making the search more contemporary, rather than simply retrospective (Day, Buckland, Furner, & Krajewski, 2014).

In a book titled 'Google's PageRank and Beyond: The Science of Search Engine Rankings', the author suggested search engine results page techniques can be applied by mathematical algorithms with ranking components such as content score and popularity score in improving business and organization rankings in the major search engines (Langville & Meyer, 2012). In a research paper titled 'A Hessenberg-type algorithm for computing PageRank Problems', the authors' described PageRank is a widespread model for analyzing the relative relevance of nodes within large graphs arising in several applications and presented a Hessenberg-type algorithm for computing PageRank Problems (Gu, 2022).

In the research paper titled 'Neuroscience Information Toolbox: An Open Source Toolbox for EEG–fMRI Multimodal Fusion Analysis', a Neuroscience Information Toolbox (NIT) has been designed and developed to provide a convenient and easy-to-use toolbox for researchers for exploring brain information in various scalp electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) studies (Li Dong *w*, 2018).

3.0 RESEARCH METHODOLOGY

An Agrolink system is shown in Figure 1. It is designed as an information toolbox, a one-stop digital platform for agri-food marketing, e-commerce, and branding in Malaysia. By leveraging marketing and branding strategies, stakeholders from suppliers to buyers are connected to the online marketplace through a searchable information toolbox. Suppliers include large-scale agriculture, contract farmers, family farming and individual farmers. Commercial establishment refer to agricultural cooperatives, government agencies, business partners, companies, and corporations. Buyers refer to consumers, end-users, and foreign buyers. Online marketplaces refer to Malaysian-owned online stores and e-commerce platforms that specialize in agri-food online stores. The diagram shown a process flow which interconnected commercial establishment, supplier, and buyer towards online marketplaces. A branding report generated to help buyers and commercial stakeholders identify the proposition delivered by the organization's marketing and brand. A link to the web store is provided to access the organization's web store.

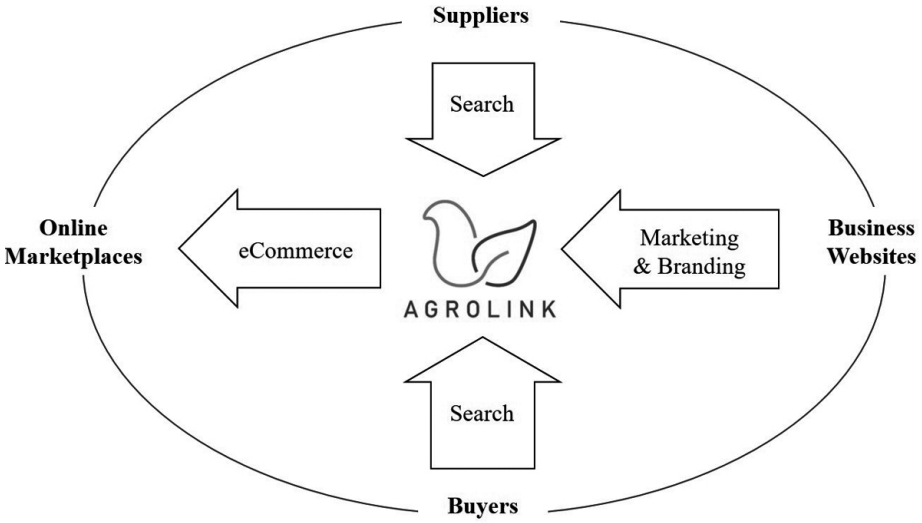


Figure 1: Agrolink economic system

The Information Toolbox (Table 1) consists of four main modules, namely, search, content, management, and forum. The search module uses techniques such as information indexing and cataloging to facilitate the storage and retrieval of information. The content module is used to manage content and resources, evaluate, and clean up. The administration module is used to support, maintain, and generate reports and statistics. Forum provides online discussions for dialogue and interaction. The programming tools including Hyper Text Markup Language, Cascading Style Sheets, Javascript, Bootstrap, Vue, Laravel, PHP: Hypertext Preprocessor, MySQL, and OAuth 2.0 are used to build the Agrolink web framework.

Table 1: The information toolbox structure consists of four main functional modules

No	Module	Sub-Modules
1	Search	Information indexing, Search Engine, Catalogue, Keyword, Label Tag, User behavior Indicator
2	Content	Content Provider, Content and Resource Management, Content Assessment, Content Cleansing
3	Management	Admin Management, Categories and Crops Management, Report and Statistic
4	Forum	Category, Post, Moderate

3.1 Information Indexing

The method of information indexing (Table 2) is described to allow users to access crop information in an organized and more structured manner. The information toolbox provides a platform for Malaysian agricultural stakeholders to freely share and contribute their knowledge and valuable information on a collaborative information toolbox platform. All contributed information is well indexed to facilitate users to search and access targeted information without searching the entire database.

The process of information indexing involves tagging information with specific tags and keywords. Information indexing tags have been identified and discussed with agricultural stakeholders. Any popular crops not listed in the information toolbox can be suggested through a feedback form.

Table 2: The method of information indexing consists of label tagging used in the index

No	Information Indexing	Labels Tagging
1	Agricultural sub-sectors	Crops, Fisheries, Livestock
2	Crops	Ginger, Star Fruit, Herb, Cucumber, Lady Finger, Broccoli, Cauliflower, Cabbage, Corn, Beans, Root, Vegetables, Melons, Chili, Eggplant, Jack Fruit, Durian, Pomelo, Pineapple, Leafy Greens, Salad, Greens, Mushroom, Coconut.
3	Fisheries	Wild Catch (Fish), Wild Catch (Crustaceans), Wild Catch (Mollusks), Aquaculture (Marine), Aquaculture (Fresh Water), Aquaculture (Crustaceans)
4	Livestock	Poultry (Broiler), Poultry (Layers)
5	Supply chains	Refer Table 3

3.2 Information Organization

The method of cataloging (Table 3) is used to organize agricultural information by creating metadata to represent information resources. The catalog divides the agri-food supply chain into five different sectors, namely production, harvesting and transportation, processing and storage, distribution, packaging and handling, and wholesale and retail (Malaysia Productivity Corporation, 2018). For each supply chain sector, more specific subsectors have been identified and categorized to enable users to locate and select the most appropriate resources. The directory provides information describing the contributors, titles, and keywords of the listed resources. This information refers to a wide range of information resources that can be accessed through the Uniform Resource Identifier. Due to the limited number of pages allowed, the level of detail is not listed.

Table 3: The method of cataloging is used to categorize the agri-food supply chain into five different sectors

No	Cataloging	Information Organization
1	Agricultural sub-sectors	Crops, Fisheries, Livestock
2	Supply Chains	Production, Harvesting and Transport, Processing and Storage, Distribution, Packaging and Handling, Wholesale and Retail
3	Crops	Refer 3.1 - 3.5
3.1	Production	About the Crop, How to Plant, Pest and Disease, Advice, Statistics, Financial Aid, Technology, Agro-Tourism, Certification, Crop Scouting, Weed Management, Organic Cultivation, Product
3.2	Postharvest	Harvest, After harvest, Transportation, Financial Aid, Technology, Product, Agro-Tourism
3.3	Processing	Uses, Financial Aid, Technology, Product, Agro-Tourism
3.4	Distribution	Financial Aid, Import and Export, Technology, Product, Market Place
3.5	Retail	Uses, Statistics, Financial Aid, Import & Export, Technology, Product, Market Place
4	Fisheries	Refer 4.1 - 4.4
4.1	Production	Fry, Farm Table Sizes, Feed Millers, Quality Control, Veterinary, Financial Aid, Technology, Certification, Product
4.2	Processing	Financial Aid, Technology, Certification, Transportation, Value Addition, Storage Facilities, Product
4.3	Distribution	Financial Aid, Technology, Uses, Statistics, Import and Export, Product
4.4	Retail	Technology, Market Place, Product
5	Livestock	Refer 5.1 - 5.4
5.1	Production	Farms, Breeds, Health and Disease, Technology, Advice @ Rules and Regulation, Product

5.2	Processing	Technology, Transportation, Certification, Packaging, Storage Facilities, Financial Aid, Product, Value Addition
5.3	Distribution	Technology, Financial Aid, Uses, Statistics, Import and Export, Product
5.4	Retail	Technology, Restaurant, Product, Market Place

3.3 Generic Website Flow of Pages

A generic site flow (Table 4) is used to guide users' access to information in a simple structure. Linear style hyperlinks provide a direct path from beginning to end, making it easy for users to follow and retrieve the resources they need. The landing page is the primary entry point for identifying user groups such as farmers, businesses, institutions, government, and the public. The secondary page shows the three main categories of the agri-food sector, namely crops, fisheries and live stocks. One of these categories can be selected by users to access the third level web page. On the third level web page, a specific sector, namely production, postharvest, processing, distribution, and retail can be selected by users. On the fourth level web page, a detail level of subsectors which is related to the selected supply chain can be selected by users to access a list of information resources.

Table 4: A generic site flow used to guide users access to information in a simple structure

No	Website Flow	Description
1	The landing page	Farmer, Supplier, Government, Institutions, Public
2	Secondary page	Crops, Fisheries, Livestock
3	Third level page	Production, Postharvest, Processing, Distribution, Retail, and the subsectors of the supply chains
4	Fourth level page	Detail level of subsectors, and a list of information resources

3.4 Ranking

The ranking criteria (Table 5) are applied, and the coding instructions for ranking calculations are developed in the information toolbox. Listing resources will be ranked using the scoring criteria and a finite sequence of coding instructions to achieve fair visibility and avoid manipulation of rankings. A standard rating scale from 1 to 5 is implemented on the star rating card as a valuable metric to fairly collect user perceptions of a listed resource. A population standard deviation formula is used to maximize accuracy as data are collected in real time and entered the algorithm instantaneously. 25 percent dispersion percentage is acceptable, which is equivalent to 0.675 standard deviations. The popularity of the listed resource is reflected by the number of viewers through user clicks. To avoid manipulation, the viewer variation must be 0.80, which means that 80 percent of the viewers must come from different viewer sites. All criteria such as total views, average rating, percentage deviation and viewer variation are combined with the rating points, resulting in rating points from high to low.

Table 5: Ranking criteria used to achieve fair visibility and avoid manipulation of rankings

No	Criteria	Details
1	Rating	From not relevant (1) to excellence (5)
2	Minimum average rating	3.50
3	Rating percentage deviation	25%
4	Minimum viewers	10
5	Viewer variation	0.80
6	Rating point	Total viewers x average rating x [1/percentage deviation] x viewer variation
7	Ranking	Rating point from high to low

3.5 Forum and Web Framework

The modules and structure of the forum (Table 6) was established to allow stakeholders to interact and initiate dialogue on agricultural topics. It provides a platform to connect Malaysian agricultural stakeholders and build an online agricultural community. An administrator acted as a moderator to moderate the discussions, maintain the quality of the forum, and keep it free from spam and irrelevant topics. The content of the forum can be viewed by visitors or guests. Registered members are allowed to post on the forum, make comments, and start new discussion topics. A file may be attached to a post, but the file size is limited.

Table 6: Modules and structure of the forum used to build an online agricultural community forum

No	Item	Details
1	Categories	General, Crops, Fisheries, Livestock, Feedback
2	Roles	Moderator, Member, Guest
3	Structure	Tree-like Three levels - Categories, Topics, Posts A Thread starting by a title, the collection of posts is displayed from the latest to the oldest
4	Posts	Member can submit messages, which consist of texts, images, and HTML tags. First post creates a thread starter, the thread can contain any number of posts.
5	Rules	Anonymous can view all topics and messages, only register member can post messages.

During the soft launched in year 2020, Agrolink information toolbox has been deployed in Tunku Abdul Rahman University College's Integrated Innovation Hub. During deployment, it also used Huawei cloud computing compute service. Laravel web application framework used to develop the Agrolink information toolbox. It is built according to the model-view-controller architectural pattern, separating the data model and business rules from the user interface. Apache software is installed as web server, MySQL is installed as a database server, and PHP Hypertext Preprocessor is used as the primary programming language. Hyper Text Markup Language, Cascading Style Sheets, Javascript, Bootstrap, and Vue are used to build the front-end user interface. The framework architecture is divided into three layers, namely, client, application, and database. OAuth 2.0 is used to provide application programming interfaces authentication. It is used to enable an external system and a chatbot built by another institution to exchange data and functionality with Agrolink information toolbox.

4.0 RESULTS AND DISCUSSION

As shown in Figure 2, the opening page of the information toolbox has been designed to identify five different user groups, namely farmers, suppliers, government, institutions, and the public. All key players or stakeholders in this complex agricultural supply chain, from farmers to end customers, are identified, with each stakeholder playing a key indicator of this information toolbox. The user groups were decided to place at the opening page, after several rounds of discussions and meetings with relevant ministry agriculture agencies, consultants, sector players, and development team members. From time to time, suggestions can be taken to be discussed in the next phase improvement and development. When stakeholders are willing to share their valuable data resources and communicate with each other through digital platforms, a connected supply chain community can be created to build a sustainable Malaysian agricultural ecosystem that helps each other in production, processes, logistics, branding, and marketing.

Figure 2: The opening page of the information toolbox

As shown in Figure 3, there are three main categories in the agri-food sector, namely crops, fisheries and livestock, which have been incorporated into the information toolbox. The crops and items under these categories are listed as indicators to guide the user to the next supply chain page. Items and products not found under the categories, but are considered as common and major commercial, agricultural products can be suggested to be added by admin through a feedback form.

Figure 3: Three main categories in the agri-food sector

As shown in Figure 4, there are five different sectors, namely production, harvesting and transportation, processing and storage, distribution, packaging and handling, and wholesale and retail. For each supply chain sector, more specific subsectors have been identified and categorized to enable users to find and select the most appropriate resources.

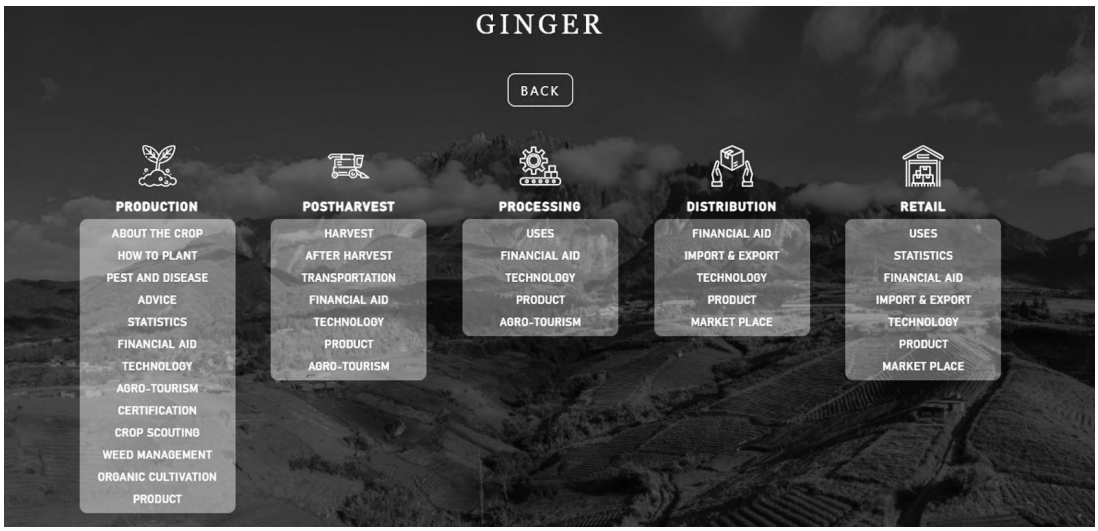


Figure 4: Five different agri-food supply chain sectors

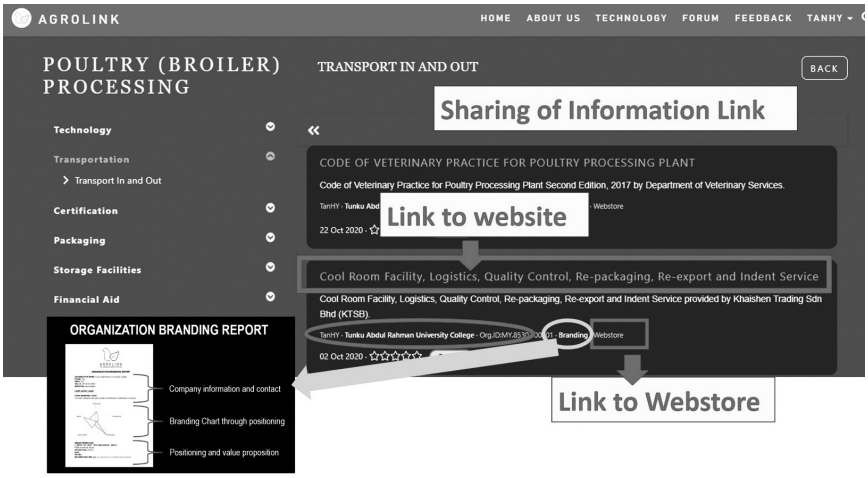


Figure 5: The resources shared by stakeholders

As shown in Figure 5, the resources shared by stakeholders include a wide range of information that can be accessed through the Uniform Resource Identifier. Contributors are listed by name and organization and a link is provided to enable access to the organization's website. An organization's brand report is generated that identifies the organization's brand positioning through a supply chain positioning chart, and the positioning and value proposition are listed in the report. The brand report helps consumers and stakeholders identify the benefits and unique selling proposition delivered by the organization's brand. A link to the web store is provided to access the organization's web store.

As shown in Figure 6, a supply chain positioning chart is generated for crops as. The chart is based on the total number of information contributions in the current information toolbox and have the potential to be expanded. The overall view of the Dynamic Supply Chain Competitive Forces Model can be outlined by the five supply chain elements, production, post-harvest, processing, distribution, and retail. It is drawn on organizational supply chain positioning to determine competitive strength. A valuable supply chain value insights can be gained from the supply chain positioning chart to inform organizational decision makers in shaping organizational brand strategy and industry competitive development.

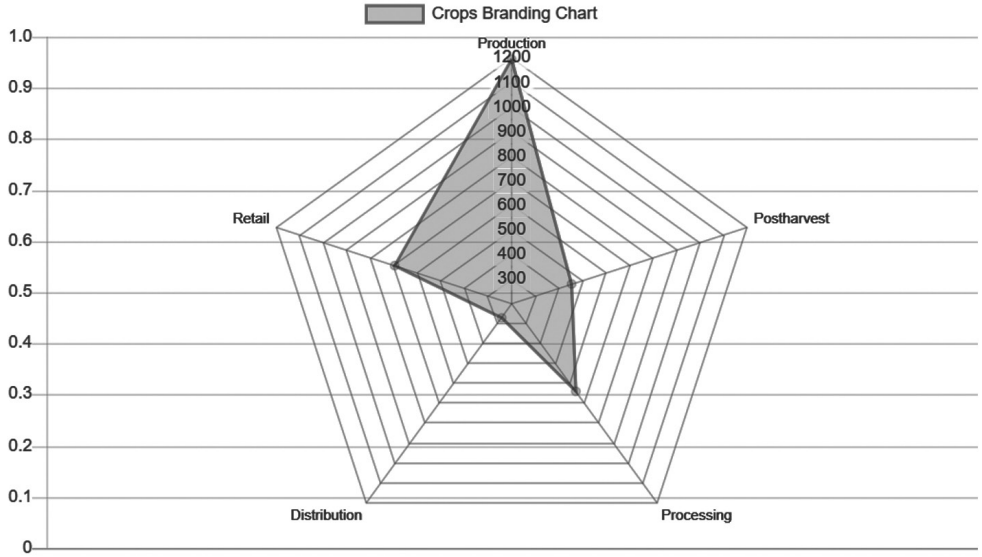


Figure 6: The supply chain positioning charts

Figure 7 presents the user response and registration statistics through registration and user data collected from the Agrolink portal in January 2021. The figure shows the relative proportions of the five user groups, with institutions at 47%, the public at 39%, farmers at 9%, vendors at 4%, and government at 1%. The Registered Organizations categorized into government agencies, industry, and higher education institutions. Number of links contributed to the three sub-sectors and to each supply chain. There were 3284 resources contributed to the crop subsector, 460 to the fisheries subsector, and 291 to the livestock subsector.

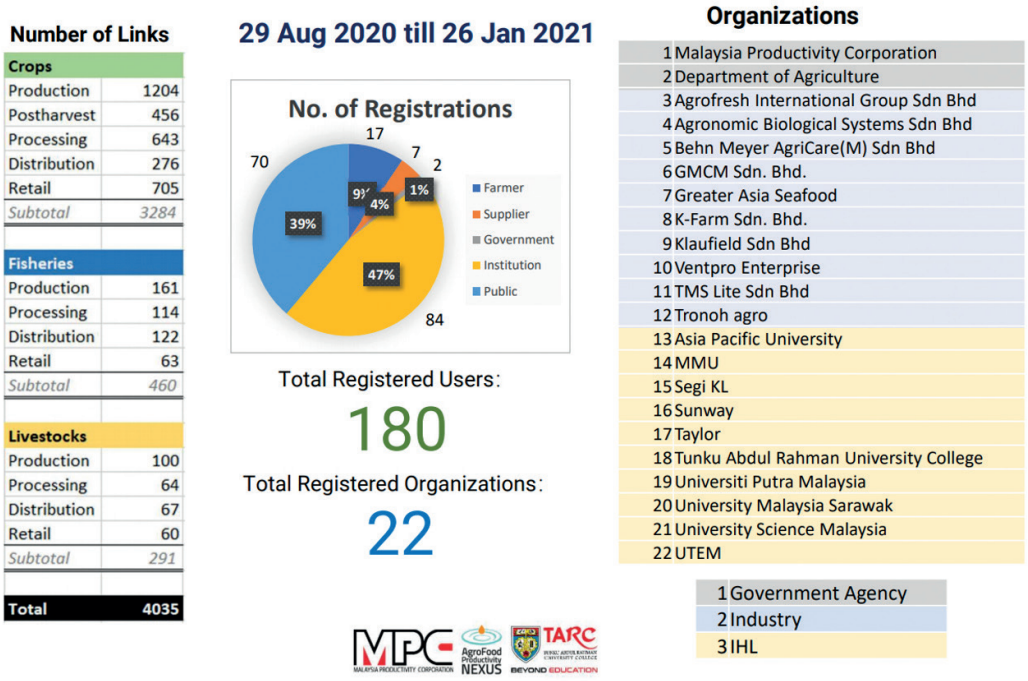


Figure 7: The user response and registration statistics

5.0 CONCLUSION

The Agrolink system aligned with the initiatives of the National Digital Blueprint by creating the Information Toolbox, a collaborative digital that connects supply chain stakeholders and shares information platform. The system is designed to fill the gap of no having any digital agricultural information toolbox to connect supply chain stakeholders in Malaysia. There are some famous shopping platforms in Malaysia such as Shopee and Lazada which are owned by Shopee Pte. Ltd. and Alibaba Group. Data sovereignty belongs to foreign companies, and Malaysian users will never use the data assets generated by the portal for business competition, operations, and productivity.

Agrolink system is deployed using Apache web server, MySQL database, and Laravel web framework. Information sharing through supply chain stake holders' connection. The information presented using data cataloging, information indexing, interconnection of information resources, and ranking in a digital information toolbox platform. A branding report generated to help buyers and commercial stakeholders identify the proposition delivered by the organization's marketing and brand. Collaboration and participation can be increased through online communication forum. OAuth 2.0 is used to provide application programming interfaces, authentication and authorization accessing an external system and application accessing resources in Agrolink information toolbox. In the initial phase, information is collected through manual input from stakeholders. Automatic connections can be deployed in the information toolbox through application programming interfaces to replace the manual approach when stakeholders digitized their operations.

Several promotion programs have been organized via online webinar. The events, dive deeper into the functions of the Agrolink information toolbox, guided how users could surf, search, find, and connect with what they hoped to achieve, assistance to create and discover the values of their companies and how they could identify their assets into the right places of the Agrolink. As a result of the promotion events, the overall response in terms of resource contribution and registration is satisfactory. However, a long-term plan and more aggressive extension efforts needed to bring in more agricultural players to build a strong and sustainable agricultural linkage among the agri-food sector in Malaysia. ■

REFERENCES

[1] Celli, F., Keizer, J., Jaques, Y., Konstantopoulos, S., & Vudragović, D. (2015). Discovering, Indexing and Interlinking Information Resources. F1000Research, 4(432). Retrieved from https://f1000research.com/articles/4-432/v2

[2] Day, R. E., Buckland, M., Furner, J., & Krajewski, M. (2014). Indexing It All: The Subject in the Age of Documentation, Information, and Data. Cambridge, MA: MIT Press.

[3] Gu, X. L. (2022). A Hessenberg-type algorithm for computing PageRank Problems. Numer Algor, 1845–1863. doi:https://doi-org.tarcez.tarc.edu.my/10.1007/s11075-021-01175-w

[4] Langville, A. N., & Meyer, C. D. (2012). Google's PageRank and Beyond: The Science of Search Engine Rankings. New Jersey: Princeton University Press.

[5] Li Dong *et al.* (2018). Neuroscience Information Toolbox: An Open Source Toolbox for EEG–fMRI Multimodal Fusion Analysis. ProQuest. doi:http://dx.doi.org/10.3389/fninf.2018.00056

[6] Malaysia Productivity Corporation. (2018). MPC 25th Productivity Report. Retrieved from https://www.mpc.gov.my/wp-content/uploads/2018/07/apr-2018.pdf

[7] Millar, M. (2015). Global Supply Chain Ecosystems. London: Kogan Page.

[8] Noor, S. (2020). AFPN Launches Agrolink Information Toolbox. Kuala Lumpur: Malaysian National News Agency. Retrieved from https://www-proquest-com.tarcez.tarc.edu.my/wire-feeds/afpn-

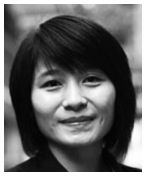
launches-agrolink-information-toolbox/docview/2439146966/se-2?accountid=38945

[9] Prime Minister Office of Malaysia. (2020). Restriction of Movement Order. Retrieved from Prime Minister's Office of Malaysia Official Website: https://www.pmo.gov.my/2020/03/movement-control-order/

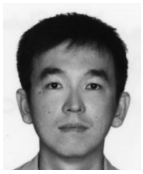
[10] Prime Minister's Department. (2021). Malaysia Digital Economy Blueprint. Putrajaya: Economic Planning Unit, Prime Minister's Department. Retrieved from https://www.epu.gov.my/sites/default/files/2021-02/malaysia-digital-economy-blueprint.pdf

[11] Stillerman, J., Frediana, T., Greenwald, M., & Manduchi, G. (2016). Data catalog project—A browsable, searchable, metadata system. Fusion Engineering and Design, 995-998.

PROFILES



VOO NYUK MEE is lecturer of Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kuala Lumpur. She received both bachelor's degree and master's degree from University of Technology Malaysia. Her specialization and area of interest are web and cloud technology.
Email address: voonm@tarc.edu.my



DR LOW YING CHIANG is senior lecturer of Faculty of Applied Sciences, Tunku Abdul Rahman University College, Kuala Lumpur. He received Diploma in Agriculture, bachelor's degree of Bioindustry Science, and PhD degree from Universiti Putra Malaysia. His specialization in Plant Pathology and area of interest are Soil and agricultural microbiology.
Email address: lowyc@tarc.edu.my

DR LIM YEE MEI received her PhD in Artificial Intelligence from De Montfort University, United Kingdom. During her tenure with TAR UC, she was appointed as Associate Dean of Department of Computer Science and Mathematics in 2017, Associate Dean of Department of ICT in 2018, the Lead of Research Centre for ICT Innovations and Creativity, and the Project Lead of TAR UC Smart Campus IoT Applications. She is currently the CEO of GMCM Sdn. Bhd., a technology company that offers Industry 4.0 services and solutions for SMEs.
Email address: yeemei@gmcmconnect.com

DR TAN HUI YIN is Associate Professor, Deputy Dean of Faculty of Applied Sciences, Tunku Abdul Rahman University College, Kuala Lumpur. She received both bachelor's degree and the Ph.D. degree from Universiti Putra Malaysia. Her specialization in bioscience, areas of interest are Gut microbiology, Genomics & Metabolism.
Email address: hytan@tarc.edu.my

DR LEE WAH PHENG is Associate Professor, Head of Centre for Postgraduate Studies and Research, Tunku Abdul Rahman University College, Kuala Lumpur. He is a pioneer and consultant in Industry 4.0. He works with a team of researchers and industry partners to develop a holistic digital solution suitable for the small and medium enterprises to start their digital transformation journey.
Email address: leewp@tarc.edu.my

AUTOMATIC IDENTIFICATION OF LIGHT VERB CONSTRUCTIONS: A REVIEW

(Date received: 13.01.2022/Date accepted: 03.04.2022)

Kathleen Swee Neo Tan[1*], Tong Ming Lim[2], Chi Wee Tan[1], Wei Wei Chew[3]

[1] Faculty of Computing and Information Technology

[2] Centre for Business Incubation and Entrepreneurial Ventures

[3] Faculty of Social Science and Humanities, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300 Wilayah Persekutuan Kuala Lumpur, Malaysia

**Corresponding author: tansn@tarc.edu.my*

ABSTRACT

Light verb constructions (LVC) are complex predicates that are present in many languages. They belong to the Multiword Expression (MWE) category known as verbal MWEs and has the canonical form of verb+noun. Examples of LVCs include give help, make decisions, and take walks. LVC identification is essential for many natural processing (NLP) applications such as machine translation, sentiment analysis, and information extraction. However, the task of LVC identification is challenging due to its characteristics such as variability, discontinuity, and ambiguity. This paper presents a review of recent work, discusses the gaps that still exist, and proposes some future work that may contribute significant progress in LVC identification.

Keywords: light verb constructions, multiword expressions, computational linguistics, natural language processing

1.0 INTRODUCTION

Multiword Expressions (MWE) are expressions that contain two or more words that are used together to convey a certain meaning. MWE identification is important for NLP tools such as part-of-speech (POS) taggers, semantic parsers and syntactic parsers, as well as downstream applications such as machine translation, emotion analysis, and question answering systems (Constant *et al*., 2017). Recently, there has been a growing interest in Light Verb Constructions (LVC), which is a type of verbal MWE (Cordeiro & Candito, 2019; Nagy T. *et al*., 2020). LVCs are complex predicates that have the canonical form of verb+noun. Examples of LVCs include *give help, make decisions,* and *take walks*. There are two particularly interesting characteristics of LVCs. Firstly, the verbal component of the LVC does not contribute much to the meaning of the LVC and are not interpreted in the literal sense. Consider the 'heavy' usages or literal meaning of the verbs used in these non-LVC examples: make implies the act of creating something (as in *make a cake*), take is an action that results in the possession of the object (e.g., *take the plate from the cupboard*), and give is the act of transferring an object to be in the possession of a subject (e.g., *gave a bouquet of flowers*). The heavy usage of these verbs is also known as *productive verbs* as their use indicate that they effect or produce some results (e.g., the creation of an object or transfer of possession). Secondly, although it is more efficient to use the synthetic verb counterpart in a sentence (e.g., the LVC *'make a review'* can actually be more efficiently replaced by its synthetic verb counterpart 'review'), there is a greater tendency to use the LVC instead due to the ease in which LVCs may be modified for greater expressiveness (Bonial & Pollard, 2020). Table 1 illustrates the use of several LVCs and their synthetic verb counterparts in sentences.

Table 1: Examples of the use of LVCs and synthetic verbs in sentences

With LVC	With synthetic verb counterpart
We *give a review* of light verb constructions in computational linguistics.	We *review* light verb constructions in computational linguistics.
Jon *took a brisk walk* around the college this morning.	Jon *walked briskly* around the college this morning.
They will *make a decision* on the new product next week.	They will *decide* on the new product next week.

The interest in LVCs has been demonstrated by work in the field of linguistics (Bonial & Pollard, 2020; Gilquin, 2019; Ong & Rahim, 2021) as well as computational linguistics in various languages (Klyueva *et al*., 2017; Maldonado *et al*., 2017; Nagy T. *et al*., 2020). The importance of LVC identification in NLP applications can be observed from recent work in the development of multilingual annotated corpora for the automatic identification of verbal MWEs which includes LVCs (Ramisch *et al*., 2018, 2020; Savary *et al*., 2017).

As an extension of our previous work (Tan *et al*., 2021b), the two main approaches in which the LVC identification task may

be framed and the types of evaluation used are presented. The remainder of this paper is organized as follows. In the next section, we present recent work related to LVC identification, followed by a discussion on observed gaps and future work that could bring novel contributions. The final section concludes the paper.

2.0 LVC IDENTIFICATION METHODS

LVC identification is the task of automatically detecting instances of LVCs in running text. The task of LVC identification is complicated by the fact that LVCs have a number of challenging characteristics which includes discontinuity (i.e. gaps between the verbal and nominal parts of the LVC such as she ***gave*** *five interesting* ***lectures***), variability (e.g., the passive form the ***lecture*** *was* ***given***), and ambiguity (e.g., the phrase *they will make the decision known to the employees* does not contain any LVCs) (Savary *et al*., 2017).

The LVC identification task may be addressed as either a classification or a sequence labeling task. For the evaluation of all MWE-types, Savary *et al*. (2017) differentiated between MWE-based evaluation and token-based evaluation. MWE-based evaluation implements strict matching in which all components of an MWE have to be correctly predicted whereas for token-based evaluation, any correctly predicted component of the MWE is counted. However, not all papers explicitly indicate the type of evaluation (i.e., MWE- or token-based) that was used. In addition, some papers provide the per-language evaluation scores while others report either the micro-average or macro-average scores across all languages. A summary and comparison of the studies on the automatic identification of LVCs in the recent years are shown in Table 2, while a chronological view of the work is shown in Figure 1. For the year 2021, the related work on LVC had been on aspects supporting the LVC identification task such as the development of annotation guidelines for LVC (Bonial, 2021), the investigation of properties for the aspectual variant of LVCs (Fotopoulou *et al*., 2021) and the study of systematic patterns for LVC families (Fleischhauer, 2021).

2.1 Classification-Based Approaches

When LVC identification is framed as a classification task, there are two main steps that need to be carried out. Firstly, LVC candidates have to be extracted. Secondly, binary classification of the extracted LVC candidates is performed using machine learning algorithms.

In the system proposed by Waszczuk (2018), a dependency tree was constructed for each sentence. For each node in the tree, the system predicted whether or not that node was an LVC based on local contextual information which included word forms, POS tags, dependency labels, lemmas, and so on. Next, segmentation to determine the LVC boundaries was carried out by constructing a hypergraph which represented all traversals of the dependency tree. The hypergraph contained a distinct hyperpath for each traversal. Using features extracted from each traversal's hyperpath, a multiclass logistic regression model was then used to determine the hyperpath with the highest probability to find the globally optimal labeling for the given dependency tree. As future work, they suggested improving the LVC segmentation part by considering the incorporation of lexicons and word embeddings in their system.

Cordeiro & Candito (2019) extracted LVC candidates based on syntactic patterns that considered LVC variations such as morphosyntactic variations, complex nominal components, and other language-specific characteristics. By extracting language-specific morphosyntactic patterns comprising the POS tag and syntactic relation between components from the LVCs present in the training dataset, the frequently occurring patterns were used to identify LVC candidates in the datasets. Binary classification of the LVC candidates was then carried out using support vector machine (SVM) and feed-forward neural network (FFN). They suggested to investigate the use of contextualized word embeddings as future work.

Nagy T. *et al*. (2020) used a decision-tree to perform classification of LVC candidates. First, dependency parsers were used to produce dependency representations of the corpus. LVC candidates were then extracted from raw text based on the syntactic relations between the verbal and nominal components of LVCs. A rich feature set that included both language-independent (i.e., statistical, lexical, morphological, syntactic, and orthographic features) and language-dependent (auxiliary verbs, gender, and agglutinative morphology) features were constructed. They found that the performance of their method depended on the quality of the dependency parsers used for LVC candidate extraction.

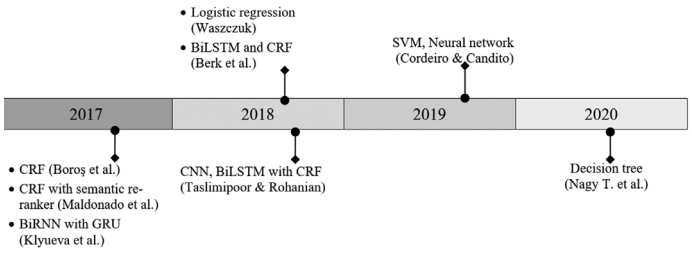


Figure 1: Chronological view of work on automatic LVC identification

2.2 Sequence Labeling Approaches

Some studies approached LVC identification as a sequence labeling problem where each word in a sentence is tagged using the Begin-Inside-Outside (BIO) tagging scheme (Ramshaw & Marcus, 1999). A word would be tagged with *B* to indicate that it is the beginning word of the LVC, *I* if the word is a component word of the LVC (i.e., inside the LVC), and *O* if the word is outside of any LVCs. The BIO tagging scheme has the benefit of being able to handle discontinuous LVCs which have components that are not adjacent to each other such as *the* ***decision*** *that was* ***made***. The use of the BIO tagging scheme is illustrated in Table 3.

Boroş *et al*. (2017) used Conditional Random Fields (CRF) to predict the transition between labels. For each word in a sentence, the lemma and POS tags for the window of words surrounding the current word served as the features. The identification of LVCs involved two steps. First, head labeling to identify the head word of the LVC (in this case the light verb) was carried out using a window size of 2. The second step was tail labeling to identify that nominal component of the LVC using a window size of 4. The authors found that compared to the single-step detection of LVC, their two-step approach increased precision by 9%.

Table 2: Summary of Papers on LVC Identification

Author(s)/Year	Approach	Method	Languages	F1-scores
Boroş *et al.* (2017)	Sequence labeling	CRF	CS, DE, EL, ES, FR, HU, IT, MT, RO, SL, SV, TR	MWE-based: 5.84% (MT) – 86.27% (RO)
Maldonado *et al.* (2017)	Sequence labeling	CRF with semantic re-ranker (SEM)	CS, DE, EL, ES, FR, HU, IT, MT, PL, PT, RO, SL, SV, TR	CRF MWE-based: 1.22% (SL) – 46.24% (PT) CRF Token-based: 4.30% (SL) – 57.08% (PT) CRF with SEM MWE-based: 1.19% (SL) – 52.67% (PT) CRF with SEM Token-based: 3.97% (SL) – 56.83% (PT)
Klyueva *et al.* (2017)	Sequence labeling	BiRNN with GRU	BG, CS, DE, EL, ES, FR, HE, HU, PL, PT, RO, SL, TR	MWE-based: 0% (BG, DE, SL) – 37% (PT) Token-based: 1% (SL) – 49% (PT)
Waszczuk (2018)	Classification	Logistic regression	BG, DE, EL, EN, ES, EU, FA, FR, HE, HI, HR, HU, IT, LT, PL, PT, RO, SL, TR	MWE-based: mAvg: 46.03
Berk *et al.* (2018)	Sequence labeling	BiLSTM and CRF	BG, DE, ES, FR, HU, IT, PL, PT, RO, SL	MWE-based: 28.55% (ES) – 74.48% (HU) Token-based: 35.66% (ES) – 81.86% (RO)
Taslimipoor & Rohanian (2018)	Sequence labeling	Convolutional, BiLSTM with CRF	BG, DE, EL, EN, ES, EU, FA, FR, HE, HI, HR, HU, IT, LT, PL, PT, RO, SL, TR	MWE-based: 6.25% (EN) – 86.15% (RO) Token-based: 11.25% (EN) – 87.41% (RO)
Cordeiro & Candito (2019)	Classification	SVM, Neural network	BG, DE, EL, EN, ES, EU, FA, FR, HE, HI, HR, HU, IT, LT, PL, PT, RO, SL, TR	MWE-based: SVM: 26% (EN) – 81 % (HU), µAvg = 63% FFN: 21% (HE) – 78% (HI), µAvg = 56%
Nagy T. *et al.* (2020)	Classification	Decision tree	DE, EN, ES, HU	50.64% (DE), 52.90 (ES), 64.72% (HU), 65.35% (EN)

F1-scores: mAvg-macro-average score across all languages; µAvg: micro-average score across all languages.
For papers with more than 4 languages, only the lowest and highest F1-scores are indicated for brevity.
Some papers did not indicate whether the evaluation was MWE-based or token-based, and some papers only reported the average F1-score across all languages.

Language codes: BG-Bulgarian, CS-Czech, DE-German, EL-Greek, EN-English, ES-Spanish, EU-Basque, FA-Farsi, FR-French, HE-Hebrew, HI-Hindi, HR-Croatian, HU-Hungarian, IT-Italian, LT-Lithuanian, MT-Maltese, PL-Polish, PT-Portuguese, RO-Romanian, SL-Slovene, SV-Swedish, TR-Turkish

Table 3: Example of a sentence with BIO tags identifying an LVC

Sentence	He will make a very difficult decision later							
Word	He	will	**make**	a	very	**difficult**	decision	later
Tag	*O*	*O*	*B*	*O*	*O*	*O*	*I*	*O*

Maldonado *et al*. (2017) used a CRF model which exploited syntactic dependency features and also included an optional semantic re-ranker as a post-processing step. Instead of developing feature sets for each of the 14 languages, the authors created a feature set for each language family based on the assumption that the morphosyntactic relationships among closely related languages would be similar. Experiments on semantic re-ranking using a regression model trained on semantic vectors was conducted for 12 languages and showed improvement in 7 of the languages. For future work, the authors plan to focus on language-specific features. In addition, they suggested looking into word embeddings as a possible way to improve the performance of the model.

Berk *et al*. (2018) proposed a bidirectional Long Short-Term Memory (LSTM)-CRF model in which the inputs to the model consisted of the POS tags and dependency relation tags. In the BiLSTM layer, the forward LSTM unit enabled the previous words to be used as features whereas the backward LSTM unit enabled the future words to be used as features. The CRF layer enabled the decoding of the sequence labels using the gappy, 1-level variant of the BIO tagging scheme proposed by Schneider *et al*. (2014).

Taslimipoor & Rohanian (2018) proposed a model comprising two convolutional layers that serve as n-gram detectors, a BiLSTM for handling long distance relationships between words, and an optional CRF layer to process dependencies among the output tags. They used pretrained Wikipedia word embeddings (Bojanowski *et al*., 2017) and binary word shape features to indicate whether the token started with an uppercase letter, was entirely in uppercase, had a # or @ as the first character, was a URL, contained a number, or was a digit. They reported that the pre-trained embeddings achieved the best performance and that the additional CRF layer did not necessarily improve the performance of the model.

The model proposed by Klyueva *et al*. (2017) was based on a bidirectional recurrent neural network (RNN) with gated-recurrent units (GRUs) that was trained using linguistic, morphological and syntactic features. Each input word was represented as a concatenation of embeddings of the word's form, lemma, and POS tag. They found that discontinuous LVCs were often not tagged. Their model did not consider embedded or overlapping LVCs.

3.0 DISCUSSION

Despite significant progress in LVC identification, there are several gaps yet to be addressed.

Large variation in the evaluation results for different languages: Table 2 shows that there is a large difference between the lowest and highest per-language F1-scores for almost all the papers that reported per-language results for the LVC identification task. One reason for this was the training corpus size differences for the languages (Berk *et al*., 2018; Cordeiro & Candito, 2019) - a larger training corpus would inevitably result in better results. Secondly, the training corpus for certain languages had lower average occurrences of LVCs (Waszczuk, 2018). This means that the machine learning models would encounter fewer examples of LVCs, which may be a hindrance to the learning process (Berk *et al*., 2018; Nagy T. *et al*., 2020). Thirdly, languages with dependency parsing tools of lower quality would inadvertently affect the quality of predictions. The difference in results across languages may also be due to the distribution of *seen* and *unseen LVCs* in the training and evaluation datasets. *Seen LVCs* refer to LVCs that occur at least once in the training dataset while those that are *unseen* were not present in the training dataset at all but appeared in the evaluation dataset (Cordeiro & Candito, 2019). Therefore, one important future work would be to explore methods for improving predictions for unseen LVCs.

Lack of tools and resources for under-resourced languages: LVCs may appear in a variety of forms due to morphosyntactic variations – for example, the nominal part of the LVC may be a complex noun phrase instead of a single noun or may even be further discontinuous whereby the verbal and nominal components are separated by many words (e.g., *the* ***walk*** *in the beautiful and dense rainforest that he had* ***taken***). This makes it particularly challenging especially because most existing work rely on the use of POS-taggers and dependency parsers to capture morphosyntactic variations. For under-resourced languages, the lack of such tools and LVC-annotated corpora provides the motivation to explore methods for LVC identification that do not depend on the use of such tools or that require a smaller annotated corpus.

Code-mixed LVCs: One growing challenge is the identification of code-mixed LVCs which are commonly used in social media. Code-mixed text includes words from two or more languages – an increasingly common phenomenon in recent times as more people are bilingual. There has been a growing interest in sentiment analysis of code-mixed text (Lo *et al*., 2017; Sasidhar *et al*., 2020; Wang *et al*., 2017) and even in Malay-English code-mixed text (Abu Bakar *et al*., 2020; Tan *et al*., 2020). The impact of code-mixed MWE identification on the emotion detection task (Tan *et al*., 2021a) and use of code-mixed LVCs (Alexiadou, 2017; González-Vilbazo & López, 2011) have also been the focus of research. To our knowledge, there has not been work on code-mixed LVC identification and therefore, this needs to be addressed to enable LVCs to be treated as a single semantic unit and avoid loss of contextual meaning that arise from the individual words of the LVC being considered as separate features.

Use of word embeddings: To overcome the problems faced in resource-poor languages and code-mixed LVC identification, the use of word embeddings should be further explored. In particular, the use of word embeddings with character n-grams to represent out-of-vocabulary (OOV) words (Bojanowski *et al*., 2017) which are often either intentionally/unintentionally misspelt words or slang words that are prevalent in social media can potentially help as they are able to reflect the semantic relationships between words. Two interesting ideas to consider is to train word embeddings using a code-mix corpus as was done for POS tagging by Bhattu *et al*. (2020), and to include LVCs and other MWEs as single tokens in embeddings.

4.0 CONCLUSION

LVCs are particularly challenging to identify due to their flexibility, ambiguity, and discontinuity. This study explored recent trends in LVC identification, a task which plays an

important role in downstream text processing tasks such as emotion analysis, machine translation, and question answer systems. In addition, existing gaps were discussed, and several promising future works were identified including investigation of methods that do not require the use of dependency parsers and POS taggers, training embeddings that represent each LVC as a single token, as well as the use of code-mixed word embeddings to facilitate the identification of code-mixed LVCs. ■

REFERENCES

[1] Abu Bakar, M. F. R., Idris, N., Shuib, L., & Khamis, N. (2020). Sentiment Analysis of Noisy Malay Text: State of Art, Challenges and Future Work. IEEE Access, 8, 24687–24696. https://doi.org/10.1109/ACCESS.2020.2968955

[2] Alexiadou, A. (2017). Building verbs in language mixing varieties. Zeitschrift Fur Sprachwissenschaft, 36(1), 165–192. https://doi.org/10.1515/zfs-2017-0008

[3] Berk, G., Erden, B., & Güngör, T. (2018). Deep-BGT at PARSEME Shared Task 2018: Bidirectional LSTM-CRF Model for Verbal Multiword Expression Identification. Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018), 248–253.

[4] Bhattu, S. N., Nunna, S. K., Somayajulu, D. V. L. N., & Pradhan, B. (2020). Improving Code-mixed POS Tagging Using Code-mixed Embeddings. ACM Transactions on Asian and Low-Resource Language Information Processing, 19(4). https://doi.org/10.1145/3380967

[5] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics, 5, 135–146. https://doi.org/10.1162/tacl_a_00051

[6] Bonial, C. (2021). Précis of Take a Look at This! Form, Function, and Productivity of English Light Verb Constructions. Colorado Research in Linguistics, 209. https://search.proquest.com/docview/1651531484?accountid=13375

[7] Bonial, C., & Pollard, K. A. (2020). Choosing an event description: What a PropBank study reveals about the contrast between light verb constructions and counterpart synthetic verbs. Journal of Linguistics, 56, 577–600. https://doi.org/10.1017/S0022226720000109

[8] Boroş, T., Pipa, S., Mititelu, V. B., & Tufiş, D. (2017). A data-driven approach to verbal multiword expression detection. PARSEME Shared Task system description paper. Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), 121–126. https://doi.org/10.18653/v1/w17-1716

[9] Constant, M., Eryigit, G., Monti, J., Plas, L. van der, Ramisch, C., Rosner, M., & Todirascu, A. (2017). Multiword Expression Processing: A Survey. Computational Linguistics - Association for Computational Linguistics, 43(4), 837–892. https://doi.org/10.1162/COLI

[10] Cordeiro, S. R., & Candito, M. (2019). Syntax-based identification of light-verb constructions. The 22nd Nordic Conference on Computational Linguistics (NoDaLiDa 2019).

[11] Fleischhauer, J. (2021). Light Verb Constructions and Their Families - A Corpus Study on German stehen unter-LVCs. Proceedings of the 17th Workshop on Multiword Expressions, 63–69. https://doi.org/10.18653/v1/2021.mwe-1.8

[12] Fotopoulou, A., Laporte, E., & Nakamura, T. (2021). Where Do Aspectual Variants of Light Verb Constructions Belong? Proceedings of the 17th Workshop on Multiword Expressions, 2–12. https://doi.org/10.18653/v1/2021.mwe-1.2

[13] Gilquin, G. (2019). Light verb constructions in spoken L2 English: An exploratory cross-sectional study. International Journal of Learner Corpus Research, 5(2), 181–206. https://doi.org/10.1075/ijlcr.18003.gil

[14] González-Vilbazo, K., & López, L. (2011). Some properties of light verbs in code-switching. Lingua, 121, 832–850. https://doi.org/10.1016/j.lingua.2010.11.011

[15] Klyueva, N., Doucet, A., & Straka, M. (2017). Neural Networks for Multi-Word Expression Detection. Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), 60–65.

[16] Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2017). Multilingual sentiment analysis: from formal to informal and scarce resource languages. Artificial Intelligence Review, 48, 499–527. https://doi.org/10.1007/s10462-016-9508-4

[17] Maldonado, A., Han, L., Moreau, E., Alsulaimani, A., Chowdhury, K. D., Vogel, C., & Liu, Q. (2017). Detection of Verbal Multi-Word Expressions via Conditional Random Fields with Syntactic Dependency Features and Semantic Re-Ranking. Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), 114–120. https://doi.org/10.18653/v1/w17-1715

[18] Nagy T., I., Rácz, A., & Vincze, V. (2020). Detecting light verb constructions across languages. Natural Language Engineering, 26, 319–348. https://doi.org/10.1017/S1351324919000330

[19] Ong, C. S. B., & Rahim, H. A. (2021). Nativised structural patterns of make light verb construction in Malaysian English. Concentric, 47(1), 93–112. https://doi.org/10.1075/consl.00024.rah

[20] Ramisch, C., Cordeiro, S. R., Savary, A., Vincze, V., Mititelu, V. B., Bhatia, A., Buljan, M., Candito, M., Gantar, P., Giouli, V., Güngör, T., Hawwari, A., Iñurrieta, U., Kovalevskaite, J., Krek, S., Lichte, T., Liebeskind, C., Monti, J., Escartín, C. P., … Walsh, A. (2018). Edition 1.1 of the PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG 2018), 222–240. https://aclanthology.org/W18-4925

[21] Ramisch, C., Savary, A., Guillaume, B., Waszczuk, J., Candito, M., Vaidya, A., Barbu Mititelu, V., Bhatia, A., Iñurrieta, U., Giouli, V., Güngör, T., Jiang, M., Lichte, T., Liebeskind, C., Monti, J., Ramisch, R., Stymne, S., Walsh, A., & Xu, H. (2020). Edition 1.2 of the PARSEME Shared Task on Semi-supervised Identification of Verbal Multiword Expressions. Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons, 107–118. https://aclanthology.org/2020.mwe-1.14

[22] Ramshaw, L. A., & Marcus, M. P. (1999). Text Chunking Using Transformation-Based Learning. Proceedings of the Third ACL Workshop on Very Large Corpora, 157–176. https://doi.org/10.1007/978-94-017-2390-9_10

[23] Sasidhar, T. T., Premjith, B., & Soman, K. P. (2020). Emotion Detection in Hinglish(Hindi+English) Code-Mixed Social Media Text. Procedia Computer Science, 171, 1346–1352. https://doi.org/10.1016/j.procs.2020.04.144

[24] Savary, A., Ramisch, C., Cordeiro, S. R., Sangati, F., Vincze, V., Qasemizadeh, B., Candito, M., Cap, F., Giouli, V., Stoyanova, I., & Doucet, A. (2017). The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), 31–47. http://multiword.sf.net/

[25] Schneider, N., Danchik, E., Dyer, C., & Smith, N. A. (2014). Discriminative Lexical Semantic Segmentation with Gaps: Running the MWE Gamut. Transactions of the Association for Computational Linguistics, 2, 193–206. https://doi.org/10.1162/tacl_a_00176

[26] Tan, K. S. N., Lim, T. M., & Lim, Y. M. (2020). Emotion Analysis Using Self-Training on Malaysian Code-Mixed Twitter Data. Web Based Communities and Social Media 2020, 181–188. https://www.elearning-conf.org/wp-content/uploads/2020/07/01_202008L022_F048.pdf

[27] Tan, K. S. N., Lim, T. M., & Tan, C. W. (2021a). A Study on Multiword Expression Features in Emotion Detection of Code-Mixed Twitter Data. 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), 1–5. https://doi.org/10.1109/IICAIET51634.2021.9573850

[28] Tan, K. S. N., Lim, T. M., Tan, C. W., & Chew, W. W. (2021b). Review on Light Verb Constructions in Computational Linguistics. International Conference on Digital Transformation and Applications 2021 (ICDXA 2021), 153–160.

[29] Taslimipoor, S., & Rohanian, O. (2018). SHOMA at Parseme Shared Task on Automatic Identification of VMWEs: Neural Multiword Expression Tagging with High Generalisation. ArXiv Preprint ArXiv:1809.03056. http://arxiv.org/abs/1809.03056

[30] Wang, Z., Lee, S. Y. M., Li, S., & Zhou, G. (2017). Emotion Analysis in Code-Switching Text with Joint Factor Graph Model. IEEE/ACM Transactions on Audio Speech and Language Processing, 25(3), 469–480. https://doi.org/10.1109/TASLP.2016.2637280

[31] Waszczuk, J. (2018). TRAVERSAL at PARSEME Shared Task 2018: Identification of Verbal Multiword Expressions Using a Discriminative Tree-Structured Model. Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018).

PROFILES

KATHLEEN TAN SWEE NEO received her BS (Computer Science and Physics) in 1992 from Campbell University, US and MSc in Management of Information Technology in 2003 from University of Sunderland, UK. She is a principal lecturer in the Department of Computer Science and Embedded Systems at TAR UC and is currently a PhD candidate in the Faculty of Computing and Information Technology, TAR UC. Her research interests include multiword expression identification and emotion analysis on code-mixed social media text, as well as big data analytics.
Email address: tansn@tarc.edu.my

DR TAN CHI WEE received BCompSc(Hons) and PhD degrees in year 2013 and 2019 respectively in Universiti Teknologi Malaysia. Currently, he is a Senior Lecturer cum Programme Leader at Tunku Abdul Rahman University College and actively involved in the Centre of Excellence for Big Data and Artificial Intelligent (CoE) and become the research group leader for Audio, Image and Video Analytics Group under Centre for Data Science and Analytics (CDSA). Dr Tan's main research areas are Computer Vision (CV), Image Processing (IP) and Natural Language Processing (NLP) and Artificial Intelligence (AI). He is an enthusiastic researcher experienced in conducting and supporting research into Image Processing. Being a meticulous and analytical researcher with Train-The-Trainer certificate of many years of educational and hands-on experience, he was invited to Université d'Artois (France) under Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) programme for collaborative research between European countries with Southeast Asian countries on motion detection and computer vision and being involved in industry project as professional consultant.
Email address: chiwee@tarc.edu.my

PROFESSOR LIM has about 10 years of industry experiences in the design, development, implementation and maintenance of commercial software from 1989 to 1999 after departing from TARC where he spent his early days with TARC as an IT lecturer from 1987 to 1989 after returning from Mississippi State University USA with a Master of Computer Science degree. He is currently the Director for CBIEV at TAR UC, Professor at FOCS at TAR UC and Head for Big Data Analytics Centre. His research interest involving Natural Language Processing, Sentiment Analysis and Code-Mixed language analysis. In the last 15 years, his work has consistently focused on organizational knowledge sharing and technology acceptance, social media analytics and social influence maximization in Sunway University and Tunku Abdul Rahman University College (TAR UC). Professor Lim has graduated more than 20 master and 2 PhD students while he was with Monash, UTAR and Sunway University.
Email address: limtm@tarc.edu.my

THE LATE DR CHEW WEI WEI obtained her BA (Honours) in Malay Studies in 1997 and Diploma in Translation in 2000 from the University of Malaya, and her MA in Translation in 2005 and PhD in English Language Skills in 2014 from Universiti Sains Malaysia. Besides her 20 years of teaching experience, she was involved in research focused on multilingualism and the use of code-mixed Malay-English in social media. In addition, she was a Course Leader for Nation Building and Languages (2013 - 2016), Chairperson for the Centre for Social Integration and Social Skills Research (2017-2019) as well as co-author for the textbooks Pendidikan Moral (2011) and Civic Consciousness and Volunteerism (2019) at Tunku Abdul Rahman University College.

PRODUCT LIFE CYCLE DATA MANAGEMENT AND ANALYTICS IN RAMI4.0 USING THE MANUFACTURING CHAIN MANAGEMENT PLATFORM

(Date received: 13.01.2022/Date accepted: 21.04.2022)

Chew Khai Min[1*], Lee Wah Pheng[2], Lee Yoon Ket[3]

[1] Faculty of Engineering and Technology, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia

[2] Centre for Postgraduate Studies and Research, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia

[3] Faculty of Engineering and Technology, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia

**Corresponding author: chewkm@tarc.edu.my*

ABSTRACT

The RAMI4.0 model consists of three axes; the Layers axis, the Hierarchy axis, and the Life Cycle Value Stream axis. This model unifies the various aspects of I4.0 to allow data generated from manufacturing and business activities to be shared effectively. The Manufacturing Chain Management software aims to provide a platform where all three axes and their associated data are tightly integrated and can be used to provide I4.0 connectivity as well as insights into the manufacturing supply chain. A data framework is proposed whereby the data from these activities can be collected and used to make more insightful decisions about an organisations value chain, value stream, and the life cycle of their product portfolio. The MCM platform is aimed at SMEs, which has lesser financial ability to invest into I4.0 technologies. Hence, the MCM platform is designed to be flexible and scalable whilst maintaining compliance with I4.0 standards. A furniture factory in Selangor, Malaysia was chosen to implement the Value Stream Mapping aspect if MCM and key production metrics such as process time, Work-In-Progress time, and manufacturing throughput were successfully recorded.

Keywords: RAMI4.0, product life cycle, value chain analysis, value stream mapping, manufacturing chain management

1.0 INTRODUCTION

Product life cycle management (PLM) is the process of managing a product from its conception through to its disposal. The effective management of a business's product portfolio allows it to stay competitive and sustainable (Cohen and Whang, 1997). In the Reference Architecture Model Industrie 4.0 (RAMI4.0), the product lifecycle is further developed and its integration with other business and production activities more fully defined. This is captured in the product lifecycle axis of the model and defines products and components in a way that allows it to be traceable throughout its lifecycle. The Bass diffusion model was introduced in 1969 and is used even today to predict the adoption of new products, thereby forming the product life cycle (PLC) curve. It has been shown that the PLC curve can be used to predict future demand. This helps organisations make better decisions regarding supply chain and inventory and is especially useful when coupled with acquisition of real-time or near real-time data (Hu *et al*., 2019).

In the context of RAMI4.0, the PLC contains two separate but related concepts, the type, and the instance. The type refers to the product as it is being developed and is related to data created in the development of the product such as CAD drawings and customer requirements. Once the product has been developed and enters production, it becomes an instance, containing data such as unique production identification or manufacturing process information. (VDI and ZVEI, 2015). This data itself provides value. For example, data on user experience can be collected from instances and used to develop an improved product type. Type data can be easily shared to stakeholders e.g., product specifications being shared to a manufacturer. The ease of collecting and distributing data allows more stakeholders to create value.

1.1 Value Chain and Value Stream

The concept of the value chain, introduced by Michael Porter in 1985, describes every business as nine generic categories of activities (Fearne *et al*., 2012). In contrast, value streams are used to conceptualise the manufacturing activities related to the product. Developed from the Toyota Production System and subsequent Lean Manufacturing tenets, value stream mapping has been shown to be effective in applying and validating implementation of lean manufacturing principles (Gurumurthy and Kodali, 2011), as well as identifying and tracking metrics of interest (Faulkner and Badurdeen, 2014).

1.2 Industry 4.0 Reference Architecture Model

The RAMI4.0 model was introduced in 2011 by Zentralverband Elektrotechnik- und Elektronikindustrie eV. (ZVEI) and Plattform Industrie 4.0 and serves as an architectural model for the implementation of Industry 4.0 (I4.0). Figure 1 shows a map of the most important aspects of I4.0, visualised as a 3-dimensional map. The three axes of the model are the Layers, Hierarchy Levels, and Life Cycle Value Stream.

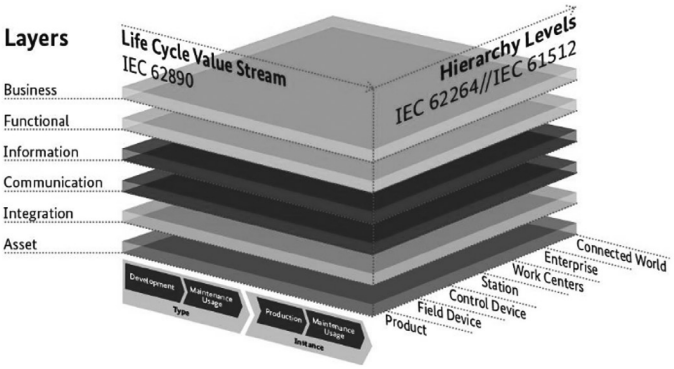


Figure 1: Map of key aspects of I4.0 (Plattform Industrie 4.0 and ZVEI)

The Life Cycle Value Stream axis is related to the development, production, and maintenance of products and services. Data created during these processes can provide value to other stakeholders. Thus, the proper management of said data is important in maintaining useability (VDI and ZVEI, 2015).

There is a large amount of data generated by the interactions between these stakeholders (DigitalEU, 2020), which is further complicated by different and competing standards. One proposed solution to this is the use of the Asset Administration Shell (AAS). Proposed and maintained by Plattform Industrie 4.0 and ZVEI, AAS is an interface that standardises the structure of information about assets. Assets include tangible and intangible resources such as a machine, component, or service. The AAS connects the asset to the wider connected I4.0 world, and ensures compatibility and interoperability along the Layers and Hierarchy axes (Plattform Industrie 4.0, 2019). The data collected is also used in the Life Cycle Value Stream axis.

With the proliferation of IoT and other information technologies, I4.0 is becoming increasingly viable, and many companies are looking to implement I4.0 in their systems. Major software vendors such as SAP and Oracle offer highly developed implementations of I4.0. However, their services are expensive, costing hundreds of thousands of ringgits or more per month (SAP, 2020). There is a market for I4.0 implementations on more modest budgets. In Malaysia, SMEs contribute 38.3% of national GDP and 66.2% of employment and make up 7.7% of the Malaysian manufacturing sector by GDP (SME Corp, 2018). This represents an opportunity for providing I4.0 consultation and software services at lower costs than more established vendors (Masood and Sonntag, 2020). While there is much research regarding specific aspects of manufacturing technology, there is a lack of research regarding the entire chain of manufacturing (Osterrieder *et al*., 2020). New strategies must be developed to help companies leverage the flexibility and customisation of I4.0 (Kumar *et al*., 2020).

2.0 MANUFACTURING CHAIN MANAGEMENT PLATFORM

The Manufacturing Chain Management (MCM) platform provides real-time data of processes in the horizontal and vertical manufacturing chain, driving value for the organisation. Figure 2 shows the vertical and horizontal manufacturing chain. The horizontal manufacturing chain includes all links in the supply chain from raw material to use and disposal. The vertical manufacturing chain includes all manufacturing and processing activities associated with the production of goods. The MCM platform provides a framework for data and information from all links to be shared with one another, fulfilling the purpose of the RAMI4.0 model.

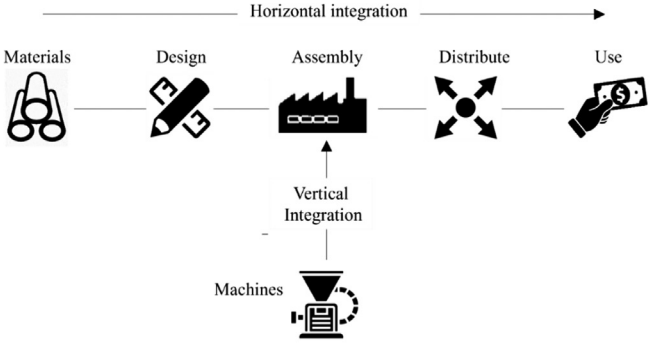


Figure 2: The horizontal and vertical manufacturing chain

The real-time nature and comprehensiveness of the data and the means that quantitative tools and methods benefit from the MCM platform. Examples include quality control methods such as Condition Monitoring, allowing manufacturing process data to be more quickly collected and analysed (Pethig *et al*., 2017).

The product lifecycle axis of the model also benefits from this framework. The product type benefits from up-to-date information in its conception and development (Suarez-Fernandez de Miranda *et al*., 2020). Information such as customer requirements and manufacturing capabilities ensures that the product is suited to the market. As an example, having access to customer and market trends means that the product can meet customer needs more easily. Traditional customer requirements capture methods such as surveys as focus groups combined with data pulled directly from the horizontal manufacturing chain can better guide the type development process. Similarly, data easily available from all links in the manufacturing chain benefits the product instance. Lead time, material quantities, quality indicators, and demand forecasts are all information that can provide for more accurate planning of production.

Qualitative methods, though requiring human input and judgement, can still benefit from real time data. Two such qualitative methods, value chain analysis and value stream mapping were selected as possible candidates for integration into the MCM product life cycle framework. Both methods will be more effective with the advantage of real-time data providing a more accurate snapshot of the current state of the organisation.

Value Stream Mapping (VSM) is the process of mapping processes that occur in the manufacturing of a product. This mapping

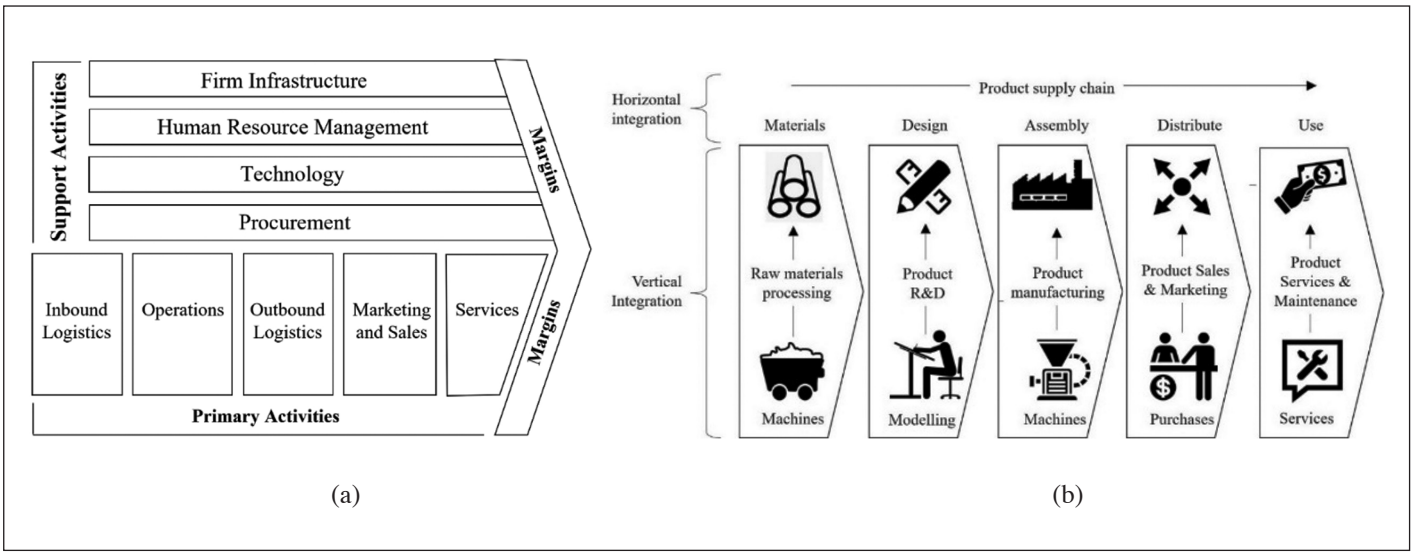


Figure 3: (a) Porter's Value Chain model for organisations and (b) integration of horizontal and vertical manufacturing chains for real-time data collection

can reveal inefficiencies in the manufacturing process, and the manufacturer can take steps to reduce wastage, thereby increasing value to the customer. The usefulness of VSM is in identifying and tracking metrics that of most interest to the company (Gurumurthy and Kodali, 2011). Examples of metrics include Work-in-Progress time, lead time, and material handling time.

The value chain is the characterisation of an organisation's business activities into nine categories, illustrated in Figure 3(a). Each of these activities create value for the organisation, and similarly to VSM, an analysis of these activities can help the organisation increase value for the customer (Koc and Bozdag, 2017). Activities along the value chain are represented in Figure 3(b), illustrating the different actions and the eventual goal of increasing margins for the company, and thusly the value to the customer.

3.0 METHODOLOGY

The Asset Administration System (AASystem) is one of the functions in the MCM Platform. The AASystem provides an AAS for all assets in the manufacturing supply chain. The Manufacturing Chain Broker (MCB) brokers the data across the different activities in the chain. This allows for complete transparency of manufacturing supply chain data, allowing it to be easily shared or processed for analytics uses. The fully integrated nature of the manufacturing supply chain data also creates a more holistic view of the system, as well as the effects of changes in one chain affecting other chains. Stakeholders across the manufacturing supply chain will be more agile, able to react quicker and more efficiently to changes (Gomez Segura *et al*., 2019).

A manufacturer of furniture in Selangor, Malaysia was chosen as a test bed for the implementation of a customised MCM system. Using a mobile application, manufacturing process data was recorded and stored in the AAS. The application also provided a platform for operators to view and assign work orders, serving as a pseudo-Manufacturing Execution System. The data collected was processed to provide performance indices of the manufacturing facility. This is further discussed in section 4.4.

4.0 RESULTS AND DISCUSSION

4.1 Data Standardisation and Representation

Figure 4 shows a representation of a generic juice product. Per the UML standard, each object has its own class name, attributes, and functions. The same asset exists on the MCM AASystem, stored as a JSON object. This can then be translated into a UML object per IEC 62890 standard. Conversion to other

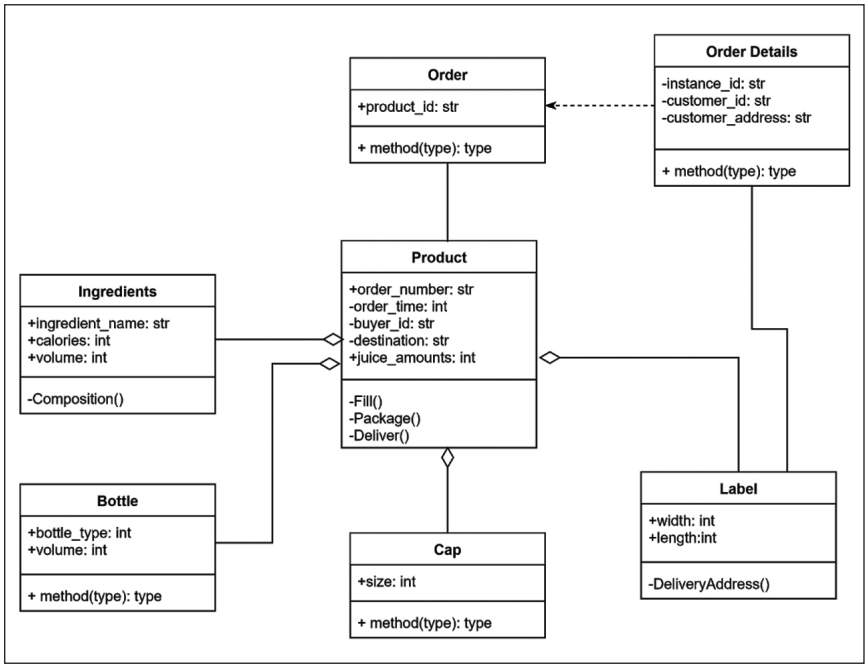


Figure 4: Generic juice product represented as UML object per IEC 62890

standards such as ECLASS and AutomationML can be achieved with similar means. The UML diagram does not only show product data, but also relationships. The connections indicate the relationship between different elements of the data framework, such as the data and functions of components that make up the fruit juice product. In this example, the Ingredients, Bottle, Cap, and Label all have their own UML object, storing unique information about that component. They are also connected to the final Product, indicating an aggregation relationship.

Another example is the relationship between the customer Order Details and the Order as well as the Label. Some but not all the data about the Product is shared with the Order and subsequently the Order Details, whilst the Order Details itself has its own unique data. Some data fields are also shared with the Label object. In this manner, the UML object diagram can display the relationship the product has with the entire vertical and horizontal manufacturing chain in an intuitive manner. This also illustrates the distinction and the link between type and instance. Both objects can be represented with the UML object diagram, and the data stored in this format will also facilitate sharing among shareholders in the integrated manufacturing chain.

The MCM platform also provides visualisation capabilities. Using these capabilities to present relevant information provided by the PLC data framework will allow the user to monitor key metrics and potentially make more informed decisions. The three key functions that the PLC data framework will provide are PLC analytics, Value chain analytics, and Value stream mapping. All three will benefit from a dashboard that summarises the generated information and recommendations.

The processing of data generated by the different transactions in the data framework is also of importance. Data volume, variety, traffic intensity and criticality must be considered whether from a hardware and software perspective. Care must be taken to ensure the stability and consistency of data processing in the manufacturing and supply chain (Raptis *et al*., 2019).

4.2 MCM Product Life Cycle Framework

Figure 5 shows the product life cycle data framework of the MCM platform. The cycle starts from the vertical and horizontal manufacturing supply chain and is concerned with increasing value to the organisation. The data framework is generic and can be used for any manufacturing entity. The three key functions are applicable for any process and manufacturing chain. It will be up to the adopter to decide which key metrics are most useful for their organisation. The generic nature of the framework

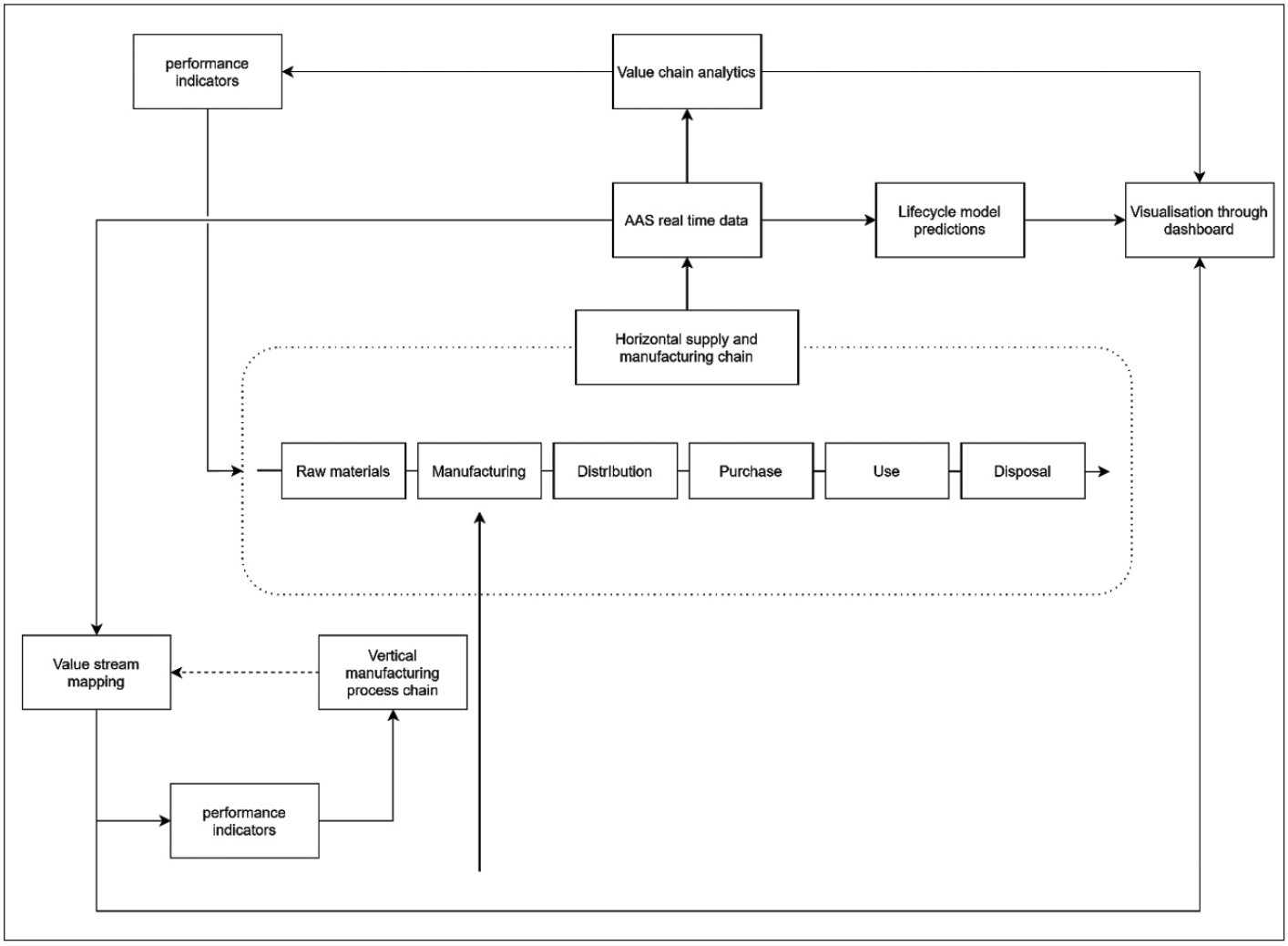


Figure 5: Manufacturing Chain Product Life Cycle Data Framework

also offers customisability and scalability. This is important in attracting organisations who are looking for effective and sustainable solutions for their I4.0 adoption (Mittal *et al*., 2018). The function of the framework is to present information about the organisation to stakeholders. The information is given in three contexts, namely PLC, value chain analytics, and value stream mapping.

An area of research to be explored is the potential for real-time analytics to positively affect the PLC, illustrated in Figure 6(a) and (b). It has already been shown that sufficient modelling can improve demand forecasts. An extension to that is to use feedback from the integrated vertical and horizontal manufacturing chain to extend the PLC. The data to drive this improvement can come from any source within the manufacturing chain. Value chain analytics also shows promise in providing continuous improvements to the horizontal manufacturing chain. Insights provided by the real-time data can be directly applied to the organisation.

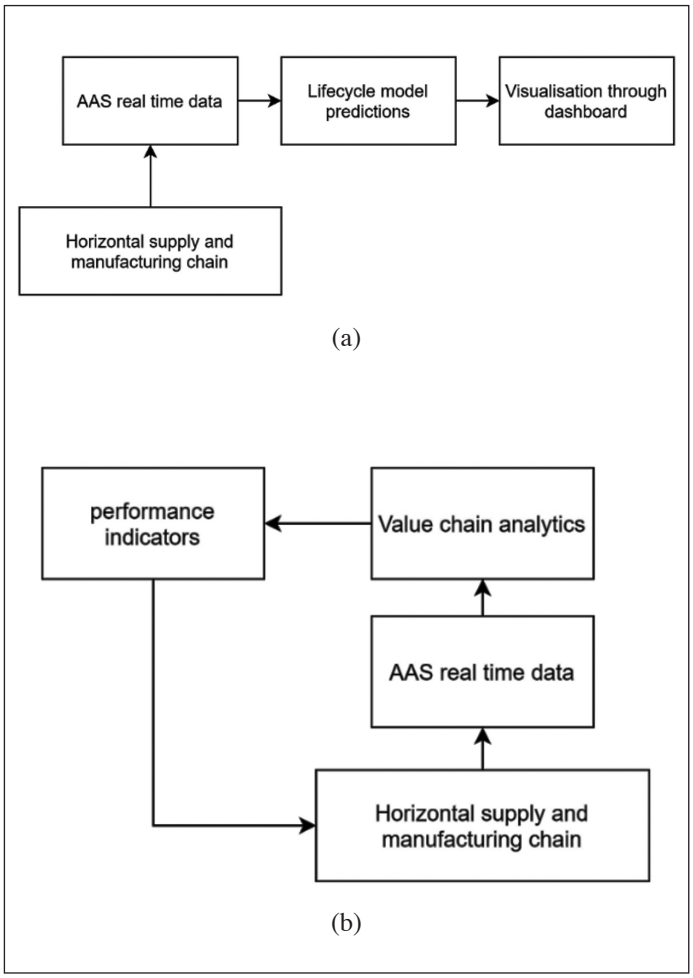


Figure 6: Product life cycle prediction framework and horizontal manufacturing chain feedback cycle

Similarly, value stream mapping can benefit the vertical manufacturing chain. VSM combined with the PLC data framework is particularly powerful, as the data framework will be able to provide a complete and real-time snapshot of the manufacturing processes. Key metrics revealed by VSM can then be used to track manufacturing performance. These actions take the form of feedback cycles, as they continuously improve the performance of the manufacturing supply chain, illustrated in Figure 7.

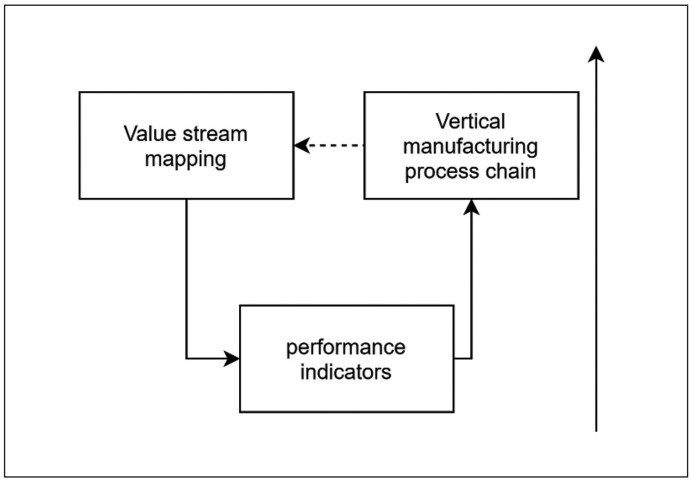


Figure 7: Vertical manufacturing chain feedback cycle

4.3 Value Stream Mapping in the Life Cycle Framework

The vertical manufacturing chain feedback cycle deals with the internal manufacturing cycle of the organisation. By tracking various metrics, the performance of different parts of the manufacturing chain can be enhanced. Indicators such as defect rate (Psarommatis *et al*., 2021), Work-in-Progress volume (Woschank *et al*., 2020) and process lead time (Chowdhury *et al*., 2016) have been shown to be effective improving production performance.

4.4 MCM Implementation and Preliminary Results

A manufacturer of furniture in Selangor, Malaysia was chosen as a test bed for use of VSM in the context of MCM to track their manufacturing process. The manufacturer uses a system of containers and tracks to transport material throughout the shop floor from station to station for processing. The following key metrics were recorded using the MCM platform customised for the manufacturer; process time for individual manufactured parts, loading time to individual containers, material usage, and Work-In-Progress time between processing stations. Using these metrics, the performance of manufacturing line and individual operators was obtained.

Figure 8 shows the planned or expected manufacturing performance of the operators versus the actual recorded manufacturing performance of the operators. The performance is indicated by manufacturing throughput, which is calculated as the components produced divided by the total time taken. Figure 9 shows the recorded working time of each individual operator, allowing the manufacturer to calculate labour costs for each part. For both cases, the increased traceability of processes is useful in keeping track of manufacturing performance and expenditure. The data shown is a portion of the full record and anonymised for confidentiality.

num.	Operator 1	Operator 2	Operator 3	components produced	planned process time	actual process time	planned throughput	actual throughput	performance measure
1	A	B		300	30	32	10	9.38	-6.25%
2	A	B		300	30	31	10	9.68	-3.23%
3	A	B		45	15	16	3	2.81	-6.25%
4	A	B		555	20	21	27.75	26.43	-4.76%
5	A	B		600	15	14	40	42.86	7.14%
6	A	B		300	15	16	20	18.75	-6.25%
7	A	B		600	30	29	20	20.69	3.45%
8	A	B		300	15	14	20	21.43	7.14%
9	A	B		600	40	42	15	14.29	-4.76%
10	A	B		150	20	18	7.5	8.33	11.11%
11	A	B		150	20	21	7.5	7.14	-4.76%
12	A	B		300	20	22	15	13.64	-9.09%
13	A	B		300	1	0.98	300	306.12	2.04%
14	C	D	E	50	1	1	50	50	0.00%
15	C	D	F	50	1	1	50	50	0.00%
16	C	D	F	45	1	1	45	45	0.00%
17	C	E	F	5	1	0.98	5	5.1	2.04%

Figure 8: Planned versus actual operator throughput

Operator 1	Operator 2	Operator 3	normal working hours	overtime working hours	Operator 1 Total wages	Operator 2 Total wages	Operator 3 Total wages	Total cost per order item	order item quantity	labour cost per piece
A	B	E	04:33:00	00:30:00	RM41.41	RM41.41	RM49.69	RM132.50	300	RM0.44
C	D		01:41:00		RM14.47	RM16.22	RM0.00	RM30.69	100	RM0.31
A	B	F	02:27:00		RM19.14	RM19.14	RM19.78	RM58.06	150	RM0.39
C	D		02:27:00		RM21.05	RM23.61	RM0.00	RM44.66	150	RM0.30
C	D		04:57:00		RM42.54	RM47.70	RM0.00	RM90.23	300	RM0.30
C	D		00:45:00		RM6.45	RM7.23	RM0.00	RM13.67	45	RM0.30
C	D		00:37:00	04:17:00	RM60.51	RM67.85	RM0.00	RM128.36	300	RM0.43
C	D		09:35:00	00:19:00	RM86.44	RM96.92	RM0.00	RM183.36	600	RM0.31
B	E	F	05:03:00		RM39.45	RM47.34	RM40.77	RM127.57	300	RM0.43
B	E	F	04:57:00		RM38.67	RM46.41	RM39.96	RM125.04	300	RM0.42
B	E	F	09:30:00	00:36:00	RM81.25	RM97.50	RM83.96	RM262.71	600	RM0.44

Figure 9: Manufacturing process time and operator wages

5.0 CONCLUSION

The aim of this data framework is to enable the sharing of data generated from the vertical and horizontal axis of the RAMI4.0 model. The real-time data will provide a snapshot of the current state of the organisation. This data will be processed through value chain analytics, value stream mapping, and product life cycle modelling to generate insights that will provide value for the organisation and customer. The visualisation and dashboard tools of the MCM platform will be used to present these insights.

It has been shown that the three presented methods are individually capable of tracking metrics to provide improvement to the company. However, an integrated platform that targets these three key axes of the RAMI4.0 model has not been proven. The MCM platform with the data framework is the first proposed solution that aims to fulfil this need.

This integrated platform will be of great value to manufacturing SMEs. They will be to obtain a clearer picture of their organisation, enabling better informed decisions. The platform will connect them to other stakeholders in the I4.0 manufacturing supply chain, providing even more opportunities and competitive advantages.

Work to validate these findings are underway with the implementation of value stream mapping and the Manufacturing Chain Management Platform being used to identify and track key performance indicators at a furniture manufacturing plant located in Selangor, Malaysia. Preliminary results are promising for the suitability of customised I4.0 solutions for manufacturing SMEs to measure the performance of their facilities. ■

REFERENCES

[1] Chowdhury, A.H., Shahriar, S., Hossen, T., Mahmud, P., 2016. Reduction of Process Lead Time Using Lean Tool - Value Stream Mapping (VSM). Appl. Mech. Mater. 860, 74–80. https://doi.org/10.4028/www.scientific.net/AMM.860.74

[2] Cohen, M.A., Whang, S., 1997. Competing in Product and Service: A Product Life-Cycle Model. Manag. Sci. 43, 535–545. https://doi.org/10.1287/mnsc.43.4.535

[3] DigitalEU, 2020. Strategy for Data | Shaping Europe's digital future [WWW Document]. URL https://digital-strategy.ec.europa.eu/en/policies/strategy-data (accessed 8.10.21).

[4] Faulkner, W., Badurdeen, F., 2014. Sustainable Value Stream Mapping (Sus-VSM): methodology to visualize and assess manufacturing sustainability performance. J. Clean. Prod. 85, 8–18. https://doi.org/10.1016/j.jclepro.2014.05.042

[5] Fearne, A., Garcia Martinez, M., Dent, B., 2012. Dimensions of sustainable value chains: implications for value chain analysis. Supply Chain Manag. Int. J. 17, 575–581. https://doi.org/10.1108/13598541211269193

[6] Gomez Segura, M., Oleghe, O., Salonitis, K., 2019. Analysis of lean manufacturing strategy using system dynamics modelling of a business model. Int. J. Lean Six Sigma ahead-of-print. https://doi.org/10.1108/IJLSS-05-2017-0042

[7] Gurumurthy, A., Kodali, R., 2011. Design of lean manufacturing systems using value stream mapping with simulation: A case study. J. Manuf. Technol. Manag. 22, 444–473. https://doi.org/10.1108/17410381111126409

[8] Hu, K., Acimovic, J., Erize, F., Thomas, D.J., Van Mieghem, J.A., 2019. Forecasting New Product Life Cycle Curves: Practical Approach and Empirical Analysis. Manuf. Serv. Oper. Manag. 21, 66–85. https://doi.org/10.1287/msom.2017.0691

[9] Koc, T., Bozdag, E., 2017. Measuring the degree of novelty of innovation based on Porter's value chain approach. Eur. J. Oper. Res. 257, 559–567. https://doi.org/10.1016/j.ejor.2016.07.049

[10] Kumar, M., Tsolakis, N., Agarwal, A., Srai, J.S., 2020. Developing distributed manufacturing strategies from the perspective of a product-process matrix. Int. J. Prod. Econ. 219, 1–17. https://doi.org/10.1016/j.ijpe.2019.05.005

[11] Masood, T., Sonntag, P., 2020. Industry 4.0: Adoption challenges and benefits for SMEs. Comput. Ind. 121, 103261. https://doi.org/10.1016/j.compind.2020.103261

[12] Mittal, S., Khan, M.A., Romero, D., Wuest, T., 2018. Mittal et al. 2018 - A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). J. Manuf. Syst. 49, 194–214. https://doi.org/10.1016/j.jmsy.2018.10.005

[13] Osterrieder, P., Budde, L., Friedli, T., 2020. The smart factory as a key construct of industry 4.0: A systematic literature review. Int. J. Prod. Econ. 221, 107476. https://doi.org/10.1016/j.ijpe.2019.08.011

[14] Pethig, F., Niggemann, O., Walter, A., 2017. Towards Industrie 4.0 compliant configuration of condition monitoring services, in: 2017 IEEE 15th International Conference on Industrial Informatics (INDIN). Presented at the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), IEEE, Emden, pp. 271–276. https://doi.org/10.1109/INDIN.2017.8104783

[15] Plattform Industrie 4.0, 2019. Details of the Asset Administration Shell Part 1. Federal Ministry for Economic Affairs and Energy.

[16] Psarommatis, F., Sousa, J., Mendonça, J.P., Kiritsis, D., 2021. Zero-defect manufacturing the approach for higher manufacturing sustainability in the era of industry 4.0: a position paper. Int. J. Prod. Res. 1–19. https://doi.org/10.1080/00207543.2021.1987551

[17] Raptis, T.P., Passarella, A., Conti, M., 2019. Data Management in Industry 4.0: State of the Art and Open Challenges. IEEE Access 7, 97052–97093. https://doi.org/10.1109/ACCESS.2019.2929296

[18] SAP, 2020. SAP Estimator Tool. URL https://www.sap.com/sea/products/cloud-platform/pricing/estimator-tool.html (accessed 5.15.20).

[19] SME Corp, 2018. SME Annual Report 2018.

[20] VDI, ZVEI, 2015. Status Report - RAMI4.0.

[21] Woschank, M., Dallasega, P., Kapeller, J.A., 2020. The Impact of Planning Granularity on Production Planning and Control Strategies in MTO: A Discrete Event Simulation Study. Procedia Manuf. 51, 1502–1507. https://doi.org/10.1016/j.promfg.2020.10.209

PROFILES

CHEW KHAI MIN has a BEng in Mechanical Engineering and a MSc in Mechanical Engineering Design from Manchester University, and has been lecturing at Tunku Abdul Rahman University College for the past 7 years. As well as teaching computer-aided design, his research interests are in mechanical design and manufacturing engineering. His experience drives him to explore the implementation of Industry 4.0 in the Malaysian manufacturing sector by helping companies collect, analyse, and use their data to provide better products and better services.

ASSOC. PROF. TS. DR LEE WAH PHENG is the Associate Professor in Tunku Abdul Rahman University College. He worked in the manufacturing industry for 10 years and more than 20 years of business and education industry experiences. Dr Lee is a pioneer and consultant in Industry 4.0. He works with a team of researchers and industry partners to develop a holistic digital solution suitable for the small and medium enterprises.

TS. DR LEE YOON KET is a principal lecturer of the Mechanical Engineering Department in the Faculty of Engineering and Technology, Tunku Abdul Rahman University College (TAR UC). He completed his PhD at Queen University of Belfast, U.K in 2004. His research interests mainly focus on application of Industry 4.0 to SMI, Autonomous system and Manufacturing planning and control.

DEEP LEARNING & HYBRID MODEL – THE FUTURE OF MEDICAL IMAGE WATERMARKING?

(Date received: 13.01.2022/Date accepted: 22.04.2022)

Chi Wee Tan [1], Yew Lee Wong [1*], Jia Cheng Loh [1], Chen Zhen Li [1]

[1] Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia

**Corresponding author: chiwee@tarc.edu.my*

ABSTRACT

The frequent usage of medical records in electronic form has made Medical Image Watermarking (MIW) relatively more significant than it used to be. MIW is very significant to preserve the completeness and integrity of the medical images. For the time being, with the trade-offs between visibility and robustness, there are no perfect algorithms for invisible watermarking. In many novels, Deep-Learning-Based Approach has been proposed to solve the trade-offs. In this study, multiple implementations of invisible watermarking techniques such as Deep-Learning-Based Approach and Non-Deep-Learning-Based-Approach are being compared. This comparative study measures the limitations and robustness on a dataset of breast ultrasound images. Eighteen extreme attacking methods were carried out on the encoded images, performance was then evaluated using peak signal-to-noise ratio (PSNR) and normalized cross correlation (NCC). Encoded images were then tested against a digital transmission channel to test its robustness. To conclude, The Deep-Learning-Based-Approach of RivaGAN showed the best robustness against multiple extreme attacks. The Non-Deep-Learning-Based-Approach of discrete wavelet transform – discrete cosine transform – singular value decomposition (DWT-DCT-SVD) has the best imperceptibility. Therefore, we confirm the feasibility of Deep-Learning-Based-Approach in Medical Image Watermarking, however more work is needed to be done to achieve perfect Deep-Learning-Based-Approach in terms of imperceptibility.

Keywords: Invisible Watermarking, DCT, DWT, SVD, RivaGAN, Deep-Learning-Based Invisible Watermarking

1.0 INTRODUCTION

Medical image watermarking has been significant in this digital era. Watermarking on medical images which is equivalent to digital signatures, is required not to compromise the quality of image. Digital image watermarking can be understood as a process of embedding and extracting signatures such as names into images that are to be distributed through digital transmission. Conventional visible image watermarking is not applicable in the use case of medical images as integrity and completeness of the photo is the utmost priority.

Invisible watermarking can be traced back to the work by Yeung *et al* which proposed a method of image verification (Yeung & Mintzer, 1997). Watermarking can be done on either spatial domain or the transform domain.

In the scenario of medical image watermarking, digital medical images transmitted over any channel may raise data integrity problems, therefore, invisible watermarking could be the solution. However, there is no perfect algorithms or solutions for invisible watermarking as trade-offs can happen between visibility and robustness when doing watermarking (Mousavi *et al*., 2014). For a watermarking technique to reach the optimum state in the use case of medical image watermarking, the techniques shall take into account robustness, imperceptibility and security. Robustness can be simply understood as the resilience of the watermarking towards any attacks while imperceptibility focuses on the quality of watermarked image after the embedment process.

With the need for a perfect algorithm that can satisfy the need of robustness and imperceptibility, research has been incorporating deep-learning-based techniques into the field of medical image watermarking. Embedded watermarks can be extracted using convolutional neural networks. However, the robustness challenge has always been hard to satisfy due to the fragility of the deep neural networks (Papemot *et. al*., 2016).

In our study, we explored the implementations of deep-learning-based invisible watermarking techniques with hybrid-based techniques. Hybrid invisible watermarking techniques has proven good performances in the past. Through our study, we hope:

- To identify and verify the robustness of deep-learning-based invisible watermarking algorithm
- To measure the watermarking effects on the medical images using the metrics of PSNR and NCC of deep-learning-based techniques and non-deep-learning-based-techniques.
- To investigate the limitations and resistance of the algorithms towards extreme attacks.
- To verify the completeness of embedded messages after digital transmission.

2.0 LITERATURE REVIEW

2.1 Digital Image Watermarking & Invisible Watermarking

Due to the rapid expansion of the internet, the distribution of digital photographs has become increasingly popular; as a result, data protection has become increasingly crucial. - (Abdulrahman, 2019). In the context of digital watermarking, the process of embedding or hiding data in another digital data, and then extracting the hidden information can be defined as invisible watermarking. (Tao *et al*., 2014) According to others, it has grown easier to tamper with medical photographs since modern picture editing software has become more widely available in the past few years (Coatrieux, 2006). In order to address these problems, invisible watermarking can be used for data concealing as well as to safeguard the integrity of data (Coatrieux, 2006). It is possible to divide the digital watermarking domain into two subdomains: the spatial domain and the frequency domain, respectively (EL-Shazly, 2004). Robustness and imperceptibility are two performance criteria that are commonly used to evaluate picture watermarking techniques; nevertheless, these two characteristics are diametrically opposed to one another (Usman *et al*., 2008). When measuring the imperceptibility of the watermark, peak signal to noise ratios (PSNR) are utilised. The image quality should not be distorted when there is a watermark present, as measured by the peak signal to noise ratios (PSNR) (Al-Haj, 2007). PSNR is commonly expressed in decibels (dB), and it is widely used in medical image watermarking (MIW) algorithms to compare their performance (Faragallah *et al*., 2021). A technique's robustness is measured by the watermark's resilience and immunity to removal attempts as well as degradation attempts (Voloshynovskiy *et al*., 2001).

2.2 Medical Image Watermarking

Telemedicine has grown in popularity over the last few decades as communication technology has advanced. Diagnostic procedures rely heavily on medical images. They can now be transmitted easily across the globe via communication channels (Pandey and Singh, 2016). However, transmission over public networks puts security, confidentiality, copyright, and integrity at risk. Medical data theft or tampering can result in incorrect diagnoses. Thus, during the transmission of medical images, security, confidentiality, and integrity are paramount concerns. In this situation, medical image watermarking (MIW) has emerged as a viable option (Hussain and Wageeh, 2013). Significant information is concealed within a cover medical image during the watermarking process, and that information should not be detected, retrieved, or modified by an unauthorized user. It is frequently used in one-to-many communication systems, whereas steganography is typically used in one-to-one communication systems (Sharma and Gupta, 2012). Watermarking medical images are classified as a reversible technique or an ROI (Region Of Interest) technique (Sonika and Inamdar, 2012). A robust and reversible watermark is required for diagnostic purposes in a health information system. The reversible watermarking technique maintains the integrity of the original medical image during recovery. If the extracted medical image is corrupted in any way, the result will be incorrect (Rohini and Bairagi, 2010).

2.3 Deep-Learning-Based Image Watermarking

Deep-Learning-Based Image Watermarking has been proved superior against other algorithsm in term of concealment and robustness. (Zhang *et. al*, 2021). Convolutional neural networks (CNN), autoencoders (AE), and generative adversarial networks (GAN), all of which are common in deep learning, have been the mainstays of research. This new deep learning architecture, known as RivaGAN, goes beyond the usual convolutional layers and algorithms. The encoder's robustness was tested and improved using two independent adversarial networks. A 32-bit watermark is embedded into a sequence of frames using this design. Any common video processing operations like cropping, scaling, and compression were shown to be robust to RivaGAN. (Zhang *et al*., 2019).

2.4 Hybrid Watermarking

The efficiency of the watermarking technique can be increased by the combination of different transformations (Assini *et al*., 2018). The hybrid watermarking of DWT-DCT-SVD was proven to be very robust because it does not embed all singular values and can be applied to create algorithms for loss image compression (Navas *et al*., 2008). The performance of the DWT-DCT hybrid watermarking was shown to be superior to the performance of the DWT method alone (Al-Haj, 2007). In comparison to DWT, the DWT-DCT significantly improved robustness especially to the linear and non-linear attacks (Abdulrahman & Ozturk, 2019). Additionally, it has been proved that the non-hybrid watermarking approach of DWT is resistant to any typical image processing processes (Lala, 2017).

3.0 RESEARCH METHODOLOGY

3.1 Dataset and Algorithms

The dataset used is a collection of breast ultrasound images among women between the ages of 25 and 75 years old which is available at Kaggle (Al-Dhabyani W *et al*., 2020). A total of 20 images were selected randomly from these 780 images with an average image size of 500×500 pixels. The chosen images were named alphabetically from “MRI_A” to “MRI_T”. Four algorithms of invisible watermarking were chosen, namely DWT, DWT-DCT, DWT-DCT-SVD & RivaGAN.

3.2 General Framework

As illustrated in Figure 1, it shows the overall flow of our study. Firstly, we will encode a watermark in string format into the original MRI images. Then we will attack those encoded images using 18 different methods. Transmission of encoded images were also done on the attacking phase. After that we will try to decode the watermark from the attacked images and calculate the PSNR and NCC value. Lastly, the result will be visualize using some chart.

3.3 Testing Criteria

On the pass rate, messages retrieved after being attacked is strictly being compared absolutely. Only if the output matches

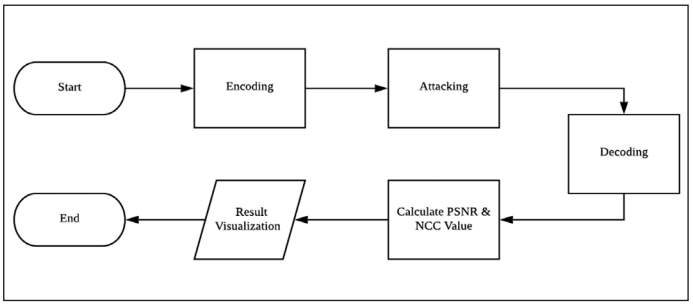


Figure 1: Medical Image Watermarking Framework

100% with the initial input can be considered as passing the test. Decoding errors were counted as failure through exceptions caught by decoder. Partial success that the output matches the input was considered as failure. On the test of transmission, watermarked photos were transmitted through WhatsApp Image, WhatsApp Document, Google Drive, Facebook Messenger and Gmail. The images received on the receiving end were put into decoder to retrieve the embedded messages. Output that matches the initial input 100% will only be considered as pass the test. On testing the implementation of the selected library, average of decoding and encoding time were done using time library in Python. Elapsed time was recorded down over 1000 iterations of the operation and mean were calculated. On measuring the relationship between characters length of embedded message and file size, different randomly generated string of different length was encoded. File size was compared on before and after encoding. On measuring the performance of each watermarking algorithm, we evaluate the image of before using the value of Peak Signal-to-Noise Ratio (PSNR) and Normalized Cross Correlation (NCC).

3.4 Experiment Environment

The testing of the implementation was done on a desktop system of such specifications in Table 1.

Table 1: Testing system

CPU	Intel Xeon E5-2650v2 @ 2.60Ghz, 8 Cores 16 Threads
RAM	16GB DDR3 1666Mhz
Operating System	Windows 10 Pro 64-bit (10.0, Build 19043
Python Version	3.8.10

4.0 RESULTS AND DISCUSSIONS

As shown in Table 2, there are 18 types of attacking methods that will be used to test the robustness of each watermarking algorithm.

As illustrated in Figure 2 & 3, RivaGAN has the highest passing rate among all the algorithms follow by DWT-DCT-SVD ranked at the second place. However, DWT-DCT has the worst performance with lowest passing cases.

The higher PSNR the better the quality of the compressed, or reconstructed image. Based on Table 3, DWT-DCT-SVD algorithm has the highest PSNR value with an average 46.83 dB that determine its criteria as best algorithm among all the algorithms.

Table 2: Attacking methods

	Kernel Settings/ Ratio / Strength
Averaging	size = 5x5
Bilateral Filtering	D = 9, sigmaColor = 75, sigmaSpace = 75
Brightness Decrease	40 %
Brightness Increase	40 %
Crop Horizontal	50 %
Crop Vertical	50 %
Gaussian Blurring	Size = 5x5
Gaussian Noise	mean=0, variance=0.01
JPG	Convert to JPG
Masks	n = 5, ratio = 0.3
Median Blurring	Size = 7
Poisson Noise	Lambda = 20
Rotate	10 degrees
Salt & Pepper	10 %
Scale Down	25 %
Scale Up	25 %
Sharpen Filtering	[-1, -1, -1], [-1, 9, -1], [-1, -1, -1]
Speckle Noise	mean=0, variance=0.01

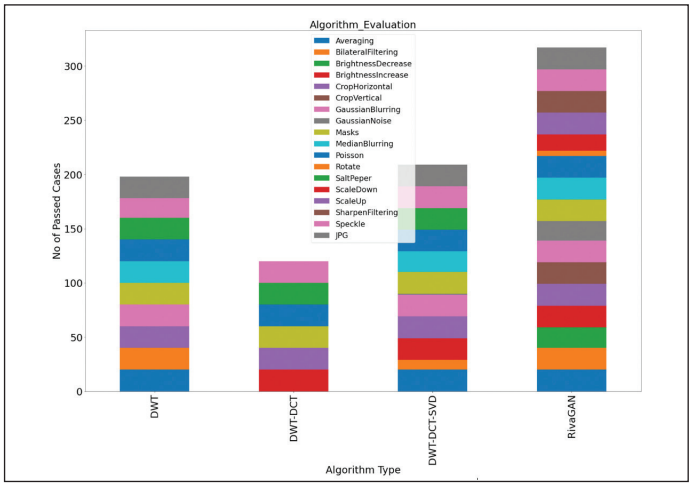


Figure 2: Algorithm Evaluation Based on Pass Rate

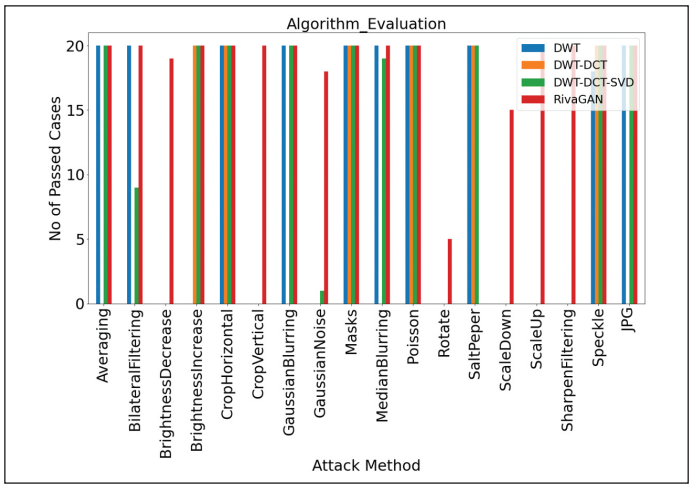


Figure 3: Attack Evaluation Based on Pass Rate

Table 3: PSNR between Original Image and Encoded Image

	DWT	DWT-DCT	DWT-DCT-SVD	RivaGAN
MRI_A	35.24	43.74	46.98	40.41
MRI_B	35.26	43.67	46.94	40.39
MRI_C	35.24	43.74	47.00	40.41
MRI_D	35.25	43.63	46.92	40.42
MRI_E	35.25	43.61	46.92	40.42
MRI_F	35.48	42.82	44.77	40.49
MRI_G	35.18	43.59	46.90	40.41
MRI_H	35.27	43.67	46.94	40.43
MRI_I	35.19	43.66	46.98	40.43
MRI_J	35.19	43.60	46.92	40.45
MRI_K	35.19	43.57	46.89	40.44
MRI_L	35.21	43.67	46.99	40.44
MRI_M	35.22	43.62	46.94	40.44
MRI_N	35.23	43.62	46.94	40.44
MRI_O	35.16	43.62	46.95	40.44
MRI_P	35.21	43.61	46.92	40.49
MRI_Q	35.36	43.74	46.96	40.45
MRI_R	35.28	43.65	46.95	40.42
MRI_S	35.23	43.60	46.93	40.41
MRI_T	35.20	43.59	46.90	40.43

Table 4: NCC between Original Image and Encoded Image

	DWT	DWT-DCT	DWT-DCT-SVD	RivaGAN
MRI_A	0.9973	0.9992	0.9998	0.9992
MRI_B	0.9969	0.9990	0.9998	0.9990
MRI_C	0.9974	0.9992	0.9998	0.9992
MRI_D	0.9973	0.9992	0.9998	0.9992
MRI_E	0.9963	0.9989	0.9997	0.9986
MRI_F	0.9962	0.9983	0.9992	0.9988
MRI_G	0.9965	0.9989	0.9997	0.9989
MRI_H	0.9976	0.9993	0.9998	0.9993
MRI_I	0.9967	0.9990	0.9998	0.9993
MRI_J	0.9975	0.9993	0.9998	0.9992
MRI_K	0.9978	0.9993	0.9998	0.9993
MRI_L	0.9978	0.9993	0.9998	0.9993
MRI_M	0.9978	0.9993	0.9998	0.9993
MRI_N	0.9996	0.9990	0.9998	0.9990
MRI_O	0.9956	0.9987	0.9969	0.9987
MRI_P	0.9973	0.9992	0.9998	0.9992
MRI_Q	0.9968	0.9990	0.9998	0.9990
MRI_R	0.9969	0.9991	0.9998	0.9990
MRI_S	0.9962	0.9988	0.9997	0.9988
MRI_T	0.9972	0.9991	0.9998	0.9991

The higher NCC value the better the degree of similarity between two compared images. Based on Table 4, all the algorithms have similar performance on NCC value. However, DWT-DCT-SVD is the best performance with highest NCC value among all the algorithms.

As illustrated in Table 5, all algorithms managed to achieve full passes for every transmission method. However, DWT-DCT algorithm failed to achieve full passes as it failed to transmit across WhatsApp Image. It can be believed that this failure was caused by the compression of WhatsApp.

Table 5: Encoded Algorithm vs Transmission Platform

	DWT	DWT-DCT	DWT-DCT-SVD	RivaGAN
WhatsApp Image	✓	×	✓	✓
WhatsApp Document	✓	✓	✓	✓
Google Drive	✓	✓	✓	✓
Facebook Messenger	✓	✓	✓	✓
Gmail	✓	✓	✓	✓

Based on the Table 6, RivaGAN has the restrictions of 4 characters length, while other algorithms have no character length limit. Besides, every algorithm implementation exhibits perfect behaviours toward case sensitive and special characters. On the Chinese characters, all the algorithms fail to encode and decode.

Table 6: Algorithm Implementation Benchmarking

	DWT	DWT-DCT	DWT-DCT-SVD	RivaGAN
Character Length Limit	×	×	×	✓ (4)
Case Sensitive	✓	✓	✓	✓
Special Characters	✓	✓	✓	✓
Chinese Characters	×	×	×	×

When comparing the length of character embedded into the image using Algorithm A (DWT), Algorithm B (DWT-DCT) and Algorithm C (DWT-DCT-SVD), as shown in Figure 4, the file size exhibited a big fluctuation for the first few 20 bytes. It can be observed that DWT went below the original file size when encoded with messages.

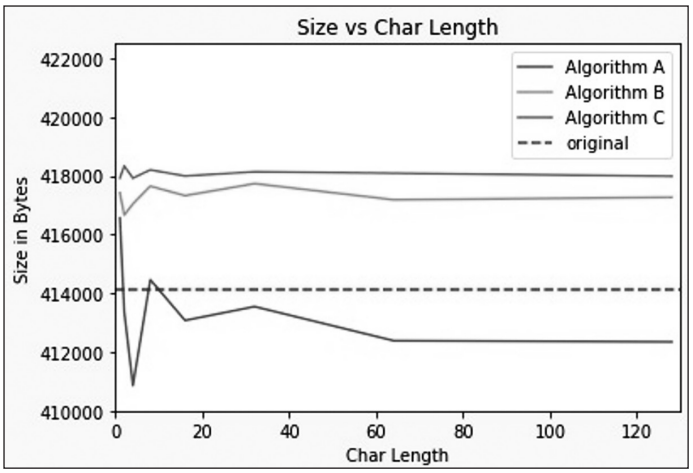


Figure 4: File Size vs Embedded Characters Length

5.0 LIMITATIONS

As for the limitation of our study, RivaGAN was initially developed for video invisible watermarking. Then the implementation was redeveloped and ported to image watermarking. Therefore, the RivaGAN library we are using has limitations on the number of characters of 32bit allowed in the embedding process.

Other than that, the hybrid methods' implementations we are currently using are the publicly available open-source library from GitHub, as such, the implementation method might have a disparity with the original algorithm or research. Future studies can explore how the algorithm is implemented according to the formula to ensure consistency of the outcome.

6.0 CONCLUSIONS

Through our study, the deep-learning-based method of RivaGAN does exhibit the state-of-the-art robustness as claimed in the paper of RivaGAN authors (Zhang *et al*., 2019). Through our tests, we can confirm the feasibility of deep-learning-based invisible watermarking techniques as claimed by Vukotic (Vukotic *et al*., 2020).

Many extreme attacks were performed on RivaGAN's watermarked images and it was still able to pass all tests as it exhibited strong robustness as compared to other non-deep-learning-based watermarking techniques.

To satisfy the requirements of medical image watermarking, the algorithm shall have the high value of PSNR, NCC and exhibit a strong robustness. It was found that RivaGAN fails to surpass the PSNR and NCC value of DWT-DCT-SVD. This can be attributed to the nature of RivaGAN which is created specifically for video invisible watermarking.

The non-deep-learning-based hybrid algorithm of DWT-DCT-SVD showed the best criteria of Imperceptibility as it topped the PSNR value of 47.00 on comparing original image and encoded image. DWT-DCT-SVD also showed the best NCC value among the algorithms.

With the feasibility of deep-learning-based invisible watermarking methods being confirmed through its strong robustness, more research is needed to refine the algorithm in terms of reaching higher PSNR and NCC as compared to the non-deep-learning-based methods.

7.0 ACKNOWLEDGEMENTS

The authors would like to thank Tunku Abdul Rahman University College (TAR UC) for providing financial support and technical support when completing this study. ■

REFERENCES

[1] Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020 Feb; 28:104863. DOI: 10.1016/j.dib.2019.104863.

[2] Al-Haj, A. (2007). Combined DWT-DCT Digital Image Watermarking. Journal of Computer Science, 3(9), 740–746.

[3] Abdulrahman, A. K., & Ozturk, S. (2019). A novel hybrid DCT and DWT based robust watermarking algorithm for color images. Multimedia Tools and Applications, 78(12), 17027–17049.

[4] El-Shazly, E. H. M. (2004). Digital Image Watermarking in Transform Domains. Minufiya University.

[5] Khare, P., & Srivastava, V. K. (2020). A Secured and Robust Medical Image Watermarking Approach for Protecting Integrity of Medical Images. Transactions on Emerging Telecommunications Technologies, 32(2).

[6] Kuang, L.-Q., Zhang, Y. and Han, X. (2009). A Medical Image Authentication System Based on Reversible Digital Watermarking. 2009 First International Conference on Information Science and Engineering.

[7] Lala, H. (2017). Digital image watermarking using discrete wavelet transform. International Research Journal of Engineering and Technology (IRJET), 4(01).

[8] Sverdlov, A., Dexter, S., & Eskicioglu, A. M. (2005). Robust DCT-SVD domain image watermarking for copyright protection: embedding data in all frequencies. In 2005 13th European Signal Processing Conference (pp. 1-4). IEEE.

[9] Tao H, Chongmin L, Zain JM, Abdalla AN (2014) Robust image watermarking theories and techniques: a review. J Appl Res Technol 12(1):122–138

[10] Voloshynovskiy, S., S. Pereira and T. Pun, 2001. "Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks," Comm. Magazine, 39(8): 118-126

[11] Vukotić, V., Chappelier, V., & Furon, T. (2020). Are Classification Deep Neural Networks Good for Blind Image Watermarking? Entropy, 22(2), 198.

[12] Yeung, M. M. (1998). Invisible watermarking for image verification. Journal of Electronic Imaging, 7(3), 578.

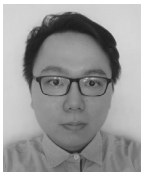
[13] Zhang, K. A., Xu, L., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Robust invisible video watermarking with attention. arXiv preprint arXiv:1909.01285.

[14] Zhang, L., Li, W., Ye, H. (2021). A blind watermarking system based on deep learning model. 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).

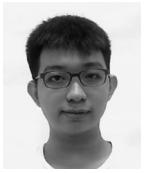
PROFILES

DR TAN CHI WEE received BCompSc(Hons) and PhD degrees in year 2013 and 2019 respectively in Universiti Teknologi Malaysia. Currently, he is a Senior Lecturer cum Programme Leader at Tunku Abdul Rahman University College and actively involved in the Centre of Excellence for Big Data and Artificial Intelligent (CoE) and become the research group leader for Audio, Image and Video Analytics Group under Centre for Data Science and Analytics (CDSA). Dr Tan's main research areas are Computer Vision (CV), Image Processing (IP) and Natural Language Processing (NLP) and Artificial Intelligence (AI). He is an enthusiastic researcher experienced in conducting and supporting research into Image Processing. Being a meticulous and analytical researcher with Train-The-Trainer certificate of many years of educational and hands-on experience, he was invited to Université d'Artois (France) under Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) programme for collaborative research between European countries with Southeast Asian countries on motion detection and computer vision and being involved in industry project as professional consultant.
Email address: chiwee@tarc.edu.my

WONG YEW LEE is a graduate of Bachelor of Computer Science (Honours) in Data Science. He received his bachelor's degree from Tunku Abdul Rahman University College in 2022. As Yew Lee believes that the advancement of technology and innovation is a need, he is looking to expand his research in Computer Vision, Natural Language Processing and Artificial Intelligence.
Email address: wongyewlee-wm19@student.tarc.edu.my



LOH JIA CHENG graduated from Tunku Abdul Rahman University College (TAR UC) in 2022 with Bachelor of Computer Science (Honours) in Data Science. His research interests include Natural Language Processing, Data Analysis & Visualisation and Artificial Intelligence.
Email address: lohjc-am17@student.tarc.edu.my



LI CHEN ZHEN graduated from Tunku Abdul Rahman University College (TAR UC) in 2022 with a Bachelor of Computer Science with Honours in Data Science. His research interests include computer vision, natural language processing, and machine learning to extract value from data. He believes technology interaction is a significant part of our daily lives with the advancements in technology. He hopes to learn more about various cybersecurity domains, as he works as a cybersecurity engineer and leads development teams supporting various projects.
Email address: licz-wm19@student.tarc.edu.my

CONTENT-BASED IMAGE RETRIEVAL FOR PAINTING STYLE WITH CONVOLUTIONAL NEURAL NETWORK

(Date received: 19.01.2022/Date accepted: 07.04.2022)

W S Tan[1], W Y Chin[2*], K Y Lim[3]

[1,2] Department of Mathematical and Data Science, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia

[3] Department of Computing and Information Technology, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia

**Corresponding author: chinwy@tarc.edu.my*

ABSTRACT

With the advancement of digital paintings in online collection platform, new image processing algorithms are required to manage digital paintings saved on database. Image retrieval has been one of the most difficult disciplines in digital image processing because it requires scanning a large database for images that are comparable to the query image. It is commonly known that retrieval performance is largely influenced by feature representations and similarity measures. Deep Learning has recently advanced significantly, and deep features based on deep learning have been widely used because it has been demonstrated that the features have great generalisation. In this paper, a convolutional neural network (CNN) is utilised to extract deep and high-level features from the paintings. Next, the features were used for similarity measure between the query image and database images; subsequently, similar images are ranked by the distance between both pair features. Our experiments show that this strategy significantly improves the performance of content-based image retrieval for the style retrieval task of painting. Besides, the extracted feature to retrieve the right classes from the query image has achieved over 61% accuracy which beat the current-state-of-art results. However, the result can be further improved in future research by leveraging CNN representations visualisation approaches for a better understanding of how CNN extract features from paintings.

Keywords: Content-based Image Retrieval, Deep Learning, Convolutional Neural Network

1.0 INTRODUCTION

With the continuous expanding due to advancement in digital imaging and internet usage, online artwork collection such as WikiArt, Artsper and Mutual Art have been one of the fastest growing databases. As a result, existing algorithms are incapable of managing these large databases, necessitating the use of robust and quick approaches. Among the several domains of image processing, image retrieval has been always one of the popular approaches in recent years. Image retrieval, which involves scanning a large database for photos that are similar to the query image, was first developed in 1970 by text-based image retrieval (TBIR), in which the system accepts a query word from the user and searches for images that include the text (Rui *et al.,* 1999). However, the concept of an image is much more complex than a few words, and it often turns out not to be so effective. This is due to the subjectivity of the task compared to the meaning of its semantic content. Therefore, content-based image retrieval (CBIR) was invented in 1990. The CBIR has been applied in numerous disciplines, including medical imaging (Campbell, 1994), video processing (Karimi and Bashiri, 2011), crime prevention and other areas that need image recognition (Hwang and Lee, 2012 and Jabalemali *et al*., 2012).

Feature extraction is a critical operation in signal processing, image, video, and speech processing (Zade *et al*., 2014 and Pasandideh *et al*., 2016). It is also one of the critical components of any image retrieval system. The features of an image can be described in two different categories: At the digital level, low-level features mainly are colour-based, texture, and shape features. At the semantic level, the image can be interpreted as having at least one meaning. Unfortunately, paintings are defined digitally in today's information system, while users are more interested in their semantic concept, rather than visually similar. The semantic gap between low-level features and human concept is huge, and it is currently difficult to identify correspondences between the digital painting and semantic levels. Although it may be able to extract increasingly complicated low-level features from images, the size of the feature vector will grow, and the retrieval speed will slow as the calculation time increases. As a result, it is necessary to extract appropriate abstracted features in order

to maximize retrieval precision while minimizing retrieval time. Thus, deep learning is one of the ways that has been shown to reduce the semantic gap between low-level features and human perception (Zade *et al*., 2016) and achieve a good efficiency of image retrieval.

It is commonly known that CBIR is a system that retrieves images from an image database using visual contents. Because it can successfully address the challenges written above, this system has now become vital for image retrieval. In CBIR, visual contents are the features extracted from digital images, and its performance is strongly influenced by the features extracted and similarity measures. Due to these reasons, CNN as a successful subfield of deep learning was used to extract deep and appropriate features for CBIR to process for image retrieval in order to improve the performance of CBIR. In addition, we should not overlook the reality where the research for CBIR has been thriving and particularly strong over the past decades such as CBIR with handcrafted features (Hiremath and Pujari, 2007; Alhassan and Alfaki, 2017 and He *at al*., 2018). However, the amount of attention obtained in the search for paintings in CNN image retrieval is minimal because there is no specific mechanism for visual art interpretation. One of the reasons could be that the visual likeness of paintings can be highly variable, with broad criteria in judging the similarity ranging from a little object, texture, brushstroke, to the entire configuration of the painting itself (Seguin, 2009). To be more explicit, developing a general content-based image retrieval system is easier than developing a domain-specific application, which necessitates domain knowledge. In short, developing a specific domain image retrieval application is difficult yet rewarding research.

In this paper, the work was motivated by the advancement and the efficiency of the features extraction in CNN. Handcrafted features methods such as SIFT (Lowe,1999), SURF (bay *et al*., 2006) and GIST (Oliva and Torralba, 2016) were popular in CBIR, however, we wish to understand if we can profit more fully and flawlessly from deep CNN to increase the efficiency for features extraction in CBIR process. Moreover, creative artwork, such as fine art painting, has attracted much attention from various researchers to seek potential applications. Undoubtedly, several researchers have published numerous publications regarding paintings' characteristic recognition and retrieval task. For instance, Cetinic *et al*., (2018) introduced an approach that are similar to Tan *et al*. (2016) for addressing the fine art classification with fine-tuning CNN, where the model can classify painting's characteristic and also explored on the applicability of the model for retrieving similar paintings based on the query image in either style or content. In the following two years, Cetinic *et al*. (2020) presented another work which used CNN for learning features that are relevant for understanding properties of artistic styles described by Heinrich Wolfflin. Their evaluations suggested that the models learn to discriminate meaningful features that correspond to the visual characteristic of the artistic concepts. Two of these papers indicate that CNN could perform very well and able to measure the artistic style or content in paintings with proper settings. Gontheir *et al*. (2021) recently did a similar experiment. The authors employed approaches to show network internal representations in order to offer information about what a network learns from aesthetic imagery. They also shown that a twofold fine-tuning using a medium-sized artistic dataset may improve the classification on smaller datasets, even when the classification task changes. Besides, Chen *et al*. (2019) expanded on previous research on the use of CNNs for style categorization by observing that various layers in existing deep learning models exhibit varied feature responses for the same input picture. To fully use the input from various levels, the authors presented an adaptive cross-layer model that incorporates responses from both lower and upper layers to capture style. Sandoval *et al*. (2019) contributed by proposing a two-stage picture classification strategy to enhance style categorization. The approach divides the input image into patches and utilises a CNN model to categorise the artistic style for each patch in the first step. The CNN's probability values are then combined into a single feature vector, which is sent as an individual input to a shallow neural network model, which conducts the final classification. The suggested technique is based on the idea that separate patches act as independent evaluators for different parts of the same image, with the final model combining those evaluations to make the ultimate judgement. As is typical in this research, there was some misunderstanding between historically related styles. In short, we conclude that differentiating visual styles remains a difficult task. However, it can be observed that with the rapid development of deep learning framework, without using conventional CBIR methods such as handcrafted features, but instead build a domain-specific CBIR could get a better classification of painting's style and also better accuracy in similar image retrieval-based models fine-tuned for style recognition. It closes various gaps in prior methodologies and pave a new direction for domain specific CBIR which provides valuable additional information into classifier decision-making processes.

The rest of this paper is organized as follows. Section 2 presents the proposed approach and then evaluate the solution through experiments and the application of using proposed approach for similarity measure in Section 3. Finally, Section 4 presents the conclusion and future direction of this work.

2.0 PROPOSED METHODOLOGY

2.1 Convolutional Neural Network Model Configuration

The overall structure of CNN was a modified version of VGG16 (Simonyan and Zisserman, 2014) where it has five convolutional layers, three max-pooling layers, and followed by Global Average Pooling layer (GAP) (Lin *et al*., 2013). GAP is a procedure that computes the average output of each feature map, decreasing the total number of parameters in the model, and preparing the model for the final classification layer. The intention of replacing fully connected layers from VGG16 with GAP was to reduce the parameters and lower the risk of overfitting to the training data set. Each convolutional layer yields 64, 128, 256, 512, and 512 feature maps, respectively. Filter size of 3 x 3 was use throughout the whole net, which are convolved with the input with only stride 1. Then, the pooling layers with a size of 2 x 2 and step 2 for down sampling. The activation function will be rectification linear unit (RELU) in all weight layers except

the last output layer, which will utilize the softmax function as activation and operate as a multi-class classifier to predict the painting categorization as shown in Figure 1. However, in order to measure the similarity of the paintings, the output layer is removed after training and the features will be extracted from the GAP layer.

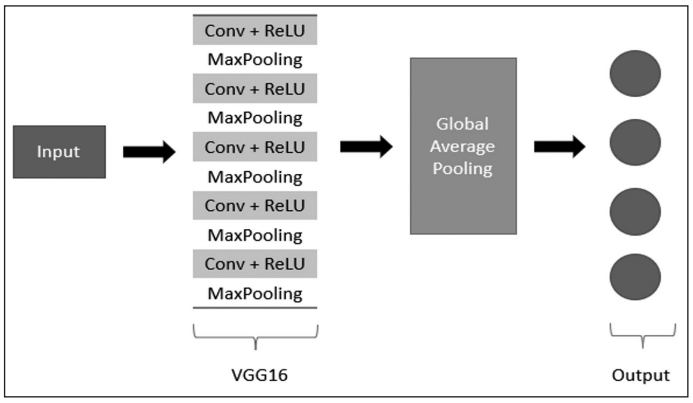


Figure 1: Architecture of CNN

2.2 Dataset

The first data source is Imagenet Dataset which used in Imagenet Large-Scale Visual Recognition Challenge (ILSVRC), in the fine-tuning process, we pre-trained the network using this dataset. It consists of 1.2 million object images with roughly 1000 images in each of the 1000 categories. In all, there are about 1.2million training images, 50,000 validation images and 150,000 testing images. The second data source WikiArt (Saleh and Elgammal, 2015), which is now the largest online available collection of digital paintings. The WikiArt paintings dataset contains over 80,000 fine-art paintings by more than 1,000 artists, it includes artworks from a wide period of time, ranging from the fifteenth century to modern times, and its particular focus on the 19th and 20th centuries, as well as contemporary art. The collection contains 27 different art styles and 45 different genres. WikiArt is also a well-organized collection that incorporates a diverse variety of metadata such as artist name, style, genre, nationality, and so on. Meanwhile, with a total of around 83,000 of samples in the dataset was split into training, validation, and testing with a ratio of 70%, 15% and 15% respectively.

2.3 Experimental Set-up

2.3.1 Input Layer and Preprocessing

The input data with a dimension of 224×224×3 where 224×224 is the width and height of the image and 3 is the number of channels which is RGB colour image. The preprocessing of input image will be subtracting the mean value of RGB over the Wikiart dataset for each pixel. No data augmentation was applied.

2.3.2 Training Details

The model is trained using stochastic gradient descent (SGD) with a batch size of 64 samples. The rest of the parameters are set as momentum of 0.9, decay rate of 0.00001 and the initial learning rate of 0.0001. The weight initialization was taken from the pre-trained VGG16 model, where it was trained with over 1.2 million images for object recognition. Since object recognition and painting's style classification have the same data consistency and share the same data type. The learnt features from object recognition can be easily transfer to the new domain images. This could help in reducing the computational cost for retraining from scratch.

2.3.3 Method for Similarity Measure

After the training process, the trained model with the painting dataset will be used to extract the features from each image. The softmax last layer is removed, and the GAP output feature will be stored to measure the similarity between images (refer to Table 1). In particular, features extracted from GAP is 512 feature vectors. Based on retrieved feature vectors, the distance between feature vectors was calculated using the k-NN brute-force approach, and Euclidean distance measure is utilised as a distance metric to calculate the painting similarity. The general formulation for points given by Cartesian coordinates in n - dimensional Euclidean space is as follows:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \quad (1)$$

Table 1: Related parameters of Convolutional Neural Network

Type	Size/Stride	Output Size
Conv1	3×3/1	64 × 224 × 224
MaxPool1	2×2/2	64 × 112 × 112
Conv2	3×3/1	128 × 112 ×112
MaxPool2	2×2/2	128 × 56 × 56
Conv3	3×3/1	256 × 56 × 56
MaxPool3	2×2/2	256 × 28 × 28
Conv4	3×3/1	512 × 28 × 28
MaxPool4	2×2/2	512 × 14 × 14
Conv5	3×3/1	512 × 14 × 14
MaxPool5	2×2/2	512 × 7 × 7
GlobalAvrPool	-	512
Softmax	-	27

2.4 CNN Architecture Details

2.4.1 Global Average Pooling

Global average pooling (GAP) (Lin *et al*., 2013) is a pooling operation designed to replace the conventional architecture of CNN that uses fully connected layers as the standard configuration. Fully connected layers usually consist of too many parameters, and this has led to slow training speed of the network. By replacing a fully connected layer with GAP, it not only can eliminate the use of parameter, but also is able to avoid overfitting. Instead of adding fully connected layers on top of the feature maps as in conventional CNN model in GAP, the average of the entire pixels of each feature map is taken, and the resulting vector is fed directly into the softmax layer for classification.

2.4.2 VGG16 Pre-Training Model

As shown in Figure 1, the first five convolutional layers which is the VGG16 (Simonyan and Zisserman, 2014) pretrained model that acts as the base model in the proposed cascade CNN architecture. The unique characteristic of VGG16 is that instead of having a large number ofconvolution filters, the authors focused on having convolution layer of 3 x 3 filter with stride 1 and followed by max pooling layer of 2 x 2 filter with stride 2. This base model is responsible on learning the low-level features that can better adapt to various problems and high-level features for domain specific problems. Also, pretrained networks for VGG are available freely on the internet, the weights can be downloaded and used for transfer learningwhere it can shift the learnt features from one domain (object recognition) to the new domain (paintings classification).

3.0 EXPERIMENTAL RESULT AND DISCUSSION

The result in Table 2 shows the proposed fine-tuning model has a competitive result as compared to the current state-of-art methods (Tan *et al*., 2016 & Cetinic *et al*. 2018) without any additional mechanisms. By comparing the current state-of-art result with the proposed model with further fine-tuning, the proposed model implementation with retraining achieved 55.6% of accuracy for classifying painting's style. In addition, proposed model has lesser parameter with approximate of 44-millions which resulted in reducing the computational cost compared to Tan *et al*. (2016) about 61-millions parameter by just replacing the fully connected layer with global average pooling layer. As a result, it was conjecturing that the proposed model can be further improved with different classifying approach and applying different data pre-processing can lead to a huge boost of performance as shown in the work of (Lecoutre *et al*., 2017). In section 3.2, different approach of classifying painting's style was further explored, and it was able to beat the current-state-of art.

Table 2: Comparison of CNN fine-tuned results on Wikiarts dataset with model pretrained with ImageNet Dataset

Reference	Methods	Number of classes	Accuracy (%)
Our model	Proposed Model (VGGNet)	27	55.6
Tan *et al.* (2016)	CNN fine tuning (AlexNet)	27	54.5
Cetinic *et al.* (2018)	CNN fine-tuning (CaffeNet) – Hybrid model (with best fine-tune scenario)	27	57

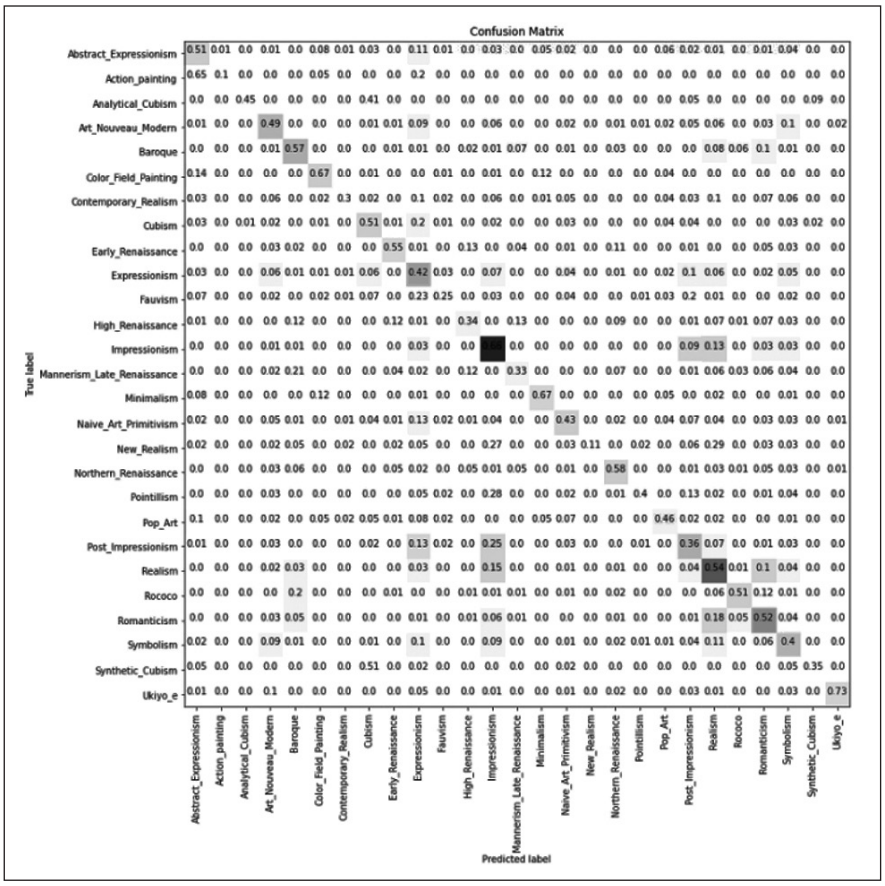


Figure 2: Confusion Matrix for style classification

Following the style classification result and comparison with the existing state-of-the-art, a further examination of the style classification was carried out by looking into the per-class classification performance that merited attention. Figure 2 depicts the WikiArt dataset's confusion matrix for each classification class. It can be observed that there are several classes perform relatively better as compared to other classes due to their distinct visual appearance, such as Ukiyo-e (73%), Minimalism (67%) and Colour Field Painting (67%). Ukiyo-e shows in general the best result as it is a type of art that flourished in Japan that has a very special characteristic. Secondly, the proposed CNN can distinctly classify the Impressionism with 66% accuracy from the other styles. The high accuracy might be due to the high number of training data with approximately 13-thousands images in the Impressionism category. This is consistent with the finding of Goodfellow *et al*. (2016) that a neural network requires around 5,000 labels per class to achieve human-level classification performance. Action paintings (10%) was wrongly classified as Abstract Expressioniosm (51%), this was because action painting was evolved in the 1940s and 1950s during a time of unrest following World War II which can be seen as both of the styles are belong to the same groups. For the misclassification on the dataset, normally two groups of styles share the common conceptual ground. For example, 25% of Post Impressionism was wrongly classified as its elder brother, Impressionism. 51% of Synthetic Cubism was classified as Cubism which Synthetic Cubism was known as later phase of Cubism started from 1908-1912. New realism was wrongly classified nearly 30% to Realism

which come from the same root as it was new form of realism developed at the beginning of 20th century. This is similar to Rococo (51%) and Baroque (57%) as these two styles are historically related. The misclassification of style also explained that it have a poor performance from the classification task as cascade model generally struggle on differentiating historically related styles. It can be concluded that artwork style is not only associated with common visual properties but contextually dependant concept.

3.1 Measure Similarity with Proposed Model

CNN models fine-tuned for style identification were used to retrieve images with similar style or content. As shown in Figure 3, each query image with four of the most similar images were retrieved. We can see from these examples that the suggested CNN fine-tuned model for style recognition focuses more on style attributes like brushwork or amount of detail. Despite some incorrectly obtained class image, it can nevertheless retrieve similar painting in terms of content by including certain items and similar compositions. In addition to the result above, we conjecture those further improvements in style-specific classification performance will result in greater distinguishability between style-similar images. Therefore, in order to validate this hypothesis, further investigation of the model features is performed to study the effect of before and after fine-tuning by transfer learning in Section 3.2. It is well known that the ImageNet dataset was used to train various pre-trained models (VGG also pre-trained with the ImageNet dataset). As a result, in most cases, they provide an excellent starting point for similarity computations. However, if these models were adjusted to suit the specific problem, they would find similar images even more accurately.

3.2 Comparison of CBIR Performance Before and After Fine-Tuning with Specific Domain Knowledge

Similar process mentioned from Section 2.3.3 which extract the features and retrieve the similar images, but from this experiment, we identify the worst-performing categories, fine-tune, and then see how the accuracy change. For every image in the WikiArt dataset, it uses the brute-force approach to determine the closest neighbours for each image in the

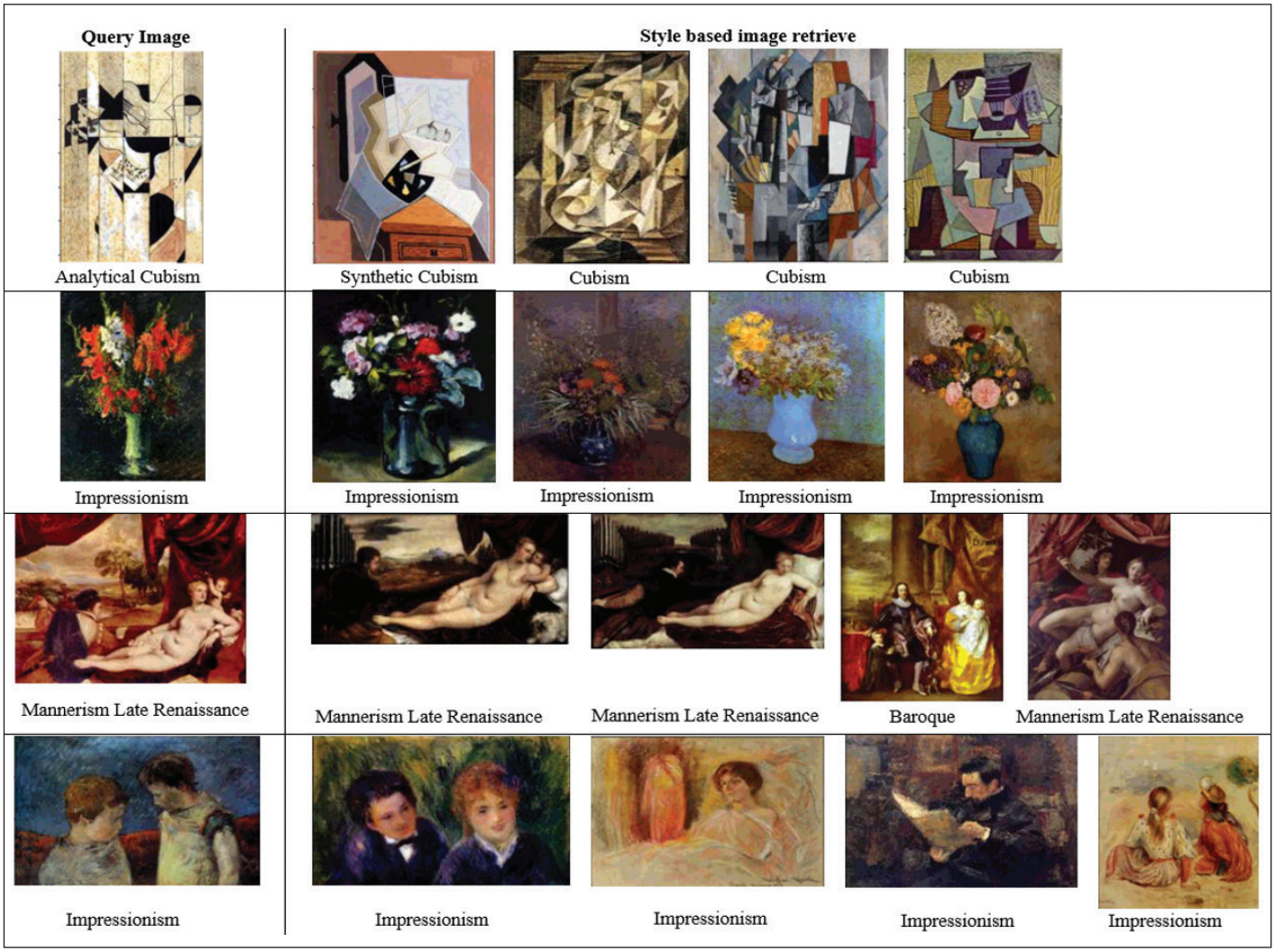


Figure 3: Examples of paintings with style label retrieved as most similar to the query image when using the fine-tuned proposed CNN model as feature extractors

dataset and then returns the top-10 classes with the lowest accuracy. The analysis would provide an overview on how fine adjustment affects the results.

Table 3: Top 10 Lowest-Accuracy Classes

No.	Class (Before fine-tuning)	Retrieval accuracy (%)	Class (After fine-tuning)	Retrieval accuracy (%)
1	New Realism	11.58	Fauvism	39.92
2	Fauvism	21.6	New Realism	40.56
3	Mannerism Late Renaissance	23.8	High Renaissance	48.12
4	High Renaissance	24.89	Mannerism Late Renaissance	50.4
5	Pointillism	25.51	Action Painting	50.82
6	Rococo	29.07	Post Impressionism	52.57
7	Post Impressionism	30.66	Expressionism	52.81
8	Early Renaissance	32.45	Synthetic Cubism	54.1
9	Action Painting	32.84	Contemporary Realism	55.48
10	Baroque	34.16	Symbolism	55.84
Average Correct Prediction Accuracy (%)		**39.2**		**61.33**

With the extracted feature vectors before fine-tuning model, it can be observed from Table 3 that the retrieval accuracy is quite poor as the lowest accuracy was only 11.58% while the highest accuracy in the Top-10 least accuracy classes was at 34.16%. This result shows that the model suffered from discriminating the correct classes when retrieving similar images. Using these feature vectors in applications such as image retrieval systems may be a bad idea because obtaining a clean plane of separation between classes may be difficult. It is hardly surprising that the retrieval accuracy performed so poorly in this nearest-neighbour-based categorization task due to the learned features being based on the natural images. In contrast, after retraining with domain dataset, the outcome is intriguing; the Top-10 least accurate classes have some changes, and retrieval accuracy has skyrocketed. Previously, the feature vectors from the model before fine-tuning achieved an overall correct prediction accuracy of only 39.2%. The new feature vectors after fine-tuning deliver a whopping 61.33% accuracy.

From Table 4, the prior works classification accuracy was again act as a benchmark. As we compared the result with prior work, it shows that with our approach could outperforms the current state-of-art reported for the WikiArt dataset. In Tan *et al*. (2016) paper, the authors achieved the best result with 54.5% by fine-tuning the Alexnet network which also pre-trained with ImageNet dataset. On the other hand, Cetinic *et al*. (2018) achieved an even better result with 57% by implementing different domain-specific weight initialization and different training settings. However, with our approach where basically extract the feature vectors from fine-tuned model and further classified with nearest-neighbour approach led to a better performance in overall. To summarise the discussion, the hypothesis expressed in the previous section was valid in which additional increases in style-specific classification performance will result in higher distinguishability across style-similar images. As a result, we may conclude that domain-specific initialization and task-specific fine-tuning can have a considerable impact on obtaining CBIR performance.

Table 4: Comparison of results to prior works on the style classification task with new feature extraction method

References	Methods	Accuracy (%)
Proposed model	Proposed Model (VGGNet)	61.33
Tan *et al.* (2016)	CNN fine-tuning (AlexNet)	54.5
Cetinic *et al.* (2018)	CNN fine-tuning (CaffeNet) – Hybrid model (with best fine-tune scenario)	57

4.0 CONCLUSIONS

In this work, we presented a study using CNN as a feature extractor for measuring similarity between painting's styles. We successfully applied the extracted feature to retrieve the right classes from the query image that achieve over 61% accuracy. This improvement is mainly due to the idea of transfer learning and the importance of retraining. As suggested by our experiments, CNN retraining is required to build a specific domain CBIR that can outperform general CBIR and is suitable for measuring the similarity of painting and feasible for use in online art galleries. However, the inclusion of a larger painting dataset should allow the model to learn more from scratch rather than via transfer learning. As a result, we intend to expand the dataset so that we may fully retrain the deep learning models. We also plan to deepen our multidisciplinary collaboration in the future by doing research on the importance of the findings to specific art history study areas. Investigate how a deep neural network may be used to extract high-level and semantically significant components that can be utilised to discover new knowledge patterns and meaningful connections between individual artworks. Increase the knowledge and interpretability of deep learning models, on the other hand, by leveraging CNN representations visualisation approaches such as activation maximisation, saliency maps, and class activation maps, as well as other visualisation techniques for a better understanding of how CNN extract features from paintings

5.0 ACKNOWLEDGMENTS

This project is funded by CMG Holdings Sdn Bhd, we would also like to express our gratitude towards our collaborator Cashierbook for their useful suggestions and special thanks to the reviewers for their feedback in this paper. ■

REFERENCES

[1] Alhassan, A.K. and Alfaki, A.A., 2017, January. Color and texture fusion-based method for content-based image retrieval. In 2017 International Conference on Communication, Control, Computing, and Electronics Engineering (ICCCCEE) (pp. 1-6). IEEE.

[2] Bay, H., Tuytelaars, T., and Van Gool, L., 2006, May. Surf: Speeded up robust features. In European conference on computer vision (pp. 404-417). Springer, Berlin, Heidelberg.

[3] Campbell, J.F., 1994. Integer programming formulations of discrete hub location problems. European Journal of Operational Research, 72(2), pp.387-405.

[4] Cetinic, E., Lipic, T. and Grgic, S., 2018. Fine-tuning convolutional neural networks for fine art classification. Expert Systems with Applications, 114, pp.107-118.

[5] Cetinic, E., Lipic, T. and Grgic, S., 2020. Learning the Principles of Art History with convolutional neural networks. Pattern Recognition Letters, 129, pp.56-62.

[6] Chen, L. and Yang, J., 2019, October. Recognizing the style of visual arts via adaptive cross-layer correlation. In Proceedings of the 27th ACM international conference on multimedia (pp. 2459-2467).

[7] Gonthier, N., Gousseau, Y. and Ladjal, S., 2021, January. An analysis of the transfer learning of convolutional neural networks for artistic images. In International Conference on Pattern Recognition (pp. 546-561). Springer, Cham.

[8] Goodfellow, I., Bengio, Y. and Courville, A., 2016. Deep learning. MIT press.

[9] He, T., Wei, Y., Liu, Z., Qing, G., and Zhang, D., 2018, January. Content-based image retrieval method based on SIFT feature. In 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 649-652). IEEE.

[10] Hiremath, P.S. and Pujari, J., 2007. Content Based Image Retrieval based on Color, Texture, and Shape features using Image and its complement. International Journal of Computer Science and Security, 1(4), pp.25-35.

[11] Hwang, YH, and Lee, YH, 2012. Uncapacitated single allocation p-hub maximal covering problem. Computers & Industrial Engineering, 63(2), pp.382-389.

[12] Jabalameli, M.S., Barzinpour, F., Saboury, A., and Ghaffari-Nasab, N., 2012. A simulated annealing-based heuristic for the single allocation maximal covering hub location problem. International Journal of Metaheuristics, 2(1), pp.15-37.

[13] Karimi, H. and Bashiri, M., 2011. The hub covers location problems with different types of coverage. Scientia Iranica, 18(6), pp.1571-1578.

[14] Lecoutre, A., Negrevergne, B. and Yger, F., 2017, November. Recognizing art style automatically in painting with deep learning. In Asian conference on machine learning (pp. 327-342). PMLR.

[15] Lin, M., Chen, Q. and Yan, S., 2013. Network in network. arXiv preprint arXiv:1312.4400.

[16] Lowe, D.G., 1999, September. Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1150-1157). IEEE.

[17] Oliva, A., and Torralba, A., 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision, 42(3), pp.145-175.

[18] Pasandideh, S.H.R., Niaki, S.T.A. and Sheikhi, M., 2016. A bi-objective hub maximal covering location problem considering time-dependent reliability and the second type of coverage. International Journal of Management Science and Engineering Management, 11(4), pp.195-202.

[19] Rui, Y., Huang, T.S. and Chang, S.F., 1999. Image retrieval: Past, present, and future. Journal of Visual Communication and Image Representation, 10(1), pp.1-23.

[20] Saleh, B. and Elgammal, A., 2015. Large-scale classification of fine art paintings: Learning the right metric on the right feature. arXiv preprint arXiv:1505.00855.

[21] Sandoval, C., Pirogova, E. and Lech, M., 2019. Two-stage deep learning approach to the classification of fine-art paintings. IEEE Access, 7, pp.41770-41781.

[22] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

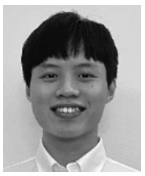
[23] Tan, W.R., Chan, C.S., Aguirre, H.E. and Tanaka, K., 2016, September. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In 2016 IEEE international conference on image processing (ICIP) (pp. 3703-3707). IEEE.

[24] Van der Maaten, L. and Hinton, G., 2008. Visualizing the data using t-SNE. Journal of Machine Learning Research, 9(11).

[25] Zade, A.E., Sadegheih, A., and Lotfi, M.M., 2014. A modified NSGA-II solution for a new multi-objective hub maximal covering problem under uncertain shipments. Journal of Industrial Engineering International, 10(4), pp.185-197.

[26] Zade, A.E., Sadegheih, A., and Lotfi, M.M., 2016. Fuzzy multi-objective linear programming for a stochastic hub maximal covering problem with uncertain shipments. International Journal of Industrial and Systems Engineering, 23(4), pp.482-499.

PROFILES



TAN WEI SHENG received the B. Sc degree in Management Mathematics with Computing from Tunku Abdul Rahman University College, Malaysia in 2019. Presently pursuing his master's degree in Mathematical Sciences at Tunku Abdul Rahman University College, Malaysia. His research interest includes Statistics, Machine Learning, Deep Learning, and other areas.
Email address: tanws-wa15@student.tarc.edu.my

DR CHIN WAN YOKE received B.Sc. and Ph.D. degrees from Universiti Teknologi Malaysia in the year 2009 and 2016, respectively. She is currently a postgraduate programme leader at Tunku Abdul Rahman University College and is actively involved in the activities organized by the Faculty of Computing and Information Technology, Center for Data Science and Analytics, and Centre for Postgraduate Studies and Research. Her main research focus is mathematics and statistics. She is a multidisciplinary researcher who is involved in several projects including data analytics, predictive modeling, optimization, statistical hydrology, painting styles classification, and social media engagement.
Email address: chinwy@tarc.edu.my

DR LIM KHAI YIN received her Master of Science (Artificial Intelligence) and PhD (Visual Analytics) in 2009 and 2017, respectively. Currently, she is a senior lecturer in TAR UC. Her research interests include computer vision, medical image processing, machine learning, and data analytics. She has also published papers both in journal and conferences relating to artificial intelligence (AI) and computer vision, specifically in fuzzy theory, neural networks, and medical imaging during her postgraduate study. She has experiences in guiding final year project (FYP) students on the topics that involves AI and computer vision. Besides that, she has been involving in some industry projects, mainly in the topics of classifying painting styles and engagement prediction in social media.
Email address: limky@tarc.edu.my

FORECASTING FACEBOOK USER ENGAGEMENT USING HYBRID PROPHET LSTM AND IFOREST

(Date received: 24.01.2022/Date accepted: 08.04.2022)

Kong Yih Hern[1], Lim Khai Yin[2], Chin Wan Yoke[3]

[1,3] Tunku Abdul Rahman University College, Jalan Genting Kelang, Setapak, 53300 Kuala Lumpur, P.O. Box 10979, 50932 Kuala Lumpur, Malaysia

[2] Department of Computing and Information Technology, Tunku Abdul Rahman University College, Penang Branch Campus

E-mail: [1] kongyh-wa15@student.tarc.edu.my, [2] limky@tarc.edu.my, [3] chinwy@tarc.edu.my

ABSTRACT

Business forecasting remains a popular topic these days. A reliable business forecast often plays a vital role in an advertising campaign. The amount of attention acquired by posting an advertisement is one of the most essential criteria to determine the effectiveness of the advertisement. The number of times that public users engage with a content signifies the amount of attention received, which was measured by user engagement. With a good forecast, the advertisement could be promoted to a larger number of people. Facebook, as the most popular social media site, is preferred by the majority of advertisers. Therefore, this study addresses Facebook user engagement by forecasting the optimum date to post an advertisement. Different forecasting models, each with its own strengths and weaknesses, are used to model time series data with various properties. The objective of this study is threefold: to investigate the accuracy of the proposed Hybrid Prophet-LSTM that combines Long Short-Term Memory (LSTM) and FBProphet (Prophet), to study the holiday impact on user engagement forecasting on Facebook brand pages, and to study the effect of implementing Isolation Forest (iForest) on the dataset and its contribution to the forecast result. Data from three popular brand pages were used in the experiments in the period of June 2018 to March 2021. The results show that the proposed hybrid model outperforms both the standalone LSTM and Prophet across the datasets. Besides, it is found that holiday effect could generally increase forecast accuracy. In general, datasets pre-processed using iForest can reduce the forecast error under specific conditions. Therefore, the optimum date for an advertisement campaign can be determined on the basis of the most anticipated user engagement, which consequently enhances the business income.

Keywords: Time Series forecasting, Hybrid forecasting, Business forecasting, Prophet, LSTM, Holiday effect

1.0 INTRODUCTION

User engagement refers to attention, interaction, perceived user control, and impression from public users (Brien and Toms, 2008). Businesses are constantly seeking innovative ways to improve the effectiveness of their advertisements. Adopting ineffective marketing techniques for a marketing campaign would not only squander corporate resources, but will also fail to get the desired results. The absence of user engagement with the advertising platform was the most common reason advertising campaigns underperforming (Goldsmith and Lafferty, 2002; Frolova, 2014). The forecasted user engagement on a given advertisement plays an essential role in maximising the effect of an advertisement. According to the study by Frolova (2014), effective advertising can considerably increase volume of sales profits, foster consumption culture, fulfil customer wants for goods, and link advertiser and consumer audience in terms of communication channels. In other words, businesses should promote at the best time possible to achieve the most responses or user engagement from the audience.

Massive volumes of data from a large number of consumers are being collected through the media, particularly social media. A variety of studies by Schoen *et al*. (2013); Srinivasan *et al*. (2013); Breitenecker (2014); Kundi *et al*. (2014); Yasuko, Etuso, Akira (2014); Li *et al*. 2015; Di Gangi and Wasko (2016); Lee *et al*., (2016); Debreceny (2019) have utilized social media data for various analysis. Researchers can study human behaviour patterns and predict user engagement using data from social media. In recent years, business forecasting has been a popular topic of study. The approach has been used to forecast time series data such as future stock movement (Sidi, 2020), traffic matrix (Azzouni and Pujolle, 2017), insurgency movement (Waeto *et al*., 2017), and user engagement (Srinivasan *et al*., 2013). However, there are some uncertainties in exploring the datasets collected in each of these applications. In this paper, these uncertainties were removed using iForest before fitting the datasets into forecasting

models. The objectives of resolving the uncertainties of the datasets are to decrease modelling error, hence increasing the forecast accuracy.

With an accurate forecasting result, user engagement for an advertisement may be easily attained. Selecting the right forecasting model is, therefore, of utmost importance. Today, businesses are using a variety of forecasting methods. In this study, forecasting experiments are conducted using Facebook data. This research employs the proposed Hybrid Prophet-LSTM by Kong *et al*., (2021) to forecast user engagement that would in turn assist businesses in making managerial decisions on the commencement of an advertising campaign.

In Section 2, applications of forecasting techniques in various fields are presented, showing how forecasting models are being used to solve various business problem. In Section 3, the details of the data set and the proposed model are explained. The results of the evaluation can be found in Section 4. Section 5 discusses the findings and conclusion for this study.

2.0 LITERATURE REVIEW

2.1 Forecasting in Businesses

Different models have been employed to analyse and solve various business problems (Polat, 2007). It is important to determine which model to use to solve a business problem.

Prediction and forecasting have recently been a popular topic. To overcome the network traffic problem, the study by Azzouni and Pujolle (2017) used the LSTM model to predict the network traffic matrix. Real-world data from the GEANT organisation network was used to test the feasibility of the LSTM model. The LSTM model is validated that could accurately predict traffic metrics. This forecast result is used to assist network operators in making decisions such as traffic accounting, short-time traffic scheduling, traffic rerouting, network design, long-term capacity planning, and network anomaly detection based on actual network traffic flows.

Yenidogan *et al*. (2018) used forecasting to tackle the difficulty of Bitcoin forecasting in a recent study. The dataset contains two years of Bitcoin exchange rates against a variety of currencies. The author employed Prophet to project future Bitcoin values, which is a critical subject for profit-seeking investors. Bitcoin values were considered successful for the future 90-day forecast with a precision of 94.5%. A credible forecast of future Bitcoin values would be valuable information for investors who want to profit from their Bitcoin investments. The articles demonstrate how a forecasting model can be used to solve a business prediction problem by providing estimated future values that can be used to help make decisions.

Another study by Li *et al*. (2015) used forecasting to address a Twitter advertising problem. The click-through rate (CTR) on the Twitter timeline was forecasted using pointwise learning, pairwise learning, and a further improvement version based on these two models. In the work, the authors proposed a model that used an improvised algorithm based on pairwise and pointwise learning to learn user impressions with the click probability. The forecast result will alter how Twitter displays advertisements to users, leading to a higher CTR from Twitter users. The outcome from the model is found to be a more successful approach than traditional computational advertising, which are sponsored search and contextual advertising. The author concluded that the proposed method could significantly enhance the users' CTR on Twitter's advertisements.

The forecasting technique could also be used on social media data for a variety of purposes. Schoen *et al*. (2013) forecast future events and developments using social media data. The events include the area of politics, finance, entertainment, market demands, health, and others. The same study by Schoen *et al*. (2013) included influenza incidence, product sales, stock market movement, and electoral results as examples of forecasting applications using social media data. As a result, user engagement is forecast to decide the optimal date to promote. A reliable forecast of user engagement could help companies make strategic decisions about how to execute a successful advertising campaign that reaches the largest number of people.

However, many of the researchers tend to neglect the effect of holiday events on the forecast result, as well as the complexity of forecasting human behaviour-related events such as user engagements in this study. To solve the above-mentioned business problem would require a hybrid methodology such as Hybrid LSTM-Prophet model (Kong *et al*., 2021).

These studies by Hummel and Sligo (1971); Saccenti *et al*. (2014) implied that both multivariate and univariate approaches should be used because the results from these two analyses are complementary. Businesses should validate the result from both multivariate and univariate approaches with numerous methods, such as referring to expert knowledge and experience in the domain, conducting experiments using other datasets, and comparing the result with other single-variable data analysis in advertising, as advertising investment is a complicated practice in the real world (Dawes *et al*., 2018). However, this study would only be scoped in univariate analysis, or single-variable business forecasting.

3.0 PROPOSED METHODOLOGY

3.1 Time Series Data

The study by Bashar *et al*., (2012) shows Facebook is the largest and the most favoured social media platform for public users. To determine which variable is important to the research, the Facebook page and post metrics (*Insight - Pages,* 2022) were examined. Three brands were arbitrarily chosen from each category of food, beverages and cosmetics. The purpose of the following sections is to forecast the daily engagement received by a certain Facebook Page in order to reach the largest number of people possible. The target variable is the daily page engagement attribute, which was crawled from Facebook. Page engagement is a daily metric derived from user actions such as clicks, responses, comments, shares, and other forms of interaction with the page. In this study, only one variable, customer page engagement, was examined and forecasted.

Three datasets from two distinct sectors were gathered. Two years of daily time series data, starting on June 1, 2018, and ending on March 31, 2021, were gathered as a dataset from the three specified pages. Malaysian Public Holiday has been included in Prophet's holiday component. The purpose of this holiday dataset is to investigate the impact of holiday effects on time series forecast results. To assess the influence of holiday effects, a comparison study is carried out.

The number of times users engage with a certain page on a daily basis is called customer page engagement. As a result, this variable is a daily data variable with a daily count of user engagement. Users' clicks, reactions, shares, comments, and other actions are used to calculate user engagement. In general, customer page engagement is a measurement of how much public users pay attention to a page.

The datasets were examined to see whether any outliers existed. The term "outlier" refers to observations that differ from the majority of the data (Rousseeuw and Hubert, 2018). These abnormal observations are also known as anomalies in a machine learning context. However, Pollet and van der Meij (2017) said that outliers may be caused by various errors such as data entry errors, experimental errors, data processing errors, measurement errors, and natural. If the outliers were not produced by any of the above-mentioned errors, these observations should be categorised as the novelties of the data. Although we ensured that the datasets do not have any outliers caused by the mentioned errors, however we conducted experiments using outlier detection and removal technique to measure and evaluate the outcome of removing these abnormal observations from the dataset and its contribution to the forecast result.

3.2 Proposed Procedure

This study uses the proposed Hybrid Prophet-LSTM by Kong *et al*., (2021) to enhance forecast accuracy. In the suggested hybrid methodology, Prophet is used as the linear model, while LSTM is used to address the residual nonlinear connection in the time series data. Prophet is used to model regular, and nonregular holiday events. Because time series data contains both linear and nonlinear structure, a nonlinear model, such as LSTM, is used to represent the residual matrix from a linear model. With its remarkable potential to address nonlinearity relationships in time series, LSTM is utilised to model nonlinear relationships in the residual matrix to produce better forecast results. To create the forecast result, the time series data were first fitted into Prophet. The residual matrix was computed using Prophet's forecasted output, and the residual matrix was then fitted into LSTM. The forecast residual is used to compute the hybrid forecast output. Finally, the output produced by the hybrid model is evaluated using several performance metrics and compared to the results of various models.

The datasets were analysed using (iForest), the outlier detection technique that we selected in this study. iForest is proved to have great performance with high efficiency with only a very small numbers of tress or sub-sampling size (Liu *et al*., 2012). Many researchers Liu *et al*., (2012); Ding and Fei (2013); Hofmockel and Sax (2018); Gao *et al*. (2019); Holmer (2019); Hariri *et al*., (2021) have been adapting iForest as a tool for anomaly detection and even creating an iForest-based approach. Liu *et al*., (2012) said that iForest takes the advantages of anomaly properties where: the anomalies are the minority and consist of fewer instances; the anomalies have values that are very different from the normal instances. In a simpler word, anomalies are less and different when compared to the normal observations. The study by Liu *et al*., (2012) demonstrates the algorithm of iForest to estimate *s*, the anomaly score of the data points. The anomaly score *s* is used to make the following assessment: normal common samples (*s* lower than 0.5), normal uncommon samples (*s* near 0.5), and outliers (*s* very close to 1).

In this study, it is observed that many unexplainable data points occurred in the time series. These data points were further analysed and it is observed that these data points were not explainable by the time series trend, seasonality, nor holiday effect. Additionally, these data points were not caused by either of the errors. Therefore, in this study, these anomalies were identified and removed using iForest, to evaluate the effect of iForest on the datasets and its impact on the forecast result.

4.0 EMPIRICAL RESULTS

4.1 Time Series Decomposition

An analysis is performed to study the individual components in the time series data. The user engagement data was analysed using decomposition feature in Prophet. A decomposition result was generated to understand various characteristics of a time series such as its trend, holiday effects, and multiple forms of seasonality.

In order to achieve a reliable forecast, it is important to capture the components in the time series. Simpler to say, the clearer the captured pattern, the better the model understands the data. The decomposition result shown in Fig. 1 is created using Dataset 1. As shown in Fig. 1, the Prophet model decomposes the time series into individual components and analyses each of these components separately. Parameters for each of these components can also be adjusted individually. Each of these graphs tells information for each component in the time series. Fig. 1(a) shows the growth trend of engagement, Fig. 1(b) shows the impact scale of holiday effects on engagement, Fig. 1(c) shows the weekly seasonality, Fig. 1(d) shows the yearly seasonality, Fig. 1(e) shows the monthly seasonality and Fig. 1(f) demonstrates the quarterly seasonality.

In Fig. 2, it shows the decomposition analysis result after the dataset is pre-processed using iForest. As the 'outliers' were detected and removed during iForest, the dataset now does not have the 'peak points' that differ from most common observations. Here we demonstrate the difference before and after using iForest. In Fig. 2(a), the growth trend reduced in total variance. The pattern of different forms of seasonality in Fig. 2(c), Fig. 2(d), Fig. 2(e), and Fig. 2(f) changed accordingly. Furthermore, in Fig. 2(d), the yearly seasonality pattern become more stably recursive compared to Fig. 1(d) that captured some noise in the time series. Although Prophet's model seems to be able to capture the pattern of the series better than using the original data, here we only compare the effect of iForest with the time series, and the contribution of iForest to the forecast result will be discussed later.

4.2 Standalone and Hybrid Prophet-LSTM Algorithm

The linear relationship in time series data was fitted using Prophet and the remaining pattern under the Prophet residual was fitted into LSTM. Five distinct methods are compared for creating a reliable forecast result and examining the impact of the holiday effect on the forecast outcome. Prophet, Prophet without holiday, LSTM, Hybrid Prophet-LSTM, and Hybrid Prophet-LSTM without holiday were among the approaches used. These methods are evaluated using different performance metrics including Weighted Mean Absolute Percentage Error (WMAPE), R^2 Score, Root mean square error (RMSE) and Mean Absolute Deviation (MAD).

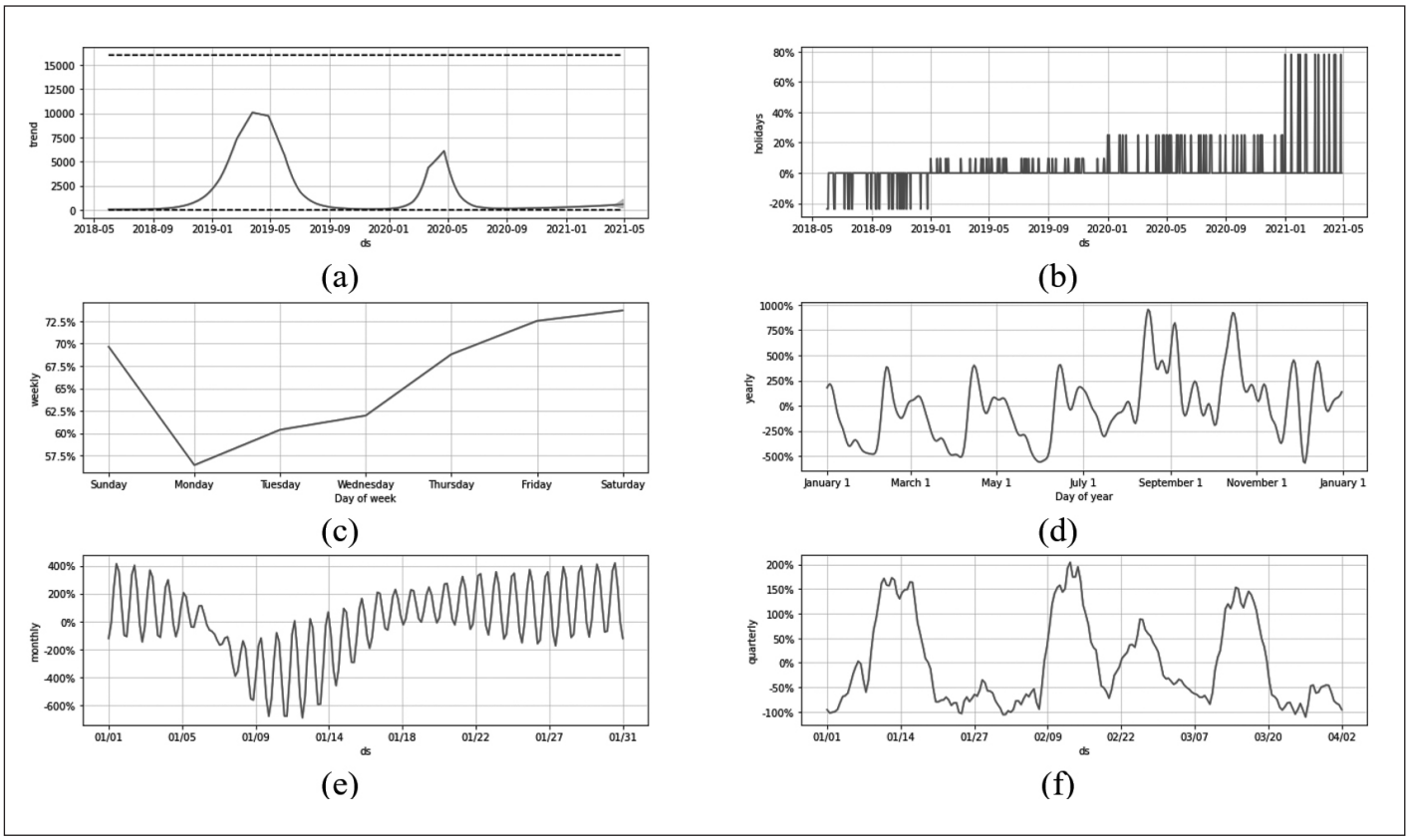


Figure 1: Individual components from Prophet's decomposition result
a) Data trend, b) Holiday effects, c) Weekly seasonality, d) Yearly seasonality, e) Monthly seasonality, f) Quarterly seasonality

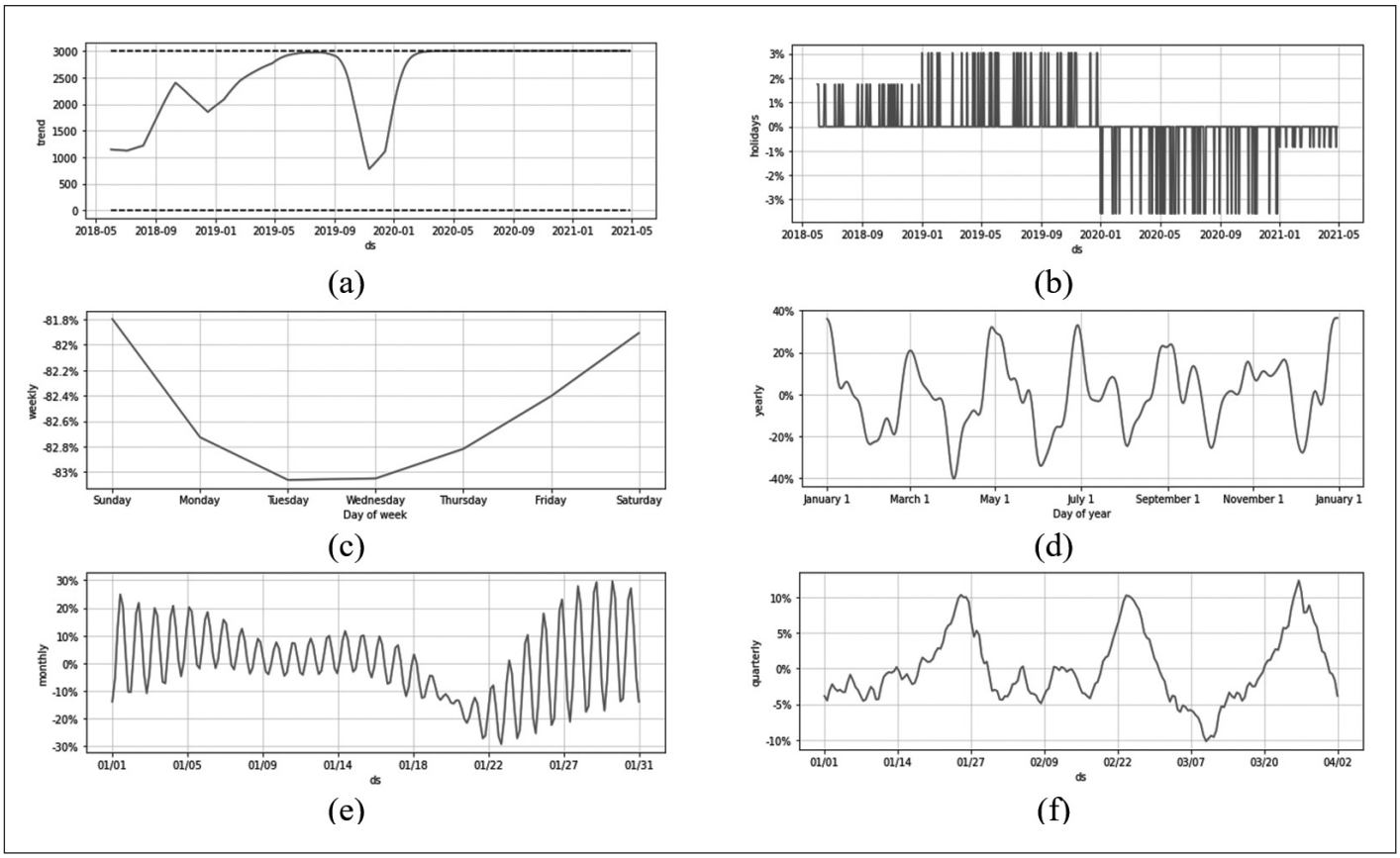


Figure 2: Individual components of Prophet's decomposition result after iForest
a) Data trend, b) Holiday effects, c) Weekly seasonality, d) Yearly seasonality, e) Monthly seasonality, f) Quarterly seasonality

Table 1 compares the results of the three models with different approaches without removing anomalies using iForest. Models 1 and 2 are the standalone Prophet and LSTM, and Model 3 is the Hybrid Prophet-LSTM. Models were also compared with and without the holiday component. Prophet has a WMAPE of 47.86 %, 21.19 % and 55.35 %, respectively, using the three datasets. By comparing the error rates of the standalone Prophet and LSTM models, the results show that Prophet's error rate is at least double or more than LSTM's error rate. LSTM outperforms Prophet by producing fewer errors and a lower overall error rate. When LSTM is compared to the hybrid model, the hybrid model outperforms the LSTM model in every aspect. By having reduced mistakes and error rates, as well as a higher R^2 value, the hybrid model outperforms the LSTM. The hybrid model is an alternative to the traditional model.

Prophet model did not demonstrate good modelling in this case since Prophet's holiday component does not significantly improve Prophet's performance. Despite the fact that holiday effects were not adequately visible in Prophet's forecasting results, the holiday component has a significant impact on the hybrid model forecast result for Dataset 3. The holiday component has a minor influence on Datasets 1 and 2, but has a considerable impact on Dataset 3 with an error reduction of 18.89 % to 31.86 % without the holiday component. Overall, the holiday component could improve the hybrid Prophet-LSTM model in producing a more accurate forecast.

Table 1 shows that LSTM outperforms Prophet when it comes to modelling user engagement time series data. As a result, a hybrid model may model both linear and nonlinear time series data with steady performance because it incorporates the strengths of both linear and nonlinear models.

During the study, it was discovered that although the Prophet linear model can detect seasonality in Datasets 1 and 3, the seasonality captured is unusual and does not show an observable pattern, which can considerably increase forecast error. When there is no observable seasonality pattern in the time series, the forecasting accuracy for these datasets is relatively low.

The forecasting models demonstrate their feasibility modelling time series data to forecast user page engagement as a result of the findings. The hybrid model, which was shown to be the best, had a forecast error range of 5.80 % to 18.89 %. Businesses can forecast page engagement using a good model, which can then be used to determine the optimal day to start an advertising campaign. Posting an advertisement on a day with higher engagement indicates that the advertisement will reach a larger group of audience.

Similarly, Table 2 uses datasets that were pre-processed using iForest. Prophet has a WMAPE of 45.28 %, 5.60 %, and 18.11 %, respectively, using the three datasets as shown in Table 1. Compared to the result shown in Table 1, Prophet generally performs better for the three datasets. It is observed that iForest can generally increase Prophet and LSTM's performance whether with holiday component included or not.

However, it is found that the Hybrid model performs terribly bad with the highest forecast error and highest RMSE and MAD in Dataset 1. The forecast error for Prophet remains the same and lower forecast error for LSTM. It is found that huge forecast errors occurred while modelling the residual matrix in the Hybrid Prophet-LSTM modelling process. Although when modelling the pre-processed Dataset 1 has a WMAPE of 45.27%, Hybrid Prophet-LSTM showed that it has the capability to reduce error even if there was high forecast error occurred when modelling using Prophet in creating residual matrix. Therefore, this can be explained by the iForest outlier removal procedure. As mentioned earlier, there were no outliers caused by any kind of errors. The reason why iForest is conducted is to experiment the effect of removing the "outlier" of the dataset and its contribution to the forecast accuracy. In Table 2, iForest generally increases the accuracy of the forecast by reducing error except for the data set using the hybrid Prophet-LSTM model in Dataset 1. This result is tally with the modelling result because both Prophet and LSTM demonstrated to capture a clearer pattern of the series than before. For the huge error that occurred in Dataset 1 Hybrid Prophet-LSTM, it can be explained that LSTM failed to model the residual matrix after removing the natural of the dataset, provided there are no outliers in the data.

Table 1: Performance metrics for Prophet, LSTM, and Hybrid models without using iForest

		Prophet	Prophet (No Holiday)	LSTM	Hybrid Prophet-LSTM	Hybrid Prophet-LSTM (no holidays)
Dataset 1	WMAPE	47.8679%	46.9547%	17.7578%	16.8768%	15.6264%
	R^2	99.9339%	99.9401%	95.8725%	99.9946%	99.9953%
	RMSE	763.306	744.791	347.279	243.283	227.217
	MAD	504.749	495.586	182.326	173.859	161.094
Dataset 2	WMAPE	21.1959%	21.4081%	6.1050%	5.8045%	5.5680%
	R^2	99.9930%	99.9928%	94.7999%	99.9995%	99.9996%
	RMSE	5088.143	5163.085	2283.617	1381.347	1298.541
	MAD	4037.848	4079.288	1160.590	935.641	855.583
Dataset 3	WMAPE	55.3595%	58.1690%	22.8360%	18.8923%	31.8675%
	R^2	99.5889%	99.4153%	58.5297%	99.9799%	99.8619%
	RMSE	328.844	337.710	783.955	112.955	283.801
	MAD	168.269	170.354	63.780	58.906	117.100

Table 2: Performance metrics for Prophet, LSTM, and Hybrid models using iForest

		Prophet	Prophet (No Holiday)	LSTM	Hybrid Prophet-LSTM	Hybrid Prophet-LSTM (no holidays)
Dataset 1	WMAPE	45.2756%	48.3359%	3.6556%	86.3346%	93.6136%
	R^2	43.8439%	38.2864%	99.1561%	0.1821%	3.1823%
	RMSE	242.941	254.679	29.793	464.702	493.059
	MAD	195.043	208.227	15.762	373.262	404.732
Dataset 2	WMAPE	5.6033%	5.6449%	1.1806%	1.1026%	1.0905%
	R^2	73.1712%	72.7941%	85.0587%	99.5299%	99.5470%
	RMSE	1192.951	1201.306	911.127	228.613	224.984
	MAD	947.188	954.237	199.254	130.310	115.753
Dataset 3	WMAPE	18.1109%	18.2126%	4.1066%	3.7666%	3.6554%
	R^2	71.5381%	71.2029%	96.6484%	99.1497%	99.2107%
	RMSE	8.431	8.480	2.899	1.770	1.716
	MAD	6.524	6.560	1.476	1.636	1.778

The proposed approach incorporates features such as a customizable calendar of events or holidays. In addition to irregular holiday events, this study models seasonality and trend components. Experiments were carried out to verify the effectiveness of the proposed model. We can see that by including the holiday components in the hybrid model, the models perform better overall. The suggested model has the smallest forecast errors and performs well on a wide range of datasets and scales of variance. Additionally, iForest could significantly reduce the forecast error when the characteristic of the dataset is well defined. It is best to seek for advice from business expert in the respective domains.

5.0 CONCLUSIONS

The Hybrid Prophet-LSTM was utilised in this work to combine linear and nonlinear models to produce improved forecast results. As a result, it can be inferred that while attempting to produce an accurate forecast, there are two critical factors to consider. The compatibility of the linear model would be the first concern. The performance of the linear model was found to have a significant impact on the hybrid forecast result. In the hybrid model, a well-performing output in the linear model would offer an exceptional outcome. The selection of features is the second factor. In this study, only one variable was chosen for forecasting, which resulted in a univariate analysis. Analysing the relationship between the dependent variable and other independent variables would be future work for this study. It can be stated that the results of this study will be possible to forecast dates with the most user engagement. However, any managerial judgment should not be made solely on the basis of this variable. This study serves the purpose of exploring the expected effect of advertising. Businesses should instead validate the result with numerous approaches, such as referring to expert knowledge and experience in the domain, conducting experiments using other datasets, and comparing the result with other single-variable data analysis in advertising, as advertising investment is a complicated practice in the real world. ■

REFERENCES

[1] Azzouni, A. and Pujolle, G. (2017) 'A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction'. Available at: http://arxiv.org/abs/1705.05690.

[2] Bashar, A., Ahmad, I. and Wasiq, M. (2012) 'Effectiveness of Social Media As a Marketing Tool: an Empirical Study', International Journal of Marketing, Financial Services & Management Research, 1(11), pp. 88–99. Available at: www.indianresearchjournals.com.

[3] Breitenecker, R. J. (2014) 'Customer engagement behaviour in online social networks – the Facebook perspective Sofie Bitter * and Sonja Grabner-Kräuter', 14.

[4] Brien, H. L. O. and Toms, E. G. (2008) 'What is User Engagement? A Conceptual Framework for Defining User Engagement with Technology', Journal of the American Society for Information Science and Technology 59(6):938-955, (April). doi: 10.1002/asi.20801.

[5] Dawes, J. et al. (2018) 'Forecasting advertising and media effects on sales: Econometrics and alternatives', International Journal of Market Research, 60(6), pp. 611–620. doi: 10.1177/1470785318782871.

[6] Debreceny, R. (2019) 'Research in Social Media : Data Sources and Methodologies Research in Social Media : Data Sources and Methodologies', (December 2017). doi: 10.2308/isys-51984.

[7] Ding, Z. and Fei, M. (2013) An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window, IFAC Proceedings Volumes (IFAC-PapersOnline). IFAC. doi: 10.3182/20130902-3-CN-3020.00044.

[8] Frolova, S. (2014) 'Svetlana Frolova THE ROLE OF ADVERTISING IN PROMOTING A PRODUCT Thesis CENTRIA UNIVERSITY OF APPLIED SCIENCES Degree Programme in Industrial Management', (May).

[9] Di Gangi, P. M. and Wasko, M. (2016) 'Social media engagement theory: Exploring the infuence of user engagement on social

media usage', Journal of Organizational and End User Computing, 28(2), pp. 53–73. doi: 10.4018/JOEUC.2016040104.

[10] Gao, R. et al. (2019) 'Research and Improvement of Isolation Forest in Detection of Local Anomaly Points', Journal of Physics: Conference Series, 1237(5). doi: 10.1088/1742-6596/1237/5/052023.

[11] Goldsmith, R. E. and Lafferty, B. A. (2002) 'Consumer response to Web sites and their influence on advertising effectiveness', Internet Research, 12(4), pp. 318–328. doi: 10.1108/10662240210438407.

[12] Hariri, S., Kind, M. C. and Brunner, R. J. (2021) 'Extended Isolation Forest', IEEE Transactions on Knowledge and Data Engineering, 33(4), pp. 1479–1489. doi: 10.1109/TKDE.2019.2947676.

[13] Hofmockel, J. and Sax, E. (2018) 'Isolation forest for anomaly detection in raw vehicle sensor data', VEHITS 2018 - Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems, 2018-March(Vehits 2018), pp. 411–417. doi: 10.5220/0006758004110416.

[14] Holmer, V. (2019) 'Hybrid Extended Isolation Forest'.

[15] Hummel, T. J. and Sligo, J. R. (1971) 'Empirical comparison of univariate and multivariate analysis of variance procedures', Psychological Bulletin, 76(1), pp. 49–57. doi: 10.1037/h0031323.

[16] Insight - Pages (2022). Available at: https://developers.facebook.com/docs/platforminsights/page (Accessed: 29 March 2020).

[17] Kong, Y. H., Lim, K. Y. and Chin, W. Y. (2021) 'Time Series Forecasting Using a Hybrid Prophet and Long Short-Term Memory Model', Communications in Computer and Information Science, 1489 CCIS, pp. 183–196. doi: 10.1007/978-981-16-7334-4_14.

[18] Kundi, F. M. et al. (2014) 'Detection and scoring of Internet Slangs for sentiment analysis using SentiWordNet', Life Science Journal, 11(9), pp. 66–72. doi: 10.6084/M9.FIGSHARE.1609621.

[19] Lee, T.-S., Shia, B.-C. and Huh, C.-L. (2016) 'Social Media Sentimental Analysis in Exhibition's Visitor Engagement Prediction', American Journal of Industrial and Business Management, 06(03), pp. 392–400. doi: 10.4236/ajibm.2016.63035.

[20] Li, C. et al. (2015) 'Click-through prediction for advertising in twitter timeline', Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015-Augus, pp. 1959–1968. doi: 10.1145/2783258.2788582.

[21] Liu, F. T., Ting, K. M. and Zhou, Z. H. (2012) 'Isolation-based anomaly detection', ACM Transactions on Knowledge Discovery from Data, 6(1). doi: 10.1145/2133360.2133363.

[22] Polat, C. (Niğde Ü. (2007) 'The Role of Forecasting and Its Potential for Functional Management: A Review from the Value-Chain Perspective', Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9(1), pp. 373–398.

[23] Pollet, T. V. and van der Meij, L. (2017) 'To Remove or not to Remove: the Impact of Outlier Handling on Significance Testing in Testosterone Data', Adaptive Human Behavior and Physiology, 3(1), pp. 43–60. doi: 10.1007/s40750-016-0050-z.

[24] Rousseeuw, P. J. and Hubert, M. (2018) 'Anomaly detection by robust statistics', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(2), pp. 1–14. doi: 10.1002/widm.1236.

[25] Saccenti, E. et al. (2014) 'Reflections on univariate and multivariate analysis of metabolomics data', Metabolomics, 10(3), pp. 361–374. doi: 10.1007/s11306-013-0598-6.

[26] Schoen, H. et al. (2013) 'The power of prediction with social media', Internet Research, 23(5), pp. 528–543. doi: 10.1108/IntR-06-2013-0115.

[27] Sidi, L. (2020) 'Improving S&P stock prediction with time series stock similarity'. Available at: http://arxiv.org/abs/2002.05784.

[28] Srinivasan, B. V. et al. (2013) 'Will your facebook post be engaging?', International Conference on Information and Knowledge Management, Proceedings, pp. 25–28. doi: 10.1145/2512875.2512881.

[29] Waeto, S., Chuarkham, K. and Intarasit, A. (2017) 'Forecasting time series movement direction with hybrid methodology', Journal of Probability and Statistics, 2017. doi: 10.1155/2017/3174305.

[30] Yasuko, K., Etuso, G. and Akira, I. (2014) 'P REDICT F ACEBOOK I MPRESSIONS ADOPTING A MATHEMATICAL MODEL OF THE H IT', 2(1), pp. 63–68.

[31] Yenidogan, I. et al. (2018) 'Bitcoin Forecasting Using ARIMA and PROPHET', UBMK 2018 - 3rd International Conference on Computer Science and Engineering, (February 2019), pp. 621–624. doi: 10.1109/UBMK.2018.8566476.

PROFILES

KONG YIH HERN is a Master student, studies in Tunku Abdul Rahman University College, Kuala Lumpur. He has a Diploma and Bachelor's degree in Computer Science and Software Engineering. His specialization is in forecasting and predictive analysis, interest in utilizing forecast tools as a solution to address business problems.
Email address: kongyh-wa15@student.tarc.edu.my

LIM KHAI YIN received her Master of Science (Artificial Intelligence) and PhD (Visual Analytics) in 2009 and 2017, respectively. Currently, she is a senior lecturer in TAR UC. Her research interests include computer vision, medical image processing, machine learning, and data analytics. She has also published papers both in journal and conferences relating to artificial intelligence (AI) and computer vision, specifically in fuzzy theory, neural networks, and medical imaging during her postgraduate study. She has experiences in guiding final year project (FYP) students on the topics that involves AI and computer vision. Besides that, she has been involving in some industry projects, mainly in the topics of classifying painting styles and engagement prediction in social media.
Email address: limky@tarc.edu.my

CHIN WAN YOKE received B.Sc. and Ph.D. degrees from Universiti Teknologi Malaysia in the year 2009 and 2016, respectively. She is currently a postgraduate programme leader at Tunku Abdul Rahman University College and is actively involved in the activities organized by the Faculty of Computing and Information Technology, Center for Data Science and Analytics, and Centre for Postgraduate Studies and Research. Her main research focus is mathematics and statistics. She is a multidisciplinary researcher who is involved in several projects including data analytics, predictive modeling, optimization, statistical hydrology, painting styles classification, and social media engagement.
Email address: chinwy@tarc.edu.my

MANUSCRIPT PREPARATION GUIDELINES FOR IEM JOURNAL AUTHORS

The aim of publishing the Journal of the Institution of Engineers Malaysia (or IEM Journal) is to promote the advancement of science, engineering, and technology; disseminate new and current knowledge; share novel findings among practising engineers, researchers and other interested colleagues. Hence, the IEM Journal covers a wide range of practical and diversified engineering disciplines, including publishing papers on any subjects relevant to the engineering and technology of today. As in other journals, all paper submissions to the IEM journal will be peer-reviewed by professionals.

Submission of a contribution is taken to manifest the fact that the submission has not been submitted, accepted, published, or copyrighted elsewhere. To avoid publication delays, please send all manuscripts to the Editor (via the Online Journal Submission) and observe the following guidelines. Each paper is independently peer-reviewed.

Types of Papers

IEM Journal will accept any submissions that fall within the three types of papers shown below:

- **Research Paper**

Significant research and development or applications in any field of engineering and/or technology. Submissions should be about 8-14 pages.

- **Review Paper**

Articles which summarises the state-of-the-art of a specific area of research. Submissions should be about 10-20 pages.

- **Brief Paper**

A concise description of new technical concepts or applications within the scope of the journal. Submission should be about 4 pages

Together with the paper submission, at least four (4) reviewers complete with their contact details (official postal and email address, and telephone number) are to be proposed by authors. Reviewers should be the experts in their research areas.

Plagiarism Policy

IEM Journal strictly prohibits any form of Plagiarism.
If the manuscript has been presented, published, or submitted for publication elsewhere, please inform the Editor. Our primary objective is to publish technical materials not available elsewhere.

Format and Elements of Submitted Texts

a) Initial Submission

For initial submission and reviewing, ALL papers are accepted regardless of the formatting used.

b) Final Submission

Upon final submission of paper, authors are required to follow the manuscript instruction guidelines and paper template as presented below.

Please prepare your main text document in Microsoft Word and PDF format, text should be single line spaced, line numbered and pages should be numbered. You can download the paper template here.

Please note that the style that you submit your paper in (e.g. any additional italics or bold fonts, bullet points, etc.) may be changed on publication to accommodate our publication style.

Manuscript Style

Language:

- The language of the IEM Journal is in English. However, a paper in Bahasa Melayu is also accepted. For accepted papers, an abstract in English and Bahasa Melayu must be included.
- The manuscript should be able to be readily understood by an engineer and researchers alike and should avoid any colloquialisms.
- The terms, including nomenclature and abbreviations, and style should be consistent throughout your journal paper. If you collaborate with other writers, please communicate clearly with them.
- Avoid referring directly to the names of individuals, organisations, products, or services unless essential to the comprehension of the manuscript. Gratuitous flattery or derogatory remarks about a person or organisation should not be included.
- Symbols and Units: SI and derived units should be used, if possible.
- Abbreviations: the use of internationally recognised abbreviations is allowed in the text. The abbreviations should be defined on first use. Abbreviations should not be used in the title.

Manuscript Guide

The following is a detailed manuscript preparation guide for articles submitted to IEM Journal; however, they can, in the most part, be used as a basis for other article types amending to concur with the page limit and premise of the formats, as appropriate.
The manuscript should be typewritten using single-spacing, font of 12 Times Roman; on one side of sheet only and in a single column format.

Title

Titles are limited to 150 characters, including spaces. Please avoid the use of any abbreviations, acronyms, or formulae. Titles should clearly reflect the content of the manuscript and any search terms that readers may use should be considered and incorporated.

List of Authors Name

List down all the names of authors (who has contributed to the paper) and the respective affiliations. From the list of authors, place an asterisk (*) next to the corresponding author's name. Provide an official email address of the corresponding author. **Please DO NOT include your personal telephone number on the title page.**

Abstract

Provide an informative 100 to 250 words abstract at the head of the manuscript. This should be a concise reflection of the aims, findings, conclusions and any interesting or important results. Carefully incorporate any terms that may be used by potential

interested readers to improve the article's discoverability online (search engine optimisation). The abstract should contain no reference. Abbreviations that are not commonly used should be defined (for the benefit of the non-specialist reader) at first use.

Keywords

These are used for indexing. Please include between 3-5 keywords.

List of Notations

Please provide a list of symbols and definitions used in the text. This will ease our readers.

Introduction

A concise summary of current background knowledge, with reference to relevant previous works in the field should be presented. Please also describe the objectives of and justification for the work contained in the submitted manuscript.

Main Text

The methods and processes applied to investigate and achieve the objectives should be communicated in sufficient detail that readers could repeat the work successfully. The results should be reported clearly and interpreted accurately and analysed thoroughly. Figures/tables can be used to support these results.

It is important that all research articles include a section at the end of the main text that highlights the novelty of the results to the engineering field and any potential applications.

All sections should be numbered in Arabic such as 1, 2, etc. with the title in capitals. Sub-sections should be numbered such as 1.1, 2.3, etc. Numbered all equations in round brackets ( ) flush to the right. The equation should be in the centre.

Style for Illustrations (Tables and Figures)

Try to include the illustrations in between the text. Each illustration must be numbered such as "Figure 1, Figures 2-3, etc." and have a meaningful caption at the bottom. For tables, the caption must be at the top.

On graphs, show only the coordinate axes, or at most the major grid lines, to avoid a dense hard-thread result.

All lettering should be large enough to permit legible reduction of the figure to column width. Typing on figures is not acceptable. Photographs should be glossy prints, of good contrast and gradation and any reasonable size.

Conclusions

A concise summary of the results of case studies or research project papers and the lessons learned should be discussed. If necessary, please elaborate the applicability / relevance of your article to readers in other countries.

Research journals must discuss the practical relevance and potential applications of the engineering work described. This is important to readers working in engineering related practice.

Please also include relevant references to demonstrate how previous engineering research work has been used. These references could be standards, codes or relevant past journal papers.

Appendices

Additional information, such as tables or mathematical derivations can be included. These will be included in the article.

Acknowledgements

Please provide acknowledgement details to those persons or organization that contributed to the paper. Additional details required by funding bodies can be included.

References

The references to other literature that you have cited in your main text should be based on the APA style of referencing (Author, Date) as described below. Please refer https://apastyle.apa.org/ for a detailed guide on the referencing style.

- Single author: (Author, Year) or Author (Year)
- Two authors: (Author 1 & Author 2, Year) or Author 1 and Author 2 (Year)
- Three and more authors: (Author 1 et al., Year) or Author 1 et al. (Year)

In the text, the author and year of the reference should be put in parentheses immediately after the work referred to, for example 'Controlled tests on the Millennium Bridge (Chapman et al., 2005; Murray & Geddes, 1987; Wilby et al., 2011) during which…..' or 'as mentioned by Lim et al. (2018), the NCA derived from the agricultural wastes…'

All references must be listed, in full, at the end of the paper in alphabetical order, irrespective of where they are cited in the text.

In the reference section, the references should be written in full, as follows:

Books: Author 1 surname, author 1 initials, & Author 2 surname, author 2 initials. (Year of publication). Book title. Publisher, City, Country. doi reference. For example:

Kobayashi, K., Khairuddin, A. R., Ofori, G. & Ogunlana, S. (Eds.). (2009). *Joint ventures in construction*. Thomas Telford, London, UK.

Owen, G. & Totterdill, B. (2008). *The dispute board hearing. In Dispute Boards: Procedures and Practice*. Thomas Telford, London, UK.

Journal, magazine and newspaper articles: Author 1 surname, author 1 initials, & Author 2 surname, author 2 initials. (Year of publication). Paper title. Journal title, Volume (Issue number), First page-Last page. doi reference. Unpublished papers and theses should not be cited as they are not readily available.

Lim, J. L. G., Raman, S. N., Lai, F. C., Zain, M. F. M., & Hamid, R. (2018). Synthesis of nano cementitious additives from agricultural wastes for the production of sustainable concrete. *Journal of Cleaner Production*, 171, 1150-1160. https://doi.org/10.1016/j.jclepro.2017.09.143.

Soon, F. C., Khaw, H. Y., Chuah, J. H., & Kanesan, J. (2018). Hyper-parameters optimisation of deep CNN architecture for vehicle logo recognition. *IET Intelligent Transport System*, 12(8), 939-946. https://doi.org/10.1049/iet-its.2018.5127.

Conference proceedings: Author 1 surname, author 1 initials, & Author 2 surname, author 2 initials. (Year of publication). Paper title. Proceedings title, Volume (Issue number), First page-Last page. doi reference. Unpublished conference proceedings (i.e. that were only given to delegates) should not be cited as they are not generally available.

Chuah, J. H., Khaw, H. Y., Soon, F. C., & Chow, C. (2017). Detection of Gaussian noise and its level using deep convolutional neural network. *Proceedings of the TENCON 2017 - 2017 IEEE Region 10 Conference*, 2447-2450. https://doi.org/10.1109/TENCON.2017.8228272.

Unpublished material should not be included in the Reference list.

Please refer to https://apastyle.apa.org/ for a detailed guide on the referencing style. Authors should strictly adhere to the referencing style specified herein.

Mathematical Equations

Only relevant equations should be included in the main text and should be numbered. An equation editor program can be used to type in a formula.

Figures and tables caption list: Please supply a figure caption list at the end of your journal paper. Figures and tables must be put in the text in consecutive order. All figures must have a brief title accompanied with a short description.

Brief Profile

At the end of the manuscript, each author should provide a brief profile (less than 150 words), together with recent photographs (preferable less than 3 MB).

Corresponding Authors

We only permit one corresponding author per submission.

Conflict of Interest

Conflict of interest occurs when an author (or the author's institution) has personal or financial relationships that inappropriately influence the statements in the publication. Authors should ensure that publications will be written in an unbiased, ethical and responsible manner. The authors working on any sponsored engineering work or publications should declare such work under Conflict of Interest during submission.

Submission of Paper

Authors are required to submit via the IEM Online Journal Submission (OJS) through the following link:

https://iemjournal.com.my/index.php/iem/about/submissions.

Together with the manuscript, the corresponding author should enclose a cover letter containing the significance of the paper, the postal and email address, and telephone number for correspondence. In the cover letter, kindly provide four (4) names as reviewers. The selected reviewers must have the relevant knowledge and research experience. Provide the reviewers postal and email addresses and telephone numbers (if possible).

Further Enquiries

For further enquires, please contact Secretariat at the address shown below:

The Institution of Engineers, Malaysia
Bangunan Ingenieur, Lots 60 & 62,
Jalan 52/4 Peti Surat 223 (Jalan Sultan),
46720 Petaling Jaya, Selangor Darul Ehsan

Tel: 03–79684001/2 Fax: 03–79577678
E-mail: pub@iem.org.my or iemjournal@gmail.com


2021
ICOXA


IEM
The Institution of Engineers, Malaysia