The 26th International Annual Symposium
on Computational Science and Engineering

PROCEEDING

ANSCSE 26th

Modern Computational Approaches Towards Sustainable
and Decarbonized Society

20-22 JULY 2023

At A-One The Royal Cruise Hotel Pattaya,
Chonburi, Thailand

https://anscse26.wordpress.com   ANSCSE   anscse26@gmail.com

Welcome Message:

Dr. Supawadee Namuangruk

President of Computational Science and Engineering Association (CSEA), Thailand

E-mail: supawadee@nanotec.or.th

Dear Colleagues,

It is a great honor to extend a warm invitation for you to attend the ANSCSE26, the 26th International Annual Symposium on Computational Science and Engineering, to be held on July 20-22, 2023. This year the symposium is organized by the Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), Computational Science and Engineering Association (CSEA), and National Nanotechnology Center (NANOTEC). ANSCSE has always been one of the greatest gatherings of computational science and engineering researchers. After 26 years, we have seen many signs of progress and many exciting research being conducted in this area and the vital role of computational science in real social problems becomes clearer and clearer day by day.

One of the great spirits of ANSCSE is the live discussion among fellow international researchers on the advancement in all fields of computational science and engineering. After a few days of intense discussion on our research, the organizers kindly arranged an excursion to Nong Nooch Tropical Garden & Cultural Village, Pattaya. Finally, thank you in advance for sharing your research, thoughts, and ideas in ANSCSE26 and I look forward to meeting all of you.

Best wishes,

Dr. Supawadee Namuangruk

President of Computational Science and Engineering Association (CSEA), Thailand

Welcome Message:

Prof. Dr. Poom Kumam

ANSCSE 26 Chairman

E-mail: poom.kum@kmutt.ac.th

It is such a great honor for KMUTT to be hosting the 26th ANSCSE this year. We are very pleased that the events of such an exciting field of science are brought to our university. As computational method becomes increasingly more important in Science and Engineering, the need to collaborate and exchange ideas and experiences become more important as well. ANSCSE has a long history of being the major forum for computational scientists to communicate their ideas to their fellow scientists. Recently, the use of computational techniques has been extended to a broader domain of problems beyond science, such as the attempt to understand economy, human society, social networking. Especially, computational techniques are also utilized to hit global problems and goals, such as, renewable energy and pollutions, which leads the thematic area of this version of ANSCSE as the 'Modern Computational Approaches for Decarbonized Society'. This year, we return to have our first full onsite conference since the COVID-19 pandemic, with many interesting international keynote speakers, workshops, and also presentations by many exciting researchers. I am confident that with the experience of the organizing team from KMUTT and strong academic committees from CSEA, we will have a great and fruitful conference. Finally, thank you for coming and we look forward to seeing all the participants in Pattaya.

Best wishes,

Prof. Dr. Poom Kumam

ANSCSE 26 Chairman

Center of Excellence in Theoretical and Computational Science (TaCS-CoE),
Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), Thailand.

ANSCSE26 Chair and Co-Chair:

Poom Kumam (Chair) TaCS-CoE, King Mongkut's University of Technology Thonburi (KMUTT), Thailand

Thana Sutthibutpong (Co-Chair) King Mongkut's University of Technology Thonburi (KMUTT), Thailand

Steering Committees

Vudhichai Parasuk	Chulalongkorn University, Thailand
Supa Hannongbua	Kasetsart University, Thailand
Supawadee Namuangruk	NSTDA, Thailand

International Scientific Committees/Invited Speakers

Nicu Sebe	University of Trento, Italy
Enrique Zuazua	Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
Thorsten Dickhaus	University of Bremen, German
Yoel Je Cho	Gyeongsang National University, South Korea
Ovidiu Bagdasar	University of Derby, United Kingdom
Yasuteru Shigeta	University of Tsukuba, Japan
Deva Priyakumar	International Institute of Information Technology, India
Daniel Escudero	KU Leuven, Belgium
Momoji Kubo	Institute for Materials Research, Tohoku University, Japan
Akira Nakayama	The University of Tokyo, Japan
Norio Yoshida	Nagoya University, Japan
Hisashi Okumura	Institute for Molecular Science, Japan
Ryuhei Harada	University of Tsukuba, Japan
Vannajan Sanghiran Lee	University of Malaya, Malaysia
Antonio Liotta	University of Bozen-Bolzano, Italy
Talha Anwar	King Mongkut's University of Technology Thonburi, Thailand
Cathal Gurrin	Dublin City University, Ireland
Piyabut Burikam	Chulalongkorn University, Thailand
Ninnat Dangniam	The Institute for Fundamental Study, Thailand
Areeya Chantasri	Mahidol University, Thailand
Adisak Boonchun	Kasetsart University, Thailand
Annop Ektarawong	Kasetsart University, Thailand
Manaswee Suttipong	Chulalongkorn University, Thailand
Nuttapon Yodsin	Silpakorn University, Thailand
Anchalee Junkaew	NSTDA, Thailand
Jirasak Wong-ekkabut	Kasetsart University, Thailand
Charin Modchang	Mahidol University, Thailand
Treenut Saithong	King Mongkut's University of Technology Thonburi, Thailand
Teeraphan Laomettachit	King Mongkut's University of Technology Thonburi, Thailand
Panupong Mahalapbutr	Khon Kaen University, Thailand
Yuttasart Nitipaichit	National Telecom Public Company Limited, Thailand
Suthep Suantai	Chiang Mai University, Thailand
Booncharoen Sirinaovakul	King Mongkut's University of Technology Thonburi, Thailand
Supavadee Aramvith	Chulalongkorn University, Thailand

Scientific Chair Sessions

Session – MST: *Mathematics and Statistics*

Narin Petrot   Naresuan University, Thailand
Wiyada Kumam   Rajamangala University of Technology Thanyaburi, Thailand
Parin Chaipunya King Mongkut's University of Technology Thonburi, Thailand

Session – PHQ: *Computational Physics and Quantum Information Science*

Tanapat Deesuwan   King Mongkut's University of Technology Thonburi, Thailand
Sikarin Yoo-Kong   The Institute for Fundamental Study, Thailand
Pakpoom Reunchan Kasetsart University, Thailand

Session – CHE: *Computational Chemistry, Materials, and Nanotechnology*

Siriporn Jungsuttiwong Ubon Ratchathani University, Thailand
Nawee Kungwan   Chiang Mai University, Thailand

Session – BIO: *Computational Biology, Bioinformatics, Biochemistry, and Biophysics*

Thanyada Rungrotmongkol Chulalongkorn University, Thailand
Rungtiva Palangsuntikul   King Mongkut's University of Technology Thonburi, Thailand
Theerapong Puangmali   Khon Kaen University, Thailand

Session – HPC: *High Performance Computing, Computer Science, and Computer Engineering*

Supakit Prueksaaroon   Thammasat University, Thailand
Rardchawadee Silapunt King Mongkut's University of Technology Thonburi, Thailand
Phond Phunchongharn King Mongkut's University of Technology Thonburi, Thailand

Session – AIS: *Artificial Intelligence for Science and Engineering*

Rabian Wangkeeree   Naresuan University, Thailand
Nirattaya Khamsemanan Sirindhorn International Institute of Technology, Thailand
Cholwich Nattee   Sirindhorn International Institute of Technology, Thailand

Session – CFD: *Computational Fluid Dynamics*

Nopparat Pochai King Mongkut's Institute of Technology Ladkrabang, Thailand
Usa Humphries   King Mongkut's University of Technology Thonburi, Thailand

Session – IEEE: *Computational Engineering and Image Processing*

Phatiphat Thounthong   King Mongkut's University of Technology North Bangkok, Thailand
Thittaporn Ganokratanaa King Mongkut's University of Technology Thonburi, Thailand

Workshop: *Applications of Network Theory on Computational Biology and Biochemistry*

Thana Sutthibutpong   King Mongkut's University of Technology Thonburi, Thailand
Teeraphan Laomettachit King Mongkut's University of Technology Thonburi, Thailand

Teerasit Termsaithong   King Mongkut's University of Technology Thonburi, Thailand

TaCS-CoE Meeting and Workshop TRF (IEEE & Engineering)

Parin Chaipunya   King Mongkut's University of Technology Thonburi, Thailand
Songpon Sriwongsa   King Mongkut's University of Technology Thonburi, Thailand
Phatiphat Thounthong King Mongkut's University of Technology North Bangkok, Thailand

TABLE OF CONTENTS

Identifying Human Fingerprints by Using Optical Coherence Tomography Image with Deep Learning

Papawit Nongkhunsan[a], Panomsak Meemon[a], Ittipon Fongkaew[a, 1]

[a]*School of Physics, Institute of Science, Suranaree University of Technology, 111 University Avenue, Muang, Nakhon ratchasima, 30000, Thailand.*

Abstract

   Optical coherence tomography (OCT) has generated depth images of living tissue in various fields. In this research, we propose a novel approach to improve the security and performance of traditional fingerprint scanners using OCT technology for internal finger biometrics. Our system utilizes the internal structure of the fingertip, such as sweat glands, epidermis, and dermis, to construct a 3D finger model. We extract features from the internal fingertip image using pre-trained deep learning models, including Inceptionv3, VGG16, Xception, ResNet50, and a custom model. A voting method is employed to ensure the system's robustness. We collected a dataset of internal fingertips from twelve unique fingers, representing twelve users, with around 1,000 images for each finger. We split 60% of the data for training and the rest for validation. Our experimental results show that the proposed method achieves an accuracy and prediction rate of approximately 99%, as evaluated using the confusion matrix. These results suggest that deep learning-based security systems utilizing internal fingertip data may be able to replace traditional verification systems with higher security. However, the effectiveness of the proposed method must be further validated and tested in real-world scenarios. Additionally, future work may include integrating this technology with in-house OCT hardware for better practical application.

Keywords: Optical coherence tomography; Deep learning; Voting method

1. Introduction

   In recent years, biometric technology has gained popularity to protect personal information and maintain privacy in numerous businesses. Fingerprint identification is one of the most prevalent biometric techniques. Traditional fingerprint identification methods, however, are susceptible to cyberattacks, particularly those involving forged fingerprints. The vulnerabilities in the technology allow hackers to obtain unauthorized access to sensitive information [1]. Consequently, developing more sophisticated and secure fingerprint identification systems is essential. Conventional fingerprint technology captures external fingertip information using capacitive [2], optical reflection [3], or ultrasonic sensors to generate a template that is compared to the input data when a user tries to access the system [4]. This approach leaves the system susceptible to spoofing attacks by presentation attack instruments (PAIs) that can manipulate the surface of the finger [1]. To protect personal information in the digital age, it is necessary to enhance the security and accuracy of fingerprint identification systems in light of the growing use of biometric technology. Optical coherence tomography (OCT) technology has emerged as a promising solution to the security problems associated with conventional fingerprint identification systems. Similar to ultrasound imaging but with a higher resolution, OCT provides cross-sectional images of living tissues and biological substances using non-harmful near-infrared light [5]. The two most common OCT implementations are TD-OCT and SD-OCT, with the latter being faster and more efficient [6, 7]. Numerous fields, including dermatology [8, 9], ophthalmology [10], biology [11, 12], aquatic toxicology [13, 14], and material characterization [15, 16], have benefited from the application of OCT technology.

---

[1]* Corresponding author. Tel.: +66-4422-3705; fax: +66-4422-4185.
   E-mail address: ittipon@sut.ac.th

Recent research indicates that optical coherence tomography (OCT) technology can significantly improve the security and accuracy of fingerprint identification systems. Nevertheless, traditional methods have flaws, and fake fingerprints can circumvent the technology. The risk stems from the fact that conventional systems only account for the finger's surface, which is the area that presentation attack instruments (PAIs) can manipulate. PAIs are inexpensive and readily available substances such as glue, play-doh, gelatin, etc. These toys can be used to commit fraud in a traditional high-security system with an approximate 70% success rate [17]. Therefore, developing more sophisticated technologies for detecting and preventing such attacks is essential. The Optical Coherence Tomography (OCT) technology is more effective against attacks involving non-authentic fingerprints, with an actual detection rate (TRD) of 99.73 percent, Chugh et al. [18] reported that OCT could distinguish between spoofed and actual fingers.

In addition, Akbari et al. have investigated the recognition of fingerprints using OCT fingerprint images. Their research demonstrated that the algorithm used to recognize OCT fingerprints could be enhanced by automating fingerprint recognition [19]. Moola et al. have studied traditional fingerprints using OCT and reported that when OCT technology is integrated into the device, the system performs better than normal [20]. To take advantage of the potential of OCT technology, this research aims to develop more secure authentication and verification systems less vulnerable to cyberattacks. We will use SD-OCT to collect and analyze image data in order to identify unique fingerprint images with a 99% degree of accuracy using deep learning techniques and computer vision. We will also perform a classification algorithm on multiple classifications using Python programming. Using OCT images and deep learning techniques, we aim to improve the security and accuracy of fingerprint identification systems in the digital age and safeguard sensitive data.

2. **Materials and Methods**

2.1 OCT system and data collection

The SD-OCT utilized in this study had a central wavelength of 840 nm and was generated by a superluminescent light-emitting diode, as shown in Figure 1. (SLED). The wave delivers light to the beam splitter, which then separates the light into a reference beam and a sample beam. The reference beam was reflected along the same path as the beam splitter after being directed to a reference mirror. The sample beam is transmitted to the objective lens, which focuses on the sample. Additionally, the objective lens combines the light reflected from the sample's structure. An interference signal was produced when both beams reflected off a beam splitter. A specialized, high-speed spectrometer was utilized to detect the interference signal in the spectral domain. To receive the depth signal, the spectral interference was Fourier transformed. By laterally scanning the focused beam with a Galvanometer mirror, we could acquire a cross-sectional image of 3D data. Each cross-section has a 1000x1000 pixel resolution. The capture area is approximately 4x4 mm, and the sampling distance between each pixel is about 4 mm. The imaging speed is around 25 frames per second.
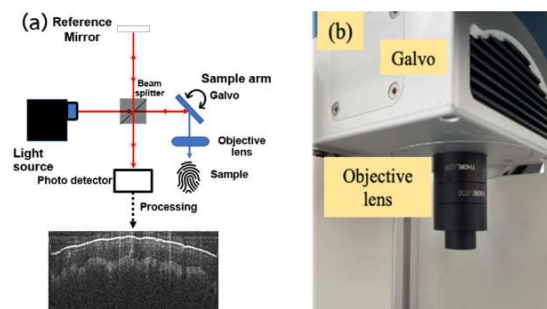

Figure 1. (a) A diagram of the SD-OCT and (b) a sample head for imaging.

2.2 Capturing and collecting internal data from fingertip

We captured internal fingerprint images using in-house OCT hardware with an imaging resolution of 1000x1000 pixels, corresponding to a 4 mm x 4 mm fingerprint area. Figure 2 shows an example of a B-scan (slice) image. Images reveal that the internal structure of the finger consists primarily of three layers: 1) epidermis, 2) dermis, and 3) subcutaneous. The surface finger or stratum corneum is the uppermost layer of the epidermis, which contains sweat glands [21]. The viable epidermis, which represents the innermost layer of the fingertip, is the subsequent layer. In the dermis layer, the papillary and reticular dermis were visible. As a result, when a user sustains a minor injury, all depth information could serve as a mother template and contain sufficient data for security, even if the finger surface is contaminated because internal information also retains the ridges and valleys. Sweat glands and pores could improve the accuracy of fingerprint identification [22]; this application of OCT properties is a breakthrough and superior to conventional fingerprints, as the latter cannot reach this layer.
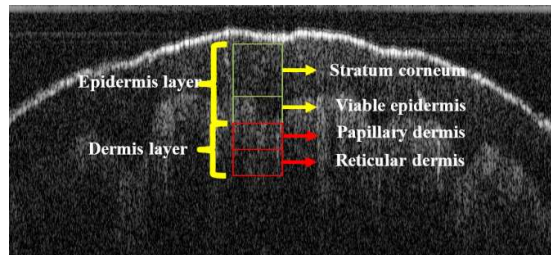
Figure 2. B-scan image with the dept information of fingertip.

According to the aforementioned images, the captured images demonstrated promising performance at multiple depths. Consequently, images are commonly referred to as layer images; this made the oct-captured technique unique because they contain internal information about the finger that should result in a more secure system. As a result, we collected 30 unique fingerprints from three SD-OCT participants; due to the slow scanning speed, each finger must remain still for approximately one minute. From the 3D OCT volume data, we will then extract the 2D fingerprint image. The OCT scanning procedure involves projecting a light beam onto the tissue and measuring the reflected light to generate an image. In the case of internal fingerprint imaging, light is projected onto the fingertip, and the reflected light captures the finger's internal structure, including the sweat glands, epidermis, and dermis. The resulting OCT image can be utilized to extract fingerprint identification features, providing a potentially more secure alternative to conventional fingerprint scanners.

2.3 Image augmentation

Image augmentation is used to randomly adjust the dataset, thereby increasing the dataset's diversity for a more accurate model and mitigating the overfitting issue. The outcome of the prediction should be better. Image processing enlarges a small dataset; examples of image processing include random cropping, adjusting contrast, rotation, and noise adjustment. The effect of image enhancement is enhancing the performance of the model [23]. An example of the image being enlarged is shown in Figure 3.

Figure 3 Image processing which A)raw data, B) contrast adjust with noise filter(median filter), C) and D) Heat mapping

Figure 3 depicts the image processing of an OCT fingerprint, where A is the raw data, and B is the processed image with noise (median filter) and contrast adjustments. C and D are heat maps used to comprehend the information on the various sides and how networks will extract features from the images. As c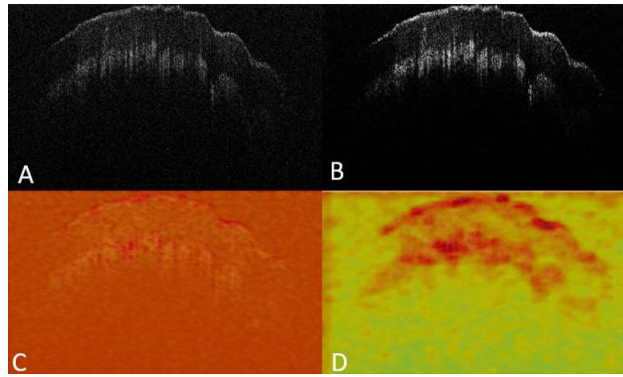an be seen, image processing would provide significantly more information about internal fingertip data. Figures 3C and 3D show that the interior surface has a high density; we could use this region as a biomarker for fingerprints [24]. With all the interior raw and processed data details, it is possible to predict what computers and convolutional neural networks (CNNs) will look for in this data. After analyzing the OCT-obtained internal fingerprint images, we observed increased information density from the surface to the interior.

Deep learning techniques can improve the accuracy of fingerprint recognition by incorporating this information. Figure 3A depicts the unprocessed OCT image of the fingerprint. It is difficult to extract useful characteristics from it. As demonstrated in Figure 3B, we applied a median filter to the image to reduce noise and extract more distinct features. Using the OCT data, we also generated a heat map of the fingerprint, as shown in Figures 3C and 3D. The heat map depicts areas of high and low image intensity. We can observe that high-density regions are present from the fingerprint's surface to its interior. This data confirms that it can be used to create a more precise and secure fingerprint identification system. We use heat mapping to ensure that OCT fingerprint images contain sufficient internal information for identification. Only The images were processed with a median filter (3B), and our CNN was fed the raw data (3A). High levels of noise and interference in the raw OCT images can negatively impact the performance of the CNN. To solve this issue, we used a median filter to reduce noise and preserve the fine details of the fingerprint valleys, ridges, and other information.

2.4 Deep Learning and Classification Model

Our deep learning strategy utilized convolutional neural networks in this work. The dataset was split into two sets: a training set and a validation set, with 60% of the data assigned to the training set and 40% to the validation set. For the multiclass labelling of 12 fingers, one-hot encoding was used to refer to 12 unique users. We downsized the images and converted them to arrays to conserve computational resources. The initial layer of our model was a convolutional layer that extracted image features. Next, the data was sent to the max pooling layer, which reduced the data's size while retaining only the most essential details or values. We utilized the dropout layer to prevent overfitting [25]. The activation function rectified linear unit (ReLu), was essential to the design of our neural network architecture.

2.4.1   Architecture of model

In this work, the model was created from five different architectures. In the first design, we will design our custom network architecture for predicting the fingerprint. In the other models, we will use the complex architecture from ImageNet that InceptionV3, Restnet50, VGG16, and Xception will use for

---

training the model, respectively [26, 27]. Then, we will ensure user confidence in the system by combining all three models. For the design parameter depicted in Figure 4, a convolutional layer with 128 features and a kernel size of 3x3 was extracted. Therefore, the ReLu activation function was utilized in this layer [28]. The pooling layer utilized max-pooling to collect the 2x2 essential data. A learning rate of 0.0001, a batch size of 128, and 1000 epochs were used during training. In this work, the models were compiled using the Adam optimizer, a batch size of 64, and 500 training iterations. After each training cycle, the learning rate was decreased by a factor of 10 from 0.1 to 0.000001 in order to promote convergence. However, due to out-of-memory (OOM) issues, it could not further reduce the learning rate. Due to OOM constraints, the maximum batch size was limited to 64 units. Due to OOM issues, the input image size was also restricted to 128x128x3. It was impossible to reduce the model's size any further because doing so would have necessitated the removal of essential components and negatively impacted its performance.
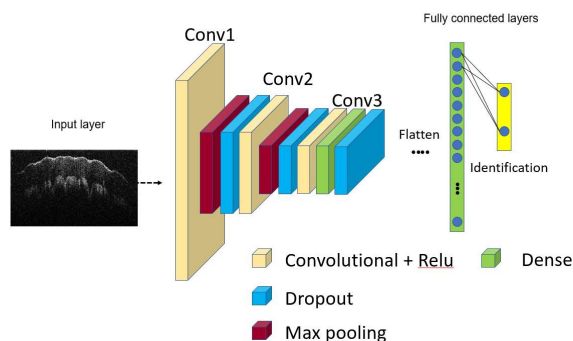

Figure 4 Neural network architecture diagrams of a model

Google developed the InceptionV3 architecture as the third iteration of its deep learning convolutional neural network architecture for image classification and identification applications. InceptionV3 was used to train around one thousand classes using an ImageNet picture dataset containing over one million images. This InceptionV3 has a high classification efficiency; hence it was chosen for fingerprint training [27]. The following model selected is the Xception model. The Xception architecture is based on the Inception architecture, which extracts information from input images using a combination of 1x1 and 3x3 convolutional filters. Xception, however, replaces the 3x3 filters with depthwise separable convolutions, which involve two steps: first, applying a single convolutional filter to each input channel (i.e., the "depthwise" part), and then concatenating the results and applying a 1x1 convolutional filter (i.e., the "separable" part) [26]. Next, the ResNet50 model, a deep residual network, was developed and trained on the ImageNet dataset, which consists of millions of photographs of various objects. The massive dataset from which our pre-trained model has acquired a wealth of features can serve as the basis for our fingerprint classification problem. The pre-trained ResNet50 model will be fine-tuned by training it on a smaller dataset of OCT images of human fingerprints. This fine-tuning procedure will update the pre-trained model's weights [29]. Significant image classification dataset ImageNet was previously used to train the deep convolutional neural network model with the VGG16 architecture. Transfer learning modifies a once-trained model for a new task by adding data, in this case, OCT pictures, into the weights of its layers. Several fully connected layers follow several stacks of convolution and pooling layers in VGG16's architecture. The convolution layers collect spatial and channel-wise data, whereas the pooling layers lower the spatial dimension of the feature maps while preserving the essential data. The layers with complete connectivity create predictions depending on the input image. The output of the VGG16 model may be connected to a fully connected layer containing several neurons equal to the number of classes in the target task, in this case, the number of classes for various fingerprints. The class of an input image is predicted based on the output of the fully linked layer. Therefore, rather than starting from scratch, the modified VGG16 model can be trained end-to-end using the OCT data, thereby increasing its accuracy for the fingerprint detection task [30].

2.5 Ensemble model

We proposed the ensemble model for more precise or reliable forecasting. Using a weighted average or some other mathematical function, an ensemble model combines the aforementioned distinct models to provide a single output. Typically, the ensemble model is utilized in applications where the objective requires the highest achievable level with limited computational resources and data. The ensemble model we employed is a voting method classifier, a voting ensemble in which each model makes a prediction, and a majority vote determines the final prediction. For instance, if five models forecast A, B, and A, the final prediction would be A. In our scenario, there are twelve classes for each finger. Figure 5 depicts the diagram of the voting ensemble model. We proposed the ensemble model for more precise or reliable forecasting. Using a weighted average or some other mathematical function, an ensemble model combines the aforementioned distinct models to provide a single output. Typically, the ensemble model is utilized in applications where the objective requires the highest achievable level with limited computational resources and data. The ensemble model we employed is a voting method classifier, a voting ensemble, in which each model makes a prediction, and a majority vote determines the final prediction. For instance, if five models forecast A, B, and A, the final prediction would be A. In our scenario, there are twelve classes for each finger. Figure 5 depicts the diagram of the voting ensemble model.

Figure 5 diagram of voting ensemble method

	Parameters	Initial input size	Initial learning rate	Batch-size	Epoch	Optimizer
Networks						
Inception-V3	128x128x3	0.00001	64	500	Adam
Xception	128x128x3	0.00001	64	500	Adam
VGG-16	128x128x3	0.00001	64	500	Adam
Resnet-50	128x128x3	0.00001	64	500	Adam
Custom model	128x128x3	0.00001	64	500	Adam

Table 1 The parameter settings for four TL networks and a custom network.

The parameters utilized by the networks are detailed in Table 1. Set the initial input size and batch size to optimize the use of computational resources. The learning rate is then modified until the learning curves converge, ensuring the model can generalize well to the current dataset.

The model weights were modified during training using backpropagation to minimize the categorical cross-entropy loss function. The models were created utilizing the Adam optimizer, 64-person batches, and 500 training iterations. After each training cycle, the learning rate was reduced by 10 from 0.1% to 0.000001% to promote convergence. It was unable to cut the learning rate further due to out-of-memory (OOM) difficulties. Due to OOM restrictions, the batch size could not exceed 64, and the maximum image size was 128x128x3. It was impossible to reduce the size any further because doing so would have resulted in losing important components and diminished the model's performance.

2.6 Evaluation model

Using model evaluation, we can determine how well the model can solve our problem. In addition, various metrics can be used to evaluate the performance of a DL model. These metrics depend on the task the model was designed for and can include accuracy, precision, recall, and F1 score. Then we will use it to calculate and visualize the model's performance through a learning curve and confusion matrix [31].

3. **Results**

3.1 Learning curve

This work uses images of twelve internal fingertips of three people that capture from OCT scans to identify the person using deep learning techniques. We create a dataset for around 12,108 impressions. Then we randomly split 60 percent for the training set and 40 percent for the validating set for each model. The learning curve of each model will show in Figure 6 below.
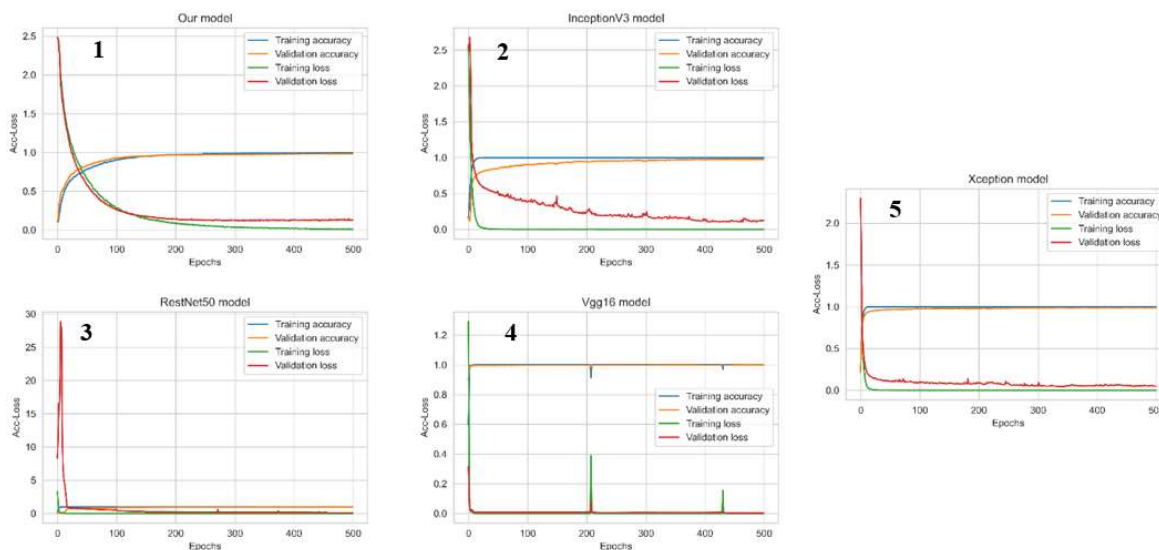

Figure 6 the learning curves

The learning curve is important when assessing a machine learning model's performance. It gives a broad picture of the link between the model's accuracy and loss and the total number of training epochs. The training accuracy and loss are metrics used to evaluate how well the model performs during training on the training dataset. The training accuracy is the percentage of correctly classified images in the training dataset, while the training loss measures how well the model is able to predict the correct label for each image in the training dataset. The model's accuracy generally grows while the loss decreases as the number of epochs rises. This pattern might not always hold, and the learning curve can be used to spot problems with the model, including overfitting, underfitting or lack of convergence. To optimize the performance of the model and adjust its hyperparameters, it is crucial to analyze the learning curve.

Figure 6 displays the result of training the data. First is the custom model result (Figure 6.1); training and validation accuracy started at low values and increased as the number of epochs increased. The training and validation loss, on the other hand, started high and decreased as the number of epochs increased. The initial high loss could be due to the random initialization of the model's weights. The weights were optimized as the model continued to train, resulting in a decreased loss. The validation accuracy also increased along with the training accuracy, indicating that the model was not overfitting to the training data. Additionally, the validation loss started higher than the training loss and remained slightly higher throughout training, which could be because the validation data was not used for weight optimization.
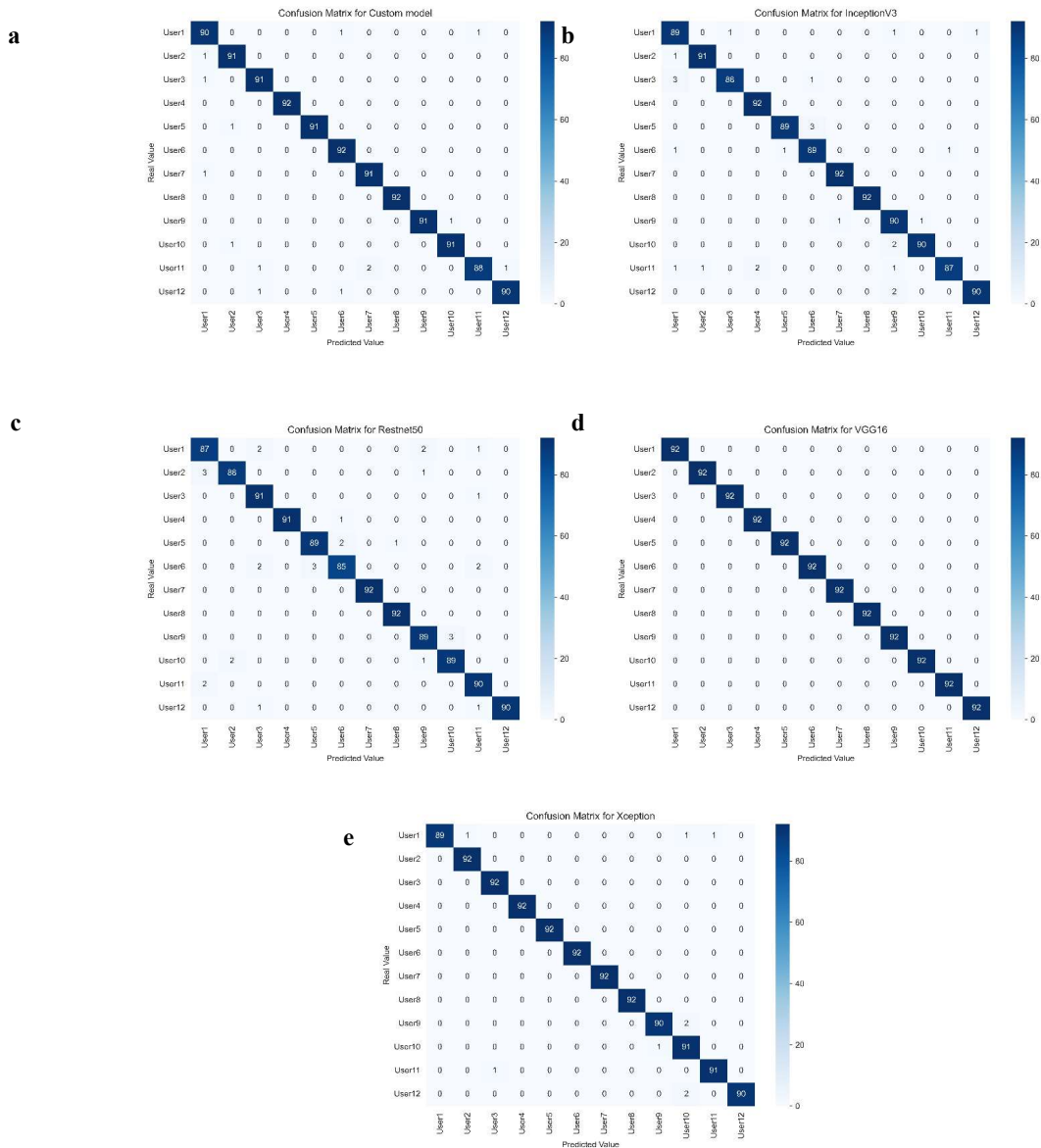
Based on the learning curve for the InceptionV3 model in (Figure 6.2), we observed that the accuracy and validation accuracy started at a higher initial value compared to the first model (Figure 6.1). The validation accuracy for this model was relatively consistent and had a slight dip towards the end. The training and validation loss showed a similar pattern as our model, where it started high and converged towards zero with higher epoch values. However, the final training loss was lower in this model, which indicates that the model could fit the data more accurately. Overall, the learning curve for the second model showed that the model was able to perform well on the data and achieve high accuracy with low loss values. Figure 6.3 illustrates the learning curve for the third model, ResNet50, which began with low training and validation accuracy and significant training and validation loss. The model quickly incorporated additional data, and by epoch 8, it had an accuracy of 0.99 and a loss of 0.034. The subsequent rapid increase in training and validation losses implies that the model may have begun to overfit. Despite this, the model completed the training phase with a high degree of accuracy. With a final score of 0.97 and a validation loss of 0.10, validation accuracy maintained a high level of consistency. Despite some evidence of overfitting, the ResNet50 model performed well in classifying fingerprint images accurately. Figure 6.4 depicts the learning curve for the VGG16 model. From the beginning epochs, both the training and validation accuracy grow dramatically while the loss decreases. At epoch 207, the model reaches its highest level of accuracy, with a training accuracy of 0.91, a loss of 0.39, a validation accuracy of 0.98, and a validation loss of 0.07. After the peak, the accuracy does not increase appreciably, however, the validation loss grows substantially at later epochs, indicating overfitting. The model achieved great accuracy on both the training and validation sets, with a final validation accuracy of 0.99 and a final training and validation loss of 0.002. In Figure 6.5, the learning curve for the Xception model shows a slow start with low accuracy and high loss in the beginning. However, as the number of epochs increases, the model converges, and the training and validation accuracy improves significantly while the loss decreases. At the end of the training, the model achieves high accuracy on both the training and validation sets with a shallow loss, indicating that the model performs well and generalizes to unseen data.

Overall, the learning curves of the five models show a similar pattern of convergence where both the training and validation losses decrease with increasing epochs. In contrast, the training and validation accuracies increase. However, the specific shapes of the curves vary between the different models, likely due to differences in their architectures. The ResNet50 and Xception models start with relatively low accuracies and high losses but eventually achieve high accuracies and low losses after many training epochs. On the other hand, the other models start with relatively high accuracies and lower losses but still improve with more training. Overall, all of the models achieve very high validation accuracies and low losses, which indicates their effectiveness for classification tasks.

3.2 Confusion matrix

The confusion matrix is a table that contrasts expected and actual classes to demonstrate the performance of a deep-learning model. It is a tool that evaluates the accuracy of a model and identifies improvement opportunities. When describing model performance in academic articles, particularly when dealing with multiclass classification challenges, the confusion matrix is useful. The five confusion matrices provided correspond to the performance of five models trained on a 12-class classification issue. Each matrix is a 12 by 12 table that indicates the number of accurate and inaccurate predictions made by the model for each class. In each matrix, the diagonal values reflect the number of accurate forecasts, while the off-diagonal values represent the number of inaccurate guesses. The confusion matrix for each of the five models is displayed in Table 2(a-e).

Table 2 Confusion matrix table that shows the performance of the 5 models a) Individual model, b) InceptionV3 model, c) ResNet50 model, d) Vgg16, and e) Xception model.

a — Confusion Matrix for Custom model

Real \ Pred	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	User11	User12
User1	90	0	0	0	0	1	0	0	0	0	1	0
User2	1	91	0	0	0	0	0	0	0	0	0	0
User3	1	0	91	0	0	0	0	0	0	0	0	0
User4	0	0	0	92	0	0	0	0	0	0	0	0
User5	0	1	0	0	91	0	0	0	0	0	0	0
User6	0	0	0	0	0	92	0	0	0	0	0	0
User7	1	0	0	0	0	0	91	0	0	0	0	0
User8	0	0	0	0	0	0	0	92	0	0	0	0
User9	0	0	0	0	0	0	0	0	91	1	0	0
User10	0	1	0	0	0	0	0	0	0	91	0	0
User11	0	0	1	0	0	0	2	0	0	0	88	1
User12	0	0	1	0	0	0	1	0	0	0	0	90

b — Confusion Matrix for InceptionV3

Real \ Pred	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	User11	User12
User1	89	0	1	0	0	0	0	0	1	0	0	1
User2	1	91	0	0	0	0	0	0	0	0	0	0
User3	3	0	86	0	0	1	0	0	0	0	0	0
User4	0	0	0	92	0	0	0	0	0	0	0	0
User5	0	0	0	0	89	3	0	0	0	0	0	0
User6	1	0	0	0	1	88	0	0	0	0	1	0
User7	0	0	0	0	0	0	92	0	0	0	0	0
User8	0	0	0	0	0	0	0	92	0	0	0	0
User9	0	0	0	0	0	0	1	0	90	1	0	0
User10	0	0	0	0	0	0	0	0	2	90	0	0
User11	1	1	0	2	0	0	0	0	1	0	87	0
User12	0	0	0	0	0	0	0	0	2	0	0	90

c — Confusion Matrix for Restnet50

Real \ Pred	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	User11	User12
User1	87	0	2	0	0	0	0	0	2	0	1	0
User2	3	86	0	0	0	0	0	0	1	0	0	0
User3	0	0	91	0	0	0	0	0	0	0	1	0
User4	0	0	0	91	0	1	0	0	0	0	0	0
User5	0	0	0	0	89	2	0	1	0	0	0	0
User6	0	0	2	0	3	85	0	0	0	0	2	0
User7	0	0	0	0	0	0	92	0	0	0	0	0
User8	0	0	0	0	0	0	0	92	0	0	0	0
User9	0	0	0	0	0	0	0	0	89	3	0	0
User10	0	2	0	0	0	0	0	0	1	89	0	0
User11	2	0	0	0	0	0	0	0	0	0	90	0
User12	0	0	1	0	0	0	0	0	0	0	1	90

d — Confusion Matrix for VGG16

Real \ Pred	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	User11	User12
User1	92	0	0	0	0	0	0	0	0	0	0	0
User2	0	92	0	0	0	0	0	0	0	0	0	0
User3	0	0	92	0	0	0	0	0	0	0	0	0
User4	0	0	0	92	0	0	0	0	0	0	0	0
User5	0	0	0	0	92	0	0	0	0	0	0	0
User6	0	0	0	0	0	92	0	0	0	0	0	0
User7	0	0	0	0	0	0	92	0	0	0	0	0
User8	0	0	0	0	0	0	0	92	0	0	0	0
User9	0	0	0	0	0	0	0	0	92	0	0	0
User10	0	0	0	0	0	0	0	0	0	92	0	0
User11	0	0	0	0	0	0	0	0	0	0	92	0
User12	0	0	0	0	0	0	0	0	0	0	0	92

e — Confusion Matrix for Xception

Real \ Pred	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	User11	User12
User1	89	1	0	0	0	0	0	0	0	1	1	0
User2	0	92	0	0	0	0	0	0	0	0	0	0
User3	0	0	92	0	0	0	0	0	0	0	0	0
User4	0	0	0	92	0	0	0	0	0	0	0	0
User5	0	0	0	0	92	0	0	0	0	0	0	0
User6	0	0	0	0	0	92	0	0	0	0	0	0
User7	0	0	0	0	0	0	92	0	0	0	0	0
User8	0	0	0	0	0	0	0	92	0	0	0	0
User9	0	0	0	0	0	0	0	0	90	2	0	0
User10	0	0	0	0	0	0	0	0	1	91	0	0
User11	0	0	1	0	0	0	0	0	0	0	91	0
User12	0	0	0	0	0	0	0	0	0	2	0	90

From Table 2a, Custom model: The large number of accurate predictions (90 or more) for the majority of classes demonstrates the custom model's high overall accuracy. Some things could be improved in the predictions, particularly for users 1, 6, and 11, where the model misclassified one or two data. For user 4, the model has the highest accuracy, with every prediction coming true. Most of the InceptionV3 (Table 2b) model's predictions were accurate, showing a high level of overall accuracy. Each of users 1, 3, 5, and 11 has a number of erroneous predictions, indicating that the model has difficulty with these people. Users 4 and 7 have the most accurate model, with all accurate forecasts. For Table 2c ResNet50 model, the majority of predictions were accurate, suggesting a good level of general accuracy. The model has difficulty with users 1, 3, 5, and 10, as each of these classes has some incorrect predictions. Users 4, 7, and 8 have the most accurate model, with all accurate forecasts. In addition, Table 2d demonstrates that the VGG16 model has a high overall accuracy, with the majority of forecasts being true. The algorithm has difficulty with users 1, 5, and 11, as each of these categories has several erroneous predictions. For user 4, the model's accuracy is the highest, since all predictions are accurate. The Xception model has the highest overall accuracy for Table 2e and produces accurate predictions for the majority of courses. There are a few mistakes in the forecasts for users 1, 2, 3, and 11. The algorithm misclassified one or two data points as belonging to a different user. For user 4, the model's accuracy is the highest, since all predictions are accurate.

The confusion matrix, in short, offers a helpful summary of each model's performance, showing each strategy's advantages and disadvantages—the result after testing five models with unseen 1104 images. The overall classification performance can be enhanced using these results to inform future model or data preprocessing adjustments. In summary, precision, recall, and F1 score are evaluation metrics calculated from a confusion matrix and used to assess a classifier's performance, illustrated in Table 3.

Table 3 Classifier report table.

Network	Accuracy	Precision	Recall	F1-score	Support
Inception-V3	97.59	97.65	97.61	97.62	1104
Xception	99.13	99.2	99.18	99.19	1104
VGG-16	100	100	100	100	1104
Restnet-50	97.1	97.18	97.1	97.12	1104
Custom model	98.65	98.66	98.69	98.67	1104
Voting method	100	100	100	100	1104

Table 3 displays the performance result of four different pre-trained models, a custom model, and a voting method. The metrics include accuracy, precision, recall, and F1-score. The models achieved high levels of accuracy, ranging from 97.1% to 100%. The VGG-16 model achieved 100% accuracy, correctly classifying all the instances in the test set. The Xception model achieved the highest accuracy of 99.13%, followed by the custom model with an accuracy of 98.65%. In terms of precision, recall, and F1-score, all models performed well, with scores ranging from 97.18% to 100%. The VGG-16 model achieved perfect scores of 100% for all metrics, while the custom model achieved the highest F1 score of 98.67%.

The classification report shows that all models, including the custom model, performed well in classifying the test set. However, it's crucial to remember that measurements like accuracy alone can be deceptive, especially when working with unbalanced datasets. That's why it's essential to consider other metrics, such as AUC and ROC curves, to better understand the overall model performance. In this regard, the pre-trained deep learning models and the custom model achieved high AUC scores, with the voting method combining their strengths to achieve perfect scores in all metrics. Therefore, using multiple models and evaluation metrics, including AUC and ROC, can lead to a more robust and reliable classification performance below in Figure 7(a-f).
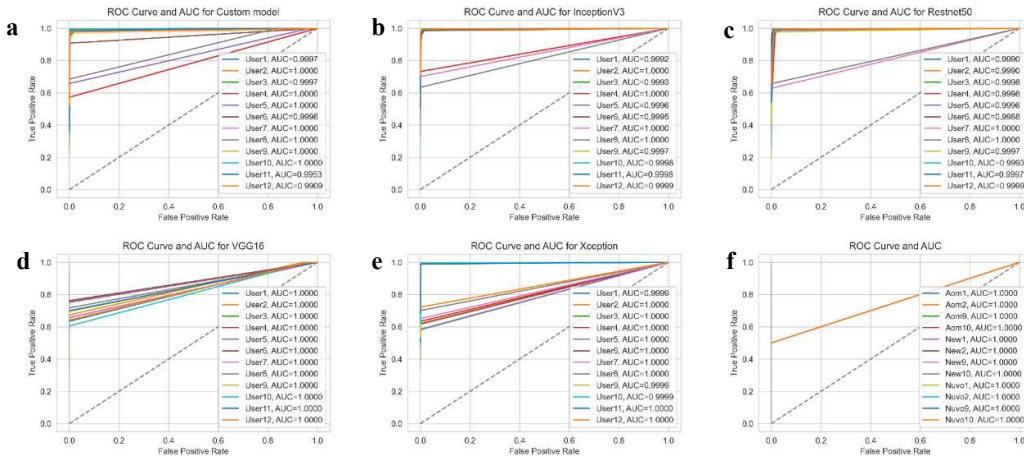
Figure 7 AUC and ROC curve that shows the performance of the 5 models a) Individual model, b) InceptionV3 model, c) RestNet50 model, d) Vgg16, e) Xception model, and f) voting method.

Specifically, our designed model (Figure 7a) for this purpose: The AUC for this model is 0.9953, showing that it distinguishes between positive and negative classifications adequately. The ROC curve illustrates that when the false positive rate (FPR) increases, the true positive rate (TPR) increases rapidly, showing that the model has high sensitivity and can identify real positives at a low cost of false positives. When the TPR reaches 1, the curve flattens, indicating that the model's classification performance at that level is optimal. In Figure 7b, the AUC for the InceptionV3 model is 0.9998, which is greater than the AUC for the custom model and demonstrates that it performs extraordinarily well at differentiating positive and negative classes. The ROC curve illustrates that while the FPR grows, the TPR climbs rapidly, confirming the model's high sensitivity and ability to recognize true positives with a low rate of false positives. When the TPR reaches 1, the curve flattens, indicating that the model's classification performance at that level is optimal, indicating that the model can perform perfectly for classification at that level. Figure 7c demonstrates that the AUC for ResNet50, which is also quite high at 0.9997, distinguishes between positive and negative classifications. The ROC curve illustrates that while the FPR increases, the TPR increases rapidly, suggesting the model's high sensitivity and capacity to recognize true positives with a comparatively low rate of false positives. When the TPR reaches 1, the curve flattens. The models for VGG16 and Xception (Figures 7d and 7e, respectively) exhibit an AUC of 1, the maximum attainable value, demonstrating remarkable ability in distinguishing between positive and negative classifications. The ROC curve illustrates the model's capacity to achieve optimum classification performance, with a TPR of 1 and an FPR of nearly 0, showing that the model has a high degree of sensitivity and specificity. In the final case voting technique (Figure 7f), the AUC is 1.0000, which is the maximum attainable value and indicates that the model distinguishes between positive and negative classifications perfectly. The high TPR values for various thresholds indicate that the model has a high sensitivity, i.e., it properly identifies a substantial number of positive cases. The low FPR values also indicate that the model has a high specificity, which indicates that it properly identifies a substantial fraction of negative situations. The outcomes indicated that the voting mechanism is highly effective for classifying data.
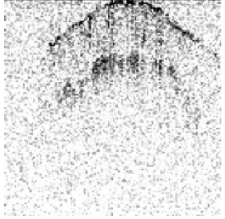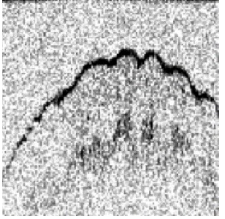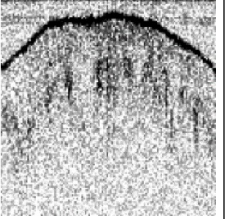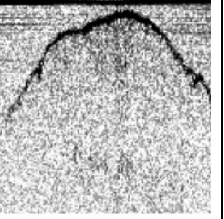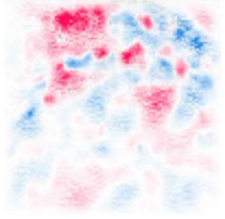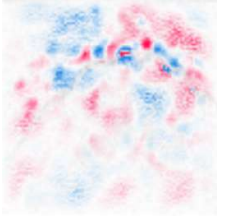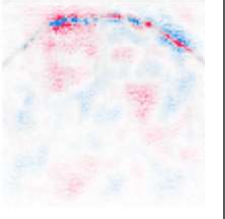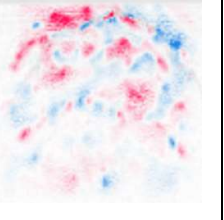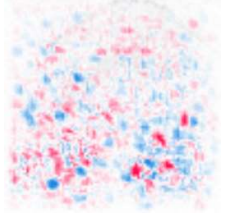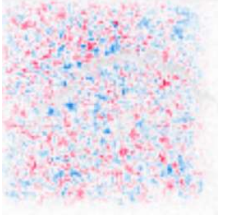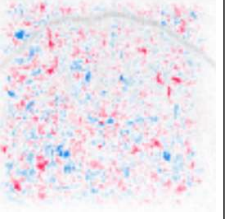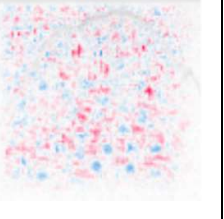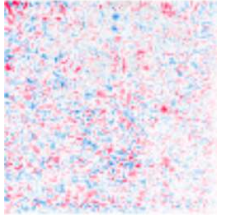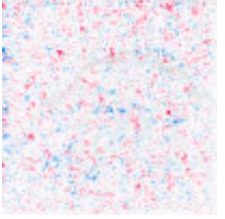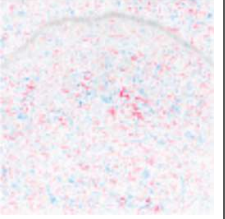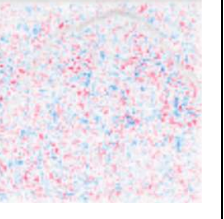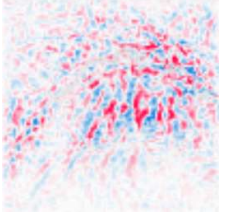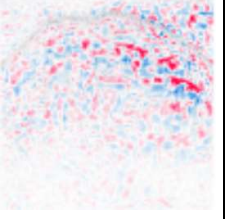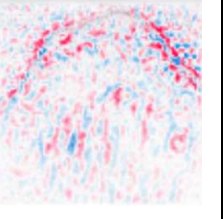
It is clear from examining the outcomes of the many models and the voting process that each model exhibits distinct performance levels, as evidenced by their AUC values and the forms of their individual ROC curves. The distinct qualities and advantages of each model are highlighted by the di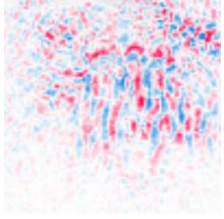sparities in the AUC values and ROC curve shapes, which may be traced to the particular behaviour of each classifier for the given task. It is important to consider not only the AUC values and ROC curves but also other performance metrics such as precision, recall, and F1-score, to gain a comprehensive understanding of

each classifier's performance. Additionally, analysing the confusion matrix can provide valuable insights into the model's overall performance and potential areas for improvement.

3.3 Model interpretation

Due to the growing use of sophisticated AI models and algorithms in crucial decision-making procedures, Explainable Artificial Intelligence (XAI) has attracted much interest in recent years [32, 33]. We use shap known as SHapley Additive exPlanations (SHAP) [34]. SHAP aims to provide understandable explanations for individual predictions made by any machine learning model. The table below will show the explainable model by random image for each model. The result of the model interpretation will show in Table 4.

Table 4 Explainable model table shows the feature selection

Original image				
Custom model				
InceptionV3				
ResNet50				
VGG16				
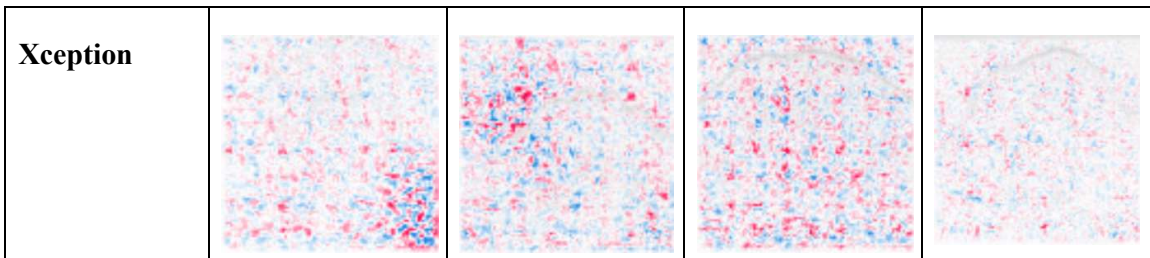
Xception	

In this study, we analysed the trustworthiness and interpretability of five distinct models: 1) a custom model, 2) InceptionV3, 3) VGG16, 4) Xception, and 5) ResNet50. We used the SHAP method to determine the significance of features and model explanations. It provided valuable insight into the decisions made by these models when presented with random, unseen data. SHAP generates colour-coded descriptions, where each colour represents the impact of a particular feature on the model's prediction. Positive values (red) indicate that the presence of the proportional feature increases the prediction probability, whereas negative values (blue) indicate that the presence of the proportional feature decreases the prediction probability. The impact on the prediction is proportional to the magnitude of the colour's intensity. The following table provides an exhaustive explanation of the procedure for selecting features. We consider four randomly selected images from the unseen dataset to illustrate how each model interprets the input data differently.

Using the SHAP method, the following analysis examines the feature selection procedure for five distinct models: models - Custom, InceptionV3, ResNet50, VGG16, and Xception. We evaluate four randomly selected images from an unseen dataset to illustrate how each model interprets the input data differently. The Custom model focuses primarily on the surface area of the fingertip, whereas the InceptionV3 and ResNet50 models exhibit a wider distribution of influential features throughout the images. In contrast, the VGG16 model focuses on particular, important image features, such as the fingerprint surface area, sweat gland areas, and sweat pores. Lastly, the Xception model exhibits a red cluster distribution similar to that of the InceptionV3 and ResNet50 models. By gaining an understanding of these feature selection processes, we can gain valuable insights into the decision-making mechanisms of each model, which may inform the development of more effective models in future research. By analyzing and comparing the interpretability results of these five models in conjunction with traditional evaluation metrics such as accuracy, precision, recall, and F1-score, as well as domain-specific metrics pertinent to the problem at hand, we can gain a deeper understanding of their behaviour. This holistic approach will ultimately aid in guiding the selection of the most suitable model for a given application and enhancing model performance through the engineering of targeted features.

4. Discussion and conclusion

In this study, a voting ensemble deep learning model was developed for fingerprint classification using optical coherence tomography (OCT) images. The model comprised five models, including an individual model designed by the authors, an InceptionV3 model, a Restnet50 model, a VGG16 model and an Xception model. The model achieved strong performance, as evidenced by the high AUC values on the ROC curves for each class and model. The model achieved an overall AUC of 1 on the validation set and a holdout test set, demonstrating its ability to generalize to unseen data. Using OCT images as input for the ensemble model allowed for capturing high-resolution images of fingerprints, contributing to the model's good performance. The model's performance was further visualized through ROC curves, which showed that the model could effectively distinguish between the different classes of internal fingerprints.

The results of this study indicate that the voting ensemble model based on a convolutional neural network utilizing OCT images is a promising approach for fingerprint classification tasks. However, there is still room for performance enhancement with this model. The model's generalization ability and overall performance may be enhanced by additional hyperparameter tuning and the use of a larger and more diverse training dataset. Using the AUC values on the ROC curves can be a useful metric for assessing the

performance of the model in such tasks. Observed differences in AUC values and ROC curve shapes between models highlight the importance of selecting the optimal model for a given task while also considering other performance metrics and classifier properties. As demonstrated by the exceptional AUC value of 1.00 achieved in this instance, it is possible to enhance classification performance by combining the strengths of multiple models using the voting method. This study contributes to the development of effective and trustworthy fingerprint classification methods, which may have applications in forensics and biometric identification.

Acknowledgements

References

1. Biggio, B., et al., Security evaluation of biometric authentication systems under real spoofing attacks. IET Biometrics, 2012. 1(1).
2. Lee, J., et al., A 600-dpi capacitive fingerprint sensor chip and image-synthesis technique. IEEE J. Solid State Circuits, 1999. 34: p. 469-475.
3. Soifer, V., et al., Optical–digital methods of fingerprint identification. Optics and Lasers in Engineering, 1998. 29(4): p. 351-359.
4. Lu, Y., et al., 11.2 3D ultrasonic fingerprint sensor-on-a-chip. 2016.
5. Huang, D., et al., Optical coherence tomography. Science, 1991. 254(5035): p. 1178-81.
6. Fercher, A., et al., Optical Coherence Tomography—Principles and Applications. Rep. Prog. Phys, 2003. 66.
7. Cense, B., et al., Ultrahigh-resolution high-speed retinal imaging using spectral-domain optical coherence tomography. Optics Express, 2004. 12(11): p. 2435-2447.
8. Lee, K.-S., et al., Cellular resolution optical coherence microscopy with high acquisition speed for in-vivo human skin volumetric imaging. Optics Letters, 2011. 36(12): p. 2221-2223.
9. Rolland, J.P., et al., Gabor Domain Optical Coherence Microscopy of Human Skin, in Advances in Dermatological Sciences. 2014, The Royal Society of Chemistry. p. 37-52.
10. Wojtkowski, M., et al., Three-dimensional retinal imaging with high-speed ultrahigh-resolution optical coherence tomography. Ophthalmology, 2005. 112(10): p. 1734-46.
11. Davis, A.M., et al., In vivo spectral domain optical coherence tomography volumetric imaging and spectral Doppler velocimetry of early stage embryonic chicken heart development. J Opt Soc Am A Opt Image Sci Vis, 2008. 25(12): p. 3134-43.
12. Tearney, G.J., et al., Rapid acquisition of in vivo biological images by use of optical coherence tomography. Optics Letters, 1996. 21(17): p. 1408-1410.
13. Bellini, N., et al., The Application of Optical Coherence Tomography to Image Subsurface Tissue Structure of Antarctic Krill Euphausia superba. PloS one, 2014. 9: p. e110367.
14. Thanomsit, C., J. Saetiew, and P. Meemon, Optical Coherence Tomography as an Alternative Tool for Evaluating the Effects of Glyphosate on Hybrid Catfish (Clarias gariepinus ×fig Clarias macrocephalus). Toxicology Reports, 2022. 9.
15. Meemon, P., et al., Optical Coherence Tomography Enabling Non Destructive Metrology of Layered Polymeric GRIN Material. Scientific Reports, 2013. 3(1): p. 1709.

16. Palawong, K., et al., Polarization sensitive optical coherence tomography for materials characterization. Chiang Mai Journal of Science, 2018. 45: p. 2232-2237.
17. Chugh, T. and A.K. Jain, OCT Fingerprints: Resilience to Presentation Attacks. ArXiv, 2019. abs/1908.00102.
18. T. Chugh, A.K.J., OCT Fingerprints: Resilience to Presentation Attacks. ArXiv, 2019.
19. Akbari, N., Automation of Fingerprint Recognition Using OCT Fingerprint Images. Journal of Signal and Information Processing, 2012. 03: p. 117-121.
20. Moolla, Y., et al., Fingerprint Matching with Optical Coherence Tomography. 2015. 237-247.
21. Cimalla, P., et al., Simultaneous dual-band optical coherence tomography in the spectral domain for high resolution in vivo imaging. Optics express, 2009. 17: p. 19486-500.
22. Jain, A.K., Y. Chen, and M. Demirkus, Pores and ridges: high-resolution fingerprint matching using level 3 features. IEEE Trans Pattern Anal Mach Intell, 2007. 29(1): p. 15-27.
23. Perez, L. and J. Wang, The Effectiveness of Data Augmentation in Image Classification using Deep Learning. 2017.
24. Almahdi, R. and H. Ragb, Fused Deep Convolutional Neural Networks Based on Voting Approach for Efficient Object Classification. 2019.
25. Srivastava, N., et al., Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 2014. 15: p. 1929-1958.
26. Chollet, F., Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: p. 1800-1807.
27. Szegedy, C., et al., Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: p. 2818-2826.
28. Agarap, A.F., Deep Learning using Rectified Linear Units (ReLU). 2018.
29. He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
30. Simonyan, K. and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, 2015. abs/1409.1556.
31. Li, Q., et al., Performance Evaluation of Deep Learning Classification Network for Image Features. IEEE Access, 2021. PP: p. 1-1.
32. Barredo Arrieta, A., et al., Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 2020. 58: p. 82-115.
33. Gilpin, L.H., et al. Explaining explanations: An overview of interpretability of machine learning. in 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA). 2018. IEEE.
34. Lundberg, S.M. and S.-I. Lee, A unified approach to interpreting model predictions. Advances in neural information processing systems, 2017. 30.

DEEP LEARNING FOR COCONUT TREE DECTECTING WITH SATELLITE IMAGERY

Chaiyasit Tanchotsrinon[1*], Chanon Boonkangwan[1], Wongnaret Khantuwan[1], Noppadon Khiripet[1]

[1]*Knowledge Elicitation and Archiving Laboratory (KEA),
National Electronics and Computer Technology Center (NECTEC),
112 Phahonyothin Road, Khlong Nueng, Khlong Luang District, Pathumthani, 12120, Thailand*

*Corresponding address (E-mail: chaiyasit.tan@nectec.or.th, +66-25646900 ext. 2253)

Abstract

In 2019, Thailand exported fragrant coconuts worth over two billion baht, which was still insufficient to meet the global market demand. Due to the significant demand for coconut products, it is crucial to monitor the cultivation areas periodically to effectively plan and estimate production for processing facilities. Aerial photography can be used to analyze the unique pattern of the perennial coconut plant, and the central coconut-growing region of Ratchaburi in Thailand was selected as the study site for the coconut cultivation areas. To efficiently detect coconut trees in broad areas, modern technology such as Graphics Processing Units (GPUs) can significantly decrease processing time. The You Only Look Once (YOLO) deep learning model was employed to identify coconut trees, and its output was compared with Land Development Department (LDD) data obtained through the THAGRI platform. Our methodology was effective in detecting coconut palms in the region, provided that the trees were not planted too closely together. However, the model's efficiency decreased when the coconut trees were densely planted. In addition to identifying the coconut cultivation site, the YOLO model also identified individual coconut trees, which can significantly decrease the amount of time needed to plan coconut cultivation locations.

Keywords: Coconut; Deep learning; Image processing; Object detection; YOLO

1. Introduction

The high export value of coconut, which surpasses two billion baht annually, indicates that it is a significant economical crop in Thailand. Because coconuts may be processed into a variety of goods, including coconut cream, coconut oil, coconut sugar, charcoal, and more, there is also an ongoing trend toward a rise in the market value of coconut exports over time. Consumers like the fruit because it may be used as an ingredient in a variety of food products, including coconut water. In this scenario, consideration is focused on the fragrant coconut type, as according to market statistics, demand is increasing, and supply is insufficient to satisfy that demand. For processing facilities to efficiently manage their supply chain while maintaining efficient monitoring of aromatic coconut crops, access to up-to-date data is necessary. It is a consequence of the reality that they are unable to identify the surrounding coconut plantations' precise locations. An updated mapping platform is required for producing agricultural maps for agricultural enterprises in various regions. This technology will make it possible to track and monitor coconut crops accurately, ensuring that the supply chain is efficiently maintained. The traditional method was made up using field surveys carried out by human laborers to gather information on the coconut plantations, which was subsequently updated on the maps. A significant amount of time and resources was required to complete this operation. These restrictions made it difficult to always maintain the data's novelty. Therefore, it's important to consider and put forward alternative solutions, like real-time mapping systems that make use of innovative technologies such as aerial photographs, remote sensing, and geographic information systems (GIS). Due to the adoption of these technologies, coconut plantations can possibly be identified and monitored in less costly and more efficient manners, provided their agriculture

maps are consistently updated and precise. However, in particular regions or for particular crops, human involvement remains necessary to ensure data accuracy. Using a data collaboration platform, the Department of Land Development's latest data on coconut plantations can be accessible, with THAGRI assisting to manage the necessary elements. Since the majority of the data on this platform is not open source, it offers effective access to data for a variety of organizations. Therefore, rights management is essential for both data owners and users who want to benefit from the data. On fragrant coconuts, the aforementioned preliminary information is accessible. From 2018, the data on coconut plantation cultivation have not been updated. In order to assist those who require the information for management and organizational goals, a preliminary classification of the coconut plantations is necessary.

The province of Ratchaburi was chosen for this study because it is well known for having the highest-ranking production of fragrant coconuts in Thailand. According to the preliminary hypothesis, coconuts' leaves differ significantly from those of other kinds of plants. It is possible that one may use this unique characteristic to distinguish coconuts from other kinds of plants. Large-scale satellite imagery processing demands a lot of resources, especially when it involves classifying various objects in the images. It could be achieved to reduce a significant amount of time by using GPU based processing as an additional technique that assists achieve time spent processing optimization. Deep learning techniques have made considerable strides in computer vision, and they have been needed in developing of techniques for object detection in images. It has improved significantly in its ability to recognize objects very precisely. This technique's resource-intensive nature is a downside, either. Deep learning tasks can require an enormous quantity of computing power. However, the time required for processing has been dramatically reduced thanks to the use of GPUs. This has provided it feasible to develop workflows for reliable identification of coconut trees across enormous regions. The following has led to the use of image processing in updating coconut cultivation maps. YOLO (You Only Look Once) is a type of convolutional neural network used for deep learning object classification. Ultralytics was the organization who developed it. The main advantage of YOLO is its real-time efficiency, which lends it appropriate for utilization in large-scale scenarios. Yolo has been used to detect various kinds of objects, and version 5 of the platform has been successfully used by [1] to precisely identify cherry trees. The identification of coconut trees was the major objective of this exploration. [2] implemented RPN to carry out the identification of coconut trees in the area. The test's position is on a closed island, but the detection results are very accurate. Processing time is not given, and there is a shortage of object variety. While [3] use the random forest approach to assist identify coconut trees. using high-resolution images taken by UAVs. [4] similar method classifies coconut trees using images with high resolution from an UAV while switching the model using Yolov5 as an alternative. Although [5] has tested the yolov5 for small object detection, it is accurate quick to detect objects. Obviously, image resolution is this aspect that has consequences for the model's accuracy. However, for realistic applications, it's reasonable to sacrifice resolution for slightly less accuracy; hence, this article shifted to the most recent Yolo version. Additionally, our chosen location contains various kinds of objects, and drop-down spatial resolution images were used for testing. A total of 21790 images—196, 5400, and 16194—were applied in the testing areas. The performance measurement of the experiment's findings is necessary due to the scope of this region. To assess the model's performance, it was divided into 40 manually labelled images. Along with all 21790 actual images that have been unlabelled. The experiment demonstrated that the results for coconut tree detection were acceptable via the model constructed from only eighty images. It still has certain shortcomings such will be discussed afterwards, compared to the entire number of images that were tested. The negative aspect of YOLO is the fact that it has difficulties detecting plenty of surrounding things[6], which is similar to the issue with poor spatial quality images when trying to recognize trees.

2. Materials and methods

In this experiment, Fragrant coconut tree detection has been classified by deep-learning models. Groups of coconut trees would be group together to detect coconut plantations. The workflow of this system can be applied to large-scale area. Figure 1 illustrates the process used for developing the prototype for detecting coconuts.
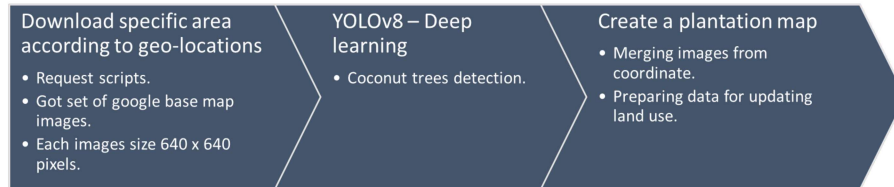
Download specific area according to geo-locations	YOLOv8 – Deep learning	Create a plantation map
• Request scripts. • Got set of google base map images. • Each images size 640 x 640 pixels.	• Coconut trees detection.	• Merging images from coordinate. • Preparing data for updating land use.

Fig. 1. Workflow of the prototype for coconut detection.

2.1. Hardware specifications

We decided to perform the workflow we designed for large-scale situations by using the DGX Workstation for our experimentation. It must be able to operate at least at the provincial level, requiring the processing of a significant amount of data and the use of massive computer resources, to assure its applicability in real-world circumstances. The workflow that was designed for this exploration could be operated on a lower-resource computer, although with longer processing times. In addition, the model training process consumed the most resources in the workflow. Table 1 contains a list of the DGX Workstation's specifications.

Table 1: Hardware specifications.

	DGX stations V100
Central Processing Unit (CPU)	1 x Intel Xeon E5-2698 v4 (20 core 40 threads, 2.2 – 3.6 GHz)
Memory (RAM)	252 GB
Storage types	SSD
Graphic Processing unit (GPU)	4 x Nvidia Tesla V100 (16 GB per device)
Operating system	Ubuntu 18.04.6 LTS

2.2. Software requirements specification

All of our experiments are based on open-source programming since it offers simple accessibility and a wide range of components. Following are the specifics: to set up multiple conditions for working on our exploration, we employed containers as based on cschranz/gpu-jupyter from Docker-Hub. It contains Python version 3.10.5, Pytorch version 1.12.0, and CUDA Toolkit version 11.7, along with Ultralytics YOLOv8. Additionally, we can use Python scripts directly for task scheduling and automation in the chosen environment or in an interactive development environment through Jupyter Notebooks for displaying results.

2.3. Dataset

Low-resolution satellite imagery could result in a reduction in the quantity of information present in the objects of interest due to identification of objects requires high-resolution satellite images for accurate processing. Consequently, they are inadequate for jobs involving object detection. High-resolution satellite pictures from Google Earth base maps, which are updated approximately once a year, have been selected as the processing input. Depending on the location, spatial resolution ranged between 0.15 and 15 meters.

Coconut is a long-term crop, so it does not require updating frequently like other agricultural products like rice, sugar cane, etc. The dataset has been divided into two parts, the first of which contains 80 images with manually marked positions for training, and the second of which has 40 images to evaluate the performance of the recently developed model. Additionally, unlabelled data with more than 20,000 images is available for testing the models against actual data. In order to evaluate the overall effectiveness of data management for agriculture map updates, the specifics of the location information for this data are described in Table 2. Moreover, each image consists of 640 x 640 pixels.

Table 2: Locations of datasets.

Dataset		
Training and Testing sets	Randomly chosen in Ratchaburi province, Thailand. It contains for training equal to 80 images and 40 images for testing, totalling 120 images.	
Unlabelled Testing sets - latitude and longitude extents		
First area	[13.5386463,100.027675]	[13.549708,100.0387367]
Second area	[13.468948721856737, 99.85694659614347]	[13.531915321856737, 99.91736049614347]
third area	[13.5574593, 100.0001742]	[13.628084, 99.9993233]

2.4. Deep learning model (YOLOv8)

Python played a key role in our design to develop a computational object search structure—in this case, a coconut tree. The deep learning framework has been selected. It can detect items regardless of their size and shape, and it can detect a variety of objects because there is no requirement to locate appropriate features for object detection. In this instance, more modifications are needed to recognize more things. Ultralytics YOLOv8 was designed for real-time object detection, classification, and segmentation. By optimizing resource usage, it makes deployment easier on a variety of hardware platforms.

3. Experimental results and Discussion

In this investigation, an approach using a pre-trained YOLOv8n model was created to update coconut plantation areas. The system had been set up in according to Ultralytics' suggested configurations. A collection of Eighty labeled images were used as the training set for the YOLOv8n model, which was created specifically for the identification of coconut trees. Making use of the LabelImg tool, a Python-based graphical user interface created with Qt, coconut trees in the photos were manually labelled. The model was evaluated for performance and restricts by testing set of 40 photos of coconut trees that were already labelled. The YOLOv8n model successfully identified 2,514 coconut trees from the mentioned testing set. It missed 1,072 coconut trees and made 114 false detections of objects which weren't coconut trees. The accuracy outcome is 67.95%. The dense clustering of coconut trees in some plantation regions, which made it challenging for the model to precisely distinguish individual trees, seems responsible for its decreased accuracy. However, considering the experiment evaluates the overall detection performance, it is important to consider those numbers carefully in terms of an object-based evaluation. Because of this, the accuracy value is less than expected. The prototype model performs better than the reported accuracy when evaluated on real-world data, as shown in the figure below, proving that it is not as inaccurate as the accuracy statistic claims. The statistics make clear that there are relatively few false detections, which are mainly due to the model's inability to accurately identify coconut trees that are closely grouped. This is especially challenging since the model has issues reliably identifying individual coconut trees when the trees are tightly planted next to each other. The resolution of the satellite imagery that was used as well as

how difficult it is for humans to identify coconut trees in heavily densely populated areas might have had influenced this issue. The creation of this prototype starts with a request for a Google satellite base map to be downloaded from a specific region. The developed model will be used to gather images in order identify coconut trees. A new map is produced by sorting and combining all the images, which can then be updated for agricultural purposes. Each image is 640*640 pixels in size, and the model detects coconut trees in less than 20 milliseconds per image. It should be noted that only one graphics card was used to obtain this performance. These aspects make this prototype appropriate for use in practical situations. The figure below shows a comparison between coconut cultivation area from LDD and the result from the model.

Fig. 2. (a) Coconut cultivation area - Ground truth from LDD. (b) Each red box represented coconut tree from the model.

4. Conclusions and Future work

Although the model's accuracy is not as high as manually counting images, especially with a test set sample size of 40 images, it is essential to randomize the images without bias to achieve the ideal performance while measuring results. Another problem is that building a model using only 80 photos might fail to account for all of the variances in coconut trees, which has a big impact on the model's accuracy. However, the model has been tested on a broader collection of test photos—more than 20,000 images with a total resolution of not less than eight million pixels and a studied area of more than ninety square kilometres which resulted in extremely excellent performance and processing times. This is appropriate for future expansion. The disadvantage of this prototype system is that, even though it doesn't happen often, it can mistakenly classify other objects as coconut trees when the coconut trees in the cultivation area are too close to one another. These incorrect facts are, however, eliminated when looking at the bigger picture of the identified cultivation area shown in the extraction of coconut trees. There are numerous ways to approach this problem in the future, including, among others, modifying the deep learning models' structural components to substitutes or utilizing additional techniques to define cultivation boundaries.
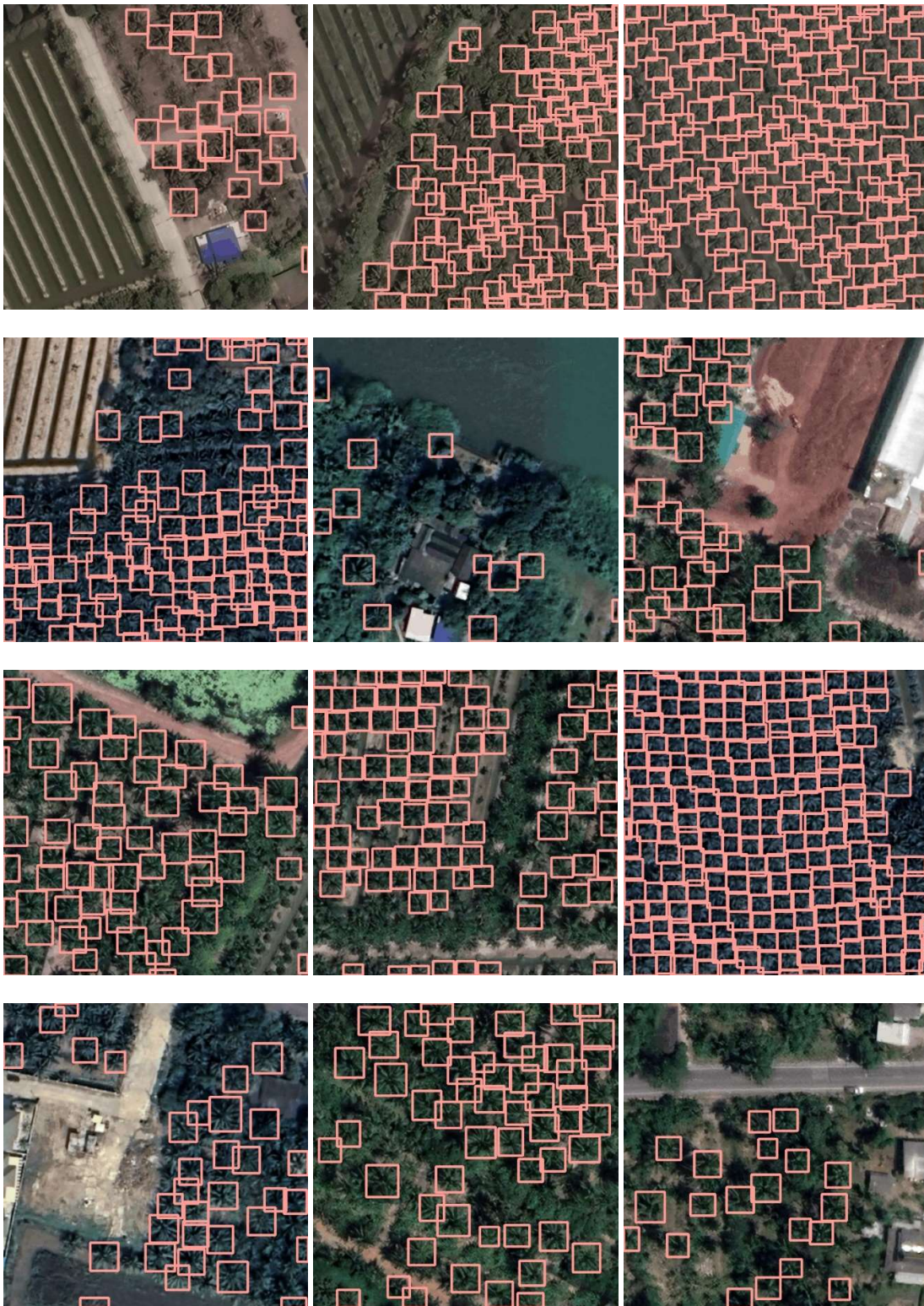
Fig. 3. Unlabelled testing images.

Fig. 4. Unlabelled testing images.

References

1.   Ozer T, Akdogan C, Cengiz E, Kelek MM, Yildirim K, Oguz Y, et al. Cherry Tree Detection with Deep Learning. In: 2022 Innovations in Intelligent Systems and Applications Conference (ASYU) [Internet]. Antalya, Turkey: IEEE; 2022 [cited 2023 Jun 10]. p. 1–4. Available from: https://ieeexplore.ieee.org/document/9925332/

2.   Zheng J, Wu W, Yu L, Fu H. Coconut Trees Detection on the Tenarunga Using High-Resolution Satellite Images and Deep Learning. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS [Internet]. Brussels, Belgium: IEEE; 2021 [cited 2023 Jun 10]. p. 6512–5. Available from: https://ieeexplore.ieee.org/document/9555008/

3.   Vargas-Munoz JE, Schibli D, Tuia D. Towards Efficient Correction of Coconut Tree Detection Errors. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium [Internet]. Kuala Lumpur, Malaysia: IEEE; 2022 [cited 2023 Jun 10]. p. 5065–8. Available from: https://ieeexplore.ieee.org/document/9883076/

4.   Kurniawan IF, Aneiba A, Hussain A, Idrissi M, Dunggio I, Asyhari AT. Large-scale Tree Detection through UAV-based Remote Sensing in Indonesia: Wallacea Case Study. In: 2022 8th International Conference on Information Management (ICIM) [Internet]. Cambridge, United Kingdom: IEEE; 2022 [cited 2023 Jun 10]. p. 110–5. Available from: https://ieeexplore.ieee.org/document/9845198/

5.   Yu X, Kuan TW, Zhang Y, Yan T. YOLO v5 for SDSB Distant Tiny Object Detection. In: 2022 10th International Conference on Orange Technology (ICOT) [Internet]. Shanghai, China: IEEE; 2022 [cited 2023 Jun 10]. p. 1–4. Available from: https://ieeexplore.ieee.org/document/10008164/

6.   E Hani U, Munir S, Younis S, Saeed T, Younis H. Automatic Tree Counting from Satellite Imagery Using YOLO V5, SSD and UNET Models: A case study of a campus in Islamabad, Pakistan. In: 2023 3rd International Conference on Artificial Intelligence (ICAI) [Internet]. Islamabad, Pakistan: IEEE; 2023 [cited 2023 Jun 10]. p. 88–94. Available from: https://ieeexplore.ieee.org/document/10136679/

Streamflow Forecasting by Artificial Neural Networks in Different Climatic Zones

Muhammad Waqas[a,b], Usa Humphries*[c], Phyo Thandar Hlaing[a,b], Angkool Wangwongchai[c]

[a]The Joint Graduate School of Energy and Environment (JGSEE), King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
[b]Center of Excellence on Energy Technology and Environment (CEE), Ministry of Higher Education, Science, Research and Innovation, Bangkok, Thailand
*[c]Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, 126 Pracha Uthit Rd., Bang Mot, Thung Khru, Bangkok 10140, Thailand; (usa.wan@kmutt.ac.th; Tel.: +66-2470-8822)

Abstract

Hydrological simulation is critical in planning and developing a water resource system. Hydrological models are often used to plan, manage, and forecast various variables in the hydrological cycle. In developing nations such as Pakistan, we can modify present prediction approaches and inflow modeling of reservoirs. A suitable approach for forecasting streamflow depends entirely on the data and study region. As a result of decreasing reliance on input data, data-driven solutions are becoming increasingly popular. This study applied various Artificial Neural Network (ANN) models to anticipate daily streamflow, and the results were compared to observed data for a performance evaluation. The research was conducted in regions with different climates: The Garhi Habibullah and Narran. Four Input combinations were used in this study using past daily flow or discharge, precipitation, and maximum and minimum temperature. Statistical measures like the coefficient of determination (R^2), root mean square error (RMSE), mean square error (MSE), and Normalized root means square error (NRMSE) were used to evaluate the ANN models. General Regression Neural Networks (GRNN) performed well at both stations. At Garhi Habibullah station, the outcomes of statistical evaluation parameters of input combination ($P_A + T_{Amax} + T_{Amin}$) were 0.99, 0.51, 1135, and 5.71 during training, while testing 0.99, 0.00, 0.262, and 0.181 were found respectively. However, at Narran station, the outcomes of these parameters with the input combination ($P_i + T_{imax} + T_{imin}$) were 0.91, 1.23, 20190, and 41.74, while testing 0.96, 0.72, 9247, and 38.90 were found, respectively. Other ANNs models also performed well, but the outcomes of the GRNN model in respect of statistical parameters were more precise.

Keywords: Artificial Intelligence; Neural Networks; Streamflow, Forecasting, Hydrological Simulation semicolons

1. Introduction

Forecasting reservoir influx systematically is a vital topic in engineering, which involves integrating and planning water resources such as hydroelectric power, water supply, flood control, and drought management [1, 2]. Numerous analyses have shown that Pakistan's water crisis worsens due to system mismanagement and planning deficiencies. As a result, changing current procedures and policies is necessary to address the described situation [3, 4]. In developing nations such as Pakistan, we can modify present prediction approaches and streamflow modeling of reservoirs, which can be crucial in water resource planning and management decisions. Typically, the suitable approach for forecasting and modeling inflows depends entirely on the data and research region. As a result of decreasing reliance on input data, data-driven solutions are becoming increasingly popular [5]. Artificial neural networks (ANNs) are currently employed in research to solve significant problems such as prediction, optimization, and pattern recognition [6]. ANN structure comprises layers referred to as the input, output, or hidden layers. Each layer has connected elements (neurons or nodes). A neuron is a component that performs various operations, including receiving input data, assigning weights to those data, adding the products of the importance and the respective input data, and applying a bias to it. The resulting value is then altered by a function called the activation function. The output of these procedures is referred to as the product shown in Fig. 1. Individual neurons in each layer are connected so that there is a connection between them and the neurons in the subsequent layer via weights that indicate their degree of interdependence and are not connected to neurons in the same layer [7]. The ANN's design is critical since it determines the type of

ANN to employ and its configured components [7, 8]. Because the architecture of the network determines the pattern of connections, the number of weights to be assigned, and how data is communicated through the network, the choice of architecture and design network significantly impacts the ANN's efficacy. It is thus viewed as the primary consideration when developing the ANN model. Additionally, constructing the ANN model is the most time-consuming part[8, 9]. Concerning previous studies on streamflow modeling, this study was designed so that different employee ANNs could achieve the primary goals of this research: 1) to calibrate and validate the ANNs for the modeling of the hydrological process and 2) to determine the optimal input combination for the applied ANNs.
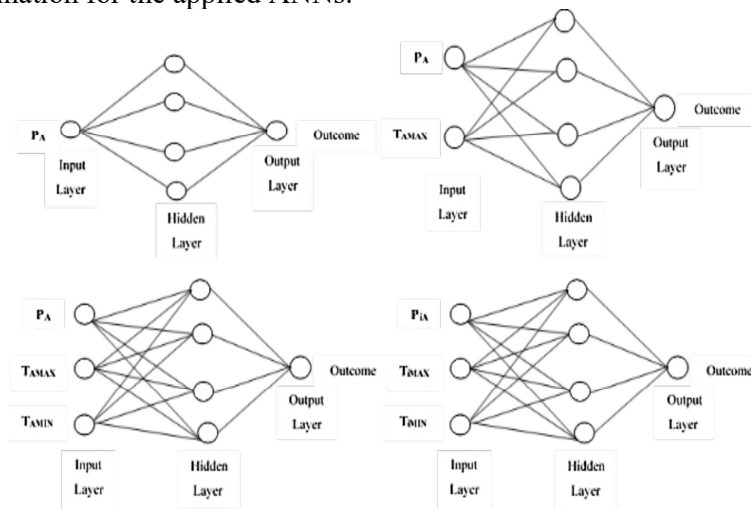
Fig. 1. ANNs architecture by different dataset inputs employed in this study.

2. Methodology

2.1. Study Area

This study evaluated different assessment models on two regions, namely Narran and Garhi Habibullah, situated in the Jhelum River basin, Western Himalayas, which is the primary catchment of the Indus River System shown in Fig. 2. The statistics for both stations are shown in Table 01.

Fig. 2. Mangla Watershed and the Narran and Garhi Habibullah Stations

Table 1. Statistics of stations

Stations	Latitude	Longitude	Elevation (m)	Std.	Median	Mean	C_v	C_s
Narran	34.9	73.7	2400	52.39	20.76	45.88	114.2	1.44
Garhi Habibullah	34.4	73.4	820	96.55	55.99	99.41	97.13	1.35

2.2. Input Dataset

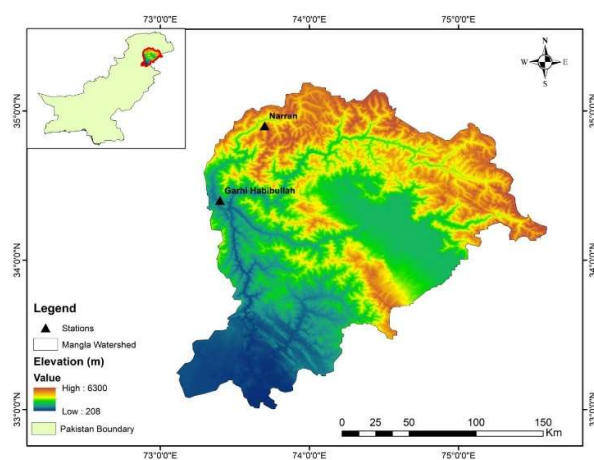The ANNs techniques need various meteorological and hydrological datasets such as daily flow or discharge, daily precipitation, and daily maximum and minimum temperatures. As earlier mentioned in the introduction, these hydrological and meteorological datasets were employed as input for all ANNs techniques in this study to achieve our key objective of evaluating the daily discharge of both stations. Four different Input Dataset combinations were employed for the calibration and validation of ANNs models; Firstly, the average daily rainfall of the station (P_A); secondly, the average daily rainfall of the station (P_A) and average daily maximum temperature of the station (T_{Amax}); thirdly, average daily rainfall of the station (P_A), the average daily maximum temperature of the station (T_{Amax}), and average daily minimum temperature of the station (T_{Amin}) and last one comprises on the average daily rainfall of all the stations present in the catchment (P_i), average daily minimum temperature of all station (T_{imin}), and the average daily maximum temperature of all the station (T_{imax}). These input combinations are 1) P_A, 2) P_A + T_{Amax}, 3) P_A + T_{Amax} + $T_{Amin,}$ and P_i + T_{imax} + $T_{imin.}$

2.3. Evaluation Criteria

The ANNs technique's performance was assessed using statistical parameters. Different statistical parameters were engaged for the performance evaluation of the applied ANNs techniques; i) Co-efficient of determination (R^2), ii) Normalized root means square error (NRMSE), iii) Root Mean Square Error (RMSE), and iv) Mean square error (MSE) was used.

$$R^2 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n[\Sigma x2 - (\Sigma x2)][\Sigma y2 - (\Sigma y2)]}} \qquad (1)$$

$$NRMSE = \frac{\sqrt{\frac{\Sigma_{i=1}^{N}(Qobs - Qpre)^2}{N}}}{\sigma} \qquad (2)$$

$$MSE = \frac{\Sigma_{i=1}^{N}(Qobs - Qpre)^2}{N} \qquad (3)$$

$$RMSE = \sqrt{\Sigma_{i=1}^{i}(Qpre - Qobs)^2 / N} \qquad (4)$$

Sometimes, facts about the model's outputs should be revealed for improved examination, description, or graphical demonstration of simulated and observed flows.

2.4. General Regression Neural Networks (GRNNs)

GRNN is an adaptation of RBFNNs (same as kernel regression)[10]. The fundamental difference between the GRNN and a conventional FFBP-ANN is that the GRNN's design is fixed for a specific input-output dataset (described in further detail below). The FFBP-ANN, on the other hand, includes determining the optimal number of hidden layers and nodes. Only one parameter must be optimized in GRNN: the RBF kernel spread parameter, r, which is used to compute the correlation of input parameters. Therefore, an optimum value of the correction factor should be sought to offer the optimal degree of

smoothness while enabling the GRNN to generalize adequately to out-of-sample inputs. The GRNN approach can estimate any continuous function mapping for any input-output data set and directly produces function estimates from training input data [11]. The schematic layout of GRNNs is shown in Fig. 3.
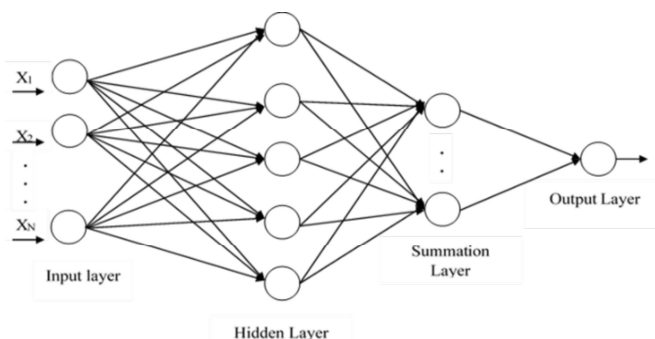

Fig. 3. Schematic diagram of GRNNs

2.5. Multilayer Perceptron Neural Network (MLPNN)

The hydrological process is a soft computational technique in which we stimulate the ANNs on a large scale. It includes neurons grouped in clusters, called layers, and is associated with weight see Fig. 4. The construction of a simple neural network comprises three layers: an input layer, an output layer, and a hidden layer. The input layer represents the data that can pass through them, and the system employs a neural network to provide the outer layer. Each neuron can accept several entries from the weighted network. Additionally, inputs are provided during the process, creating the arguments for the transmission purpose, such as hyperbolic, logistic, or linear targeted parameters that can give the output neuron's output [12].
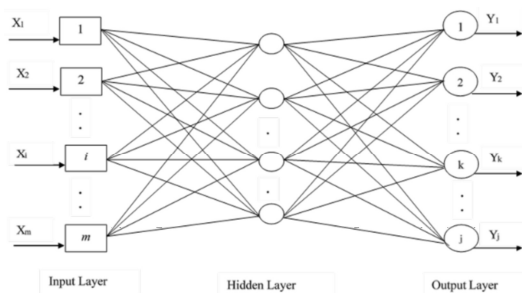
Fig. 4. Schematic diagram of MLPNN

2.6. Radial basis function neural networks (RBFNNs)

Reference [13] proposed the RBFNNs. It is a form of feed-forward neural network acquired through administrative training. Its output nodes are derived from the linear combination of radial functions calculated by the hidden nodes layers [14]. RBFNN is a feed-forward network with three layers typically utilized as a function of linear transfer for the units of the outcome or as a function of nonlinear transfer for the hidden units. The input layer of RBFNNs is composed entirely of nodes connected via weighted networks in the hidden layers. Hartman and colleagues established that the RBFNN neural network model best approximates all functions[15, 16]. This RBFNN neural network comprises three layers: an insertion, a secrete, and an outcome layer. The network structure is depicted in Fig. 5.

Fig. 5. Schematic Diagram of RBFNNs

3. Results and Discussions

The observed discharge assessed the calibration and validation of all applied ANNs techniques. Various input combinations are employed fo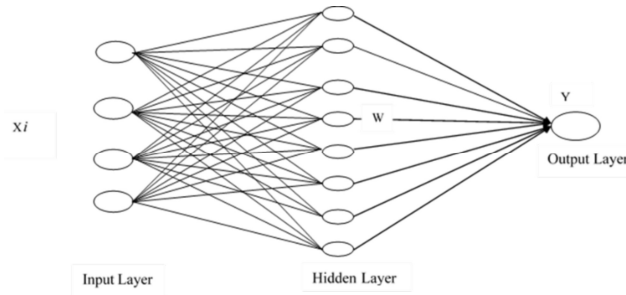r the ANNs architecture, as represented in Fig. 1. This study used three models (GRNN, MLPNN, and RBFNN) and four different input combinations. Their performance was evaluated by statistical parameters mentioned previously. The best ANN architecture was engaged for both stations. All input combinations were used in Table 2 and Table 3 on the Narran and Garhi Habibullah stations. These models were trained and tested simultaneously until more efficient and accurate results were obtained. Then all three applied ANNs techniques were assessed based on performance evaluation criteria. At Narran and Garhi Habibullah stations, the GRNN model performed efficiently during the training and testing of models. The Performance evaluation parameters R^2, MSE, RMSE, and NRMSE results are 0.91, 1.23, 20190, and 41.74 during training, and 0.96, 0.72, 19247, and 38.90 were found in the testing of models, respectively. These results were found on the fourth input combination ($P_i + T_{imax} + T_{imin}$) shown in Table 2. At Garhi Habibullah station, the outcomes of statistical evaluation parameters of the third input combination ($P_A + T_{Amax} + T_{Amin}$) were 0.99, 0.51, 1135, and 5.71, while testing 0.99, 0.00, 0.262, and 0.181 were found more accurate of GRNN model respectively.

Other ANNs models (MLPNN and RBFNN) also performed well, but the outcomes of the GRNN model in respect of statistical parameters were more accurate. The reason behind this better performance was the one-pass algorithm of the GRNN model. Only a single simulation and total dependence on general regression enable the GRNN model to give a more precise output. On the other hand, linear combination and backward propagation made MLPNN and RBFNN less precise. The results in Tables 2**Error! Reference source not found.** and 3 revealed that the GRNN performance is excellent during training and testing.

Table 2. Performances of ANN models. MLPNN, GRNN, and RBF at Narran Catchment.

River	Model	Input Combination	Training				Testing			
			R^2	MSE	RMSE	NRMSE	R^2	MSE	RMSE	NRMSE
Narran	MLPNN	P_A	0.80	0.01	1622	13.99	0.81	0.02	723	9.38
		$P_A + T_{Amax}$	0.75	0.02	1079	9.31	0.83	0.05	475	6.17
		$P_A + T_{Amax} + T_{Amin}$	0.70	0.015	900	7.76	0.75	0.02	453	5.87
		$P_i + T_{imax} + T_{imin}$	**0.87**	**0.00**	**24308**	**50.25**	**1.00**	**0.14**	**21277**	**43.0**
	GRNN	P_A	0.82	0.22	1578	13.61	0.77	0.02	720	9.34
		$P_A + T_{Amax}$	0.86	0.02	709	6.11	0.85	0.06	563	4.86
		$P_A + T_{Amax} + T_{Amin}$	0.70	0.06	563	4.86	0.92	0.00	329	4.27
		$P_i + T_{imax} + T_{imin}$	**0.91**	**1.23**	**20190**	**41.74**	**0.96**	**0.72**	**19247**	**38.90**
	RBFNN	P_A	0.70	0.00	1585	13.67	0.75	0.00	691	8.96
		$P_A + T_{Amax}$	0.75	0.00	1060	9.14	0.88	0.00	432	5.60
		$P_A + T_{Amax} + T_{Amin}$	0.78	0.00	1059	9.12	0.90	0.00	360	4.67
		$P_i + T_{imax} + T_{imin}$	**0.84**	**0.00**	**19643**	**40.61**	**0.92**	**0.00**	**18312**	**37.01**

Table 3. Performances of ANN models. MLPNN, GRNN, and RBF at Garhi Habibullah

River	Model	Input Combination	Training				Testing			
			R^2	MSE	RMSE	NRMSE	R^2	MSE	RMSE	NRMSE
Garhi Habib ullah	MLPNN	P_A	0.87	0.02	4412	22.19	0.88	0.00	0.25	0.175
		$P_A + T_{Amax}$	0.93	0.09	2467	12.41	0.92	0.00	0.262	0.181
		$P_A + T_{Amax} + T_{Amin}$	**0.97**	**0.14**	**2049**	**10.31**	**0.99**	**0.00**	**0.262**	**0.181**
		$P_i + T_{imax} + T_{imin}$	0.96	0.01	1841	9.26	0.97	0.00	0.261	0.179
	GRNN	P_A	0.55	0.40	4322	21.74	0.70	0.01	0.256	0.176
		$P_A + T_{Amax}$	0.98	0.78	1445	7.27	0.99	0.00	0.262	0.181
		$P_A + T_{Amax} + T_{Amin}$	**0.99**	**0.51**	**1135**	**5.71**	**0.99**	**0.00**	**0.262**	**0.181**
		$P_i + T_{imax} + T_{imin}$	0.97	0.35	1484	7.47	0.98	0.01	0.261	0.180
	RBFNN	P_A	0.87	0.00	4216	21.20	0.60	0.04	0.245	0.168
		$P_A + T_{Amax}$	0.91	0.00	1699	8.55	0.92	0.00	0.255	0.176
		$P_A + T_{Amax} + T_{Amin}$	**0.97**	**0.00**	**1601**	**8.05**	**0.99**	**0.01**	**0.243**	**0.168**
		$P_i + T_{imax} + T_{imin}$	0.96	0.00	1587	7.99	0.98	0.01	0.240	0.165

4. Conclusion

The study compared ANNs (GRNNs, MLPNNs, and RBFNNs) to understand the hydrological simulation of different climatic parameters in the Narran and Garhi Habibullah Basin. Different combinations of datasets were utilized for training and evaluating ANNs. After training and testing, modeled and observed data were evaluated using R2, NRMSE, RMSE, and MSE. The Performance evaluation results were more accurate on the fourth input combination (Pi + Timax + Timin). At Garhi Habibullah station, the outcomes of statistical evaluation parameters of the third input combination (PA + TAmax + TAmin) were 0.99, 0.51, 1135, and 5.71, while testing 0.99, 0.00, 0.262, and 0.181 were found more accurate of GRNN model respectively. Other ANNs models (MLPNN and RBFNN) also performed well, but the outcomes of the GRNN model in respect of statistical parameters were more accurate.

Acknowledgements

References

[1] Sit, M., et al., A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, 2020. **82**(12): p. 2635-2670.
[2] Waqas, M., et al., Assessment of advanced artificial intelligence techniques for streamflow forecasting in Jhelum River Basin. *Pakistan Journal of Agricultural Research*, 2021. **34**(3): p. 580.
[3] Baloch, M.H., et al., Hybrid energy sources status of Pakistan: An optimal technical proposal to solve the power crises issues. *Energy Strategy Reviews*, 2019. **24**: p. 132-153.
[4] Solangi, Y.A., et al., Evaluating the strategies for sustainable energy planning in Pakistan: An integrated SWOT-AHP and Fuzzy-TOPSIS approach. *Journal of Cleaner Production*, 2019. **236**: p. 117655.
[5] Peña-Arancibia, J.L., et al., Climate change and reservoir sedimentation implications for irrigated agriculture in the Indus Basin Irrigation System in Pakistan. *Journal of Hydrology*, 2021. **603**: p. 126967.
[6] Jain, A.K., J. Mao, and K.M. Mohiuddin, Artificial neural networks: A tutorial. *Computer*, 1996. **29**(3): p. 31-44.
[7] Yesilnacar, M.I., et al., Neural network prediction of nitrate in groundwater of Harran Plain, Turkey. *Environmental Geology*, 2008. **56**: p. 19-25.
[8] Hydrology, A.T.C.o.A.o.A.N.N.i., Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, 2000. **5**(2): p. 115-123.
[9] Curteanu, S. and H. Cartwright, Neural networks applied in chemistry. I. Determination of the optimal topology of multilayer perceptron neural networks. *Journal of Chemometrics*, 2011. **25**(10): p. 527-549.
[10] Okkan, U. and H.Y. Dalkılıç, Reservoir Inflow modeling with artificial neural networks: the case of Kemer Dam in Turkey. 2011.

[11] KIŞI, Ö.J.H.S.J., Generalized regression neural networks for evapotranspiration modelling. 2006. **51**(6): p. 1092-1105.

[12] RAJKUMAR, K.V., K.J.J.o.T. SUBRAHMANYAM, and A.I. Technology, A HYBRID ACO-CS BASED OPTIMIZED KNN CLASSIFIER ALGORITHM FOR RAINFALL DETECTION & PREDICTION. 2021. **99**(13).

[13] Broomhead, D.S. and D. Lowe, Radial basis functions, multi-variable functional interpolation and adaptive networks. *Royal Signals and Radar Establishment Malvern (United Kingdom)* 1988.

[14] Hush, D.R. and B.G.J.I.s.p.m. Horne, Progress in supervised neural networks. 1993. **10**(1): p. 8-39.

[15] Hartman, E.J., J.D. Keeler, and J.M.J.N.c. Kowalski, Layered neural networks with Gaussian hidden units as universal approximations. 1990. **2**(2): p. 210-215.

[16] Park, J. and I.W.J.N.c. Sandberg, Universal approximation using radial-basis-function networks. 1991. **3**(2): p. 246-257.

Machine learning-based communication power adaptation
for connected truck platooning

Sirapob Keeratichandecha[a],*and Rardchawadee Silapunt[a]

[a]Department of Electronic and Telecommunication Engineering, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok 10140, Thailand

Abstract

A vehicle-to-vehicle communication (V2V) is a key technology in autonomous and connected vehicles. For a connected vehicle platoon, the communication power utilization is still a key challenge, particularly for long-distance transport. We acknowledge that the historical feedback information from data exchanged within the platoon is advantageous for power optimization. This paper therefore presents a communication power optimization based on Machine Learning (ML) in New Radio (NR) V2V in truck platooning. We performed simulations of the V2V links among truck platooning in the absence of cellular network and infrastructure assistance by a network simulator named MilliCar based on the 3GPP standard. Feedback information among the platoon was also collected for developing the ML models. We compared 3 popular ML classification techniques, namely Random Forest (RF), k-Nearest Neighbors (KNN), and Radial Basis Function Kernel - Support Vector Machine (RBF). Afterward, a power adaptation with the optimal ML technique called ML-based power adaptation (ML-PA) was proposed. The comparison showed that RF classification technique was slightly better than RBF and KNN for the prediction of transport block reception status of NR-V2V in the platoon. The ML-PA performance was evaluated by comparing with the default power transmission scenario and non-ML power adaptation (PA) scenario. The simulation results indicated that ML-PA developed by RF saved the overall transmitting power in the V2V compared to the default and PA scenarios by 28% and 24% respectively. The ML-PA also reached the maximum power reduction of 56% and 15% while limiting block error rate (BLER) in desired communication range.

Keywords: Platoon; V2V; 5G; NR; Power adaptation; Machine learning

1. Introduction

Developments of vehicular communication technologies play an important role in the emerging era of autonomous vehicles. The V2V in 5G, the fifth generation of mobile technologies, is a part of vehicle-to-everything communication (V2X) developed by 3GPP. The 5G NR-V2V was designated as PC5 interface and also called Sidelink in a physical layer. It was designed to enhance several applications and services in global level of industries and general use. In transportation, autonomous truck platooning has been enabled by the technology. A lot of benefits are provided by the platoon such as saving fuel costs, improving safety by removing driver reaction time and human errors, and reducing CO_2 emissions. The technology allows vehicles to connect and exchange necessary information in and outside the vehicle group. The vehicle platooning therefore requires capabilities for reliable and low latency inter-vehicular communication to readily decide and control its operations.

Several studies were conducted to improve the communication technology and application. Some studies focused on network and resource management such as [1] studied on vertical handover decision of vehicle platooning to cope with quality of service degradation in communication, [2] studied various network topologies for V2V to maintain inter-vehicular distance in vehicle platooning while considering disturbance and system response, and [3] studied radio resource allocation scheme in LTE-V2V to minimize tracking error in vehicle platooning, while some studies focused on technical configurations such as [4] studied on antenna configurations and propagation loss characteristics in V2V for vehicle platooning. However, communication power optimization in 5G NR-V2V for vehicle platooning is still a challenge and rarely explored. In this work, we therefore explored the communication power optimization

using information exchanged in the platoon and Machine Learning (ML). We proposed to optimize the power consumption in truck platooning by a ML based power adaptation. This paper is structured as following sections. Section 2 presents our system model studied in this work. Section 3 describes the study methodology. Our study results and discussion are presented in section 4. Lastly, the conclusion of our study is revealed in section 5 of this paper.

2. System model

The system model and scenarios studied in this work are described as a framework in this section. There are descriptions of the truck platoon setup, V2V transmitting power considered in this paper, and simulating scenario for V2V in the truck platoon as follows.

2.1 Platoon setup

In this work, we considered the V2V links in a truck platoon operating in dedicated mode which was a device-to-device communication (D2D) without cellular networks and infrastructure assistances. The platoon consisted of a platoon leader (PL) and 3 platoon members (PMs). All PMs were forming as a group by communicating with the PL and PL also communicated to each member as groupcast communication. Only V2V links between PL and each PM were considered as shown in Fig. 1. All trucks in the platoon moved in the same direction and lane with variables of velocity v and inter-vehicular distance d. Considering the positions of trucks and channel conditions, a V2V link between PL and the first PM (PM1) was always line-of-sight (LOS) and the others were always non-LOS path blocked by vehicles (NLOSv).
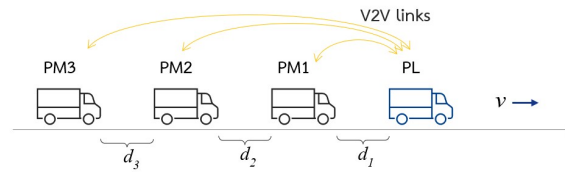

Fig. 1. setup of V2V links in a truck platoon.

2.2 Consideration of V2V transmitting power

A V2V transmitting power range of $17 - 33$ dBm was studied. The default transmitting power was set to 23 dBm for all trucks. The power would be differently set as 3 power schemes. The first one was a fixed transmitting power with the default power level. The second power scheme was a simple power adaptation or power control algorithm (PA). A basic concept of PA was to adapt the transmitting power of trucks' V2V in the platoon based on Signal to Interference plus Noise Ratio (SINR). The difference of perceived SINR ($SINR_{PER}$) and target SINR ($SINR_{TAR}$) that usually depends on system design was considered. When the perceived SINR was less than the target, the transmitting power for next transmissions (P_{T_next}) would be increased as

$$P_{T\_next} = P_T + SINR_{TAR} - SINR_{PER} \ \ (dB) \tag{1}$$

On the other hand, the power would be decreased for high perceived SINR scenarios to reduce power consumption. For the third power scheme, it was proposed to optimize the power consumption using ML together with SINR. It is called ML-based power adaptation (ML-PA). ML-PA's algorithm is described in section 3 of this paper.

2.3 Simulating scenarios

Four different scenarios of a truck platoon with V2V communication were studied. They were assumed to be activities of joining and leaving the vehicle group and to vary different parameters to collect data. In scenario 1, we studied cases that all trucks moved with the same velocity and inter-vehicular distance. The

distances in the platoon were set equally and all trucks started moving together from standstill. The platoon constantly accelerated from 0 km/h and the simulation ended when it reached 120 km/h. The distances were set to 10 m for the first simulation of this scenario and then the simulation was repeated by varying the distances to reach 1000 m with the same movement for each simulation. In scenario 2, three following trucks joined the PL and left the group. Scenarios 3 was similar to scenario 2. The last 2 trucks joined the group of PL and PM1 then left the group. Likewise, the last truck joined the group of PL PM1 and PM2 then left the group in scenario 4. All the 4 scenarios were simulated with the default power scheme for collecting dataset and performance evaluation reference in this work. We also called it the default simulations. Afterward, the dataset would be used for ML process and the other power schemes would be also applied for performance evaluation and comparison. The 4 scenarios are simply visualized in Fig. 2 (a) scenario 1, (b) scenario 2, (c) scenario 3, and (d) scenario 4.
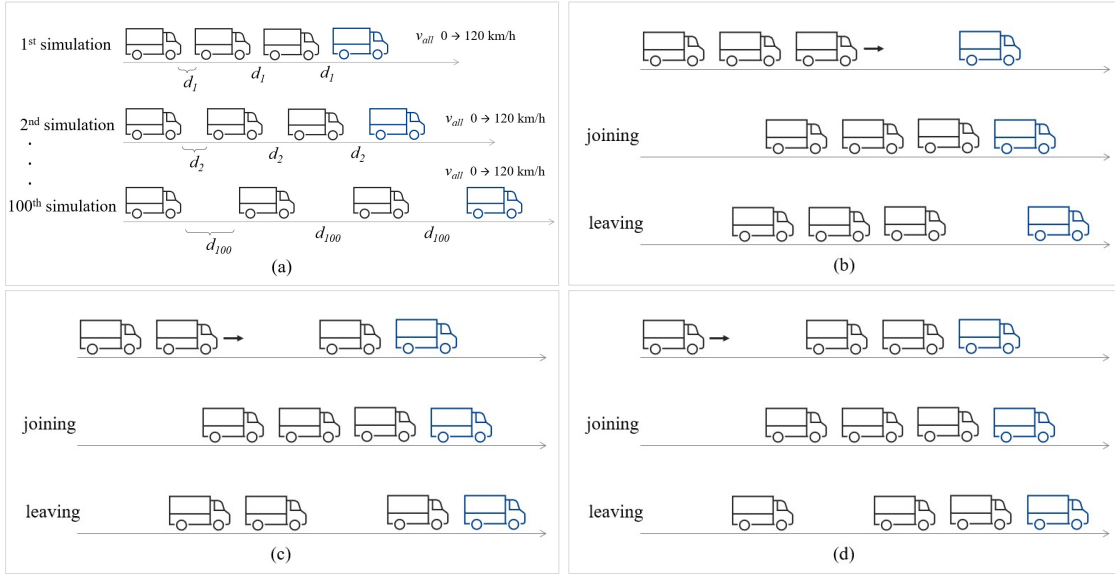

Fig. 2. truck platoon simulation in (a) scenario 1; (b) scenario 2; (c) scenario 3; (d) scenario 4.

3. Methodology

3.1. Simulations of V2V in truck platooning

The simulations of V2V links among truck platoon were performed with Millicar which is a network simulator for 3GPP standard-based 5G-V2X [5]. The simulation and parameter settings are shown in table 1 which mostly conformed to the original simulator and V2X evaluation methodology in [6]. We assumed that information exchanged between PL and PMs included *SourceID, DestinationID, TxPower, TxSpeed, RxPower, RxSpeed, RelativeSpeed, Distance,* and *perceived SINR*. After the simulations of all studied scenarios were performed with the default transmitting power, the exchanged information was collected as a dataset and then would be used to develop a predictive model by ML. The specific information was also considered as ML features.

Table 1. Simulation and parameter settings

Parameter	Power Scheme		
	Default P_T	PA	ML-PA
Transmitting power	23 dBm	17-33 dBm	17-33 dBm
V2V operating frequency	5.9 GHz		
Bandwidth	40 MHz		
Subcarrier spacing	60 kHz		
Pathloss model	$PL = 32.4 + 20 \log_{10}(d) + 20 \log10(f_c)$		
Shadowing distribution	Log-normal		
Shadow fading std	3 dB		
Additional blockage loss for NLOSv	4 dB		
Packet size	100 bytes		
Packet Interval	100 ms		
Vehicle size	Length = 13.0 m, Width = 2.5 m, Height = 3.0 m		
Antenna pattern	4x4 isotropic		
RLC mode	Un-Acknowledge Mode (UM)		
HARQ	No		

3.2. Development of a predictive model

In this work, an objective of the predictive model was to predict the communication results which is status of transport block (TB) receptions in the platoon. The status could be "success" or "fail" for each transmission. We therefore could optimize the power consumption in V2V by adapting or controlling level of transmitting power when we predicted if the TBs were delivered successfully or not. Three binary classification techniques, namely Random Forest (RF), k-Nearest Neighbors (KNN) and Radial Basis Function Kernel - Support Vector Machine (RBF), were studied. The most suitable technique would be used to develop the model and ML-based power adaptation as ML-PA afterwards. From the default simulations, a dataset of 149,443 records was provided. For the first ML step, the dataset was split to a training set and a test set in amount of 80% and 20% respectively. Secondly, we proceeded feature selection to select important features which were significant for prediction of the communication results. Next, three predictive models with the 3 ML techniques were developed and compared. After that, the optimal model was selected for ML-PA.

ML-PA was developed by applying the predictive model to the simple power adaptation (PA) algorithm described in section 2. Concept of ML-PA algorithm is shown in Fig 3. In this power scheme, the power was increased and decreased depending on predicted results by the predictive model and also SINR. It performed like a TB delivery failure preventer and a power-saver in the operation of V2V.
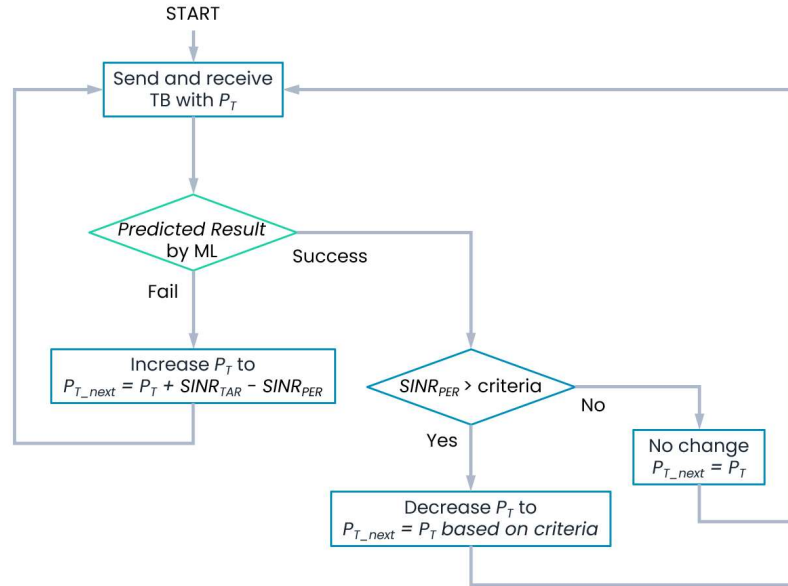
Fig. 3. machine learning-based power adaptation algorithm.

At this stage, all 3 different power schemes were presented, namely the default power scheme, PA, and ML-PA. Afterwards, all the 3 power schemes would be applied to the same simulating scenarios to evaluate and compare the performance.

4. Results and discussion

Our comparison of the predictive model performance indicated that RF model was the optimal model, in terms of accuracy and cross validation score, for the prediction of TB reception status in V2V. The comparison chart of predictive models is shown in Fig. 4. The performances of the models were not significantly different from each other. However, it resulted that RF was slightly better than RBF and KNN with the best accuracy of 96.72%, therefore, the model with RF algorithm was chosen. RBF was in the second place with accuracy of 96.63% and KNN was in the third place with accuracy of 96.59%.
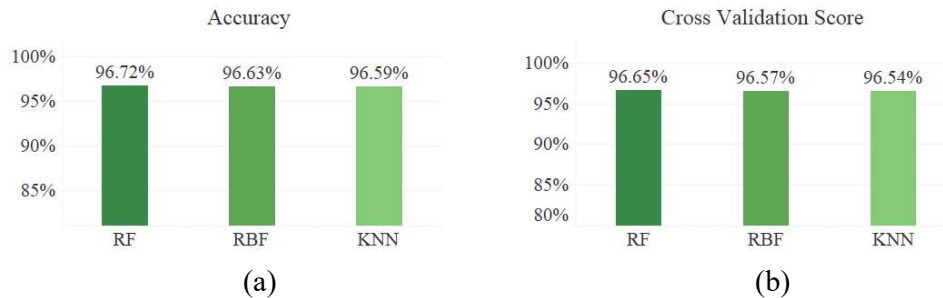
Fig. 4. predictive model's performance comparison of (a) accuracy; (b) cross validation score.

After the chosen model was applied as ML-PA to the simulations, V2V system performance analysis was conducted by comparing to the default simulation and PA simulation. To illustrate the results of simulations, the performance comparisons in aspect of transport block error rate (BLER) and transmitting power consumption varied by the communication distance are presented in Fig. 5. The results showed that average BLER increased by effects of propagation loss, random noise and interference. In overall comparison, ML-PA had the smallest BLER with a bit difference from the default simulation and PA. In terms of V2V power consumption, ML-PA obviously consumed less power than the others in

communication range within 350 m. In longer communication ranges, ML-PA and PA continually increased the power consumption to limit BLER that naturally increased by the distance. As a result, ML-PA and PA were better than the default power scheme in terms of BLER particularly in further distance. Moreover, ML-PA consumed significantly less power than PA, even though both power schemes had close BLER performance.
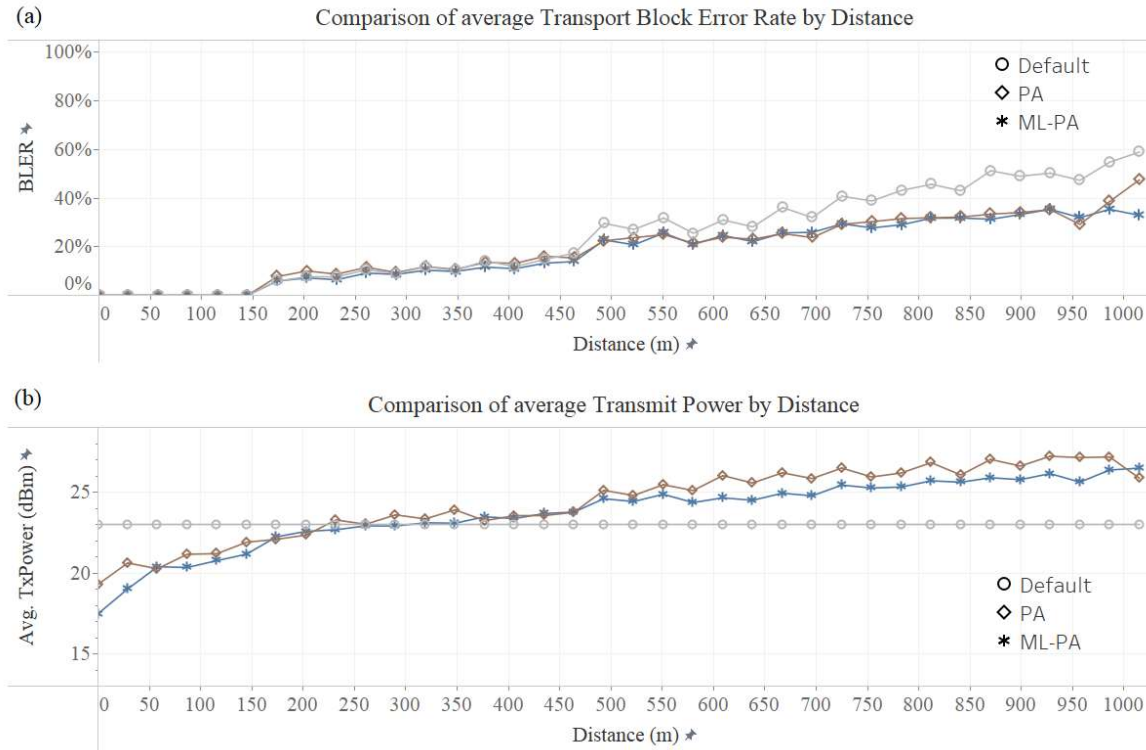

Fig. 5. comparisons of average (a) transport block error rate by communication distance; (b) transmitting power by communication distance.

To numerically evaluate the performances, the performance analysis was conducted in 2 ranges of communication, namely communication ranges of 0-350 m and 350-1000 m to clearly illustrate average BLER and power consumption in each power scheme. The BLER and transmitting power in each communication range were averaged and compared. The differences of V2V transmitting power used in ML-PA and PA compared to the default simulations are considered. Fig. 6 shows the comparisons of BLER and %difference of transmitting power in communication ranges of (a) 0-350 m and (b) 350-1000 m. In the communication range within 350 m, all the 3 power schemes resulted similarly in terms of average BLER. The average BLER of all simulations were limited to 3.33% in overall scenarios, 6.34% in scenario 1, and closely to 0.0% in scenarios 2-4. In aspect of transmitting power consumption, ML-PA transmitted less power than the default power scheme and PA by 28.82% and 24.08% respectively. In scenario 1, ML-PA transmitted less power than the default power scheme by 9.55% with average BLER of 5.16% while PA transmitted more power than the default power scheme by 25.79% with average BLER of 6.34%. In scenario 2, ML-PA and PA reduced similar transmitting power of 32.17% and 31.64% respectively while limiting the BLER to 0.00%. The maximum power reduction of 56.67% was reached by ML-PA in scenario 4 with satisfied BLER. In the communication range of 350-1000 m, the overall BLER and transmitting power drastically increased as a result of scenario 1 which all trucks were equally separated. The 2 power adaptation schemes increased the power to limit BLER in all scenarios. The comparison in Fig. 6 (b) also clearly shows that ML-PA was better than PA at limiting the average BLER and the V2V power consumption.

(a)

Scenario	avg. BLER (%)			%difference of avg. TxPower compared to Default		
	Default	PA	ML-PA	Default	PA	ML-PA
Grand Total	2.96%	3.33%	2.74%	0.00%	-4.74%	-28.82%
1	5.62%	6.34%	5.16%	0.00%	25.79%	-9.55%
2	0.00%	0.00%	0.00%	0.00%	-31.64%	-32.17%
3	0.00%	0.04%	0.00%	0.00%	-37.98%	-53.24%
4	0.00%	0.00%	0.00%	0.00%	-41.69%	-56.67%

(b)

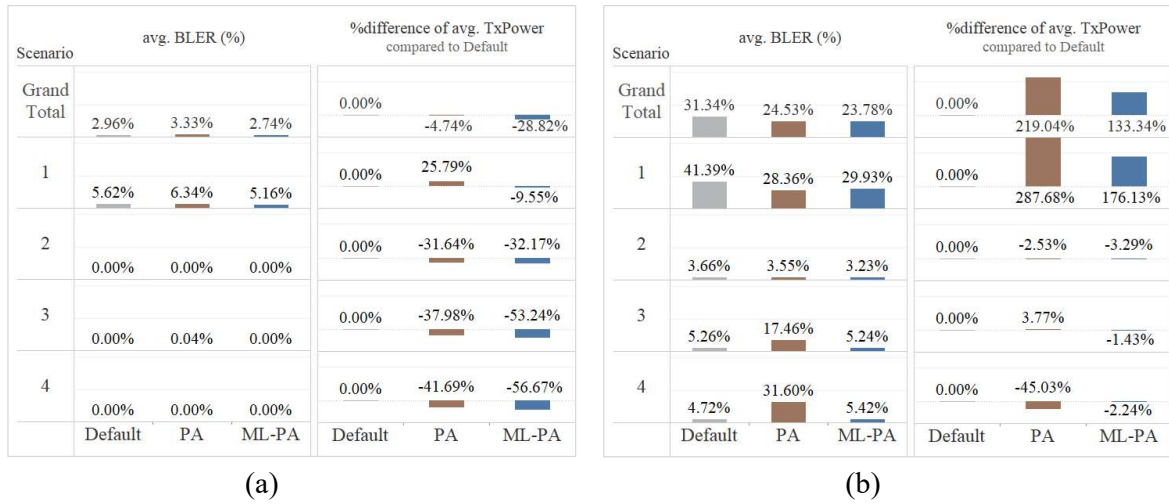| Scenario	avg. BLER (%)			%difference of avg. TxPower compared to Default		
	Default	PA	ML-PA	Default	PA	ML-PA
Grand Total	31.34%	24.53%	23.78%	0.00%	219.04%	133.34%
1	41.39%	28.36%	29.93%	0.00%	287.68%	176.13%
2	3.66%	3.55%	3.23%	0.00%	-2.53%	-3.29%
3	5.26%	17.46%	5.24%	0.00%	3.77%	-1.43%
4	4.72%	31.60%	5.42%	0.00%	-45.03%	-2.24%

Fig. 6. comparisons of transport block error rate and percent difference of transmitting power in communication ranges of (a) 0-350 m; (b) 350-1000 m.

5. Conclusions

This paper studied a power optimization based on Machine Learning (ML) in New Radio (NR) V2V in truck platooning. Three ML algorithms namely Random Forest (RF), k-Nearest Neighbors (KNN), and Radial Basis Function Kernel - Support Vector Machine (RBF) were compared. RF, the optimal ML algorithm, was developed into a ML-based power adaptation (ML-PA) and applied to simulations of V2V in truck platooning afterward. To evaluate performance of V2V system with different power schemes, average transport block error rate (BLER) and transmitting power were considered. Our study showed that the ML algorithm could optimize V2V power consumption in truck platooning. ML-PA significantly saved power by reaching maximum of 56.67% while limiting the BLER to 0.00% in close communication range.

Acknowledgements

References

[1] Khoder R, Naja R, Ismail S, Mouawad N, Tohme S. Vertical Handover Decision using Machine Learning in Vehicular Platooning. *2021 3rd IEEE Middle East and North Africa COMMunications Conference (MENACOMM)*, Agadir, Morocco, 2021, p. 60-64, doi: 10.1109/MENACOMM50742.2021.9678250.

[2] Prayitno A, Nilkhamhang I. V2V Network Topologies for Vehicle Platoons with Cooperative State Variable Feedback Control. *2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, Bangkok, Thailand, 2021, p. 1-4, doi: 10.1109/ICA-SYMP50206.2021.9358435.

[3] Mei J, Zheng K, Zhao L, Lei L, Wang X. Joint Radio Resource Allocation and Control for Vehicle Platooning in LTE-V2V Network. in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, p. 12218-12230, Dec. 2018, doi: 10.1109/TVT.2018.2874722.

[4] Tomimoto K, Serizawa K, Miyashita M, Yamaguchi R, Fukusako T. Antennas and Propagation Technologies of V2V Communications for Platooning. *2020 14th European Conference on Antennas and Propagation (EuCAP)*, Copenhagen, Denmark, 2020, p. 1-4, doi: 10.23919/EuCAP48036.2020.9135759.

[5] Drago M, Zugno T, Polese M, Giordani M, Zorzi M. MilliCar – An ns-3 Module for MmWave NR V2X Networks. *12th Workshop on ns-3, June 2020.*

[6] 3GPP. TR 37.885 V15.3.0 (2019-06) Study on evaluation methodology of new Vehicle-to-Everything (V2X) use cases for LTE and NR. *Technical Report*, 2019.

CLASSIFICATION OF SENTINEL-2 IMAGES FOR THAILAND'S PLANTATION DETECTION

Wongnaret Khantuwan[1*], Chaiyasit Tanchotsrinon[1], Li-Yu Chang[2], Noppadon Khiripet[1]

[1]Knowledge Elicitation and Archiving Laboratory (KEA),
National Electronics and Computer Technology Center (NECTEC),
112 Phahonyothin Road, Khlong Nueng, Khlong Luang District, Pathumthani, 12120, Thailand
[2]Satellite Data Processing Division
Taiwan Space Agency (TASA)
8F, 9 Prosperity 1st Road, Hsinchu Science Park, HsinChu City 300, Taiwan (R.O.C.)

*Corresponding address (E-mail: wongnaret.kha@nectec.or.th, +66-25646900 ext.2273)

Abstract

Land use and land cover data are essential for managing large-scale plantations and precision farming. However, traditional survey methods for mapping land use over large areas are expensive in terms of time, labor, and costs. Fortunately, remote sensing techniques have become more robust, sensible, up-to-date, and inexpensive due to technological advances. However, detecting crops, such as rice paddy fields and sugarcane plantations, is a significant challenge for remote sensing. Existing remote sensing methods for identifying plantation areas, such as false color composite and change vector analysis, still need to be improved in accuracy. They can only be widely applied to some regions within the country.

Additionally, these methods are only applicable for analyzing data during specific periods, which means the identification of cultivated areas is only accurate during certain stages of the growing season. Therefore, this study aims to improve a method for detecting these crops using Sentinel-2 satellite images. The study begins with surface reflectance interpretation and noise reduction, then identifies time series surface reflectance value patterns. The process involves supervised learning to differentiate between rice and sugarcane plantations based on their vegetation index change signatures. The results demonstrate that the Long Shot Term Memory (LSTM) method, which uses the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-up Index (NDBI), can classify rice paddy fields and sugarcane plantations in the testing area with an overall accuracy of 85.27% for rice paddy field detecting and 84.80% for sugarcane plantation detecting using Sentinel-2 images which is high enough for nation-wide agricultural management.

Keywords Crop Detection; LSTM; Remote Sensing; Satellite Image; Sentinel-2;

1. Introduction

Thailand relies heavily on rice and sugarcane as its most important crop in various aspects. It serves two major roles in the country's agricultural exports. Thailand, the second-largest exporter of rice and sugar globally, has achieved remarkable production levels in both crops. The Ministry of Agriculture reported a paddy production of 28 million metric tons for the primary and second crops combined in the 2020–2021 season [1]. Additionally, the production volume of sugarcane in Thailand reached approximately 66.8 million metric tons in 2021 [2]. These impressive figures solidify Thailand's position as a significant international rice and sugar market player. Given the significance of rice and sugarcane, governments have prioritized supporting farmers and addressing their needs. Over the years, various policies have been implemented to assist farmers, including schemes that guarantee prices, pledging initiatives, and financial aid to cover production costs.

Land use data in Thailand is essential and crucial information encompassing various aspects. This includes reporting land use statistics, areas of economically important crop cultivation, and, most

importantly, providing initial data necessary for policy-oriented cultivation planning to address issues such as market oversupply and low prices. It also enables monitoring and forecasting of production outcomes in large-scale areas. Therefore, the accuracy and timeliness of surveying and data collection are paramount. However, the current methods of obtaining land use data have several limitations. It can be time-consuming, labor-intensive, and monotonous, resulting in delayed availability of the survey results within the required timeframe. Furthermore, due to the extensive coverage area and limited resources, the LDD can only conduct a nationwide assessment of agricultural land use, including rice fields and sugarcane plantations, every two years to reflect significant changes in land use patterns.

The current trend in detecting land-use changes involves leveraging geographic information systems and satellite image processing technologies. The goal is to develop methods and services to monitor and track land-use changes effectively. In the present era, modern Earth Observation (EO) satellites offer extensive global coverage and generate increasingly large volumes of data, which is freely accessible. This availability of open data eliminates the concern of data scarcity. Various methodologies for change detection have been proposed to address this data abundance. Change Vector Analysis (CVA) is an effective and widely used unsupervised algorithm in remote sensing research. It enables the differentiation of pixels as either changed or unchanged by analyzing time-series images. Over the years, several variants of CVA have been developed through continuous research, incorporating improvements and addressing specific challenges. For instance, the tri-temporal logic-verified change vector analysis (TLCVA) approach enhances CVA by utilizing logical reasoning and additional assistance to identify errors [3]. Another example is the extended CVA, which incorporates textural change information alongside the traditional spectral-based CVA, enabling the detection of changes in rural-urban fringe areas [4].

New approaches are being explored to address the problem, utilizing machine learning algorithms (MLAs) as more precise and efficient alternatives. For instance, the partial Lanczos extreme learning machine (PL-ELM) offers fast machine learning speed and has shown promise [5]. Another example is using long short-term memory (LSTM) to detect rice crops through the Sentinel-1 time series[6]. Recent reviews highlight Artificial Neural Networks (ANN) and Random Forest (RF) algorithms as the top performers in land use classification. At the same time, the non-parametric classifier Spectral Angle Mapper (SAM) consistently demonstrates high accuracy levels[7].

However, in the crop detection process, features that capture the specific characteristics of each type of plant growth are crucial. Previous studies using MODIS satellite data [8], [9] have shown that satellite indices related to plant health, such as Normalized difference vegetation index (NDVI) [10], Enhanced vegetation index (EVI) [11], and indices related to water, such as Normalized difference vegetation index (NDWI) [12] and Land Surface Water Index (LSWI) [13], can effectively classify cultivated areas. However, due to the low spatial resolution of MODIS (only 250 meters) [14], which is unsuitable for small-scale cultivation areas in Thailand, data from other satellite sources, such as Sentinel-2, should be used. Sentinel-2 [15] provides a higher spatial resolution of up to 10 meters, making it more suitable for cultivation areas in Thailand.

This article aims to study and identify features that can accurately classify cultivation areas, particularly for rice and sugarcane, using time series data from Sentinel-2 satellites. The content will cover details of Sentinel-2 data and the calculation of various indices representing plant growth, water sources, and built-up areas. The experimental design and methodology will be explained, followed by the presentation of results and discussions, showcasing the use of satellite products for machine learning to identify cultivated areas. Finally, the study findings will be summarized in the conclusion section.

2. Remote sensing data

2.1. *Sentinel-2 Imagery*

Sentinel-2 [15] is a satellite mission operated by the European Space Agency (ESA) that provides global surface imaging in various spectral bands, image resolutions, temporal frequencies, and suitable continuity for environmental monitoring and agricultural purposes, free of charge to users worldwide. The Sentinel-2 mission consists of two identical satellites (S2A, S2B) in sun-synchronous orbit, passing over Thailand

between 10 to 11 AM local time. Both satellites orbit in the same plane but on opposite sides of the Earth at a given time. They are combining S2A and S2B data results in revisiting the same area every five days (5-day revisit frequency). The acquired images have spatial resolutions of approximately 10 meters, 20 meters, and 60 meters per pixel, depending on the spectral band. These data can be applied in various ways, such as crop yield prediction, disaster impact assessment, forest monitoring, urban planning, water resource management, etc. Table 1 illustrates the data characteristics in different spectral bands, covering the visible light range to the short-wave infrared, as shown below.

Table 1. Sentinel-2 Data Classification by Spectral Bands and Image Resolution.

Sentinel-2 bands	Central wavelength (μm)	Resolution (m)
Band 1 – Coastal aerosol	0.443	60
Band 2 – Blue	0.490	10
Band 3 – Green	0.560	10
Band 4 – Red	0.665	10
Band 5 – Vegetation red edge	0.705	20
Band 6 – Vegetation red edge	0.740	20
Band 7 – Vegetation red edge	0.783	20
Band 8 – NIR	0.842	10
Band 8A – Vegetation red edge	0.865	20
Band 9 – Water vapour	0.945	60
Band 10 – SWIR – Cirrius	1.375	60
Band 11 – SWIR	1.610	20
Band 12 – SWIR	2.190	20

2.2. *LDD's agricultural land use data*

In addition to satellite imagery, another important dataset is the actual field survey data of agricultural areas. This is important for verifying the crop types cultivated in each area. The Department of Land Development collects the survey data through continuous field surveys, ensuring nationwide coverage and the most up-to-date information available. In this project, the research team utilized two sets of survey data: rice cultivation areas and sugarcane plantation areas. These datasets served as ground truth for developing models to identify rice and sugarcane cultivation areas from satellite images. Both datasets were last updated in 2019, corresponding to the same dataset available on the Agri-Map online system [16].

The following section will describe the experimental design process and how satellite imagery and agricultural land use are employed.

3. Methodology

3.1. *Experimental areas*

In developing a model for identifying cultivation areas, the researchers selected important rice and sugarcane cultivation areas in various regions of Thailand, which are significant for rice and sugarcane production in the country, covering 14,871 square kilometers. The selected regions include the provinces of Sukhothai, Kamphaeng Phet, Phitsanulok, Uthai Thani, Chai Nat, Suphan Buri, Chaiyaphum, Khon Kaen, Udon Thani, Prachinburi, Chachoengsao, and Chonburi, as shown in Fig 1.
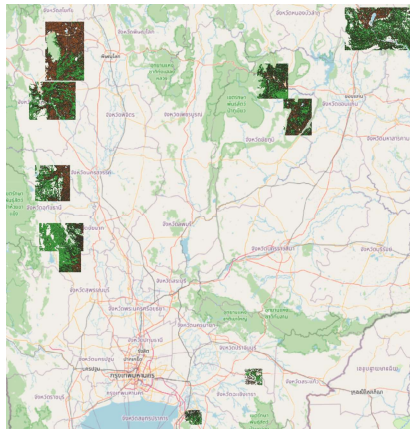

Fig. 1. Experimental areas

These selected areas are divided into three groups based on the land use dataset from the Department of Land Development, which serves as representative categories of different cultivation areas:
- Rice cultivation areas
- Sugarcane cultivation areas
- Other types of land use areas

3.2. *Satellite products*

For the Sentinel-2 satellite imagery, three additional indices, namely the Normalized Difference Vegetation Index (NDVI) [17], Normalized Difference Water Index (NDWI) [12], and Normalized Difference Built-up Index (NDBI) [18], are computed to create a time-series dataset for crop detection purposes. These three indices can be calculated using the following equations:

The Normalized Difference Vegetation Index (NDVI) is a widely used method that calculates the ratio of the difference in reflectance between near-infrared (NIR) and red wavelengths to the sum of their reflectance. This normalization helps to standardize the distribution pattern. The NDVI equation is as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

In the range of -1 to 1, NDVI values greater than 0 indicate surfaces with vegetation cover, where the reflectance in the near-infrared spectrum is higher than in the red spectrum. NDVI values close to 0 represent surfaces without vegetation cover, with similar reflectance in both spectrums. Negative NDVI values correspond to water surfaces, where the reflectance in the near-infrared spectrum is lower than in the red spectrum.

Normalized Difference Water Index (NDWI) is a measurement that quantifies the changes in water content on the Earth's surface. It is calculated by taking the difference between the reflectance values in the near-infrared (NIR) and short-wave infrared (SWIR) spectra and dividing it by their sum. The NDWI equation is as follows:

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR}$$

The resulting NDWI values range from -1 to 1. Positive NDWI values indicate areas with a higher presence of water, where the NIR reflectance is greater than the SWIR reflectance. Values closer to 1 represent surfaces covered by water bodies. On the other hand, negative NDWI values indicate areas with less water content, such as vegetation or bare soil, where the SWIR reflectance is higher than the NIR reflectance. Values closer to -1 often correspond to dense vegetation. NDWI is a helpful tool for identifying and monitoring water bodies, assessing changes in water content over time, and analyzing hydrological patterns. It finds applications in various fields, including environmental monitoring, agriculture, and water resource management.

Normalized Difference Built-up Index (NDBI) is a method used to assess the presence and density of built-up structures such as buildings, roads, and man-made infrastructure. It is calculated by taking the difference between the reflectance values in the short-wave infrared (SWIR) and near-infrared (NIR) spectra and dividing it by their sum. The NDBI equation is as follows:

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$$

The resulting NDBI values range from -1 to 1. Positive NDBI values indicate areas with a higher density of built-up structures where the SWIR reflectance is greater than the NIR reflectance. Values closer to 1 represent dense urban areas with significant human-made constructions. On the other hand, negative NDBI values indicate areas with lower density or absence of built-up structures, such as vegetation or open spaces, where the NIR reflectance is higher than the SWIR reflectance. Values closer to -1 often correspond to rural or natural areas.

The NDBI index is utilized in this research to effectively differentiate built-up areas from agricultural areas, resulting in improved accuracy in land use classification. This enables more precise identification of cultivated areas.

3.3. *Eliminating Interference in Cloud-Covered Areas*

Sentinel-2 satellite captures high-resolution images of the Earth's surface, providing valuable insights for various applications. However, cloud formations in the atmospheric layers often disrupt the data acquired by Sentinel-2. In the tropical region, Thailand experiences frequent cloud cover, which can significantly impact data collection. In some instances, the presence of clouds can persist for several weeks, resulting in an incomplete representation of Earth's surface during periods of cloud interference. Consequently, the data collected during cloud-covered periods may not accurately reflect the actual conditions on the ground.

One potential solution to overcome the challenge of cloud interference is the analysis of time-series data to identify and repair cloud-distorted signals. This approach involves applying techniques such as smoothing and interpolating to restore the missing or disrupted data caused by cloud cover. In this research, the RMMEH method [19] was selected to repair the time-series data affected by cloud interference in Sentinel-2 satellite imagery. RMMEH, specifically designed for restoring time-series data in satellite imagery compromised by cloud cover, proves to be an effective technique for data restoration.

4. Experimental results and discussions

In identifying agricultural areas, researchers randomly selected three groups of image points from significant rice and sugarcane-producing regions, as well as other land types. The image points were chosen within ten previously described regions, covering 14,871 square kilometers. This sample represented various types of cultivation, totaling 300,000 pixels or an area equivalent to 30 square kilometers.

The selected dataset was then processed to calculate the NDVI, NDWI, and NDBI indices while addressing data preparation tasks such as eliminating noise caused by cloud cover, as depicted in Fig. 2.
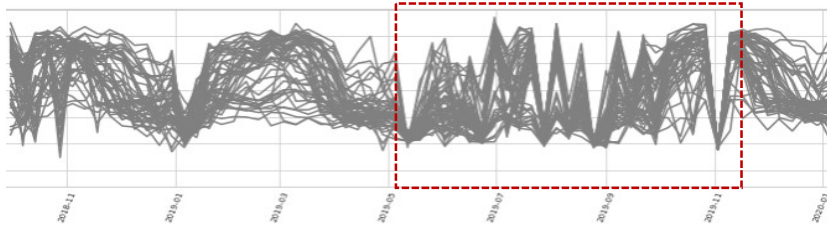

Fig. 2. Example of NDVI signal disturbed by cloud cover (Red Line)

During periods of cloud cover, the NDVI values significantly decrease, deviating from those observed during nearby periods. It is necessary to remove this interference to restore the reflectance values across different wavelengths. For agricultural areas, the reflectance values indicate plant growth, which should ideally exhibit incremental or decremental characteristics. To eliminate noise, data from nearby periods were used. In this research, a smoothing function based on the RMMEH method [19] was employed. This approach used sliding windows to consider average or median values within the window as replacements for cloud-affected data. Moving averages were then calculated on the resulting time-series data. The window size for the smoothing process in this project encompassed a five-cycle orbit or approximately one month of data, aiding in noise removal.

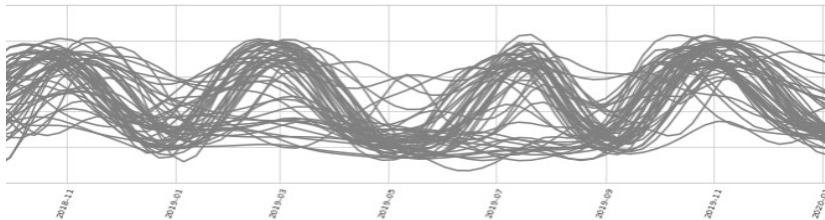After eliminating data noise, the visual representation of the noise-removed data is shown in Fig. 3.

Fig. 3. Example of NDVI signal after noise removal using the RMMEH method.

Next, a random selection of images from the prepared dataset was used to demonstrate the changes in NDVI values based on different types of cultivation, as depicted in Fig. 4.
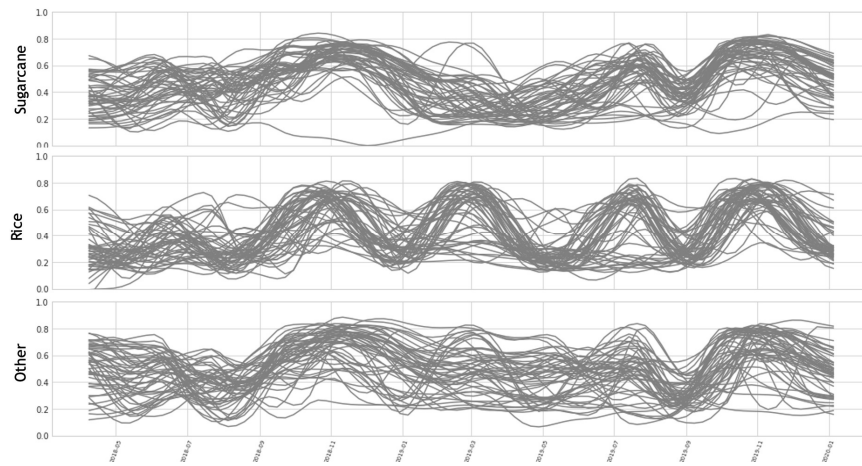
Fig. 4. NDVI variation by time and land use type: sugarcane plantation (Top), Rice paddy field (Middle), and other areas (Bottom)

The figure shows that each crop type exhibits distinct patterns and unique changes in the NDVI values. These characteristics enable machine learning algorithms to classify different types of crops effectively.

To determine cultivated areas, the researchers conducted a preliminary experiment in the Mueang District of Kamphaeng Phet Province, covering an area of 323.33 square kilometers (1935 x 1671 pixels). Two Long Short Term Memory (LSTM) models [20] by learning time-series data of NDVI, NDWI, and NDBI were developed and evaluated as follows:

- An LSTM model for identifying rice cultivation areas
- An LSTM model for identifying sugarcane cultivation areas

For the accuracy assessment of rice paddy fields and sugarcane plantations, it was compared with land use data from the 2019 land use survey conducted by the Department of Land Development, Ministry of Agriculture and Cooperatives. The assessment was performed in a test area. The results of the comparison and analysis are presented in the following figures.
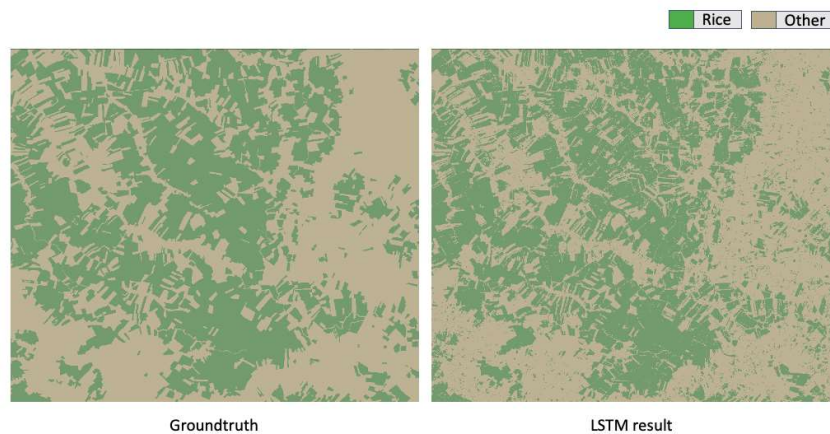

Fig. 5. Comparison of rice paddy field classification (Right) with ground truth from the Department of Agricultural Promotion (Left)
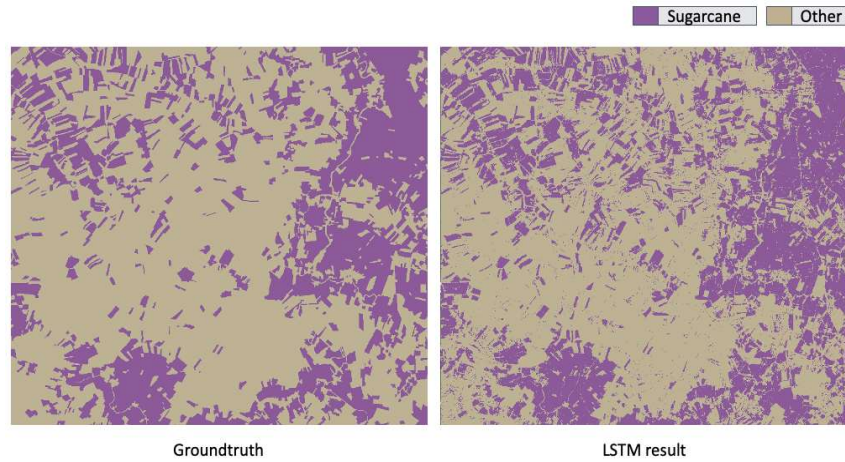
Fig. 6. Comparison of sugarcane plantation classification (Right) with ground truth from the Department of Agricultural Promotion (Left)

Calculating the areas classified as true positive, false positive, true negative, and false negative generated a confusion matrix, as shown in Fig. 7. The accuracy values are summarized in the following table.
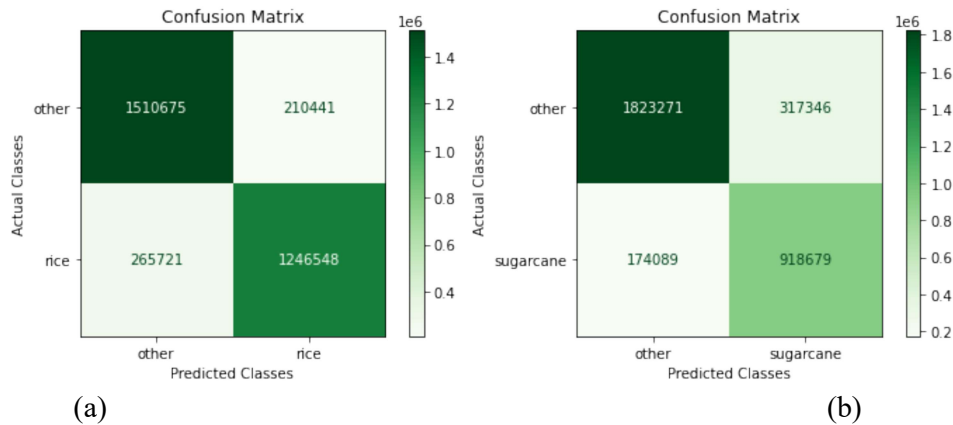

Fig. 7. (a) Comparing the rice paddy field identification accuracy using LSTM. (b) comparing the accuracy of sugarcane plantation identification using LSTM.

Table 2. Crop Identification Accuracy

	Rice paddy field	Sugarcane plantation
Image size	1935 x 1671 pixels	1935 x 1671 pixels
Total pixels	3,233,385 pixels	3,233,385 pixels
Accuracy	85.27%	84.80%
User's accuracy (precision)	85.56%	74.32%
Producer's accuracy (recall)	82.53%	84.07%

From the accuracy values, it can be observed that the model performs reasonably well in identifying rice paddy fields, with an overall accuracy of 85.27% and overall accuracy of 84.80% for sugarcane plantation detection. However, the user's accuracy (the accuracy of correct crop identification when predicted) is 85.56% for paddy field detection and 74.32% for sugarcane plantation detection. In comparison, the producer's accuracy (the accuracy of correct crop predicting when it is present) is 82.53% and 84.07%, respectively.

However, the proposed method still has limitations in several aspects. For example, the land use data from the Department of Land Development may not provide detailed at the lowest level. In other words, there may be contamination of other land uses within the designated agricultural areas, such as water sources and buildings. When using such data for comparing and evaluating the accuracy of classifying agricultural areas, it may affect the accuracy.

As for the development of machine learning models for classifying different types of cultivated areas, this research has developed models specifically for rice fields and sugarcane plantations. When considering the detected areas between rice fields and sugarcane plantations derived from both models, it is evident that there are some areas overlapped. Therefore, the next step in developing a classification method should aim to create a model that can classify multiple types of cultivated areas or consider the collaboration of multiple models to avoid overlapping results.

Furthermore, another challenge is that Thailand is located in a tropical and humid zone, resulting in cloud cover over a significant portion of the country during certain periods. This limited the ability of Sentinel-2 to capture ground images, leading to a lack of data for various experiments during specific time intervals. The researchers believe that it is necessary to study and experiment to understand the use of other sensor data, such as Synthetic Aperture Radar (SAR), which can read ground information without being affected by clouds. However, SAR data from the Sentinel-1 satellite has only two channels, which may affect the accuracy of classifying cultivated areas. Therefore, it may be necessary to consider the combined use of Sentinel-2 and Sentinel-1 data.

5.   Conclusion and future works

This research presented a remote sensing-based approach for classifying rice paddy fields and sugarcane plantations using near real-time data. The method proved to be cost-effective and efficient in various aspects. Through experimentation, it was found that calculating the NDVI, NDWI, and NDBI indices from Sentinel-2 satellite imagery can effectively identify rice fields and sugarcane plantations. The time series patterns of both crops could clearly represent their distinct growth patterns. However, since Sentinel-2 data can be affected by cloud cover, it is necessary to employ suitable methods to mitigate signal interference and enhance the quality of the time-series data. In this study, the RMMEH method was chosen to remove cloud-related interference.

For future research directions, there are several areas that can be further developed. One aspect is the development of classification methods for cultivated areas that can classify multiple types of cultivated areas using a single model or considering the collaborative work of multiple models to avoid overlapping results. Additionally, the combined use of Sentinel-2 and Sentinel-1 data should be explored to eliminate the impact of long-term cloud cover over specific regions. Furthermore, scaling up the data processing capabilities to identify larger scale, such as at the provincial or nationwide, should be considered for future applications.

References

[1] "Rice in Thailand | OEC," *OEC - The Observatory of Economic Complexity*. https://oec.world/en/profile/bilateral-product/rice/reporter/tha (accessed Jun. 06, 2023).

[2] "Thailand: production volume of sugarcane 2016-2022," *Statista*. https://www.statista.com/statistics/1041070/thailand-sugarcane-production-volume/ (accessed Jun. 06, 2023).

[3] P. Du, X. Wang, D. Chen, S. Liu, C. Lin, and Y. Meng, "An improved change detection approach using tri-temporal logic-verified change vector analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 278–293, Mar. 2020, doi: 10.1016/j.isprsjprs.2020.01.026.

[4] C. He, A. Wei, P. Shi, Q. Zhang, and Y. Zhao, "Detecting land-use/land-cover change in rural–urban fringe areas using extended change-vector analysis," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 4, pp. 572–585, Aug. 2011, doi: 10.1016/j.jag.2011.03.002.

[5] N.-B. Chang, M. Han, W. Yao, L. Chen, and S. Xu, "Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine," *Journal of Applied Remote Sensing*, vol. 4, Nov. 2010, doi: 10.1117/1.3518096.

[6] H. Crisóstomo de Castro Filho *et al.*, "Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series," *Remote Sensing*, vol. 12, no. 16, Art. no. 16, Jan. 2020, doi: 10.3390/rs12162655.

[7] S. Talukdar *et al.*, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sensing*, vol. 12, no. 7, Art. no. 7, Jan. 2020, doi: 10.3390/rs12071135.

[8] X. Xiao *et al.*, "Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images," *Remote Sensing of Environment*, vol. 100, no. 1, pp. 95–113, Jan. 2006, doi: 10.1016/j.rse.2005.10.004.

[9] T. Sakamoto, N. Van Nguyen, A. Kotera, H. Ohno, N. Ishitsuka, and M. Yokozawa, "Detecting temporal changes in the extent of annual flooding within the Cambodia and the Vietnamese Mekong Delta from MODIS time-series imagery," *Remote Sensing of Environment*, vol. 109, no. 3, pp. 295–313, Aug. 2007, doi: 10.1016/j.rse.2007.01.011.

[10] S.-H. by Sinergise, "Normalized difference vegetation index," *Sentinel-Hub custom scripts*, 2021. https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndvi/ (accessed Mar. 22, 2021).

[11] "Measuring Vegetation (NDVI & EVI)," Aug. 30, 2000. https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php (accessed Jun. 28, 2022).

[12] S.-H. by Sinergise, "NDWI Normalized Difference Water Index," *Sentinel-Hub custom scripts*, 2021. https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndwi/ (accessed Mar. 22, 2021).

[13] I. Caballero, A. Román, A. Tovar-Sánchez, and G. Navarro, "Water quality monitoring with Sentinel-2 and Landsat-8 satellites during the 2021 volcanic eruption in La Palma (Canary Islands)," *Science of The Total Environment*, vol. 822, p. 153433, May 2022, doi: 10.1016/j.scitotenv.2022.153433.

[14] "Moderate Resolution Imaging Spectroradiometer (MODIS) - LAADS DAAC." https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/modis/ (accessed Jun. 06, 2023).

[15] THE EUROPEAN SPACE AGENCY, "Sentinel-2 - Missions - Sentinel Online - Sentinel," 2022. https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed Apr. 09, 2021).

[16] Ministry of Agriculture and Cooperatives and National Electronics and Computer Technology Center, "Agri-Map - Sign in," *Agri-Map*, 2022. https://agri-map-online.moac.go.th/login (accessed Jun. 28, 2022).

[17] S. Montoya, "How to calculate the NDVI Index from a Sentinel 2 Image," *Hatari Labs*, Apr. 2017. https://www.hatarilabs.com/ih-en/how-to-calculate-the-ndvi-index-from-a-sentinel-2-image (accessed Mar. 22, 2021).

[18] Y. Zheng, L. Tang, and H. Wang, "An improved approach for monitoring urban built-up areas by combining NPP-VIIRS nighttime light, NDVI, NDWI, and NDBI," *Journal of Cleaner Production*, vol. 328, p. 129488, Dec. 2021, doi: 10.1016/j.jclepro.2021.129488.

[19] Z. Jin and B. Xu, "A novel compound smoother - RMMEH to reconstruct MODIS NDVI time series," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, pp. 942–946, Jul. 2013, doi: 10.1109/LGRS.2013.2253760.

[20] H. Tian, P. Wang, K. Tansey, J. Zhang, S. Zhang, and H. Li, "An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China," *Agricultural and Forest Meteorology*, vol. 310, p. 108629, Nov. 2021, doi: 10.1016/j.agrformet.2021.108629.

AGRICULTURAL DATA SHARING TO APPLICATIONS VIA THAGRI

Jutarat Khiripet[1]*, Wongnaret Khantuwan[1], Pornpimon Sinpayak[1] and Noppadon Khiripet[1]

[1]*Knowledge Elicitation and Archiving Laboratory (KEA),*
National Electronics and Computer Technology Center (NECTEC),
112 Phahonyothin Road, Khlong Nueng, Khlong Luang District, Pathumthani, 12120, Thailand

*Corresponding address (E-mail: jutarat.khiripet@nectec.or.th, +66-25646900 ext. 2221)

Abstract

In today's agricultural sector, data-driven decision-making has replaced intuition-based decision-making. However, due to a lack of collaboration between data and services, data is not being fully utilized, and various types of data are not being organized efficiently. This deficiency significantly hinders the development of new agricultural applications and services. We have designed and implemented the Thailand Agricultural Data Collaboration Platform (THAGRI) to encourage the sharing of agricultural data, which may not be open data. The platform facilitates the exchange of various types of data including farmland data, soil data, weather data, market locations, growth prediction data, etc. This study aims to develop a platform that enables developers and data providers from both public and private organizations to collaborate and create innovative applications. The platform, called THAGRI, was tested in this preliminary work, and its utility was demonstrated in the creation of several applications, including the Agri-Map, a national agricultural policy tool, the Din-Dee Smart Chatbot, and a crop yield prediction analysis. The results of the study indicate that THAGRI is an effective tool for facilitating the creation of novel applications in the agricultural sector.

Keywords: THAGRI; Agri-Map; data collaboration; data platform; data sharing

1. Background

Nowadays, decisions are made based on data instead of intuition, especially in agricultural sector. Following a data-driven uses facts and hard information rather than instinct to make confident decision. Due to a lack of mutual collaboration between data and services, data is not being fully utilized and insufficient organization of various types of data. A data platform must be constructed that is the lake of combing reliable information and the center of building research knowledge or service from data based on objectives of users. The several digital agricultural data platforms have been founded currently operating under a variety of governance models (Table 1) [6]. This community includes both users that contribute data or applications, as well as participants who simply provide data or who access to complete information from aggregated data. Some of those platforms are independent software companies such as Farmers Business Network [7], Farmobile [8], Granular [9] or Climate Corporation [10]. Moreover, there are many platforms for big data in agriculture which make use of the power of big data for agricultural research and development.

WAGRI [1,2,3] is intended to promote the collaboration of agricultural data and the growth of agriculture based on that data. This application allows developers and companies to both receive and provide various types of data including farmland data, weather data, soil data, historical yield data, market condition data and growth prediction system. The most important goal is to deliver those significant data to farmers and companies in the country such that they can improve work plan based on their business style.

ADC [11] is a nonprofit organization focused on connecting the data about food and agriculture. They concern about two problems. First is that too often farmers are beholden to one technology or service provider due to the difficulty in retrieving the data from that entity to use within other tools that may or

may not be competitive. Second, in general, data storage, data sharing, insufficient data documentation, and a lack of connection between observation and theory have been identified as limitations to the potential for big ag data today. To fix mentioned problem, their objectives are to educate the agricultural industry and public about the value of agricultural data and its impact or potential, to support collaborative data research, and to facilitate a data platform for non-commercial research and development based on data owner permission.

WinField United [12] is one of platform founded by America's premier agricultural business and food companies. The uses of data, contributed from this platform, are many new products and techniques to help farmers, retailers and other partners succeed. For example, R7 Field application is the forecasting tool that simulate daily crop growth to monitor predicted yields and assess optimal timing for nutrient and water. The platform collects and stores data to support creating greater value, where the total value generated will be the product of the benefit accrued by the data, the benefit itself generated by the system, and the benefit received by the user community. In additional, stakeholders will use the data that is developed by the system to achieve better results for the same cost. Instead of keeping data in private, data can be provided as a fungible good that is offered by the companies in exchange for higher-value information.

Table 1. Digital agricultural data platforms.

Data Platform	URL	Region	Source
WAGRI	https://wagri.net/en-us/	ASIA	Private
THAGRI	https://thagri.in.th/	ASIA	Private
FaST	https://fastplatform.eu	EU	Public
Climate Corporation	https://climate.com	EU	Private
New Vision Coop	https://newvision.coop	USA	Private
WinField United	https://www.winfieldunited.com	USA	Private
ADC	https://agdatacoalition.org	USA	Private
GISC	https://www.gisc.coop	USA	Private

In Thailand, to empower the use of data, *THAGRI* [4,5] is implemented as an agricultural data sharing platform of Thailand which provides agricultural data or services through an application programming interface (API). The main goals are to conduct, support, facilitate collaborative data research and support a data platform for non-commercial research and development based on data owner permission. This platform is a center of collaborating many developers and companies for sharing data and developing service through THAGRI including Thailand's National Electronics and Computer Technology Center (NECTEC), Ministry of Agriculture and Cooperatives, Land Development Department (LDD), National Agricultural Big data Center (NABC), Government Big Data Institute (GBDi) etc.

In this paper, we introduce the design and implementation of the THAGRI platform to support the above-mentioned goals. The primary function of the platform is to gather and maintain data that contributes to value creation. The overall value produced will be calculated by multiplying the benefits provided by the data, the benefits generated by the system, and the benefits received by the user community [14, 15]. The research objective is to create a digital data aggregator that is easily accessible and serves as a foundational component for a future digital platform through THAGRI. The goal is to develop an intelligent system that employs data integration and machine learning techniques to offer decision support for agriculture, with a specific focus on the industrial crop sector. There are example applications and research that have been developed by obtaining data from the THAGRI platform in Thailand such as Agri-Map [13], *Nong Din Dee* (AI Chatbot) [16], crop yield prediction model.

2. THAGRI Data and Implementation

The data for modern agricultural sector provided from THAGRI including long-range climate forecasts, crop production, local weather station data, GIS mapping shapefile or raster, industry historical data (past yields, market data) etc. The benefits of THAGRI platform are three main points. *First,* the platform can

consolidate data and can exact insights from any type and size of data sources in one place. *Second*, low time consumption is one of important reason because the users can request to access data through one platform under the conditions of data owner. Moreover, the platform provides ability to make critical business decisions based on valuable, timely insights and quicker time-to-market. *Third*, the platform comes with enterprise security and data catalog.

2.1   *API catalog*

Currently, THAGRI API Catalog publish at https://catalog.thagri.int.th which designed to help data professionals quickly find the most appropriate data for any analytical or ad hoc purpose. In Fig.1 capture THAGRI API Catalog homepage that offers 6 data groups, comprising 78 datasets, representing a total of 642 APIs) most of which come from Agri-Map.
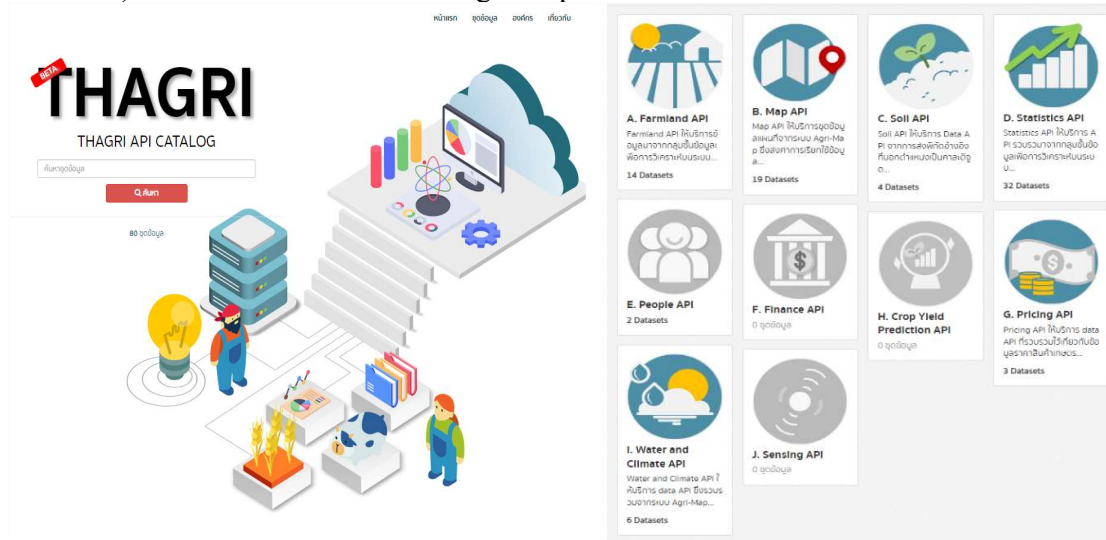

Fig. 1. THAGRI API Catalog.

The 6 groups of data services on the THAGRI API Catalog are provide the directed to API that cut off the digging connected to DB. The API services accept the request use the reference coordinates indicating the location as Latitude and Longitude in requesting detailed information about the selected API at such coordinates. Another way, the API can be called by specifying an administrative area (province, district, subdistrict) or watershed (22 watersheds). The following section is representing the 6 data groups follow as:

1. Farmland API: This provides data from the Agri-Map analytical layers, through spatial analysis, geographic data stored in vector format (Vector), including geospatial data provided by organizations.
2. Map API: This provides geospatial data services on the Agri-Map system in the form of web map tile service and point, which is under OGC standard.
3. Soil API: This API gather information layers from analytics on an Agri-Map system that has been analysed spatially with shape file related to soil series data, soil suitability and land use.
4. Statistic API: This API provides API services gathered from data layers for analysis on the Agri-Map system through spatial analysis with shape, point, as well as related statistical analysis from other agencies.
5. Pricing API: This API provides agricultural product price data. For example, information from Agri-Map, which is an analysis of prices, costs, returns of economic crops. Or information from the National Agricultural Information Center (NABC), which provides an API, the prices farmers sell at the average farm. National and regional levels since 1997 - present, etc.

6. Water and Climate API: A data API service collected from Agri-Map systems and contributing organizations such as the Hydro Informatics Institute. (Public Organization), Royal Irrigation Department, Land Development Department, etc. For weather datasets include daily and future weather data, rain forecasts. Forecasting drought or flooding in terms of data sets about water sources such as water levels in large and medium-sized reservoirs, etc. The data provided has been synthesized to be ready to be used as an API.

2.2 *THAGRI architecture*

THAGRI" is a data service in the form of Application Programming Interface (API) that can be used in conjunction with the App to create cooperation in the use of data (Data Collaboration) and data sharing (Data Sharing) between the public and private sectors accurately and precise. Moreover, THAGRI increases channel for bringing information to further develop into new innovations efficiently. The architecture of THAGRI in Fig.1 is designed to support the above requirements as follows:

A. User groups are both public and private agencies that want to use information from THAGRI for the purpose of developing work according to their objectives. Including a group that has registered as a member of THAGRI, such as the Department of Land Development, calling the API on AI Chatbot, soil, or the Government Big Data Analysis and Management Institute (ONK) calling the Agri-Map APIs data group to develop a Dashboard. solving provincial problems, etc.

B. THAGRI Platform creates data services in the form of Application Programming Interface (API) by exchanging data (Data Sharing) between groups of data owners. and user groups

C. The data owner is a public and private agency that sees the benefits of data sharing by being the person who has the right to place data items on THAGRI's data accounting system, including the right to approve data items. upon request from the user
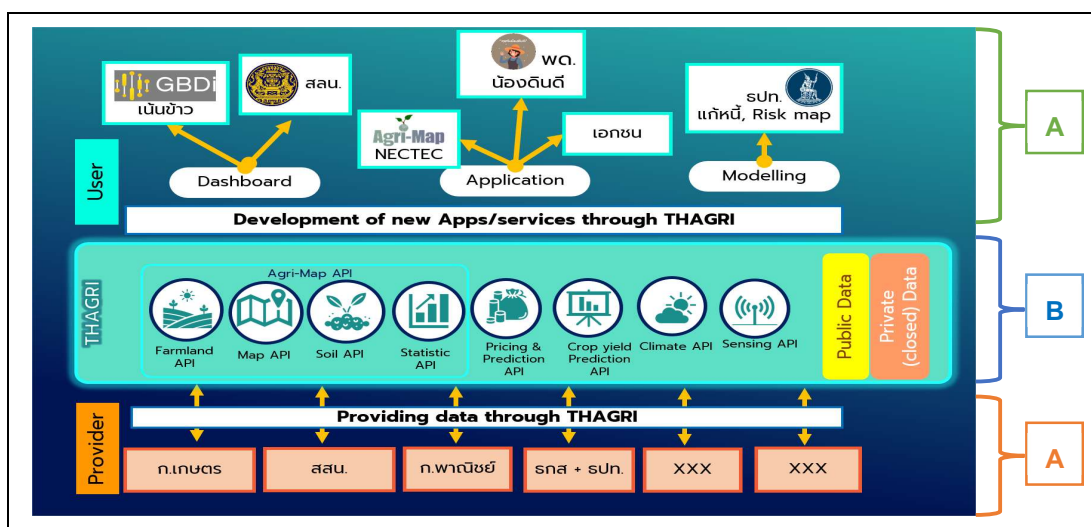
Fig. 2. THAGRI architecture.

3. Example Applications

3.1 *Agri-Map*

Agri-Map [13] is well-known GIS application for Agriculture management based on agriculture information from the Ministry of Agriculture. Moreover, the vary of data in the application provides the capability to extract information and insights at a larger scale to benefit those upstream or downstream of the farm more than farmers themselves. In addition, insight data which is delivered from the data platform enable the economic growth of the agricultural sector by that the use of digital agricultural technologies

with new decision-making tools, for example, farm management, and product pricing. Agri-Map provides an efficient way to comprehend large amounts of data quickly.

Agri-Map is a valuable tool used in agriculture for various purposes. It serves as a platform for aggregating and visualizing agricultural data, enabling users to gain a comprehensive understanding of the information at hand. By presenting data in a visually appealing and easily interpretable format, Agri-Map facilitates quick decision-making and aids in identifying patterns, trends, and insights related to agriculture.

One of the primary benefits of Agri-Map is its ability to consolidate and display diverse agricultural data sets in a unified manner. It can integrate data from multiple sources, such as weather patterns, soil conditions, crop yields, pest outbreaks, and market prices. By combining these different data points, Agri-Map provides a holistic view of the agricultural landscape, helping farmers, researchers, and policymakers make informed choices.

Agri-Map's visualization capabilities are particularly helpful in simplifying complex data sets. By transforming raw data into interactive maps, charts, and graphs, it enables users to grasp key information quickly. For example, farmers can easily identify areas with higher crop yields, analyze the distribution of pests, or visualize soil variations across their fields. Such visual representations help in optimizing resource allocation, identifying areas for improvement, and enhancing overall agricultural productivity.

While Agri-Map facilitates data analysis and visualization, it's important to note that it may not directly provide real-time or specific data upon user request. Users and companies might need to access data from other sources or platforms, and then utilize Agri-Map as a tool to consolidate and interpret that data effectively.

Overall, Agri-Map plays a crucial role in supporting evidence-based decision-making in agriculture by simplifying data visualization, enabling data integration, and providing a comprehensive understanding of the agricultural landscape.

Fig. 2. Agri-Map.

3.2 *Nong Din Dee*

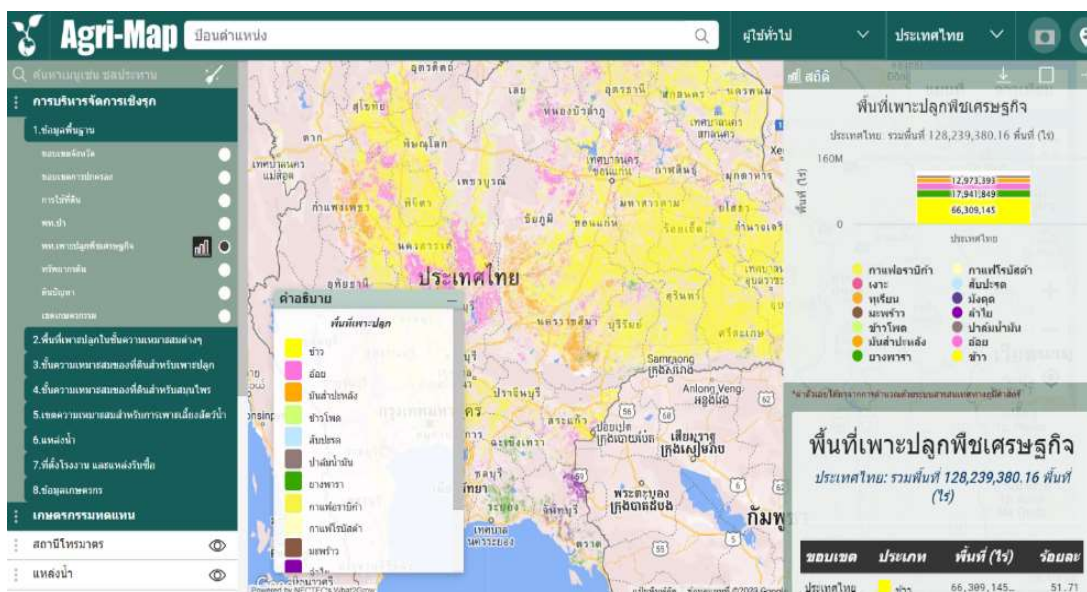Nong Din Dee, the AI chatbot, is specifically designed to provide information and assistance in the areas of soil quality, economic crops, and land use. It is equipped with a comprehensive knowledge base that covers these topics, allowing it to deliver accurate and relevant responses to user inquiries.

1) Soil Quality: Nong Din Dee can answer questions related to soil characteristics, such as soil

composition, pH levels, nutrient content, moisture levels, and soil fertility. It can provide guidance on soil testing methods, soil amendment recommendations, and best practices for soil management. Users can inquire about specific crops and their suitability for different soil types, as well as techniques for improving soil quality and addressing common soil-related issues.

2) Economic Crops: Nong Din Dee possesses information about various economic crops, including their cultivation practices, growth requirements, pest and disease management, harvesting techniques, and market trends. Users can seek guidance on selecting appropriate crops based on factors like climate, soil type, market demand, and profitability. The chatbot can provide insights on crop rotation, intercropping, and other sustainable farming practices to optimize economic crop production.

3) Land Use: Nong Din Dee can offer information on land use planning and management. It can provide guidance on land suitability assessments for different purposes, such as agriculture, residential development, or conservation. Users can inquire about zoning regulations, land productivity evaluations, and land-use optimization strategies. The chatbot can also offer insights on sustainable land management practices, land conservation methods, and the impact of land use on the environment.

Nong Din Dee's knowledge base is regularly updated to ensure that it incorporates the latest research, best practices, and industry trends in the fields of soil quality, economic crops, and land use. It aims to provide users with accurate and valuable information, assisting them in making informed decisions and optimizing their agricultural practices and land management strategies.

Fig. 3. Nong Din Dee.

3.3 *Crop yield prediction*

One of the interesting agricultural markets is yield prediction. Crop yield prediction can be accomplished by many tools such as a multiple linear regression model (MLR). This statistical technique allows us to analyze the relationship between multiple independent variables (such as weather conditions, soil properties, etc.) and the dependent variable, which in this case is the crop yield. In this case, KEA laboratory has been studied and implemented data to forecast cane yield by using MLR obtained from THAGRI data platform. The goal of this research is to examine how well different methods can predict how much sugarcane will be produced between 2019 and 2021. We will be comparing the accuracy of various algorithms, including Multiple Linear Regression (MLR). In MLR, we will use past sugarcane yields as a basis for our predictions. Furthermore, we will enhance the MLR model by incorporating additional factors like temperature, humidity, precipitation, GDD, and NDVI. These prediction models will rely on different sources of information, such as remote sensing and climate data, to effectively forecast the amount of sugarcane that will be harvested.

4. Conclusions

THAGRI has been designed and implemented to overcome challenges in agricultural big data access and utilization in Thailand. It has collected diverse agricultural data in an integrated manner and provides

a large number of APIs as a service. THAGRI enables developers to create new applications easily by accessing the provided APIs. Moreover, researchers can further gain more insights based on this information to create new analysis models. Two successful applications and one crop yield forcasting example are briefly discuss to illustrated the contribution of this platform. More applications, services and utilization of THAGRI are still in progress and will be made available publicly and commercially in the near future.

References

[1]   About the WAGRI Agricultural Data Collaboration Platform, Agricultural Data Collaboration Platform Office, Research Center for Agricultural Information Technology, National Agriculture and Food Research Organization (NARO), 2021. Accessed: 2021-07-04.
[2]   WAGRI API, https://wagri.net/en-us/wagriapi, 2017. Accessed: 2021-07-04.
[3]   Food safety and technology, https://mygfsi.com/blog/food-safety-and-technology a-look-into-japanese -agricultural-data-collaboration-platform-wagri/, 2018. Accessed: 2021-07-04.
[4]   THAGRI, https://www.nectec.or.th/news/news-pr-news/2021-mou-thagri.html/,2021. Accessed: 2021-07-04.
[5]   THAGRI News, https://thainews.prd.go.th/th/news/detail/TCATG211102172830809, 2021. Accessed: 2021-07-04.
[6]   A Case Study of a Digital Data Platform for the Agricultural Sector: A Valuable Decision Support System for Small Farmers, Agricultural Economics Research Group, Department of Management and Marketing, University of Huelva, 2022. Accessed: 2023-03-25.
[7]   Farmers Business Network. Available online: https://www.fbn.com. Accessed: 2023-03-25.
[8]   Farmobile. Available online: https://www.farmobile.com/data-engine. Accessed: 2023-03-25.
[9]   Granular. Available online: https://granular.ag/granular-insights/. Accessed: 2023-03-25.
[10]   Climate Corporation. Available online: https://climate.com/. Accessed: 2023-03-25.
[11]   ADC. Available online: https://agdatacoalition.org. Accessed: 2023-03-25.
[12]   WinField United. Available online: https://www.winfieldunited.com. Accessed: 2023-03-25.
[13]   Agri-Map. Available online: https://agri-map-online.moac.go.th/login. Accessed: 2023-03-25.
[14]   Wysel, Matthew, Derek Baker, and William Billingsley. "Data sharing platforms: How value is created from agricultural data." Agricultural Systems 193 (2021): 103241.
[15]   Jones, Charles I., and Christopher Tonetti. "Nonrivalry and the Economics of Data." American Economic Review 110.9 (2020): 2819-58.
[16]   Nong Din Dee, Available online: https://www.ldd.go.th/dindee/Download.html. Accessed: 2023-05-11.

DESIGN AND DEVELOPMENT OF THE BACK-END SYSTEM FOR AN ONLINE LEARNING ASSESSMENT PLATFORM

Watcha Sasawattakul[1], Atiwong Suchato[2], Dittaya Wanvarie[3], Naruemon Pratanwanich[4], and Nuengwong Tuaycharoen[5,*]

[1]Department of Computer Engineer, Chulalongkorn University, Bangkok, Thailand 6470271821@student.chula.ac.th
[2]Department of Computer Engineer, Chulalongkorn University, Bangkok, Thailand, atiwong.s@chula.ac.th
[3]Department of Mathematics and Computer Science, Chulalongkorn University, Bangkok, Thailand, dittaya.w@chula.ac.th
[4]Department of Mathematics and Computer Science, Chulalongkorn University, Bangkok, Thailand, naruemon.p@chula.ac.th
[5]Department of Computer Engineer, Chulalongkorn University, Bangkok, Thailand, nuengwong.t@chula.ac.th

*Corresponding author. Tel.:+66-809008362.
E-mail-address: watcha.sasawattakul@gmail.com

Abstract

The COVID-19 pandemic has had a profound global impact on people's behavior, particularly students and faculty members. To facilitate the assessment of learning activities, especially during high-demand periods like midterm and final exams, an online learning assessment platform has been proposed. This paper focuses on the design and development of the platform's back-end system, which utilizes a RESTful API with NodeJS and Typescript. A cloud architecture is proposed to ensure scalability and handle high loads during peak examination periods. The first phase of system design emphasizes the caching technique for efficient data retrieval during critical periods. Data management and reliability are handled by DynamoDB, a non-relational database. The paper also presents the mapping rules used to transfer data from relational to non-relational databases, specifically for data transfer to SQL databases. Performance testing involving 500 user requests demonstrates the system's ability to handle large loads and maintain acceptable response times. User satisfaction is evaluated using the Technology Acceptance Model (TAM), with feedback from 132 student participants. The findings, supported by hypothesis testing using a Structural Equation Model, reveal high user satisfaction with Data Reliability, Perceived Usefulness, and Intention to Use. Perceived Ease of Use also receives positive ratings. However, there is room for improvement in terms of Latency Performance. In conclusion, the proposed platform effectively supports learning assessment during challenging times, with users expressing high satisfaction. Addressing Latency Performance will further enhance the system. These findings contribute to the development of e-learning technologies and provide valuable insights for similar platforms.

Keywords: UML diagram, Software Requirement, Learning Management System, NoSQL databases, Mapping rules, Technology Acceptance Model;

1. Introduction

The COVID-19 pandemic has stimulated a widespread shift to online communication platforms, including learning assessments in universities, such as examinations and quizzes. In order to support continuous learning assessment activities, the proposal of an Online Learning Assessment Platform has emerged. This research paper focuses on the design and development of the back-end system for this platform on first phase. The back-end system must efficiently handle a heavy load during peak hours, which necessitates the implementation of a cloud architecture system.

Compared to the University's existing system known as "MyCourseville" (MCV), which primarily supports informal quizzes and exams throughout the semester, MCV has been constrained by limited

resources and cannot adequately handle formal exams such as mid-term or final exams. These formal exams are typically compressed into one or two weeks and involve all students in the university. Given the critical stakes associated with formal exams, ensuring the reliability of examination data becomes paramount.

To support formal high-stakes examinations during midterm and final exam weeks, the proposed system must efficiently handle a substantial load within a specific timeframe while ensuring high availability for users. This necessitates the implementation of a cloud architecture system capable of handling load changes in the resource system. Additionally, a cache technique is selected to manage the immense load during critical periods, specifically handling test session data that contains all questions and answers for each student. By managing the business logic in memory before transferring it to the main database, the interconnectivity between the server and the main database is minimized.

Effectively managing the massive and complex data, which includes various question types and hierarchical structures, is achieved through the use of non-structural information, particularly a non-relational database such as DynamoDB, in conjunction with caching techniques to handle the critical loads experienced during exams [1]. Given the structural differences between the main relational database and the NoSQL database, mapping rules will be applied to optimize data synchronization Field [2].

Therefore, this research paper proposes the design and development of the back-end system for the Online Learning Assessment Platform, with a specific focus on handling the load during the exam period. It also encompasses database design considerations, particularly the transfer of data from NoSQL to SQL, in order to effectively address these challenges. The system's performance is evaluated through performance testing, and the technology acceptance model (TAM) is employed to assess user acceptance and satisfaction. By adopting these approaches, the paper aims to provide comprehensive solutions to the identified challenges and contribute to the improvement of the Online Learning Assessment Platform.

2. Literature Review

2.1. Designing and Implementing of an Online Management Information System for Exam Committee [3]

This work proposes the development of a web-based Management Information System (MIS) for the exam committee unit at Koya University's Faculty of Engineering (FENG). The project adopts the waterfall methodology and utilizes JavaScript, PHP, Laravel, and MySQL for design and implementation. The system aims to streamline committee tasks and address the challenges arising from the absence of an information system. Extensive testing was conducted to ensure adherence to system requirements, webpage style and formatting, and dynamic data updating, with software testers confirming the system's satisfactory performance. The methodology employed in this paper can serve as a valuable resource for enhancing our own developed system, enabling similar functionality.

2.2. Exam Wizard: A novel e-assessment system [4]

The work presents the development of an e-assessment platform called Exam Wizard for higher education courses. Instructors can generate tests using pre-written questions stored in a database and automatically grade them after submission. The platform supports various question types, test scheduling, and a timer to enhance the testing experience for students. Exam Wizard is free and open source, making it accessible to more educational institutions. The paper highlights Exam Wizard's robustness, scalability, reusability of questions, and resumption capability, which are often missing from existing systems. The proposed system has similar functions to Exam Wizard, especially for database design to support changes

in logic. Integrating the database structure from this research could improve the proposed system's database design.

2.3. System Characteristics in Predicting E-Learning Acceptance: An Extended Technology Acceptance (TAM) Study [5]

This study contributes to the development of e-learning technologies during the COVID-19 pandemic by collecting data from e-learning students at Universiti Teknologi Mara (UiTM) Terengganu Branch in Malaysia through a self-administered questionnaire and online survey. The research employs the Technology Acceptance Model (TAM) with system characteristics to investigate the acceptance of e-learning in developing countries. The study finds that system quality and information quality significantly influence students' perceived ease of use and perceived usefulness of the e-learning system. All relationships within the TAM were found to be significant. Drawing from this research, we will apply the TAM to evaluate our MAP platform, integrating critical factors for system acceptance. This approach addresses the specific requirements and challenges of our online learning assessment platform.

2.4. Mapping rules for schema transformation SQL to NoSQL and back [2]

This paper proposes mapping rules for transforming a relational database schema to a key-value NoSQL database schema, specifically focusing on MongoDB. The mapping rules consider various types of relationships, such as one-to-one, one-to-many, many-to-many, and entity relationships in the relational database, and can also be applied in the opposite direction. The paper addresses the challenges of migrating from a relational to a non-relational database and discusses the use of the ETL (Extract, Transform, Load) method for data transformation. These mapping rules will be applied to the database design, specifically to support the caching technique for load handling. the practical implications of this paper are that it provides a new method for selecting data in non-relational databases that can improve efficiency and response time. The paper's procedures involve recognizing user-requested values and transferring them to an in-memory database for faster response.

2.5. Improved method of selecting data in a nonrelational database [1]

The experiments described in the paper demonstrate the limitations of efficiency from a computational time perspective. The experiments were performed on applications hosted by Amazon cloud service while using the non-relational database DynamoDB. This paper also proposed a cache technique using the non-relational database to handle the performance capacity during the critical period also. This can be useful for our systems that require a large volume of highly accessible data.

3. Work Methodology

To address time pressure and unclear requirements, the development of our online learning assessment platform was divided into two phases. In the first phase, We adopted the Scrum framework within the agile software development model, allowing for continuous work and effective change management. UML diagrams were selected as the primary modeling tool to comprehensively illustrate all aspects of the proposed platform [6]. This paper specifically focuses on solving the critical problem of handling the large transfer load during exams for students' roles. The design and development process [7] is presented as follows.

3.1. System Requirements [8]
Upon receiving the system requirement in textual form, the document undergoes analysis and classification to identify functional and non-functional requirements. In terms of functional requirements, the proposed online learning assessment platform can be divided into four subsystems: **MCV Integration**,

responsible for managing connections with existing learning management systems such as MCV; **Question Pools**, tasked with managing and collecting questions; **Test Session**, focused on managing test sessions; and **Test** Grading, responsible for evaluating and grading test results. The overall functionality of the system will be represented through use-case diagrams (see Figure 1) to provide a visual representation. In order to address the core problem, the subsystems that are directly related to the main issue will be marked in the diagram.
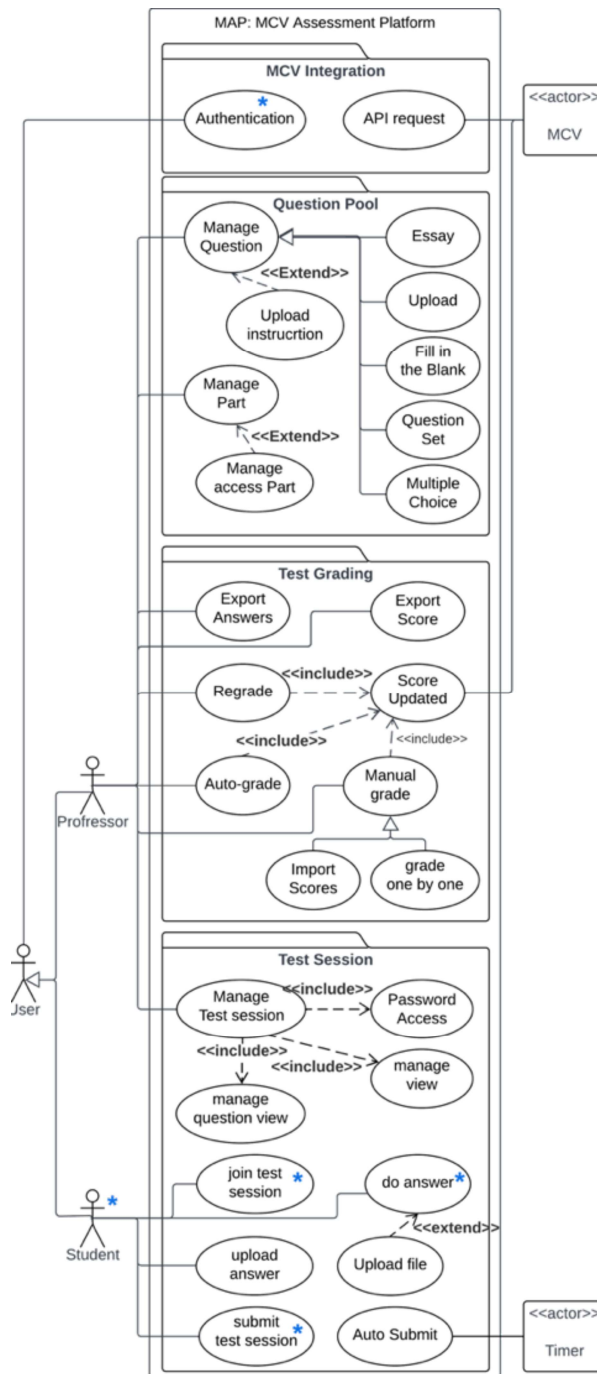

Fig. 1. Use Case Diagram for Online Learning Assessment Platform

Particularly, the functionality related to students poses a significant challenge. This paper will concentrate on the critical problem of scoped functionality, aiming to address the key characteristics necessary to solve the main issue. Specifically, *data reliability* is crucial, ensuring that all information retrieved during exams remains secure and available even if students lose their connection. Additionally, *latency performance*, represented by speed or response time during the handling of large loads, must meet user expectations

3.2. Design Phase

The requirements extracted from the previous process will now be processed in the design phase, integrated with the proposed system architecture as depicted in the deployment diagram. The proposed system was deployed on Amazon Web Services (AWS) using various services. An Amazon Elastic Compute Cloud (Amazon EC2) instance was utilized for managing the system, along with a MySQL database. The AWS App Runner service was employed for deployment, while the Amazon DB service was selected as the serverless non-relational database platform to optimize system performance under heavy load. Amazon ECS was used for automated deployment management, and Docker was integrated to facilitate the creation, testing, and deployment of the system via containers (see Figure 2).
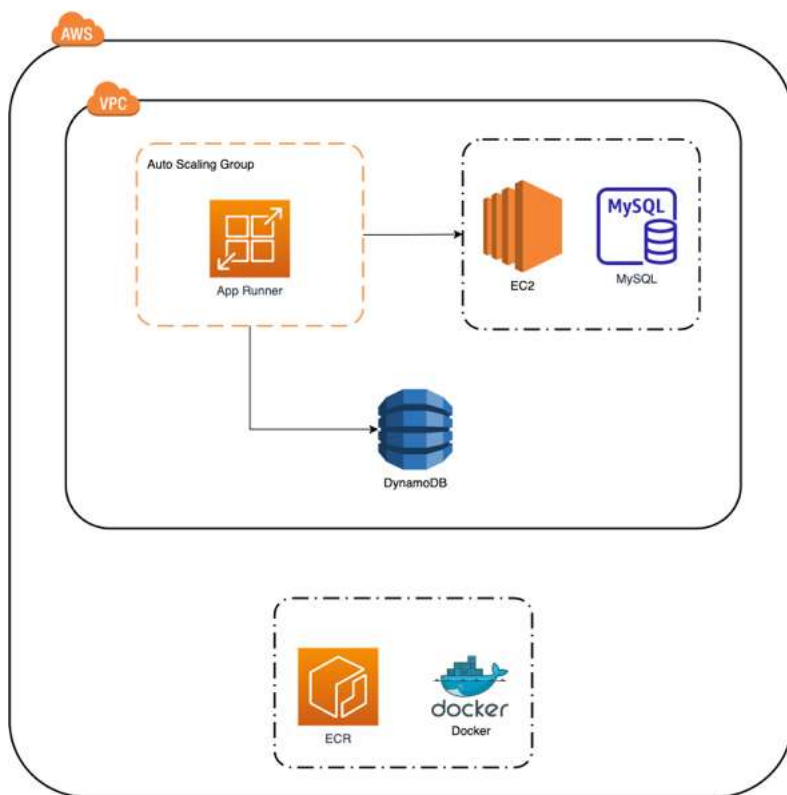
Fig. 2. Deployment Diagram for Online Learning Assessment Platform

Focusing on the design of the back-end system, which focuses on implementing the RESTful API [9] and its relationship with the database design, will be represented through UML diagrams. Keeping in mind the main objective of the development process, special attention will be given to scoping critical functionality, particularly during critical periods that directly involve student roles. To address this, we

will define the critical objects and their specific relationships with marked symbols through a class diagram (see Figure 3)
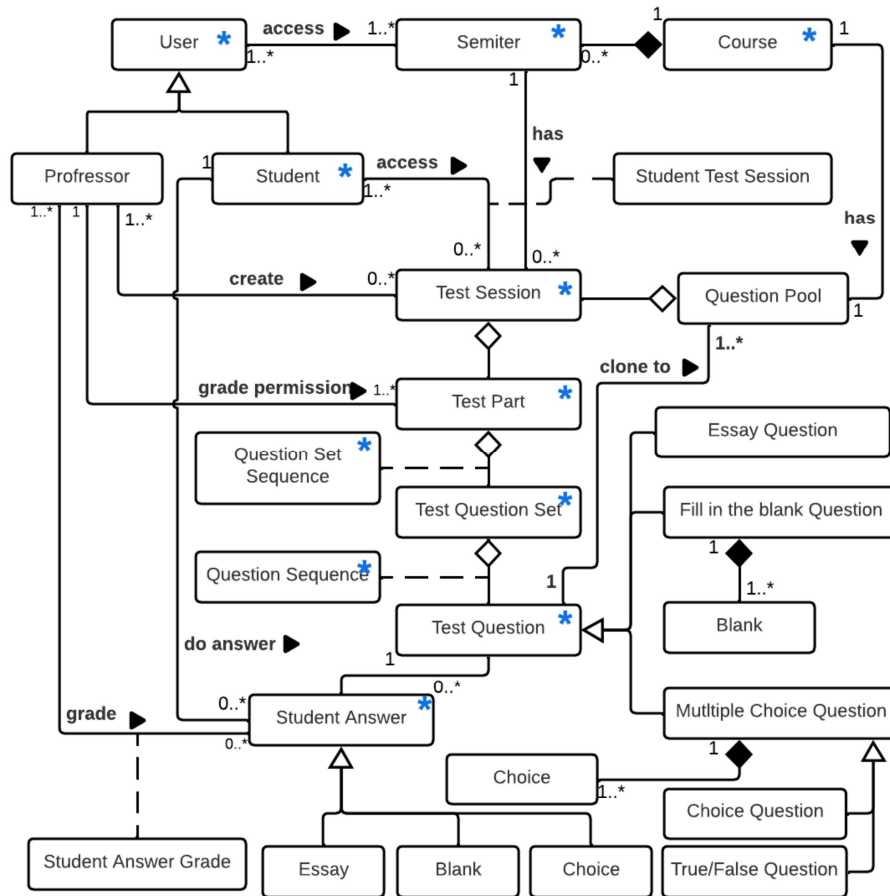
Fig. 3. Class Diagram for Online Learning Assessment Platform

To reduce latency and minimize interconnections between the server and database with the identified objects, a cache technique will be proposed. This technique aims to decrease the reliance on the main database or MySQL by storing data in an in-memory cache instead. However, the complexity of exam data, with its diverse question types, numerous exam settings, and intricate sub-structures as depicted in the ER diagram [10] (see Figure 4), poses challenges for in-memory computing. To address these challenges, a non-relational database, specifically NoSQL DynamoDB, which excels in handling non-structured or semi-structured data, will be proposed as an alternative solution with similar functionality. Additionally, DynamoDB ensures data reliability as it is an isolated service, offering resilience in the event of server downtime [1]

Fig. 4. ER diagram of identified entity for Online Learning Assessment Platform

Due to the distinct key-value structure of NoSQL databases compared to the column-based structure of SQL databases, the design of NoSQL tables needs to be caref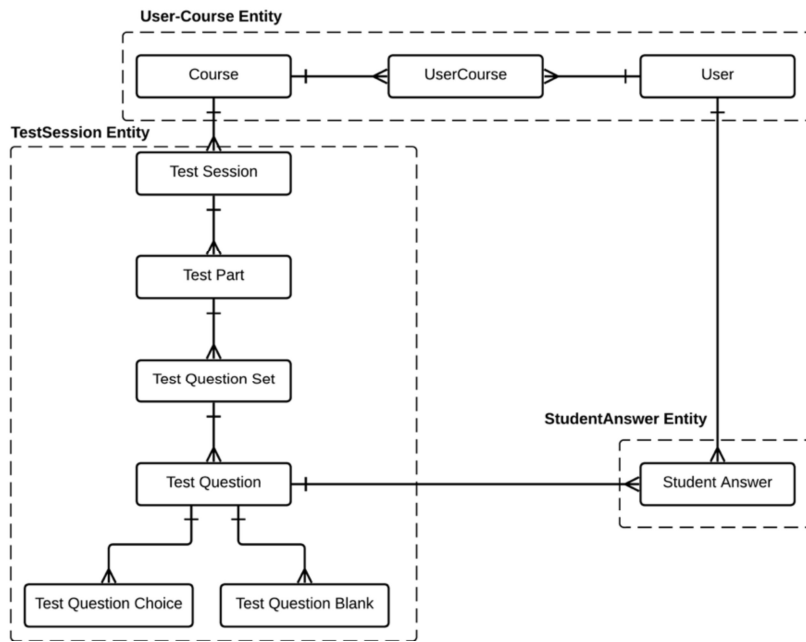ully considered and optimized to align with the proposed system. In order to achieve this, mapping rules are proposed based on the associative relationships between entities, encompassing one-to-one, one-to-many, and many-to-many relationships [2]. With a focus on the one-to-many relationship, which involves an object having multiple nested child objects, the mapping rule dictates the insertion of all nested child objects as sub-sequences within the parent objects. The design of the non-relational database for critical entities will be approached separately to implement the mapping rule design in three distinct parts.

- **User-Course Entity** - To support the high concurrency of students during exam periods and ensure efficient data access and partitioning based on roles, the one-to-many mapping approach will be applied, as illustrated in Figure 5
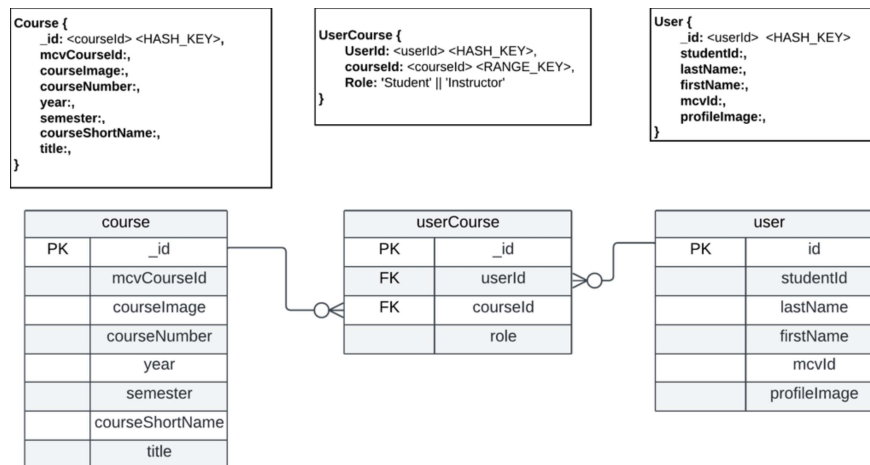
Fig. 5. Mapping Rule for User-Course Entity

- **Test-Session Entity** - The proposed technique enables effective handling of the deep nesting of test session data and has the potential to reduce interconnections with large loads, as depicted in Figure

6 The one-to-many mapping approach will be applied to consolidate all nested test session data within the same data structure, ensuring proper sequencing.
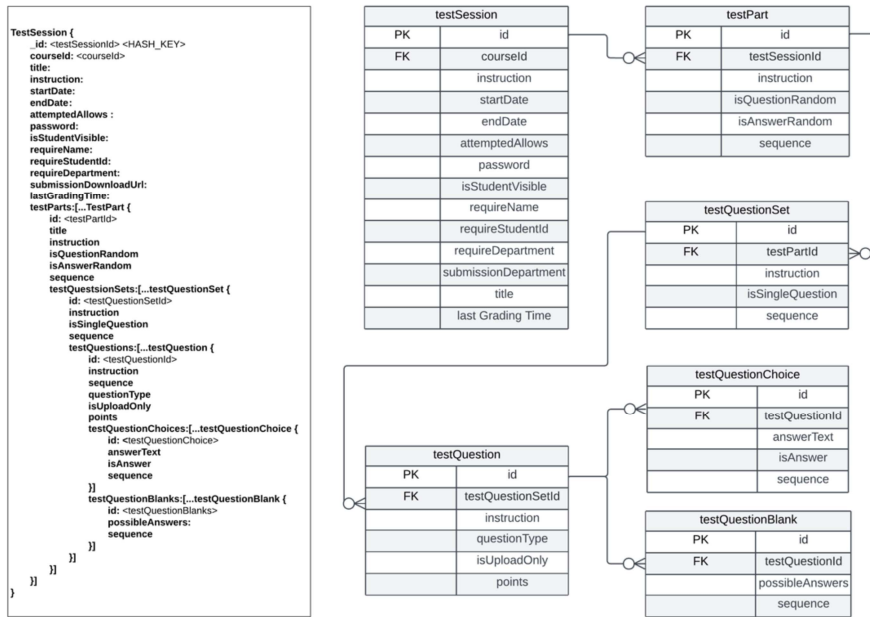
Fig. 6. Mapping Rule for Test-Session Entity

- **Student Answer Entity** - During exams, the Student Answer entity is responsible for managing all answers for each question submitted by students and ensuring that the answers are not lost, even in cases such as internet connectivity issues. The design, incorporating the applied rule, will be represented in Figure 7 As the Student Answer entity does not require data to be reversed back to the student, the non-relational database table can be optimized by reducing the fields to only include the necessary ones.
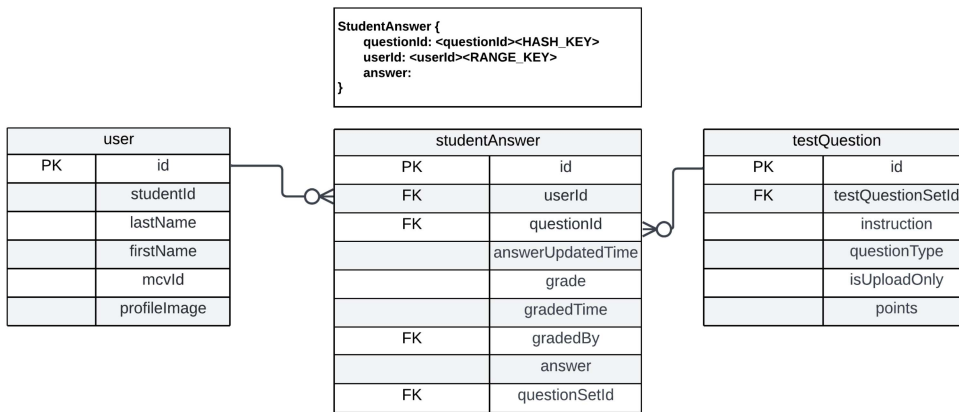
Fig. 7. Mapping Rule for Student Answer Entity

3.3. Implementing

The implementation of the proposed back-end system utilizes TypeScript, Node.js, TypeORM, GitLab, and Docker. TypeScript and Node.js are employed for API development and routing, while TypeORM manages the database and migration processes. GitLab serves as the development platform,

offering features for issue tracking, continuous integration, and collaboration. Docker is used to containerize the back-end system, facilitating deployment and scalability. This technology stack enables the creation of a modern, scalable, and robust back-end system that is easy to maintain and deploy. The implementation adheres to the application programming interface (API) specification, which controls the development process and ensures alignment with the design. The API specification references the details from the designed models, encompassing the requirement specification, RESTful API specification, and main protocol. It represents the API route details, including response messages, as illustrated in Table 1.

Table 1. API Routes list for Online Learning Assessment Platform

Method	API Route Name	Role	
		Student	*Instructor*
Path: /test-parts			
POST	Create Test Part		*
PUT	Update Test Part		*
DELETE	Delete Test Part		*
Path: /test-sessions			
GET	Get Test Session By ID	*	*
GET	Get All Test Session By Course ID	*	*
POST	Create new Test Session		*
PUT	Update Test Session		*
DELETE	Delete Test Session		*
POST	Verify Password	*	
Path: /test-questions/test-question-sets			
DELETE	Delete Test Question Set		*
Path: /test-questions			
POST	Create TF Test Question		*
PUT	Update TF Test Question		*
POST	Create Multi-Choice Test Question		*
PUT	Update Multi-Choice Test Question		*
POST	Create Essay Test Question		*
PUT	Update Essay Test Question		*
Path: /student-test-sessions			
POST	Create Student Test Session	*	
GET	Get Student Test Session By Test Session ID	*	
PUT	Record Submission	*	
GET	Get All Student Test Session By Test Session ID		*
PUT	Auto-grade By Test Session ID		*
Path: /student-answers			
GET	Count Student Answers by Parts	*	
GET	Get Student Answers by test Question Set ID	*	
PUT	Save Student Answer	*	
GET	Navigate Student Answers	*	

3.4. Testing Phase

In order to address the concerns regarding handling a large load, data reliability, and latency performance, the testing phase will encompass two methods: evaluating system performance and assessing

user satisfaction using the Technology Acceptance Model (TAM) [5, 11, 12]. These methods will be employed to measure the system's performance and gather user feedback, respectively, present as follow:

3.4.1. System's Performance

The system's performance was evaluated under normal conditions of 500 active users by conducting a performance test on selected API routes that represent different user roles. *The GET Student test session* route is used by students to retrieve test session data from the server to start an exam, while *PUT save student answer* is used to save answers during an exam. On the other hand, *POST Test Session* represents the instructor's role and is used to create an exam session along with exam session data. The expected average response times were determined based on the user's attention with a limit of 10,000 milliseconds [13].

3.4.2. User Satisfaction

The Technology Acceptance Model (TAM) is a framework used to assess user acceptance and adoption of e-learning systems, considering factors such as Intention of Use, perceived ease of use, and perceived usefulness [5, 11]. In order to represent the TAM method within our research, we will cover the following sub-topics: Model Hypothesis, Data Collection, Instrument Development, Reliability Analysis, and Structural Equation Testing. These components will be employed to evaluate user satisfaction and acceptance of our online learning assessment platform, taking into account the perceived ease of use and perceived usefulness factors as outlined by the TAM framework.

- **Model Hypothesis**: Building upon the mentioned factors, the Model Hypothesis is developed using the TAM framework as the foundation. Specifically, we identify and derive the influential values related to user satisfaction, with a specific focus on data reliability and latency performance. These factors are considered essential in assessing the acceptance and satisfaction of users within the context of our online learning assessment platform. Next, to integrate with our proposed systems and address the critical problem, we establish the Continuance of User Acceptance of the Online Learning Assessment Platform, as depicted in Figure 8. In line with this, we put forth the following hypotheses:
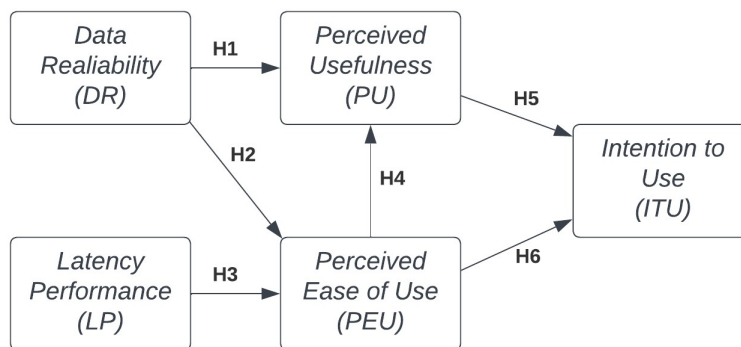
Fig. 8. The Continuance of User Acceptance Model of the Online Learning Assessment Platform

H1:   **Data Reliability** *has a positive impact on* **Perceived Usefulness**.
H2:   **Data Reliability** *has a positive impact on* **Perceived Ease of Use**.
H3:   **Latency Performance** *has a positive impact on* **Perceived Ease of Use**.
H4:   **Perceived Ease of Use** *has a positive impact on* **Perceived Usefulness**.
H5:   **Perceived Usefulness** *has a positive impact on* **Intention to Use**.
H6:   **Perceived Ease of Use** *has a positive impact on* **Intention to Use**.

- **Data Collection**: The data for this study will be collected from students who voluntarily participate in the survey and are enrolled in volunteer general education courses. The online platform will be tested by assigning the final exam activity of this course during the second semester of April 2023, requiring the system to handle a total of 300 students within a 3-hour exam time frame. The survey will be conducted among a total of 132 bachelor's degree students, with a breakdown of 94 first-year students (71%), 22 second-year students (17%), 4 third-year students (3%), and 12 fourth-year students (9%).
- **Instrument Development**: The data collection instrument for this study consists of a survey that includes both system characteristics *(Data Reliability and Latency Performance)* as external variables and TAM variables *(Perceived Ease of Use, Perceived Usefulness, and Intention to Use)* with a specific number of items. To achieve the research objectives, a questionnaire was developed using a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Participants were asked to indicate their level of agreement with each question on the provided scale. An example of the survey questions, categorized by context, is presented in Table 2. The collected data will be analyzed using the *SPSS AMOS Version 28.0* [12], which allows for testing specific hypotheses by examining the estimated coefficients (path estimates) and their significance. This analysis will enable us to assess the statistical significance of the relationships between variables, as well as the significance of the estimated paths.

Table 2. Example of Surveys question

System Characteristics	Example question
Data Reliability (DR)	*"The system rarely experiences downtime or interruptions."*
Latency Performance (LP)	*"I rarely experience delays or lags when using the system."*
Perceived Usefulness (PU)	*"The system makes it easier for me to participate in test sessions."*
Perceived Ease of Use (PEU)	*"The system is easy to navigate and use."*
Intention of Use (ITU)	*"Utilizing the MCV Assessment platform for exam management is a good idea."*

- **Reliability Analysis**: To verify the reliability of the collected data, a reliability analysis was conducted using the SPSS software, employing Cronbach's alpha coefficient. The overall table (see Table 3) summarizing the results indicated that all variables demonstrated satisfactory reliability, with Cronbach's alpha values ranging from 0.7 and above. This confirms the consistency and internal reliability of the data, ensuring that the test items are reliable for further analysis and interpretation.

Table 3. The Overall Cronbach's Alpha Table

Overall Cronbach's Alpha	Items
0.92	132

- **Structural Equation Test**: In this study, a hypothesis-test approach was employed at a 95% confidence level to evaluate the hypotheses and determine the significance of the relationships between variables. The analysis involved conducting t-tests to assess the statistical significance of each relationship. The t-test calculated the t-value, which measures the difference between a sample mean (or other test statistic) and the null hypothesis expectation, taking into account the variability of the data. Additionally, the p-value was calculated to assess the probability of obtaining a test statistic as extreme as or more extreme than the observed value, assuming the null hypothesis is true. Consistent with a 95% confidence level, the resulting t-values exceeded the critical value of 1.96, and the p-values were below 0.05, indicating statistical significance and providing support for the hypotheses [5, 11].

4. Results

To evaluate the design and development of the backend system, especially for handling high concurrent during the exam as mentioned above. The result are presented as follow:

4.1. System Performance

In performance testing, the result is shown in Table 4. The average response time result during the test load of 500 concurrent requests is lower than 10,000 milliseconds [3]. In detail, for *GET Test Session*, we found this API route have response times higher than other API route because it will transfer test session data from the server to collect in temporal memory to eliminate multiple data retrievals to only one time. From the result, it can represent the overall performance that can process in normal conditions of 500 concurrent users.

Table 4. Average Response Time Result

API Routing		Concurrent	Average Response time (milliseconds)
GET	Get Test Session By ID	500	4,977
PUT	Save Student Answer	500	1,749
POST	Create Student Test Session	500	2,162

4.2. User satisfaction

The evaluation of user satisfaction with regards to data reliability (DR) and latency performance (LP), utilizing the Technology Acceptance Model (TAM) framework, revealed significant impacts of these factors along with Perceived Ease of Use (PEU) and Perceived Usefulness (PU) on user satisfaction and intention to use (ITU) the system. Our data collection's reliability, as indicated by Cronbach's Alpha analysis, exceeds 0.7. This confirms the consistency and internal reliability of the data, ensuring the test items are dependable for further analysis and interpretation. The average satisfaction scores on a 5-point Likert scale indicated that users found Data Reliability (3. 7), Perceived Ease of Use (4.1), Perceived Usefulness (4.3), and Intention to Use (3.9) satisfactory, whereas Latency Performance (3.43) was slightly less satisfying, suggesting that system performance could be an area for improvement. The t-test results presented in Table 5 support the initial hypotheses, found that users highly value data reliability, effective latency performance, ease of use, and perceived usefulness, which significantly contribute to their satisfaction and intention to continue using the system. However, no significant relationship was found between perceived ease of use and intention to use. These findings highlight the significance of prioritizing data reliability, latency performance, ease of use, and perceived usefulness to enhance user satisfaction and adoption of the proposed system.

Table 5. T-Test Result

Hypothesis	T-value	T-critical	p-value	Information
H1: DR → PU	4.779	1.96	<0.001	Significant
H2: DR → PEU	2.181	1.96	0.029	Significant
H3: LP → PEU	4.079	1.96	<0.001	Significant
H4: PEU → PU	8.272	1.96	<0.001	Significant
H5: PU → ITU	4.447	1.96	<0.001	Significant
H6: PEU → ITU	-1.060	1.96	0.289	Not Significant

5. Conclusion

The successful design and development of the back-end system for an online learning assessment platform were achieved through the strategic use of a serverless, non-relational database - DynamoDB. This key implementation not only offered a powerful solution to manage the significant challenge of accommodating large numbers of students during exam activities but also contributed notably to enhancing the system's load capacity. DynamoDB's utility was particularly evident in its ability to effortlessly handle high-traffic loads, thanks to its built-in partitioning. This enabled us to spread data and traffic over servers to meet throughput requirements and maintain consistent performance. Further, the database design for mapping rules facilitated an efficient synchronization process, supporting the necessary partition information. The simplicity and effectiveness of this design allowed us to maximize the potential of DynamoDB, resulting in a database system that was both useful and simple to use.

Performance testing further validated the system's robustness. The average response time was less than 10,000 milliseconds even under the test load of 500 concurrent requests, thus confirming that the system's performance indeed meets the main criteria set out in the platform's proposal. In particular, the GET Test Session API route, despite having higher response times due to its role in reducing multiple data retrievals to a single instance, performed effectively under normal conditions with 500 concurrent users. These results definitively demonstrate the system's capability to handle high loads while maintaining solid performance.

In assessing user satisfaction, we employed the TAM framework, which provided important insights into user perceptions and intentions of use. The users expressed high satisfaction in aspects such as Data Reliability, Perceived Usefulness, and Intention to Use, with a notable score for Perceived Ease of Use. However, the Latency Performance was identified as an area with room for improvement. The feedback regarding Data Reliability is particularly valuable and will guide our focus in the next phase of development to further enhance this crucial aspect of the system.

In conclusion, the methodologies utilized in this study have not only proved the system's capacity to handle high loads but have also illuminated user sentiment, providing a comprehensive evaluation. The dual focus on technical performance and user satisfaction will inform both immediate system assessment and future improvements. With its versatility, the research approach can be applied to other platforms sharing similar functionalities or constraints, thus making the evaluation process integral to not just assessing system performance, but also driving ongoing improvements and support.

References

[1] R. Ceresnak, M. Kvet, and K. Matiasko, "Improved method of selecting data in a nonrelational database," in *2021 International Conference on Information and Digital Technologies (IDT)*, 22-24 June 2021 2021, pp. 59-64, doi: 10.1109/IDT52577.2021.9497640.
[2] R. Čerešňák, A. Dudáš, K. Matiaško, and M. Kvet, "Mapping rules for schema transformation : SQL to NoSQL and back," in *2021 International Conference on Information and Digital Technologies (IDT)*, 22-24 June 2021 2021, pp. 52-58, doi: 10.1109/IDT52577.2021.9497629.
[3] S. M. Abdullah, M. M. Ameen, and R. M. Abdullah, "Designing and Implementing of an Online Management Information System for Exam Committee," presented at the 2022 8th International Engineering Conference on

Sustainable Technology and Development (IEC), 23-24 Feb. 2022, 2022.

[4] D. Vomvyras, A. Andreatos, and C. Douligeris, "Exam Wizard: A novel e-assessment system," in *2019 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, 20-22 Sept. 2019 2019, pp. 1-6, doi: 10.1109/SEEDA-CECNSM.2019.8908372.

[5] N. H. M. Noh, M. T. Amron, and M. A. Mohamad, "System Characteristics in Predicting E-Learning Acceptance: An Extended Technology Acceptance Model (TAM) Study," in *2022 International Conference on Engineering and Emerging Technologies (ICEET)*, 27-28 Oct. 2022 2022, pp. 1-6, doi: 10.1109/ICEET56468.2022.10007206.

[6] A. Dennis, B. H. Wixom, and D. Tegarden, *Systems Analysis and Design: An Object-Oriented Approach with UML*. Wiley Publishing, 2015.

[7] M. A. B. Azman, N. K. Madzhi, J. Johari, and N. Buniyamin, "Design and implementation of an online education and evaluation system," in *2017 IEEE 9th International Conference on Engineering Education (ICEED)*, 9-10 Nov. 2017 2017, pp. 248-253, doi: 10.1109/ICEED.2017.8251202.

[8] J. Cleland-Huang, "Software requirements," *Software Engineering,* vol. 1, pp. 113-123, 2005.

[9] B. M. Adam, A. R. A. Besari, and M. M. Bachtiar, "Backend Server System Design Based on REST API for Cashless Payment System on Retail Community," in *2019 International Electronics Symposium (IES)*, 27-28 Sept. 2019 2019, pp. 208-213, doi: 10.1109/ELECSYM.2019.8901668.

[10] P. P.-S. Chen, "The entity-relationship model—toward a unified view of data," *ACM Trans. Database Syst.,* vol. 1, no. 1, pp. 9–36, 1976, doi: 10.1145/320434.320440.

[11] M. F. A. Rizqullah, H. H. Nuha, and R. Yasirandi, "Analysis of E-learning Indonesian Student Acceptance Behavior in Online Practicum Using TAM and Flow Theory," in *2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT)*, 22-23 Nov. 2022 2022, pp. 57-60, doi: 10.1109/ICoSEIT55604.2022.10029996.

[12] B. Yang and Y. Zhao, "Research on the Continuance Intention of MOOC Platforms Based on the Technology Acceptance Model : A Case of MOOC Learning Platforms in Chinese Universities," 2021 2nd International Conference on Artificial Intelligence and Education (ICAIE), 2021.

[13] J. Nielsen, "Chapter 5 - Usability Heuristics," in *Usability Engineering*, J. Nielsen Ed. San Diego: Morgan Kaufmann, 1993, pp. 115-163.

Cost-Sensitive Cloud Architecture for an Online Assessment Web Application

Subtawee Hanyut[1], Nuengwong Tuaycharoen[2], Naruemon Pratanwanich[3], Dittaya Wanvarie[4], and Atiwong Suchato[5]*

[1]Department of Computer Engineer, Chulalongkorn University, Bangkok, Thailand, 6470188421@student.chula.ac.th
[2]Department of Computer Engineer, Chulalongkorn University, Bangkok, Thailand, nuengwong.t@chula.ac.th
[3]Department of Mathematics and Computer Science, Chulalongkorn University, Bangkok, Thailand, naruemon.p@chula.ac.th
[4]Department of Mathematics and Computer Science, Chulalongkorn University, Bangkok, Thailand, dittaya.w@chula.ac.th
[5]Department of Computer Engineer, Chulalongkorn University, Bangkok, Thailand, atiwong.s@chula.ac.th

*Corresponding author. Tel.:+66-843786402.
E-mail-address: subtawee.h@gmail.com

Abstract

A university is required to implement online education and evaluation due to the COVID pandemic. The university already has an online learning platform with a fixed resource architecture, but it cannot accommodate the university's examination workload patterns. To address this issue, the university will need to design a new architecture to accommodate the current workload behaviours, as the number of users during the exam will be extremely high, and the workload will spike when students begin taking exams simultaneously. Utilizing a scalable API service and a scalable relational database, this study aims to develop a scalable system capable of handling the increased traffic during exam periods. The evaluation results indicate that the new scalable architecture design can handle the sudden increase in users from zero to five hundred and five hundred to one thousand with an acceptable error rate of 5 percent. However, the expense exceeds the predefined budget of 500 USD monthly, and a scalable relational database is primarily responsible for the high cost. Therefore, we have redesigned the architecture using a hybrid database that is a combination of a relational database with fixed resources and a non-relational database that is scalable and changed the scalable API services from AWS ECS, which requires at least one active instance, to AWS App Runner, which can be configured with zero active instances. This redesign has significantly reduced costs to within the predefined cost while maintaining the reliability and consistency of a relational database and improving the error rate.

Keywords: Auto Scaling; Relational Database; Non-Relational Database; Amazon Web Service.

1. Introduction

In present, each subject in various university curricula has regular student performance assessments, i.e. examinations, including both online assessment formats and in-class assessments. Online assessments pose significant challenges as the system must be able to accommodate rapidly changing usage volumes. When the assessment begins, there will be a large number of participants accessing the system simultaneously within a short period. While universities currently have online teaching and learning management system, but it is unable to handle the rapidly changing usage volumes. Therefore, it is necessary to design and create a new system as an extension of the existing system, utilizing student and course data from the current system.

Currently, the preferred approach for system installation is to use cloud-based architectures more frequently, as they offer stability and cost savings compared to deploying self-hosted infrastructure. Our university's current online teaching and learning management system is installed on the cloud architecture using the services of Amazon Web Services (AWS).

Designing a system that can accommodate a large number of users can generally be achieved by using highly efficient service instances, either single or multiple instances. However, this incurs high costs,

including installation and continuous operation. Additionally, if multiple instances are used to handle high usage volumes, even when the user numbers are relatively constant, it leads to wasteful expenditure.

Therefore, this research aims to design an online assessment system and select a cloud architecture that can automatically scale to match user volumes at different time intervals. The ability to auto-scale refers to the capability of increasing or decreasing the number of service instances automatically by defining algorithms for scaling up or down. The limitation of using an auto-scaling cloud architecture is the requirement of having at least one instance running continuously, even if there are no users. However, it is possible to choose instances with relatively lower performance to save costs and use scaling techniques to increase the number of instances when user volumes increase. Designing a system using a cloud architecture that can auto-scale helps to reduce costs compared to using highly efficient instances, either single or multiple, constantly.

Overall, the system's design using an auto-scaling cloud architecture enables significant cost savings compared to using a single highly efficient instance or multiple instances constantly, even when there is no increase in user volume.

2. Related works

2.1.   *Docker-based Web Server Instructional System [1]*

This research presents the design and development of an online learning and teaching system using Docker as the underlying technology for system creation and deployment. The research also compares Docker with Virtual Machines (VMs). Docker has advantages over VMs as it eliminates the need to manage the system environment each time a new system is installed. In contrast, VMs require setting up a new environment every time a new installation is performed or when the system is relocated. Once a Docker image is created, it can be stored on a Container Registry, which is an online repository for Docker images. This allows users to download Docker images and install them anywhere.

2.2.   *Auto-Scaling Web Applications in Hybrid Cloud Based on Docker [2]*

This research presents the design of a web application that can automatically scale in a hybrid cloud environment using Docker. The application can increase or decrease the number of Docker containers dynamically. The hybrid cloud architecture in this research involves separating the system installation into two locations which are the private cloud and the public cloud. A Load Balancer is used to manage the routing of user requests based on their IP addresses. If the IP address originates from within the university, the Load Balancer forwards the request to the private cloud. If the IP address comes from outside the university, the Load Balancer sends the request to the public cloud. Docker images for this research are stored on a private cloud service, enabling faster downloads compared to storing them on a public cloud service. The speed of downloading Docker images has an impact on the time required to scale up the number of Docker containers. Two algorithms are utilized in this research: predictive and adaptive algorithms based on usage volume at a given time. The predictive algorithm employs regression equations to calculate the number of requests and the required number of Docker containers. The adaptive algorithm sets upper limits for the resource usage of each Docker container.

2.3.   *Cost-Aware Multidimensional Auto-Scaling of Service- and Cloud-Based Dynamic Routing to Prevent System Overload [3]*

This research presents an improved architecture that incorporates self-management capabilities for multidimensional scaling, referring to automatic scaling of both size and quantity, known as Adaptive Dynamic Routers (ADR). The ADR architecture was previously developed by the research team to

mitigate the risks of operating beyond system limitations. Through 9,600 case studies, it was found that this new architectural design significantly reduced the request rate per service. When a user makes a request, the system forwards it to the gateway, which then routes it to the router and subsequently to the various services. The QoS Monitor is responsible for monitoring the usage and request volume of the services. If the usage exceeds the predefined thresholds, the QoS invokes the Dynamic Reconfigurator, which adjusts the resource allocation and quantity of the services to appropriately scale them based on the user demand. Incoming requests are processed one by one in sequence, with each service component consisting of a buffer and a processor. The buffer represents the number of pending requests, while the processor indicates the processing rate in requests per second. If the buffer size exceeds the processing rate, the system scales up the corresponding service. The components of the ADR architecture are illustrated in Fig. 1.
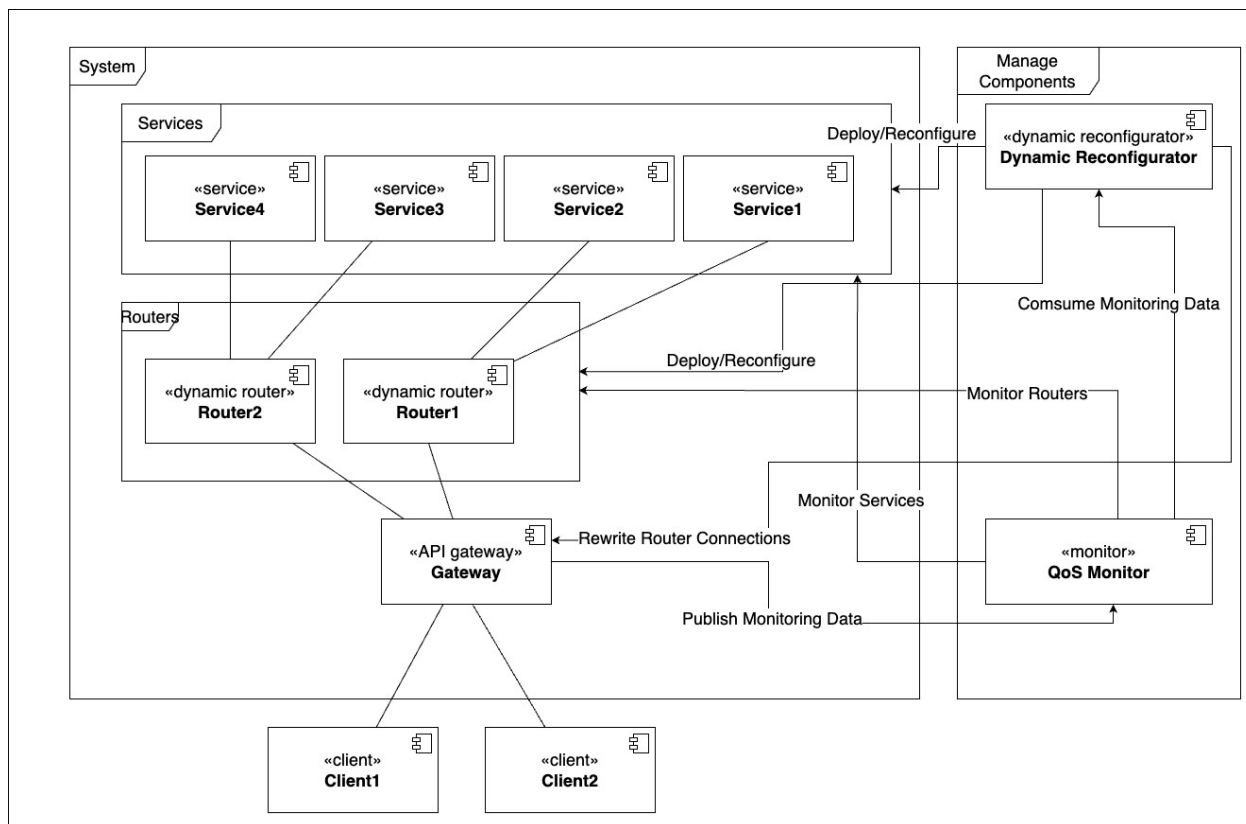

Fig. 1. Component Diagram of the ADR Architecture

Through testing 9,600 cases over a duration of 1,200 hours, it was observed that the architecture achieved a reduction in exceeding operation rates of 47.6% for routers and 61.8% for services.

2.4.   Budget in the Cloud: Analyzing Cost and Recommending Virtual Machine Workload [4]

This research study presented an analysis of real-world data obtained from Microsoft's Azure cloud service provider, which provided information on the size and lifespan of virtual machines. The researchers observed that the virtual machines were not being utilized to their full potential, leading to the possibility of incurring unnecessary costs. To address this, the research study employed regression equations to estimate the cost per hour and the total cost of virtual machines with varying compute units and memory sizes.

The experimental results applying the algorithm to the dataset demonstrated cost savings of up to 25% to 40% compared to the actual costs incurred by the virtual machine instances.
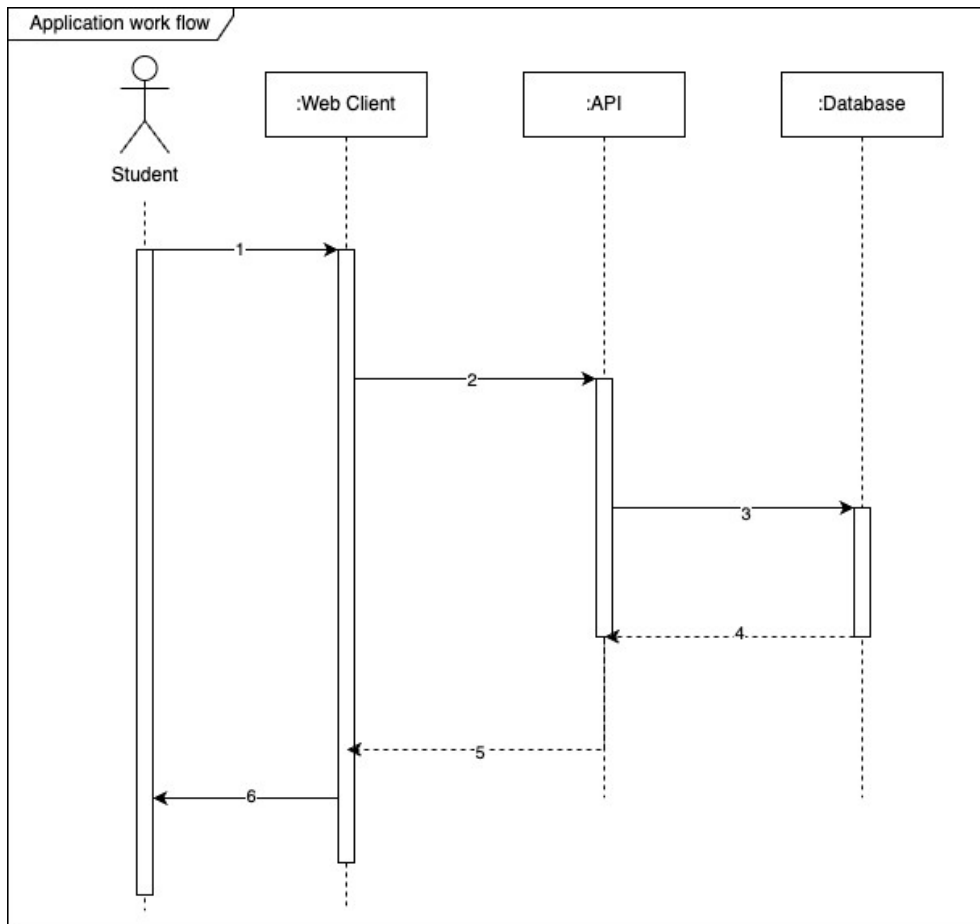
3. Application Workload Analysis

Fig. 2. Overall Application Workflow

The online assessment system is a system that allows users to create and manage assessments. Only users with the instructor role have access privileges to create assessments, while users with the role of student can only view exams and record their answers. Our concern is student workflow, which increases sharply during examination weeks and demands high reliability.

Fig. 2. illustrates the overall workflow of the online assessment system, starting from accessing the web client designed for the online assessment system in this study. Once logged in, a student can utilize various features of the online assessment system, such as viewing examination rooms, accessing exam questions, recording answers, and submitting responses. After the user performs these actions, the web client sends requests to the API service, then the API sends a database query to the database to store or retrieve data and returns the data to the API. Finally, the API sends the data back to the web client to display the results of the user's request.

In general, student performance is evaluated twice per academic semester per course, namely the midterm exam and the final exam. In one semester, the midterm exam typically spans a duration of approximately 7 days, while the final exam lasts for 14 days. Some courses may also have additional quizzes or exams in addition to the midterm and final exams. During the exam period, there is a consistently high volume of daily users. Exams are scheduled during the university's working hours,

from 8:00 AM to 4:00 PM, which amounts to 7 hours per day except for the lunch break from 12:00 AM to 1:00 PM. Outside these hours, there are either no users in the system or a significantly reduced number of users. This period is from 4:01 PM to 7:59 AM of the following day, totaling 16 hours. The number of users in the system depends on the number of students taking exams in a given time period for each course. Based on statistics from the university's current online teaching and learning management system for the most recent academic year, there are 2,909 courses in the system, and there are 131,831 registered students. The average number of students per course is 45.31.

Table 1. The average number of online examination rooms

Examination	Courses	Duration	Average course per day
Midterm	2,909	7 Days	416 (415.57)
Final	2,909	14 Days	208 (207.78)

From Table 1., the average number of online examination rooms per day during the midterm examination period is shown to be 416 rooms per day, and during the final examination period, the average number of online examination rooms per day is 208 rooms per day.

The duration of each examination session may vary. Based on statistics collected from the current online assessment system, the average duration of an examination is 3 hours. Therefore, during each day of examination, there are only 4 possible time slots which are 8:00 AM - 11:00 AM, 8:30 AM - 11:30 AM, 9:00 AM - 12:00 PM, and 1:00 PM - 4:00 PM. No examinations are scheduled during the 12:00 PM - 1:00 PM time slot, which is the lunch break.

Table 2. The number of examination rooms in each time slot during the midterm examination period

Time Period	Number of examination rooms	Number of users
8:00 AM – 8:30 AM	104	4,713 (4,712.24)
8:30 AM – 9:00 AM	208	9,425 (9,424.48)
9:00 AM – 9:30 AM	312	14,137 (1,4136.72)
9:30 AM – 10:00 AM	312	14,137 (1,4136.72)
10:00 AM – 10:30 AM	312	14,137 (1,4136.72)
10:30 AM – 11:00 AM	312	14,137 (1,4136.72)
11:00 AM – 11:30 AM	208	9,425 (9,424.48)
11:30 AM – 12:00 AM	104	4,713 (4,712.24)
12:00 AM – 1:00 PM	0	0
1:00 PM – 4:00 PM	104	4,713 (4,712.24)

Fig. 3. The user traffic in each time slot on days with midterm examinations

According to Table 2. and Fig. 3., at the time slot from 8:00 to 8:30 AM., there will be 104 examination rooms conducting exams simultaneously, accommodating a total of 4,713 examinees. Additionally, during the time slot from 8:30 to 9:00 AM., there will be an additional 104 examination rooms, resulting in a total of 208 examination rooms conducting exams simultaneously, accommodating 9,425 examinees. Moreover, during the time slot from 9:00 to 11:00 AM., an additional 104 examination rooms will be added, bringing the total number of examination rooms conducting exams simultaneously to 312, accommodating 14,137 examinees. It should be noted that during the time slot from 11:00 to 11:30 AM., the number of examinees decreases to 9,425 individuals, as the examination rooms that started at 8:00 AM. have finished their exams. Similarly, during the time slot from 11:30 AM. to 12:00 AM., the number of examinees decreased to 4,713 individuals as the examination rooms that started at 8:30 AM. finished their exams. Finally, from 1:00 to 4:00 PM., there will be 104 examination rooms accommodating 4,713 examinees.

Table 3. The number of examination rooms in each time slot during the final examination period

Time Period	Number of examination rooms	Number of users
8:00 AM – 8:30 AM	52	2,357 (2,356.12)
8:30 AM – 9:00 AM	104	4,713 (4,712.24)
9:00 AM – 9:30 AM	156	7,069 (7,068.36)
9:30 AM – 10:00 AM	156	7,069 (7,068.36)
10:00 AM – 10:30 AM	156	7,069 (7,068.36)
10:30 AM – 11:00 AM	156	7,069 (7068.36)
11:00 AM – 11:30 AM	104	4,713 (4712.24)
11:30 AM – 12:00 AM	52	2,357 (2356.12)
12:00 AM – 1:00 PM	0	0
1:00 PM – 4:00 PM	52	2,357 (2356.12)

Fig. 4. The user traffic in each time slot on days with final examinations

According to Table 3. and Fig. 4., at the time slot from 8:00 to 8:30 a.m., there will be 8 examination rooms conducting exams simultaneously, accommodating a total of 1,600 examinees. Additionally, during the time slot from 8:30 to 9:00 a.m., an additional 8 examination rooms will be added, resulting in 16 examination rooms conducting exams simultaneously, accommodating a total of 3,200 examinees. Moreover, during the time slot from 9:00 to 11:00 a.m., an additional 8 examination rooms will be added, bringing the total number of examination rooms conducting exams simultaneously to 24, accommodating a total of 4,800 examinees. It should be noted that during the time slot from 11:00 to 11:30 a.m., the number of examinees decreases to 3,200 individuals, as the examination rooms that started at 8:00 a.m. have finished their exams. Similarly, during the time slot from 11:30 a.m. to 12:00 p.m., the number of examinees decreased to 1,600 as the examination rooms that started at 8:30 a.m. concluded their exams. Finally, from 1:00 to 4:00 p.m., there will be 8 examination rooms accommodating 1,600 examinees.

Therefore, it can be concluded that during the midterm exams, the maximum number of examinees is 4,800 people, and during the final exams, the maximum number of examinees simultaneously is 9,000 people.

4. COST EVALUATION

The evaluation of cloud architecture service costs can be done by calculating the costs of each component in the architecture and the number of resources used, as well as the time required to perform tasks for each component. These values are then aggregated to determine the monthly cost of the entire system. In the components where services with the ability to scale automatically are used, the capacity to support the maximum number of users per instance must be calculated. This is done to determine the total units required to accommodate users during each time period. These calculations are then used to determine the total cost per service.
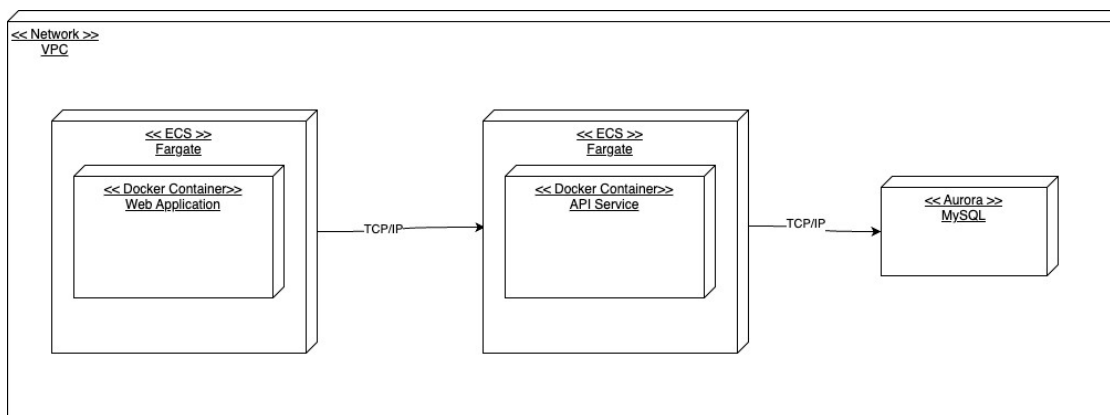
Fig. 5. Deployment diagram of the application

4.1. Performance Testing of AWS ECS

Despite the variable usage of the system, as of 2022, AWS ECS [7] is necessary always to have at least one active instance. To optimize cost, specific settings of the instance need to be configured with the minimum specifications.

The performance testing of AWS ECS, explicitly using Fargate, is conducted by simulating the number of users and sending requests to AWS ECS. The number of users is gradually increased, and the system observes the number of errors encountered in response to the users. As the system's non-functional requirements, the error rate is set to be within 5% of the total number of users. Additionally, the response time should be at most 3 seconds per request. The testing is performed using a minimum configuration setup of 1 instance.

Table 4. The performance testing results of AWS ECS using Fargate with the minimum configuration of 1 instance.

Number of Users	Errors	Error Rate (%)	The highest response time (s)
100	0	0.00	0.4
200	0	0.00	0.6
300	10	3.33	0.9
400	20	5.00	4.0

According to Table 4., the performance testing results of AWS ECS using Fargate with the minimum configuration of 1 instance show that it is capable of concurrently supporting up to 300 users with an error rate within the specified threshold.

4.2. Performance Testing of AWS Aurora

AWS Aurora [10] charges for services based on the number of resources used to process each request. It is designed to handle high user volumes, and AWS has confirmed that Aurora can support up to 100,000 transactions per second. The cost of using AWS Aurora varies according to the resources consumed. Therefore, there is no need to perform performance testing for AWS Aurora.

4.3. Service Cost calculation

The process of calculating the service cost involves calculating the total cost of the entire system for a month. The results are divided into three categories.

- The service cost for the month without exams.
- The service cost for the month with midterm exams
- The service cost for the month with final exams,

Web client and API Service have the same service cost because they use the same service and configuration. Specifically, they utilize AWS ECS using Fargate with the specific features of 0.25 virtual CPU units and 0.5 gigabytes of memory.

4.3.1. The Cost for Web Client and API Service

The monthly service cost for the web Client and API service can be calculated by multiplying the cost of the ECS with the minimum configuration by the number of instances that automatically scale according to the user volume in each time period. To enable automatic scaling for ECS, it must meet certain conditions. For example, the number of users per instance exceeds the predefined threshold, determined based on the testing conducted in section 4.1, specifically 300 users. If the number of users in the system exceeds 300, automatic scaling will be triggered to align with the number of users.

Table 5. The required number of AWS instances in relation to the total number of users.

Number of users	Number of AWS ECS Instance
300	1
400	2
700	3

This minimum configuration of AWS ECS has a virtual CPU count of 0.25 units, which incurs a service cost of $0.0506 per unit per hour. Additionally, a temporary storage capacity of 0.5 gigabytes is required, with a service cost of $0.00553 per gigabyte per hour when utilized in the Singapore region.

Table 6. The cost of the minimum configuration of AWS ECS

Components	Cost per 1 unit per hour (US Dollar)
vCPU	0.0506 (1 processor)
Memory	0.0053 (1 GB)

Therefore, the minimum configuration service cost would be equal to 0.25 * (0.0506) + 0.5 * (0.0053) = 0.12915 US dollars per hour.

Table 7. The service cost of Web Client and API Service for the month without exams

Duration	Cost (US Dollars)
1 Hour	0.12915
1 Day (24 Hours)	3.09960
1 Month (30 Days)	92.9880

Table 8. The number of instances for each time period on days with midterm exams.

Time Period	Number Of Users	Number Of Instances	Duration (Hour)
8:00 AM – 8:30 AM	4,713 (4,712.24)	16	0.5
8:30 AM – 9:00 AM	9,425 (9,424.48)	32	0.5
9:00 AM – 9:30 AM	14,137 (1,4136.72)	48	0.5
9:30 AM – 10:00 AM	14,137 (1,4136.72)	48	0.5
10:00 AM – 10:30 AM	14,137 (1,4136.72)	48	0.5
10:30 AM – 11:00 AM	14,137 (1,4136.72)	48	0.5
11:00 AM – 11:30 AM	9,425 (9,424.48)	32	0.5
11:30 AM – 12:00 AM	4,713 (4,712.24)	16	0.5
12:00 AM – 1:00 PM	0	0	1.0
1:00 PM – 4:00 PM	4,713 (4,712.24)	16	3.0

Table 9. The cost of Web Client and API Service on days with midterm exams.

Number of Instances	Duration (Hour)	Cost (US Dollars)
1	17	2.19555
16	4	8.26560
32	1	4.13280
48	2	12.3984
Total		26.99235

Table 10. The cost of Web Client and API Service for the month with midterm exams.

Day	Duration (Day)	Cost per day	Total Cost
With Midterm exams	7	26.99235	188.94645
Without midterm exams	21	3.099600	65.091600
Total cost per month			254.03805

Table 11. The number of instances for each time period on days with final exams.

Time Period	Number Of Users	Number Of Instances	Duration (Hour)
8:00 AM – 8:30 AM	2,357 (2,356.12)	8	0.5
8:30 AM – 9:00 AM	4,713 (4,712.24)	16	0.5
9:00 AM – 9:30 AM	7,069 (7,068.36)	24	0.5
9:30 AM – 10:00 AM	7,069 (7,068.36)	24	0.5
10:00 AM – 10:30 AM	7,069 (7,068.36)	24	0.5
10:30 AM – 11:00 AM	7,069 (7068.36)	24	0.5
11:00 AM – 11:30 AM	4,713 (4712.24)	16	0.5
11:30 AM – 12:00 AM	2,357 (2356.12)	8	0.5
12:00 AM – 1:00 PM	0	0	1
1:00 PM – 4:00 PM	2,357 (2356.12)	8	3

Table 12. The cost of Web Client and API Service on days with final exams.

Number of Instances	Duration (Hour)	Cost (US Dollars)
1	17	2.19555
8	4	4.13280
16	1	2.06640
24	2	6.19920
Total		14.59395

Table 13. The cost of Web Client and API Service for the month with final exams.

Day	Duration (Day)	Cost per day	Total Cost
With Final exams	14	14.59395	204.3153
Without final exams	16	3.099600	49.59360
Total cost per month			253.9089

According to Table 7., the monthly cost for the web client and API service during the exam-free month is $92.988. However, the expenses for the web client vary during months with midterm exams and months with final examinations due to fluctuating user numbers. In months with midterm exams, the cost is $254.03805 (See Table 10.), while the cost is $253.9089 in months with final exams.

4.3.2. The Cost of Database

The service cost calculation for the database, AWS Aurora MySQL, is based on the Aurora Capacity Unit (ACU) unit of measurement. ACU represents the database's resources, with 1 ACU equating to 2 gigabytes of memory. AWS has confirmed that AWS Aurora supports up to one hundred thousand transactions per second. Therefore, it is difficult to calculate the specific service cost for the database. In this study, the minimum service cost for the database will be determined by presuming a usage of 1 ACU per month, which is the minimum configuration option. The service cost for AWS Aurora is priced at $0.1 per ACU per hour in the Singapore region.

Table 14. The Cost for Database

Duration	Cost (US Dollars)
1 Hour	0.1
1 Day (24 Hours)	2.4
1 Month (30 Days)	72.0

According to Table 14., the service cost for a database with a usage of 1 ACU over a fixed period of 1 month is $72.

Table 15. Cost per month in each month's categories

Services	Cost Per Month (US Dollar)		
	Without Exams	With Midterm Exams	With Final Exams
Web Client	92.988	254.03805	253.9089
API Service	92.988	254.03805	253.9089
Database	72.000	72.000	72.000
Total	257.9760	580.0761	579.8178

According to Table 15., monthly service costs can be broken down into three categories. When there are no exams, the monthly service fee is $257.976. The monthly service cost is $580.0761 during months with midterm exams. The monthly service cost for months with final examinations is $579.8178.

5. Performance Evaluation

In this section, we will simulate system users and send requests to the API directly using JMeter [12]. Due to the fact that the web client performs no processing, it is unnecessary to assess its performance. We will conduct a performance evaluation by simulating users beginning with zero and increasing by 500 users every two minutes until we reach 2,500 users. Each user's request will be sent sequentially, meaning the next request will be sent when the previous request has received a response. This procedure will be repeated until the testing process is finished. The acceptable criteria for the test are that the error rate should be at most 5% of the number of inquiries during that time period, and the average response time must be at most 3 seconds.

Fig. 6. The Experiment Workload Simulationa

Table 16. Experimental workload simulation results

Thread	Sample	Average Response Time (ms)	Median	90% Line	95% Line	99% Line	Min response (ms)	Max Response (ms)	Error (%)	Throughput/sec
500	15,465	4,129	231	15,154	36,190	42,357	83	42,464	3.38	113.38
1,000	28,104	6,728	1,123	24,167	36,266	42,300	21	43,424	2.55	109.60
1,500	29,645	9,338	1,148	36,447	42,278	42,835	85	43,135	8.81	117.82
2,000	30,778	12,279	3,134	36,469	42,333	42,695	5	42,929	9.45	116.36
2,500	31,375	15,049	15,142	42,504	42,814	44,579	26	50,893	13.35	118.19

Table 16. demonstrates that the increase in users from 0 to 500 and from 500 to 1,000 have an error rate within the acceptable criteria of 3%, and the rest are having an error rate higher than 5%, and the average response times are higher than the acceptance criteria of 3 seconds, while the throughputs differ slightly. This is because AWS ECS can detect the number of requests with a minimum of 1 minute interval, which means the scaling process will take more than 1 minute to start executing the scaling policy and take some time to provision the new instance to be ready. It cannot scale up the instances to align with the number of users at that period. The scaling process requires pulling container images from AWS ECR [11], and we found that sometimes they fail to pull the container images due to the internal

service's network error, and it needs to re-start itself again, which will take more time than the usual scaling process and very difficult to predict the exact time needs in scaling process.

6. Architectural Re-designs

The architecture design has revealed that the system can handle variable user loads and significantly reduce costs. When there are no users in the system, it automatically scales down to the minimum size and can scale up to accommodate the number of users during specific time periods. However, the total monthly cost exceeds the predefined cost of $500, and the database is found to be excessively high compared to its actual usage. Requests entering the system are divided into two groups which are requests from students and requests from instructors. Instructor requests involve creating and editing exam rooms or exams, which are significantly fewer than student requests. This is because the number of student users is much higher than the number of instructors. Therefore, we have redesigned the database architecture by incorporating a non-relational database [5] to handle the high volume of student requests because it is faster than a relational database [6]. Student requests include entering exam rooms, viewing exams, and recording answers. The recorded answer data is processed after the exams are completed, and answer checking occurs. Hence, we store student answer data in the non-relational database and transfer it to the main database after finishing the exams. Additionally, we have changed the service for the main database from AWS Aurora MySQL to AWS EC2 MySQL. Using a scalable database with high-performance capabilities is no longer necessary due to most of the requests have been moved to a non-relational database, and the total cost of using a non-relational database AWS DynamoDB [9] and AWS EC2 MySQL is lower than using only a scalable primary database (See Table 17.).

Table 17. The comparison of database cost

Service	Cost per hour (US Dollar)	Cost per month (US Dollar)
AWS Aurora	0.1	72.00
AWS EC2 (t3.medium) + AWS DynamoDB (Read cost + Write cost)	0.05 + (0.00074+0.000148)	36.63

Additionally, we switched from AWS ECS to AWS App Runner [8] as the cloud service for the Web client and API service because AWS ECS needs at least 1 instance to run continuously, which costs money even when no users are using the system. Due to its ability to scale down to zero instances while keeping the same instance configuration of 0.25 vCPU and 0.5 GB of memory, AWS App Runner aids in the solution of this issue. Below is a breakdown of the AWS App Runner pricing (See Table 18.).

Table 18. The comparison of AWS ECS and AWS App Runner cost

Service	Cost per hour (US Dollar)	Cost per month (US Dollar)
AWS ECS	0.12915	92.988
AWS App Runner (vCPU + memory)	0.064 + 0.007	51.120
AWS App Runner (0 instances)	0.007	5.040

According to Table 18., the monthly cost for AWS App Runner is lower than AWS ECS, and AWS App Runner won't be charged the vCPU cost when idle.

However, When the AWS App Runner runs with zero instances (idle), the first incoming request will trigger the instance to be provisioned to a ready state. It will get a slower response time than sending the request when the instance is ready (See Table 19).

Table 19. The comparison of the response time of the first incoming request when AWS App Runner is in an idle and ready state.

Instance State	Response Time (s)
Idle	1.036
Ready	0.114

Table 20. Cost per month in each month's categories for the re-designed architecture

Services	Cost Per Month (US Dollar)		
	Without Exams	With Midterm Exams	With Final Exams
Web Client	51.12	139.586	139.657
API Service	51.12	139.586	139.657
Database	36.63	36.630	36.630
Total	138.870	315.802	315.944

Following the architecture redesign, the total monthly cost is reduced to $138.87 in the month without exams, $315.802 in the month with midterm exams, and $315.944 in the month with final exams. Therefore, the redesigned architecture brings the monthly cost within the predefined cost of $500.

Table 21. The experiment result of the re-designed architecture

Thread	Sample	Average Response Time (ms)	Median	90% Line	95% Line	99% Line	Min response (ms)	Max Response (ms)	Error (%)	Throughput/sec
500	10,422	5,834	873	21,629	28,773	45,969	492	46,347	1.63	84.31
1,000	26,842	7,102	814	24,599	28,660	49,674	468	63,131	0.12	103.90
1,500	27,946	10,137	1,112	28,750	42,650	63,103	461	63,177	1.22	108.50
2,000	28,436	13,613	1,607	42,575	49,655	63,134	445	63,185	2.49	106.19
2,500	28,783	16,564	15,525	43,702	63,116	63,664	452	63,831	6.00	110.00

Table 21. The performance evaluation of the re-designed architecture demonstrates that the error rates are within the acceptable criteria of 3%, except when there are 2,500 users in the system, which has an error rate of 6% and exceeds the acceptance criteria. Additionally, the average response times exceed the threshold of 3 seconds, while the throughputs differ slightly compared to the first version of the architecture design.

7.   Conclusions and Future Works

This study presents an architecture design for a cloud-based online assessment system that can accommodate variable user traffic while considering service costs. The design uses the Amazon Web Services (AWS) cloud provider's automatic scaling capabilities to reduce costs during low or minimal utilization periods. The performance testing has been done by simulating the users and sending requests to the API, starting from 0 to 2,500 users. The result shows that the error rate of an increase in users from 0 to 500 and from 500 to 1,000 are in the acceptable criteria of 5%. Still, when the users in the system are more than 1500, the average response time and error rate exceed the acceptance criteria due to the insufficiencies of ready instances. Rather than that, the monthly service fees were determined to be $257.976 during non-exam months, $580.0761 during midterm exam months, and $579.8178 during

final exam months. The monthly cost for the month with midterm exams and the month with the final exam exceeded the predefined cost of $500. Then we redesigned the architecture by using a combination of a non-relational database and a non-scalable primary database to reduce the cost and changed the cloud service for the Web client and the API service from AWS ECS to AWS App Runner, which can be run at the minimum of 0 instances. The monthly cost after the redesign is reduced to $138.87 during non-exam months, $315.802 during midterm exam months, and $315.944 during final exam months. The performance of the re-designed architecture is also improving the error rate, while the average response times are slightly different from the first version of the architecture designed.

In further research, it is advisable to study the prediction of future user numbers. This is because we have observed that sometimes, AWS ECS and AWS App Runner fail to scale up to accommodate rapidly increasing user numbers due to the scalable services requiring time to increase the number of instances, resulting in most users being unable to access the system during that period. However, the system becomes usable once the number of instances is increased to align with the number of users. Therefore, it is necessary to configure a sufficient number of running instances in advance to match the anticipated increase in user numbers. The system can estimate user numbers by calculating the total number of examinees in the upcoming exam sessions in each exam room.

References

[1] Suping Wang Ligu Zhu, Mengke Cheng "Docker-based Web Server Instructional System", 2019 IEEE/ACIS 18th. International Conference on Computer and Information Science (ICIS), 17-19 June 2019, doi: 10.1109/ICIS46139.2019.8940219.

[2] Yunchun Li, Yumeng Xia "Auto-Scaling Web Applications in Hybrid Cloud Based on Docker", 2016 5th International Conference on Computer Science and Network Technology (ICCSNT), 10-11 December 2016, doi: 10.1109/ICCSNT.2016.8070122.

[3] Amirali Amiri, André van Hoorn, Uwe Zdun, Schahram Dustdar "Cost-Aware Multidimensional Auto-Scaling of Service- and Cloud-Based Dynamic Routing to Prevent System Overload". in 2022 IEEE International Conference on Web Services (ICWS), 10-16 July 2022, doi: 10.1109/ICWS55610.2022.00063.

[4] Brian Zhang, Valencia Zhang, Michael Hum "Budget in the Cloud: Analyzing Cost and Recommending Virtual Machine Workload". 2022 International Communication Engineering and Cloud Computing Conference, 28-30 October 2022, doi: 10.1109/CECCC56460.2022.10069750.

[5] Chafia Bouanaka, "Relational Database and NoSQL Inspections using MongoDB and Neo4j on a Big Data Application", 2022 7th International Conference on Computer Science and Engineering (UBMK), 14-16 September 2022, pp. 2, doi: 10.1109/UBMK55850.2022.9919589.

[6] Ticiana Capris, Pedro Melo, Nuno M. Garcia, Ivan Miguel Pires, Eftim Zdravevski, "Comparison of SQL and NoSQL databases with different workloads: MongoDB vs MySQL evaluation". 2022 International Conference on Data Analytics for Business and Industry (ICDABI), 25-26 October 2022, pp. 4, doi: 10.1109/ICDABI56818.2022.10041513.

[7] Amazon Web Services. (n.d.). AWS Elastic Container Service (ECS). Retrieved from https://aws.amazon.com/ecs/.

[8] Amazon Web Services. (n.d.). AWS App Runner. Retrieved from https://aws.amazon.com/apprunner/.

[9] Amazon Web Services. (n.d.). AWS DynamoDB. Retrieved from https://aws.amazon.com/dynamodb/.

[10] Amazon Web Services. (n.d.). AWS Aurora. Retrieved from https://aws.amazon.com/rds/aurora/.

[11] Amazon Web Services. (n.d.). AWS Elastic Container Registry. Retrieved from https://aws.amazon.com/ecr/.

[12] JMeter. Retrieved from (n.d.). https://jmeter.apache.org/

SUGARCANE YIELD PREDICTION BASED ON SATELLITE INDICES AND CLIMATE DATA IN THAILAND

Pornpimon Sinpayak[1]* and Noppadon Khiripet[1]

[1]*Knowledge Elicitation and Archiving Laboratory (KEA),*
National Electronics and Computer Technology Center (NECTEC),
112 Phahonyothin Road, Khlong Nueng, Khlong Luang District, Pathumthani, 12120, Thailand

*Corresponding address (E-mail: pornpimon.sin@nectec.or.th, +66-25646900 ext. 2899)

Abstract

The sugarcane plant holds significant importance in Thailand's agricultural sector and plays a pivotal role in the country's economy. Apart from its use as a primary source of sugar production, the crop also provides opportunities for the creation of value-added products such as ethanol, biomass electricity, and fertilizers. However, due to the impact of climate change, the output of Thai sugarcane yield is subject to significant variations. In order to mitigate the potential risks stemming from uncertainty, an accurate forecast of sugarcane yield can serve as a valuable guide for farmers, policymakers, and other stakeholders involved in the industry. This foresight allows for informed decision-making regarding the allocation of resources, production planning, and the development of effective market strategies. This study develops a model to predict annual sugarcane yields of 46 sugarcane-producing provinces in Thailand from year 2018 to 2021 by using multiple linear regression. We consider independent variables which are the climate data such as humidity, temperature, precipitation, and growing degree days (GDD), and the satellite data including the normalized difference vegetation index (NDVI). These historical data, obtained from Thailand Agricultural Data Collaboration Platform (THAGRI), are aggregated into each province. In addition, we apply mean absolute percentage error (MAPE) to compare the performance of multiple linear regression (MLR) and the method of using average values of historical yield as a predictor. Results show that MLR outperforms other models for forecasting annual sugarcane yield of each province from year 2018 to 2021, and the average value of MAPE is 3.97% approximately. By combining statistical techniques, remote sensing, weather data, and sugarcane yield data, this work provides more reliable forecasts of sugarcane yields for each province in Thailand.

Keywords: sugarcane; yield prediction; multiple linear regression; normalized difference vegetation index; absolute percentage error

1. Introduction

Sugarcane is a vital source of sugar and oil worldwide. Thailand is currently the fourth largest producer of sugarcane and one of the leading exporters of sugar globally. Over the past decades, there has been an upward trend in agricultural yields (measured in tons per rai) globally, largely driven by technological advancements. The trend in crop yield has mostly been linked to weather conditions.

Field measurements, which are time-consuming and expensive, are often used to evaluate cane yield. However, relying solely on harvest data for cane yield does not allow policymakers to implement timely coping strategies. To address this issue, many models have been developed. Experimental data from remote systems and environmental data from local climate stations can be used for these predictions. Son et al. effectively used Normalized Difference Vegetation Index (NDVI) data to predict yield. Linear and quadratic regression models have been employed to analyse the relationship between crop yield and NDVI. However, linear models tend to be unstable in complex crop production systems. Therefore, many studies have utilized non-linear models that can provide reliable yield estimates.

Agro-meteorological models have been developed to predict cane yield, primarily based on four variables: temperature, humidity, precipitation, and GDD (Growing Degree Days). Additionally, remotely sensed vegetation indices have shown great potential as explanatory variables in yield models. The most

commonly used vegetation index is the Normalized Difference Vegetation Index (NDVI), which indicates the health of plants [1]. Higher NDVI values correspond to healthier plants.

The aim of this study is to compare the performance of different algorithms in predicting sugarcane yield from 2019 to 2021. These algorithms include Multiple Linear Regression (MLR) with historical sugarcane yield used as predicted values. Additionally, MLR is employed to incorporate external factors such as temperature, humidity, precipitation, GDD, and NDVI. These prediction models utilize various data sources, such as remote sensing and climate data, to successfully forecast cane yield.

2. Data preparation

The study was conducted in 46 provinces in Thailand where sugarcane is planted as a major crop. The sugarcane yields monthly data in Thailand from year 2018 to 2021 obtained from the Office of Cane and Sugar Board (OCSB). In this work, natural climate data and vegetation index including temperature (T), humidity (H), precipitation (P), GDD and NDVI, will have a certain impact on crop yield. Therefore, these five factors, monthly data obtained from Cropwatch explorer, can be used to estimate crop yield in long or short term with enough and useful data based on the calculation of Pearson's correlation coefficient.

In the study area, the suitable cane varieties and specific agronomic traits, climate data, satellite index and cane yield were collected to form the dataset for building the prediction models. See Table 1.

Table 1: Statistics of collected data from year 2018 to 2021.

	Variables	Synonym	Min	Max	Mean	Std.
Agronomic traits	Growing degree days (GDD)	GDD	931.1	2751.5	1546.4	211.78
Climate data	Humidity (g/m^3)	H	21.87	86.17	72.02	8.16
	Temperature (C°)	T	22.68	34.46	28.05	2.42
	Precipitation (mm)	P	4.77	2333.63	523.24	443.79
Satellite index	Normalized difference vegetation index	NDVI	0.37	0.8	0.63	0.09
Cane yield	Cane yield (ton/rai)	y	5.45	11.53	8.75	1.77

The following steps are data handling:

• Outliers of yield data of 46 provinces in Thailand are replaced by its reliable value which are calculated from fixing yield of sub-districts in the province with outliers.

• For each year, all variables are considered as quarter and normalized into [0, 1].

• To test the data distribution is normal, the Shapiro-Wilk's W test is selected when p-value > 0.05. The results of testing all data are reject the null hypothesis.

3. Methodology

Time series data analysis means analyzing the available data to find out the pattern or trend in the data to predict some future values which will, in turn, help more effective and optimize business decisions. There are two popular time series forecasting model, multiple linear regression and multiple linear regression with stepwise selection, which can handle time series data.

3.1 Multiple linear regression

Multiple linear regression (MLR) is a popular technique which can be applied to predict a sugarcane yield at t^{th} year denoted y_t^*, using a set of independent variables x_{it_q} which is stand for a factor for the i municipality and t_q^{th} quarter. MLR model is described by

$$y_t^* = B_0 + \sum_{i=1}^{k} B_i x_{it_q} + \mathcal{E} \tag{1}$$

where B_0 is the y-intercept, B is the regression coefficient of the independent variable and $\mathcal{E}$ is error associated with the prediction.

3.2 Performance Comparison

3.2.1 Mean absolute percentage error (MAPE)

The mean absolute percentage error is a measure of prediction accuracy of a forecasting method in statistics.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - y_t^*}{y_t}\right| \tag{2}$$

where N is number of time series of dependent variables y.

3.2.2 The relation coefficient (R^2) and the root mean square error (RMSE)

The root mean square error is a frequently used measure of the differences between values predicted by a model or an estimator and the values observed.

$$RSME = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - y_t^*}{y_t}\right| \tag{3}$$

The relation coefficient (R^2) is used to evaluate the model performances.

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t - y^*)^2}{\sum_{t=1}^{N}(y_t - \bar{y})^2} \tag{4}$$

3.2.3 Percentage error (P)

The percent error is the difference between estimated value and the actual value in comparison to the actual value and is expressed as a percentage.

$$P = \left|\frac{y_t - y_t^*}{y_t}\right| \times 100 \tag{5}$$

For this work, we set three criterions to represent which province has high accuracy in each model including best fit, moderate fit and worst fit. The specific thresholds are shown as Table 2:

Table 2: Percentage error criterion

	Percentage error (P) criterion of yield prediction
high accuracy	$0 \leq P < 5$
moderate accuracy	$5 \leq P < 20$
low accuracy	$20 \leq P < 100$

4. Results

In this part, we compare the results of predicting sugarcane yield using two models in 46 provinces in Thailand from the crop year 2018 to 2021. We then calculate the average predictions for each province and compare them with the annual yield estimates for Thailand, which are aggregated at the province level. We measure the comparison in terms of the Mean Absolute Percentage Error (MAPE). The results indicate that the Multiple Linear Regression (MLR) model performs better than other models for all crop years (refer to Figure 1). The average MAPE for the MLR model is approximately 3.97% (refer to Figure 2). The high performance of the MLR model in predicting yield suggests that climate and satellite indices have an effect on sugarcane yield.
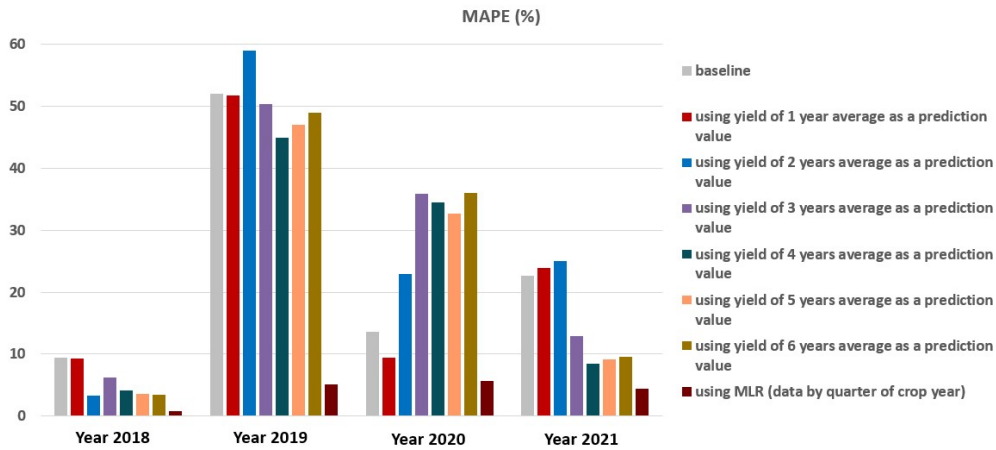
Fig. 1. The bar chart shows MAPE value for different yield prediction models.
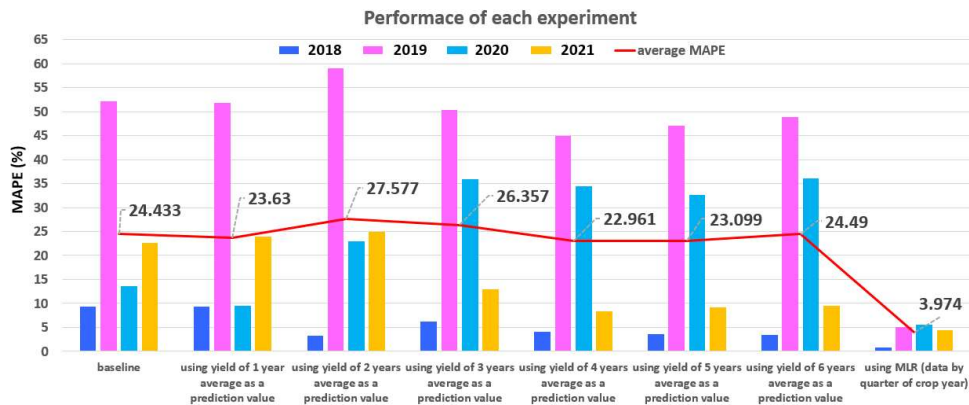
	baseline	using yield of 1 year average as a prediction value	using yield of 2 years average as a prediction value	using yield of 3 years average as a prediction value	using yield of 4 years average as a prediction value	using yield of 5 years average as a prediction value	using yield of 6 years average as a prediction value	using MLR (data by quarter of crop year)
Year 2018	9.37	9.34	3.33	6.22	4.11	3.57	3.38	0.84
Year 2019	52.07	51.80	58.99	50.32	44.88	47.04	48.93	5.08
Year 2020	13.59	9.46	22.98	35.94	34.44	32.62	36.05	5.60
Year 2021	22.69	23.91	24.99	12.92	8.39	9.14	9.58	4.36
average MAPE	24.43	23.63	27.57	26.35	22.96	23.09	24.49	3.97

Fig. 2. MAPE performance comparison

Based on the data presented on Figure 3, the observed versus predicted sugarcane yield for each model are explored and predicted yield from MLR is the most correlative with observation value when R^2 value is 0.9 approximately. Regardless of yield prediction performance, the external factors including

humidity, temperature, precipitation GDD and NDVI are the most effect to sugarcane yield. This because MLR outperform the other method which is contain only time series of sugarcane yield.

The analyses of percentage error shows that MLR tend to present low percentage error for forecasting yield and consequently high accuracy. From Figure 4, blue, orange and red represent high, moderate and low accuracy, respectively. For provinces with low accuracy (Figure 4) from year 2019 to 2021 are Lopburi, Singburi, Saraburi, Phetchaburi, Uttaradit, Uthai Thani and Phrae.
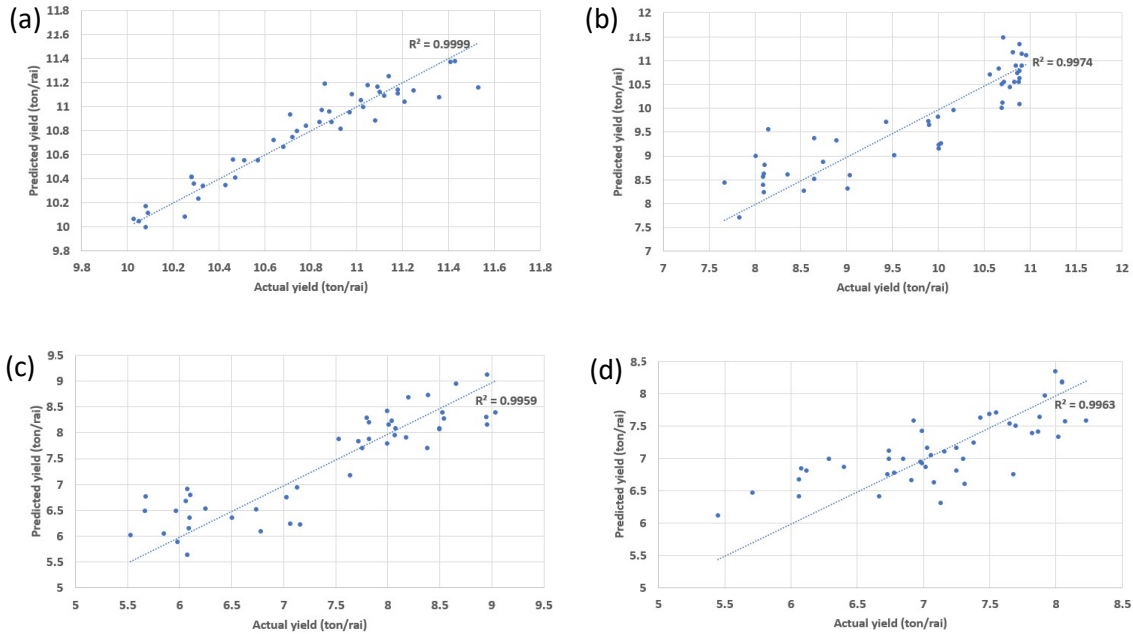
Fig. 3. The scatter plot compares between predicted yield and actual yield of MLR model. (a) year 2018 (correlation = 0.955, RMSE = 0.122); (b) year 2019 (correlation = 0.759, RMSE = 0.435); (c) year 2020 (correlation = 0.895, RMSE = 0.477); (d) year 2021 (correlation = 0.896, RMSE = 0.503).
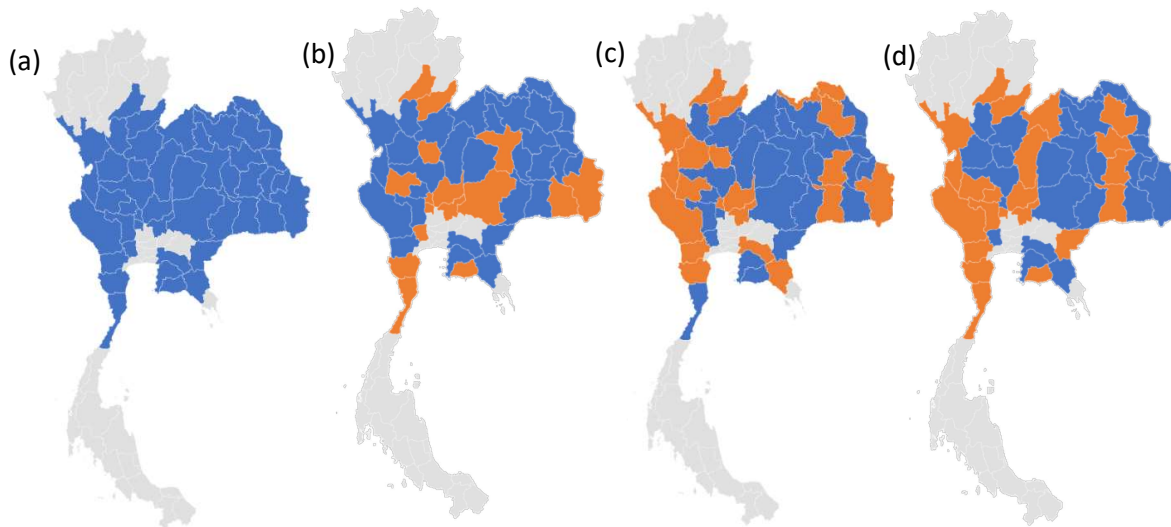
Fig. 4. Three maps show percentage error (P) of MLR yield forecasting model, denoted that blue = high accuracy and orange = moderate accuracy. (a) year 2018; (b) year 2019; (c) year 2020; (d) year 2021.
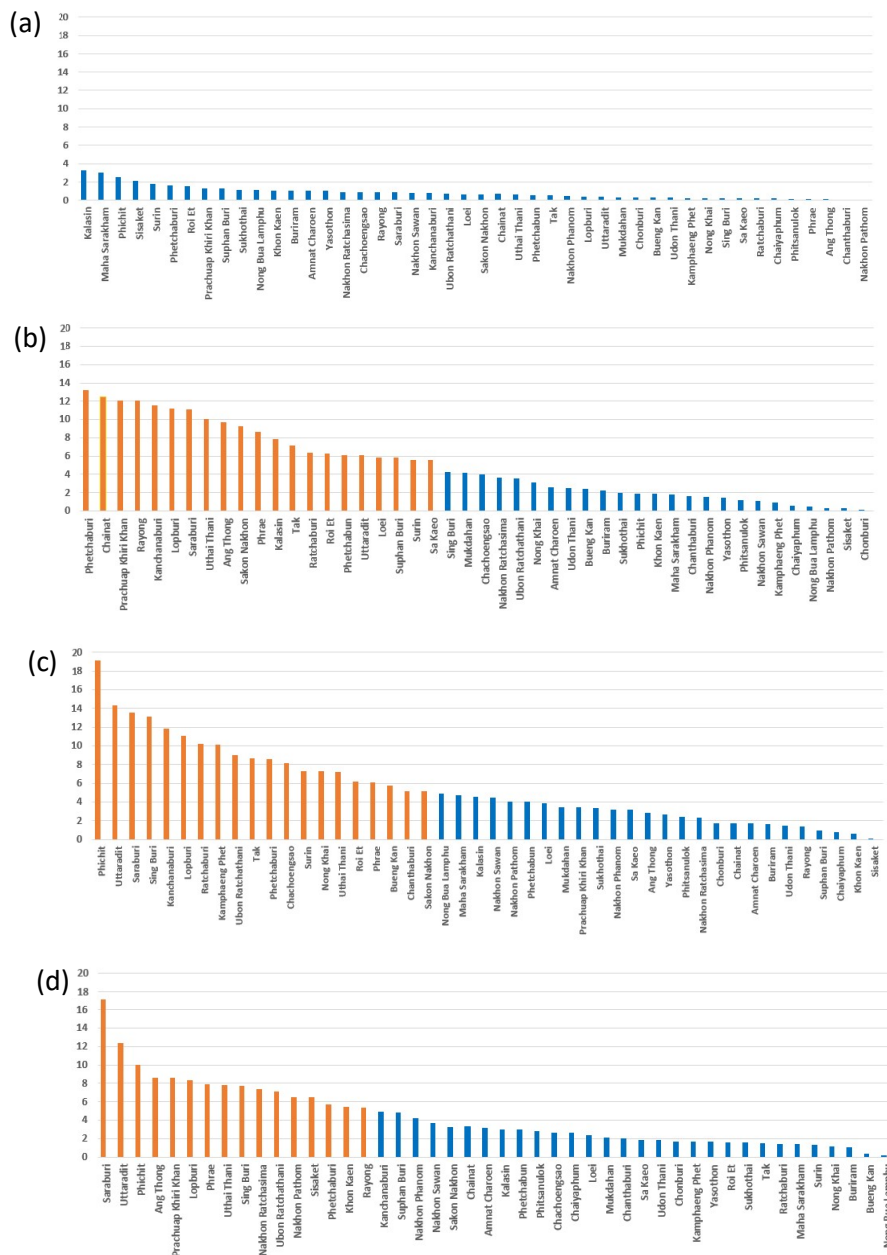
Fig. 5. The bar charts show actual yield and predicted yield from MLR model sorted by percentage error from highest to lowest. (a) year 2018; (b) year 2019; (c) year 2020; (d) year 2021.

5.   Discussion

This paper presents several methods for crop yield prediction using climate data and remote sensing data. It allows for real-time forecasting throughout the year and is applicable worldwide, especially for developing countries where field surveys are hard to conduct. We are the first to use modern representation learning ideas for crop yield prediction, and successfully learn much more effective features from raw data than the hand-crafted features that are typically used. To understand how our model is utilizing the input data, we provide an analysis inspired by stepwise selection. More specifically, we consider the effect of randomly permuting the values of a specific feature over the entire data. Next, we compare the result from all models and conclude that the quarter data as an input to MLR which performs the greatest accuracy comparing to a baseline model.

Acknowledgements

References

[1] Huang, Jing, et al. Analysis of NDVI data for crop identification and yield estimation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7.11 (2014): 4374-4384.

[2] Cao J, Zhang Z, Tao F et al (2021) Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. Agric for Meteorol 297(108):275.

[3] Ceglar A, Zampieri M, Gonzalez-Reviriego N et al (2020) Time-varying impact of climate on maize and wheat yields in France since 1900. Envir Res Lett 15(9):094039.

[4] W. L. Bai and F. P. Zhang, Estimation of winter wheat yield in Guanzhong area of shanxi province using SPOT VGT/ NDVI, Resour. Dev. Market, vol. 28, pp. 483–485, Jun. 2012.

[5] Gao Geng. Estimation drift of multiple linear regression and its determination method [J] statistics and decision, 2018, 14: 31-34.

[6] G. Papadavid, D. G. Hadjimitsis, L. Toulios, and S. Michaelides, Mapping potato crop height and leaf area index through vegetation indices using remote sensing in Cyprus, *J. Appl. Remote Sens.*, vol. 5, no. 1, p. 3526, Jan. 2011.

PROVINCE SCALE RICE YIELD PREDICTION USING STATISTIC METHOD

Chanon Boonkangwan[1*], Chayanee Kongsubchat[1], Jutarat Khiripet[1] and Noppadon Khiripet[1]

[1]Knowledge Elicitation and Archiving Laboratory (KEA),
National Electronics and Computer Technology Center (NECTEC),
112 Phahonyothin Road, Khlong Nueng, Khlong Luang District, Pathumthani, 12120, Thailand

*Corresponding address (E-mail: chanon.boo@nectec.or.th, +66-25646900 ext.2225)

Abstract

Over half of the world's population consumes rice, which is also an important source of income for many farmers. Therefore, crop yield prediction is useful for informing the expected yield accurately to help make decisions about planting and harvesting. While traditional yield forecasting systems typically rely on limited ground survey data and techniques such as supervised machine learning and statistical models to predict rice yield based on weather conditions, recent advances in free-to-use multi-time and multi-spectral remote sensing technology have provided a valuable tool for supporting these systems through accurate monitoring and estimation of pre-harvest crop yields. In order to leverage the benefits of this technology for rice yield prediction, climate data, observed rice yields, and remotely sensed data must be combined over an adequate time frame. Additionally, corresponding records of crop yields in the province should be taken into account for the development and implementation of provincial rice-yield predictions. Here, Multiple linear regression (MLR) and Ordinary Least Squares (OLS) are used in modelling to predict a response variable based on predictor variables. OLS is a specific MLR technique that estimates the model parameters (coefficients) by minimizing the sum of the squared errors between the predicted and actual values of the dependent variable. The weather conditions such as humidity, temperature, rainfall and Normalized Difference Vegetation Index (NDVI) are needed to use as predictors when predicting rice yield. Rice dataset was divided into 2 seasons that are in-season rice (na pi) and off-season rice (na prang). Each season has 6 types - Thai Hom Mali (in area), Thai Hom Mali (outer area), white rice, Pathum Thani 1, glutinous rice and others paddy where each type has different characteristics. The model's performance was evaluated using mean absolute percentage error (MAPE). The results show that we can predict in-season rice with high degree of correspondence with actual yield levels. On the contrary, the prediction of off-season rice using the available data poses a greater challenge, although the outcomes still exhibit a satisfactory level of performance.

Keywords: rice; yield prediction; statistical model; remote sensing; NDVI;

1. Introduction

Rice is a vital staple food for more than half of the world's population and plays a significant role in the income of many farmers [2]. The spread of COVID-19 has caused uncertainty in Thai Rice trade flows. But after the outbreak, the world demand for rice continues its dramatic growth. Thai rice is expected to keep rising on the back of growing demand in many countries. Principal rice exporting countries worldwide 2022/2023, Thailand is the world's second-largest rice exporter after India. As is global demand for Thai rice remains strong, which has sustained prices and made them competitive [2]. Accurate prediction of rice crop yields is crucial for making decisions to planting and harvesting. While traditional yield forecasting systems have relied on limited ground survey data and techniques such as supervised

machine learning and statistical models **Error! Reference source not found.**4], recent advancements in free-to-use multi-time and multi-spectral remote sensing technology have provided valuable tools for enhancing these systems through accurate monitoring and estimation of pre-harvest crop yields [5].

This study aims to predict provincial rice-yield effectively. The remote sensing predictor variables used in this research include weather conditions such as humidity, temperature, rainfall, and the Normalized Difference Vegetation Index (NDVI). It is necessary to use climate data over an appropriate time frame to maximize accuracy. To provide more specific,rice dataset was divided into 2 seasons that are in-season rice (na pi) and off-season rice (na prang). Each season has 6 types- Thai Hom Mali (in area), Thai Hom Mali (outer area), white rice, Pathum Thani 1, glutinous rice and others paddy where each type has different characteristics, such as planting methods, growth requirements, and environmental tolerances. These differences can affect the yield of each type of rice and can have an impact on the accuracy of the prediction models. By dividing rice into six types, the prediction models can be specific characteristics of each type of rice, thereby increasing the accuracy of the predictions. Additionally, separating rice into different types can be useful for farmers and policymakers in developing strategies to improve crop yield and manage resources effectively.

Multiple linear regression (MLR) and Ordinary Least Squares (OLS) are common methods used in statistical modeling to predict a response variable (dependent variable) based on one or more predictor variables (independent variables) [1]. MLR allows for analysis relation between multiple independent variables and a dependent variable, assuming a linear relationship. This method is particularly useful when multiple variables influence the response variable. OLS is a specific MLR technique that estimates the model parameters (coefficients) by minimizing the sum of the squared errors between the predicted and actual values of the predictor variable. Additionally, these methods provide a way to identify which variables have a significant impact on the response variable and can be used to create a predictive model for future yield prediction

The performance of the predictive models is evaluated using metrics such as mean absolute percentage error (MAPE) [6]. The study compares the predictions against the baseline, which is assumed to be the average yield of the country from the previous year for each type.

2. **Methodology**

2.1 *Problem description and overall process*

In the agriculture industry, crop yield prediction accuracy is significant challenge and important for farmers to make decisions about planting, harvesting, and marketing their crops. In this study, we used Multiple Linear Regression (MLR) and Ordinary Least Squares (OLS) methods to create accurate models for predicting the yield of six different types of rice- Thai Hom Mali (in area), Thai Hom Mali (outer area), white rice, Pathum Thani 1, glutinous rice and others paddy. Comparison against the baseline, which assumed to be the average yield of the country from the previous year.

In the agriculture industry, crop yield prediction accuracy is significant challenge and important for farmers to make decisions about planting, harvesting, and marketing their crops. In this study, we used Multiple Linear Regression (MLR) and Ordinary Least Squares (OLS) methods to create accurate models for predicting the yield of six different types of rice- Thai Hom Mali (in area), Thai Hom Mali (outer area), white rice, Pathum Thani 1, glutinous rice and others paddy. Comparison against the baseline, which assumed to be the average yield of the country from the previous year.

The overall process of this work is shown in **Error! Reference source not found.**. The data used in the study was preprocessed to remove any missing values or outliers then calculated as average, minimum, and maximum values to represent quarterly values to predict yield per year for each province. The data collection time for each province was considered and two models were developed. The *first model* based on data collected at the start of planting, while the *second model* based on data collected when the largest area of the province began planting.
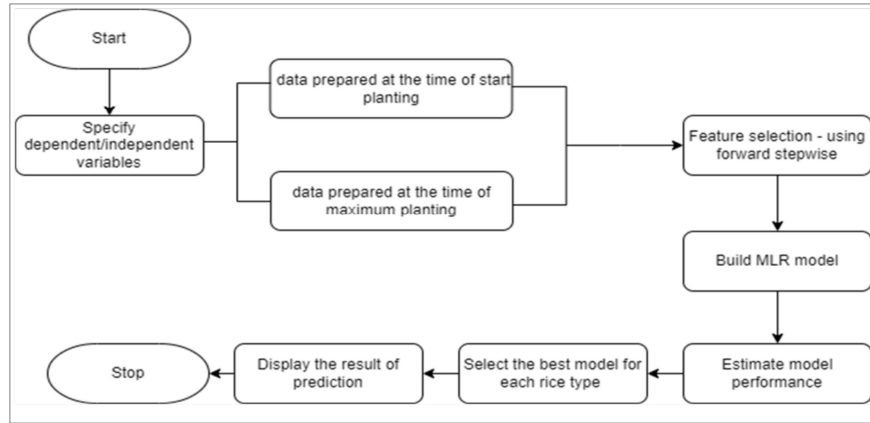
Fig. 1. Model workflow.

We used the average, minimum, and maximum values of NDVI, rainfall, temperature, and humidity for each province to represent quarterly values.

These values were used as predictor variables in the models. We also included the previous year's yield as a predictor variable, as it is known to have a significant impact on the current year's yield.

The results of the analysis were evaluated using the Mean Absolute Percentage Error (MAPE) metric and compared against the baseline, which was assumed to be the average yield of the province from the previous year. The comparison between the two models and the baseline was performed to determine the most accurate prediction of rice yield.

We applied our model to a real case of rice in Thailand. The 4 crop year data on the rice yield from crop year 2018/2019 - 2021/2022 was provided by Department of Agriculture (DOAE) [8], Humidity data was collected from the Hydrology and Irrigation Information (HII) [9]. Rainfall, temperature, and Normalized Difference Vegetation Index (NDVI) data were collected from the CropWatch API [10].

For this section, we divided problems into two cases for each type with different data time frame. First, we assume that collection time of remote sensing data was the quarter of start to plant for each province. In another case, collection time of remote sensing data was the quarter of the most area was plant.

2.2 Data Preparation

First, we used data on rice productivity from Department of Agriculture Extension (DOAE) that have 6 type of rice consists of Thai Hom Mali (in area), Thai Hom Mali (outer area), white rice, Pathum Thani 1, glutinous rice and others paddy. It contains.

- Amount of farmer's household
- Area of rice per month   (Unit: Rai)
- New plant area of rice   (Unit: Rai)
- Damaged area of rice   (Unit: Rai)
- Rice harvesting area   (Unit: Rai)
- Produce harvested rice   (Unit: Kg)
- Average of produce harvested rice (Unit: Kg/Rai)
- Average price that farmer sold (Baht/kg)

DOAE's data is data for each month in each province for each type of rice crop year 2018/2019 - 2021/2022. We explore data and we see many zero value from rice harvesting area column because farmer doesn't harvest on that month. If we see zero value from rice harvesting area column, we won't use data on that month but find month that start to plant and start to harvest.

Table 1 Example data of of-season others paddy.

Area	Area of rice per month (Unit: Rai)	New plant area of rice (Unit: Rai)	Damaged area of rice (Unit: Rai)	Rice harvesting area (Unit: Rai)	Produce harvested rice (Unit: Kg)	Average of produce harvested rice (Unit: Kg/Rai)	Average price that farmer sold (Baht/kg)
All Area	6,157,129.44	6,157,129.44	0	3,348.75	6,117,854.24	4,548,286,362.75	743.44
Bangkok	50,305.25	50,305.25	0	0	50,305.25	37,425,850.00	743.98
Nov 20	23,015.25	23,015.25	0	0	0	0	-
Dec 20	42,677.25	19,662.00	0	0	0	0	-
Jan 21	43,397.25	720.00	0	0	95.00	76,000.00	800.00
Feb 21	43,397.25	0	0	0	18,368.00	13,860,050.00	754.58
Mar 21	43,397.25	0	0	0	19,013.75	14,400,500.00	757.37
Apr 21	50,305.25	6,908.00	0	0	5,920.50	4,253,700.00	718.47
May 21	50,305.25	0	0	0	0	0	-
Jun 21	50,305.25	0	0	0	0	0	-
Jul 21	50,305.25	0	0	0	0	0	-
Aug 21	50,305.25	0	0	0	6,908.00	4,835,600.00	700
Sep 21	50,305.25	0	0	0	0	0	-
Oct 21	50,305.25	0	0	0	0	0	-

So, we plot data from average of produce harvested rice in boxplot to see some interesting data.
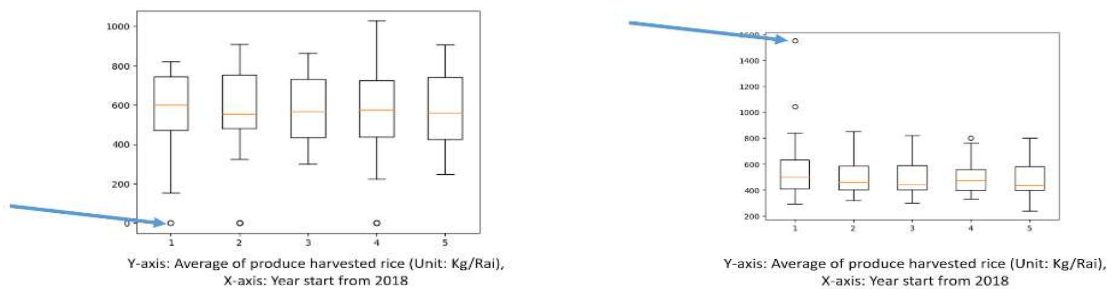
Fig. 2. (a) Average of produce harvested rice boxplot of in-season others paddy (b) Average of produce harvested rice boxplot of in-season glutinous rice

Include, province data that don't plant on that year, we also ignore that province on that year. After we plot the boxplot, we see some outlier's data from the graph. We need to fix it if data is the minimum outliner, so we change it to a minimum of boxplot (Q_0) from Fig.2 (a). If data is the maximum outliners, we change it to a maximum boxplot (Q_4) from Fig.2. (b).

After that, we need to know when the farmer starts harvest and maximum harvest from *Table 1*. The start harvest month has defined by looking at the first data is not zero of New plant area of rice (Unit: Rai) column and the maximum harvest month has defined by looking at the maximum value of Produce harvested rice (Unit: Kg). At the same time, we need to find actual yield by using data of Average of produce harvested rice (Unit: Kg/Rai) in first column is province. At the end, we got the perfect data in ***Error! Reference source not found.*** for analysis in next step.

Table 2 Data after preparing.

Province Name	Actual year	Actual yield	Start harvest month	Max harvest month	Start plant	Max plant	Max plant month	Max plant year	Start plant month	Max plant quarter	Start plant year	Start plant quarter
Bangkok	2562	781.00	2	2	201811	201811	11	2018	11	4	2018	4
Kanchanaburi	2562	901.95	3	7	201811	201903	03	2019	11	1	2018	4
Kalasin	2562	690.76	4	4	201812	201901	01	2019	12	1	2018	4

2.3   *Analytics*

In this study, MLR and forward stepwise OLS techniques were used for rice yield prediction. MLR is a statistical modelling for analysing relationship between a dependent variable and multiple independent variables, assuming a linear relationship. In this study, MLR was employed to develop a predictive model for rice yields based on various independent variables from Fig.3, such as weather conditions (humidity, temperature, rainfall) and the Normalized Difference Vegetation Index (NDVI).
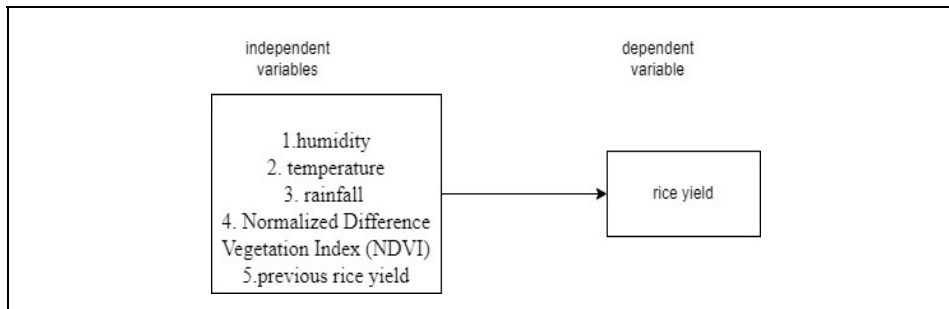
Fig. 3 The variable of the multiple linear regression

Forward stepwise OLS approach was applied to refine MLR model and select the most significant predictor variables. This is a variable selection technique that starts with an empty model and sequentially adds predictor variables based on their contribution to the model predictive power. This stepwise process continues until no further variables significantly improve the model performance or until a predetermined stopping criterion is reached.

The general formulation of MLR model can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$$

Where Y represents dependent variable (rice yield), β_0 is intercept, β_1, β_2, ..., β_p are coefficients associated with predictor variables X_1, X_2, ..., X_p, and ε represents error term.

MLR model aims to estimate values of coefficients (β_0, β_1, β_2, ..., β_p) that minimize sum of squared errors between predicted and actual values of the rice yield.

Forward stepwise OLS procedure involves iteratively adding the most significant predictor variable to the model based on statistical criteria such as the p-value. The process continues until the additional variables no longer significantly improve model performance.

To evaluate the performance of the predictive models, the mean absolute percentage error (MAPE) was employed as an evaluation metric. MAPE measures the average percentage difference between the predicted and actual values of the rice yield. Lower MAPE values indicate higher prediction accuracy.

The general formulation of MAPE can be expressed as:

$$\text{MAPE} = \frac{1}{n} \sum \frac{Y_i - Y^i}{Y_i} \times 100$$

Where: Y_i represents the actual value of the dependent variable for the ith observation, Y^i represents the predicted value of the dependent variable for the ith observation, n is the total number of observations.

3. Result

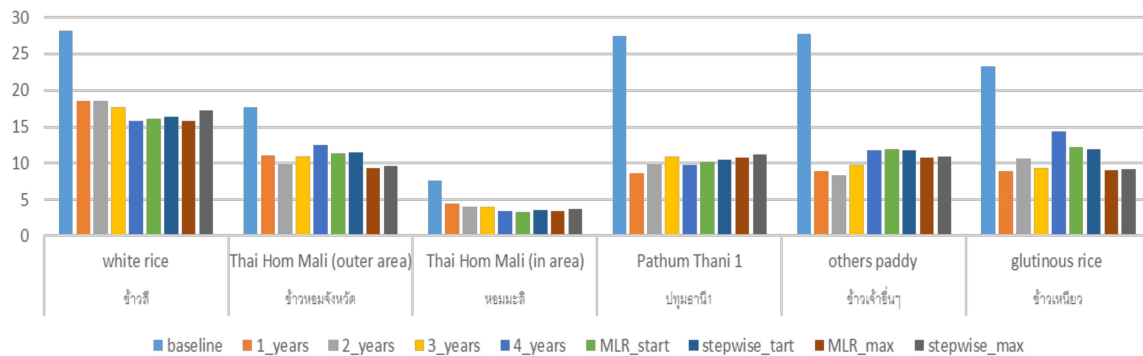From our method, we will separate into 3 results that is in-season rice, off-season rice, and both.

Fig. 1. MAPE of in-season rice

Fig. 1, shows us the accuracy for all models in each type of rice. Each type of rice will get different accuracy and it has a different best model. The lowest accuracy is white rice, it gets 15.79% with the model of average 4 years before. Most of the best models give an error between 5-10%, Thai Hom Mali (outer area) with 9.39% by MLP with maximum productivity, glutinous rice with 8.94%, and Pathum Thani 1 with 8.66% by using data from the latest year, and others paddy with 8.41% by the model of average 2 years before. But the most efficiency is Thai Hom Mali (outer area) with 3.32% by MLP with start crop model.
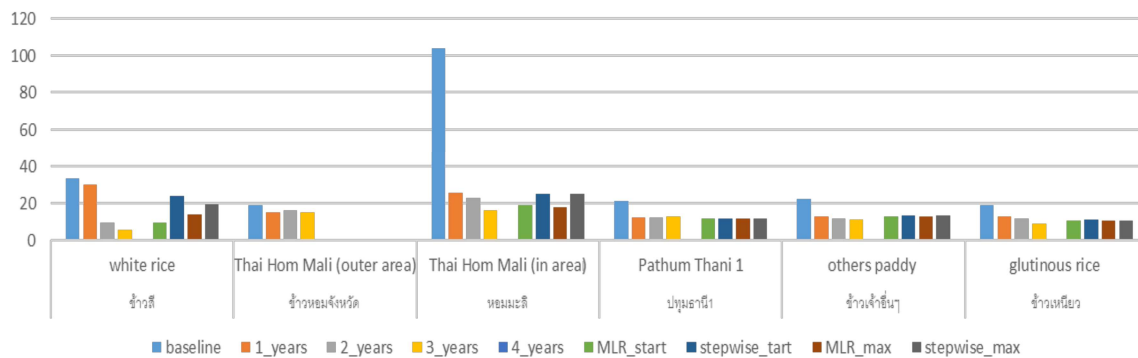
Fig. 2. MAPE of off-season rice

In the same way shown in Fig. 1 , most of off-season rice prediction will get more error because off-season rice some crop will not plant every year. The lowest accuracy is Thai Hom Mali (in area) with 16.27% by using average 3 years before. Next MAPE is about 10%, Pathum Thani 1 with 11.78% by using MLP with maximum productivity, others paddy with 11.53% and glutinous rice with 9.23% by using average 3 years before. White rice looks like it the most efficiency prediction with 5.90% by using average 3 years before but Thai Hom Mali (outer area) get no error by using all MLP model. We assume data that we have from Thai Hom Mali (outer area) is too simple for using Multiple Linear Regression, but it can only use Linear Regression.
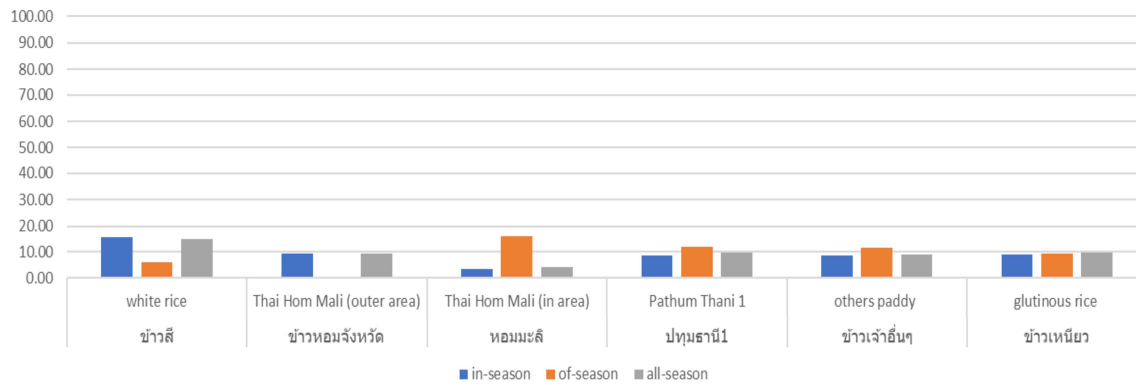
Fig. 3. MAPE of all-season rice

The key to this project is next year's prediction of all-season rice. Fig. 3 shows us the MAPE of all-season rice, we will see every type of rice after combining in-season and off-season together, MAPE is always below 15%. Because the amount of data in the in-season has bigger than the data of-season, so MAPE must follow in-season prediction. 14.94% is the MAPE of white rice, it is the most MAPE of all types of rice. MAPE of other is 9.79% from Pathum Thani 1, 9.49% from glutinous rice, 9.15% Thai Hom Mali (outer area), 9.06% from others paddy and the most efficiency is Thai Hom Mali (in area) with 4.28%. Therefore, the MAEP could be proposed for crop yield predictions in the study all region before the harvesting period.

4.   Conclusion

This research performed of MAPE predictions for rice crop yields in scale of province up to country level. Moreover, in-season rice prediction is more efficient than off-season rice prediction because of the amount of data and the continuity of data. For example: in-season rice has many data and continuity because the farmer will plant rice every year, but of-season is not. So, the accuracy of in-season rice is better. Moreover, popularity of each type of rice also affects its accuracy. In the future, we can use this model with another plant and scale up to predict all kind of plants.

References

[2]   Joachim VB, Ed and María SB. The changing economics and politics of rice: implications for food security, globalization, and environmental sustainability. *IRRI reports* 2005:7-20.
[3]   Exports of rice surge to 2.8m tonnes. [Internet]. Bangkok Post. (2023, May 21). Retrieved from https://www.bangkokpost.com/business/2574899/exports-of-rice-surge-to-2-8m-tonnes
[4]   Piyaporn S, Viwan K, Patchana S and Nop S. (2019). Forecasting Jasmine Rice Yield in Nakhon Ratchasima Province. *Science and Technology Research Journal Nakhon Ratchasima Rajabhat University* Vol.4 No.2, 25-37.
[4]   Norawat L and Nantachai K. (2015). Agricultural Yields Forecasting by Time Series Methods. Thai Industrial Engineering Network Journal, 1(1), 7–13. Retrieved from https://ph02.tci-thaijo.org/index.php/ienj/article/view/179093
[5]   Nguyen-Thanh Son, et al. (2022). Field-scale rice yield prediction from Sentinel-2 monthly image composites using machine learning algorithms. Ecological Informatics on ScienceDirect Vol.69 doi:https://doi.org/10.1016/j.ecoinf.2022.101618
[6]   Aiken LS., Stephen GW, and Steven CP. "Multiple linear regression." Handbook of psychology (2003): 481-507.
[7]   De Myttenaere, Arnaud, et al. "Mean absolute percentage error for regression models." Neurocomputing 192 (2016): 38-48.
[8]   Department of Agricultural Extension. (n.d.). Retrieved from https://production.doae.go.th/

[9]   Institute of Remote Sensing and Digital Earth Chinese Academy of Sciences. (n.d.). CropWatch. Retrieved from http://cloud.cropwatch.com.cn/user/login
[10] Hydro Informatics Institute. (n.d.). Retrieved from https://www.hii.or.th/

MODELING COVID-19 TRANSMISSION IN THAILAND: THE ROLE OF VACCINATION, NPIS, AND TREATMENT

Alhassan Ibrahim[1*], Usa Wannasingha Humphries[1*], Parinya Sa-ngiamsunthorn[1], Isa Abdullahi Baba[1,2]

[1]*Department of Mathematics, Faculty of Science, King Mongkut's University of Technology, Thonburi (KMUTT), 126 Pratcha Uthit Road, Bang mod, Thrung Khru, Bangkok 10140, Thailand*
[2]*Department of Mathematic Science, Bayero University, Kano, Nigeria*

* Corresponding author. Tel.: 0640436014, E-mail address:alhassan.i@mail.kmutt.ac.th
* Corresponding author. Tel.: 0863454892, E-mail address:usa.wan@kmutt.ac.th

Abstract

We present a comprehensive study of the transmission dynamics of COVID-19 by developing and analyzing a mathematical model. Our research focuses on investigating the qualitative properties of the model, specifically, the existence of equilibria. We identify two equilibrium points: disease-free equilibrium (DFE) and endemic equilibrium (EE). We establish that EE is globally asymptotically stable when $R_0 > 1$. To validate the applicability of our model, we used real COVID-19 data from Thailand and assessed its fitting performance. These results underscore the crucial role of vaccination campaigns in mitigating the spread of the disease. However, our research also uncovered a significant challenge in the form of vaccine hesitancy within certain populations, which hampers progress in containing the disease. By conducting a sensitivity analysis, we demonstrate that the effective utilization of non-pharmaceutical interventions (NPI) and the avoidance of public places are vital control measures to prevent future outbreaks of COVID-19 in Thailand.

Keywords: COVID-19; Stability; Sensitivity; Equilibrium points; Vaccine

1.   Introduction

Designing efficient control techniques for COVID-19 requires a comprehensive understanding of its transmission mechanics. Including vaccine reluctance in transmission models can be crucial for accurately representing vaccine coverage and its effect on disease spread [1]. By accounting for vaccine hesitancy, researchers and policymakers can assess potential barriers to achieving the desired vaccination rates and better evaluate the effectiveness of control measures [2, 3]. In this study, we provide a compartmental model that allows us to categorize susceptible individuals into different groups based on their vaccination status and susceptibility to the virus. This categorization enabled us to track the interactions and transmission patterns among these groups. By explicitly considering vaccine hesitancy, we can explore the impact of hesitancy on the overall transmission dynamics and effectiveness of vaccination campaigns.

2. Methods

2.1 Structure   In this study, we present a mathematical model that tracks the movement of individuals through eight (8) different states:

   I.   Susceptible (S): This group represents individuals who are vulnerable to contracting COVID-19.
   II. Vaccinated (V): Individuals in this compartment received the COVID-19 vaccine and were protected against the disease or acquired some level of immunity.
   III. Vaccine-hesitant V_H: This group comprises individuals who are hesitant or unwilling to

receive the COVID-19 vaccine, are assumed to be more vulnerable and acquire vaccination at a reduced rate [4].

IV. Exposed (E): Individuals in this compartment are exposed to the virus but are not yet infectious.

V. Asymptomatic infection (I_a): This group represents individuals who are infected with the virus but do not exhibit any symptoms.

VI. Symptomatic infection (I_s): Individuals in this compartment are infected with the virus and display symptoms of COVID-19.

VII. Hospitalized (H): This group included individuals who had acquired hospitalization due to a severe illness caused by COVID-19.

VIII. Recovered (R): Individuals in this compartment have recovered from the infection and are currently immune.

$$\frac{dS}{dt} = (1 - \rho)\Pi - (\lambda + \theta_1 + \theta_2 + \mu)S$$
$$\frac{dV}{dt} = \rho\Pi + \theta_1 S + \tau V_H - (\lambda + \kappa + \mu)V$$
$$\frac{dV_H}{dt} = \theta_2 S - (\lambda + \tau + \mu)V_H$$
$$\frac{dE}{dt} = \lambda S + \lambda(1 - \kappa)V + \lambda V_H - (\eta + \mu)E$$
$$\frac{dI_a}{dt} = \eta(1 - \omega)E - (\sigma_1 + \mu) I_a$$
$$\frac{dI_s}{dt} = \eta\omega E + \sigma_1 (1 - \varpi) I_a - (\sigma_2 + \delta_1 + \mu) I_s$$
$$\frac{dH}{dt} = \sigma_2 (1 - \phi) I_s - (\sigma_3 + \delta_2 + \mu)H$$
$$\frac{dR}{dt} = \sigma_1 \varpi I_a + \sigma_2 \phi I_s + \sigma_3 H + \kappa V - \mu R \qquad (1)$$

where,

$$\lambda = \frac{\beta(1 - \varepsilon)(I_a + \upsilon_1 I_s + \upsilon_2 H)}{1 + \alpha(I_a + I_s + H)} \qquad (2)$$

The mathematical equations in (1) are based on population dynamics and epidemiological concepts. These equations were used to model the spread of COVID-19 in a population and examine the effects of different scenarios and treatments. The model allows people to move between different compartments, such as susceptible, infected, and recovered compartments, based on factors such as transmission rates, vaccination rates, and the length of each stage of the infection. The model also considers the impact of vaccination campaigns, vaccine reluctance, and the course of the illness from the asymptomatic to symptomatic phases. In addition, the model includes a compartment for hospitalized people, which allows it to consider the potential impact on healthcare facilities. A list of all the parameters involved in this model is provided below.

Table 1. Model Parameters and Meaning

Parameter	Meaning
Π	Recruitment rate
ρ	Newly recruited Vaccinated
τ	Vaccine-hesitant individuals to Vaccinated

θ_1	Rate of Vaccination of Susceptible individuals
θ_2	Rate of Vaccine-hesitancy
κ	Vaccine efficacy
μ	Natural Fatality rate
ε	Rate of NPI usage
β	Contact rate
α	Rate of Vaccine Compliance
υ_1	Infectiousness of Is
υ_2	Infectiousness of H
H	The transition from Exposed to Infectious class
Ω	The rate at which infected individuals show symptoms
σ_1	Escape rate of individuals in Ia
σ_2	Escape rate of individuals in Is
σ_3	Recovery rate of individuals in H
Π	Recovery rate of individuals in Ia
Φ	Recovery rate of individuals in Is
δ_1	Death caused by COVID-19 in Is class
δ_2	Death caused by COVID-19 in H class

We find and alayzed the qualitative features of the model equilibrium point: the disease-free equilibrium (DFE) and the endemic equilibrium (EE). The model's basic reproduction number (R_0) is calculated as follows:

$$R_0 = \frac{\Pi \beta \eta_2 (1 - \varepsilon)(m_6 m_7 (1 - \omega) + \vartheta_1 m_7 \varphi_3 + (1 - \phi)\sigma_2 \vartheta_2 \varphi_3)}{m_1 m_2 m_3 m_4 m_5 m_6 m_7},$$

Where $m_1 = \theta_1 + \theta_2 + \mu, m_2 = \kappa + \mu, m_3 = \tau + \mu, m_4 = \eta + \mu, m_5 = \sigma_1 + \mu, m_6 = \sigma_2 + \delta_1 + \mu, m_7 = \sigma_3 + \delta_2 + \mu, \varphi_1 = \mu^2 \rho + (\rho \tau + \rho \theta_2 + \theta_1)\mu + \tau(\theta_1 + \theta_2), \varphi_2 = (1 - \rho)(k_3 + \theta_2)m_2 + (1 - \kappa)\varphi_1, \varphi_3 = (1 - \varpi)(1 - \omega)\sigma_1 + \omega m_5$.

2.2. Data collection and model validation

We used the COVID-19 data from Thailand from February to June 2023, provided by the WHO, to calibrate the model to real-world data. We collected relevant data during this time period, such as the number of confirmed cases. These data helped us better understand the pandemic dynamics and current trends in Thailand. As part of the model validation process, we compared the model's output with the observed data from the same time period. We fine-tuned the model parameters by analyzing the similarities and differences between the simulated and actual data. Our goal was to minimize the

discrepancy between the observed and simulated values to ensure that the model accurately reflected real-world situations.

2. Findings and Discussions

Concerted efforts should be made to reduce vaccine hesitancy and to combat the COVID-19 pandemic. This involves vaccinating individuals and disseminating accurate information regarding vaccine benefits, safety, and effectiveness. Trust can be built by collaborating with credible local authorities and medical experts, while targeted education campaigns can address concerns and debunked vaccine myths among specific groups. Simultaneously, a sensitivity analysis demonstrated the effectiveness of non-pharmaceutical interventions (NPIs), such as avoiding crowded areas and halting viral transmission. NPIs, including social distancing, mask-wearing, and handwashing, are important control measures to limit contact and reduce disease spread. Adhering to NPIs remains crucial for public health protection and the prevention of future outbreaks.

3. Conclusion

In summary, our study provides significant insights into the spread of COVID-19. We have identified and examined two equilibrium points: the disease-free equilibrium (DFE) and the endemic equilibrium (EE). Notably, EE demonstrates global asymptotic stability when $R_0 > 1$. To validate the efficacy of our model, we utilized authentic COVID-19 data obtained from Thailand, thereby ensuring its accuracy and reliability.

Undoubtedly, the paramount importance of prioritizing vaccination campaigns cannot be overstated in effectively combatting the rapid global spread of COVID-19. Vaccination has proven to be a pivotal factor in mitigating transmission rates and reducing the incidence of severe infection cases. Throughout our research, we have discovered a concerning variable: vaccine hesitancy prevalent within specific communities. This hesitancy poses a significant obstacle to progress in curbing the detrimental effects of the pandemic.

To address this issue, intentional education and communication interventions must be implemented, tailored to effectively reduce reluctance towards vaccinations and promote higher uptake rates across diverse demographics. By acknowledging these critical insights, we aim to contribute to the collective efforts to overcome the challenges posed by COVID-19, fostering a healthier and safer global community. In addition, our sensitivity analysis emphasizes the efficacy of non-pharmaceutical interventions (NPIs) and the significance of avoiding public areas as crucial measures to control and prevent future COVID-19 outbreaks in Thailand. NPIs, including practices like mask-wearing, physical distancing, and proper hand hygiene are paramount.

In summary this study provides important insights into COVID-19 transmission and the effectiveness of public health measures. Vaccination campaigns significantly reduce transmission; however, vaccine hesitancy remains a challenge. Non-pharmaceutical interventions, such as social distancing and wearing masks, are effective. A comprehensive approach is needed to effectively manage the pandemic, including widespread vaccination, addressing vaccine hesitancy, and strict adherence to non-pharmaceutical interventions. These measures can prevent future outbreaks and minimize their impact on public health.

References

[1] Usherwood, Th, Zachary La, and Vikas Sr. A model and predictions for COVID-19 considering population behavior and vaccination. *Scientific reports* 2021; 11.1:1-11. DOI:

[2] Kwok Kin, Kin-Kit Li, Wan Ln, Arthur Ta, Samuel Ye, Shui Sh. Infuenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: A survey. *International journal of nursing studies* 2021; 114:103854. DOI:

[3] Lyer Ga, Vivek Na, and David So. Vaccine hesitancy and monetary incentives. *Humanities and Social Sciences Communications* 2022; 9.1. DOI:

[4] Santavicca Ta, Cindy Ng, Rochelle Fr, Diana Mi, Anna Le, Cecile Ro. COVID-19 vaccine hesitancy among young adults in Canada. *Canadian Journal of Public Health* 2023; 114. 10-21. DOI:

[5] Lee Le, and Brian R. Generating optimal control simulations of musculoskeletal movement using OpenSim and MATLAB. *PeerJ* 2016; e1638. DOI:

A COMPARTMENT MODEL OF CANNABIDIOL PHARMACOKINETICS AFTER SUBLINGUAL ADMINISTRATION, AND APPLICATION

Thanachok Mahahong[1] and Teerapol Saleewong[1*]

[1]Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok, 10140, Thailand

Abstract

In this study, we present a compartmental model for investigating the pharmacokinetics of cannabidiol (CBD) after sublingual administration. The goals were to study the behavior of CBD concentration in the body after sublingual administration and use the results as a guide to determine the appropriate dosage and duration of CBD after sublingual administration to avoid harm to the body. The compartment model represents the body and describe the pharmacokinetics of the CBD in the body, it is transformed into ordinary differential equations (ODEs) by applying the law of mass action to describe the rate of change in CBD concentration in each compartment. Following that, the CBD concentration in each compartment was simulated using the exact solutions of the ODE system. To produce the simulation curves, the exact solutions and a sample of the actual data were written in Microsoft Excel, and the optimal rate constants were obtained by curve fitting to create a simulated curve that closely matched the actual data. In addition, we applied the simulation results to study the pharmacokinetics of CBD from cannabis sublingual oil of Chaophraya Abhaibhubejhr hospital formulation. The study found that our model was able to accurately describe the concentration of CBD in plasma as compared to the actual data and can also predict the concentration of CBD in other tissues. So, our model can be used as a tool to guide the determination of dosage and duration of CBD products administration, such as CBD sublingual oil, in actual treatment.

Keywords: Cannabidiol (CBD), Compartment model, Pharmacokinetics, Sublingual administration.

1. Introduction

1.1. *Background and rational of research*

Currently, many countries around the world have supported the legal use of cannabis for medical and research purposes. In Thailand, it was announced in the Royal Gazette on February 18, 2019 [1], that cannabis can be used for medical purposes. As a result, cannabis has been used medically to treat diseases that cannot be treated with modern drugs. However, research into the biological processes and effects of medical cannabis is still in its early stages.

Medical cannabis, often known as medical marijuana, is an herbal medication derived from cannabis plants used to treat specific symptoms or ailments. Cannabis contains about 100 active chemicals, referred to as "cannabinoids," which are unique to the cannabis plant [2]. One of these cannabinoids is cannabidiol (CBD), a non-psychoactive compound extracted from the cannabis plant. CBD has gained increasing attention in recent years due to its potential treatment for various symptoms, such as anti-seizure, anti-inflammatory, and analgesic properties, including its role in preventing nerve deterioration [2]. Sublingual administration is a convenient and effective route for CBD-based products. After administration, CBD is absorbed through the mucous membrane under the tongue into the capillaries in the lining of the tongue, bypassing first-pass metabolism in the liver. It enters systemic circulation directly before being distributed to other tissues. Additionally, CBD is excreted from the body without causing any internal chemical alterations, in the form of urine or feces [2].

The study of the kinetics of drug absorption, distribution, metabolism, and excretion is called pharmacokinetics [3]. It aids in understanding drug behavior and the relationships between drug absorption, distribution, and elimination rates inside the body, as well as in creating the appropriate

treatment response. There are numerous methodologies used to study pharmacokinetics, and one commonly utilized method is compartmental modeling. A compartment model is a mathematical model of the body or a portion of the body used to study the kinetics of physiological or pharmacological processes by dividing the body into compartments [4]. This approach facilitates comprehension of the biological processes underlying the kinetic behavior of drugs or chemicals administered to the body.

Since the kinetic process of CBD after sublingual administration is not well understood, and there is a lack of mathematical models to describe the pharmacokinetics of CBD in the body, we propose a compartment model in this work. The model aims to describe the pharmacokinetics of medical cannabis containing CBD after sublingual administration. It takes into consideration the rate of absorption, distribution of CBD into different body compartments, and elimination of CBD through excretion. The compartment model will be converted to ordinary differential equations (ODEs) to describe the rate of change in CBD concentrations in each compartment, utilizing the law of mass action. Mathematical methods will then be employed to solve the exact solutions of the ODE system. Subsequently, the CBD concentration in each compartment will be simulated using the exact solutions and optimized rate constants obtained through curve fitting, aiming to achieve a graph that closely represents the actual data of CBD concentration. The model will be validated by comparing the simulation results with the actual data. Additionally, we will apply the simulation results to investigate the pharmacokinetics of CBD products used in Thailand. An example of such a CBD product is the cannabis sublingual oil formulated by Chao Phraya Abhaibhubejhr Hospital [7], which we will investigate in this study. If our model is correct, it will serve as a valuable tool for describing and predicting the pharmacokinetics of CBD products after sublingual administration.

1.2. *Objective of Research*

The objectives of this study were to investigate the pharmacokinetics of medical cannabis containing cannabidiol (CBD) after sublingual administration in terms of the relationship between concentration and time, and to utilize the results as a guideline for determining the appropriate dosage and duration of sublingual CBD administration.

2. **Methodology**

2.1. *Actual data*

The actual data we consider is the concentration of CBD in plasma after 10 mg sublingual administration [5]. This data was obtained from a study conducted on twelve healthy volunteers (six males and six females) who were administered CBD through sublingual sprays over a period of 720 minutes (12 hours). Table 1 presents the mean plasma concentration of CBD over time.

Table 1. Mean CBD concentration in plasma after 10 mg CBD sublingual administration.

Time (min)	Concentration (ng/ml)	Time (min)	Concentration (ng/ml)
0	0.00	165	1.26
15	0.06	180	1.23
30	0.82	210	0.96
45	1.00	240	0.72
60	1.30	270	0.67
75	1.55	300	0.55
90	1.60	330	0.38
105	1.73	360	0.33
120	1.79	480	0.22
135	1.53	720	0.11
150	1.36		

2.2. Compartmental Modeling

In this section, we will create a compartment model to describe the pharmacokinetics of CBD after sublingual administration. We divide the body into five physiologically significant compartments. The first compartment is the absorption compartment (located under the tongue), the second compartment is the central compartment (representing the blood or systemic circulation), the third compartment is for rapidly equilibrating tissues such as the brain, lungs, liver, and kidneys, the fourth compartment is for slowly equilibrating tissues such as skin, bone, and fat. Additionally, we design the effect site compartment to describe the CBD concentration at the target site, as depicted in Fig. 1.
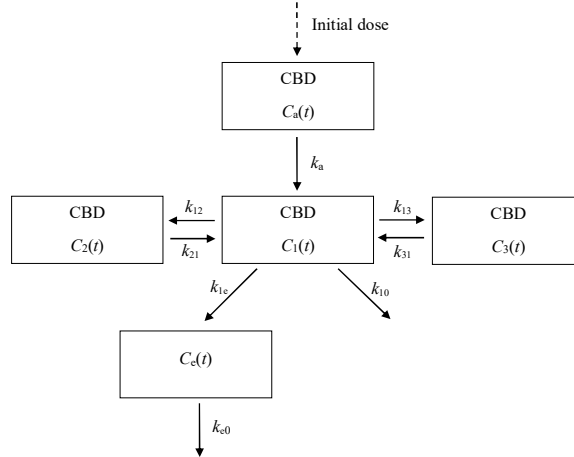
Fig. 1. A compartment pharmacokinetic model of CBD after sublingual administration

Let $C_a(t)$ represent the concentration of CBD in the absorption compartment, $C_1(t)$ represent the concentration of CBD in the central compartment, $C_2(t)$ represent the concentration of CBD in the rapidly equilibrating tissue compartment, $C_3(t)$ represent the concentration of CBD in the slowly equilibrating tissue compartment, and $C_e(t)$ represent the concentration of CBD in the effect compartment. Additionally, k_a is absorption rate constant, k_{12}, k_{21}, k_{13}, and k_{31} are the distribution rate constants of CBD in each compartment, while k_{10} and k_{e0} are the elimination rate constants of CBD from the central compartment and effect compartment, respectively.

Based on the compartment model shown in Fig. 1, this study will examine the case where the redistribution rate of CBD from the rapidly equilibrating tissue compartment and the slowly equilibrating tissue compartment to the central compartment is either equal or $k_{31} = k_{21}$. As a result, we obtain the following set of ordinary differential equations, which describe the change in CBD concentration in each compartment after sublingual administration using the law of mass action:

$$\frac{dC_a(t)}{dt} = -k_a C_a(t); \qquad\qquad C_a(0) = C_0, \tag{1}$$

$$\frac{dC_1(t)}{dt} = -(k_{10} + k_{12} + k_{13} + k_{1e})C_1(t) + k_{21}C_2(t) + k_{31}C_3(t); \quad C_1(0) = 0, \tag{2}$$

$$\frac{dC_2(t)}{dt} = k_{12}C_1(t) - k_{21}C_2(t); \qquad\qquad C_2(0) = 0, \tag{3}$$

$$\frac{dC_3(t)}{dt} = k_{13}C_1(t) - k_{21}C_3(t); \qquad\qquad C_3(0) = 0, \tag{4}$$

$$\frac{dC_e(t)}{dt} = k_{1e}C_1(t) - k_{e0}C_e(t); \qquad\qquad C_e(0) = 0, \qquad\qquad (5)$$

where C_0 is the initial concentration of CBD.

3. Results

In this topic, we will present the exact solutions of Eqs. (1) - (5) and demonstrate the simulation results of the pharmacokinetics of CBD after a 10 mg sublingual administration. Furthermore, we will apply the results to a CBD oil product.

3.1. Exact solutions

From Eqs. (1) - (5), using mathematical methods, we obtain the following exact solutions:

$$C_a(t) = C_0 e^{-k_a t}, \qquad\qquad (6)$$

$$C_1(t) = \frac{-k_a C_0 (k_a - k_{21}) e^{-k_a t}}{(k_a + \lambda_1)(k_a + \lambda_2)} + \frac{k_a C_0 (\lambda_1 + k_{21}) e^{\lambda_1 t}}{(\lambda_1 + k_a)(\lambda_1 - \lambda_2)} + \frac{k_a C_0 (\lambda_2 + k_{21}) e^{\lambda_2 t}}{(\lambda_2 + k_a)(\lambda_2 - \lambda_1)}, \qquad (7)$$

$$C_2(t) = \frac{k_a k_{12} C_0 e^{-k_a t}}{(k_a + \lambda_1)(k_a + \lambda_2)} + \frac{k_a k_{12} C_0 e^{\lambda_1 t}}{(\lambda_1 + k_a)(\lambda_1 - \lambda_2)} + \frac{k_a k_{12} C_0 e^{\lambda_2 t}}{(\lambda_2 + k_a)(\lambda_2 - \lambda_1)}, \qquad (8)$$

$$C_3(t) = \frac{k_a k_{13} C_0 e^{-k_a t}}{(k_a + \lambda_1)(k_a + \lambda_2)} + \frac{k_a k_{13} C_0 e^{\lambda_1 t}}{(\lambda_1 + k_a)(\lambda_1 - \lambda_2)} + \frac{k_a k_{13} C_0 e^{\lambda_2 t}}{(\lambda_2 + k_a)(\lambda_2 - \lambda_1)}, \qquad (9)$$

$$C_e(t) = -\frac{k_{1e}}{k_{e0}} C_1(t), \qquad\qquad (10)$$

where

$$\lambda_1 = -\frac{1}{2}\left(k_{10} + k_{12} + k_{13} + k_{1e} + k_{21}\right) + \frac{1}{2}\sqrt{\left(k_{10} + k_{12} + k_{13} + k_{1e} + k_{21}\right)^2 - 4\left(k_{10}k_{21} + k_{1e}k_{21}\right)},$$

$$\lambda_2 = -\frac{1}{2}\left(k_{10} + k_{12} + k_{13} + k_{1e} + k_{21}\right) - \frac{1}{2}\sqrt{\left(k_{10} + k_{12} + k_{13} + k_{1e} + k_{21}\right)^2 - 4\left(k_{10}k_{21} + k_{1e}k_{21}\right)}.$$

3.2. Simulation Results

First, we will estimate the initial concentration (C_0) of CBD in the absorption compartment after 10 mg sublingual administration using the concept of bioavailability. Bioavailability (F) represents the fraction of the administered dose of a drug that reaches the systemic circulation unchanged [6]. Thus, we can estimate the initial concentration in the absorption compartment as follows:

$$\text{Initial concentration of CBD in absorption compartment}\,(C_0) = \frac{\text{Maximum concentration of CBD in plasma}}{\text{Bioavailability of sublingual CBD}}$$

where the bioavailability (F) of CBD after sublingual administration will be approximate 35% (0.35) [7], and the maximum concentration of CBD in plasma is 1.79 ng/ml. Therefore, we obtain the estimated initial concentration, C_0 is 5.11 ng/ml.

Table 2. The optimal rate constants of CBD pharmacokinetics after 10 mg sublingual

Parameters	Value	Unit
k_a	0.0097	min^{-1}
k_{10}	0.0011	min^{-1}
k_{12}	0.1356	min^{-1}
k_{21}	1.0000	min^{-1}
k_{13}	0.0435	min^{-1}
k_{1e}	0.0115	min^{-1}
k_{e0}	0.0150	min^{-1}
λ_1	-0.0106	min^{-1}
λ_2	-1.1809	min^{-1}

By utilizing the exact solutions of $C_a(t)$ and $C_1(t)$, as presented in Eq. (6) and Eq. (7), respectively, along with the estimated initial concentration (C_0) and the optimal rate constant parameters obtained through curve fitting between the exact solutions and the actual data (as shown in Table 2), we can derive the simulation results of CBD in the absorption compartment and plasma, as displayed in Table 3.

Table 3. Actual and simulated CBD concentration in plasma after 10 mg CBD sublingual administration.

Time (min)	Actual CBD concentration in plasma (ng/ml)	Simulated $C_a(t)$ (ng/ml)	Simulated $C_1(t)$ (ng/ml)
0	0.00	5.11	0.00
15	0.06	4.42	0.55
30	0.82	3.82	0.94
45	1.00	3.30	1.20
60	1.30	2.85	1.38
75	1.55	2.46	1.47
90	1.60	2.13	1.52
105	1.73	1.84	1.52
120	1.79	1.59	1.49
135	1.53	1.37	1.44
150	1.36	1.19	1.37
165	1.26	1.03	1.30
180	1.23	0.89	1.22
210	0.96	0.66	1.05
240	0.72	0.49	0.88
270	0.67	0.37	0.73
300	0.55	0.28	0.60
330	0.38	0.21	0.49
360	0.33	0.15	0.39
480	0.22	0.05	0.15
720	0.11	0.00	0.02

3.3. *Application to CBD product*

In this section, we will examine the pharmacokinetics of CBD from cannabis sublingual oil formulated by Chaophraya Abhaibhubejhr hospital after sublingual administration.

Product basic information [8]:
Product name: Cannabis oil drops under the tongue 5 ml CBD
Component: CBD 10% W/V (100 mg/ml) (2.94 mg CBD/drop)
Packing size: 5 ml/bottle
Product form: Cannabis oil drips under the tongue
Implement: Use drops under the tongue in the initial size of 1 drop/day and increase the size slowly
Indications: Parkinson's disease and syndrome that does not respond to modern medicine
Manufacturer: Chaophya Abhaibhubejhr hospital, Prachin buri, Thailand.

Here, we will utilize the optimal rate constants from Table 2 to investigate the pharmacokinetics of CBD from cannabis sublingual oil formulated by Chaophraya Abhaibhubejhr hospital. In this study, we will consider an initial dose of 1 drop, equivalent to 2.94 mg CBD, per day. According to the product's basic information, the concentration of CBD oil is 100 mg/ml. Therefore, the concentration per 1 drop, or the initial concentration (C_0), is also 100 mg/ml. By using the exact solutions shown in Eqs. (6) - (10), we obtain the simulation results as shown in Table 4.

Table 4. The simulated CBD concentration after 1 drop CBD oil (2.94 mg CBD) sublingual administration

Time (min)	Absorption compartment (mg/ml)	Plasma (mg/ml)	Rapidly distribution tissues (mg/ml)	Slowly distribution tissues (mg/ml)	Effect site (mg/ml)
0	100.00	0.00	0.00	0.00	0.00
15	86.42	10.71	1.37	0.44	8.20
30	74.68	18.29	2.42	0.78	14.00
45	64.53	23.51	3.15	1.01	18.00
60	55.77	26.89	3.62	1.16	20.59
75	48.19	28.84	3.90	1.25	22.08
90	41.64	29.70	4.02	1.29	22.74
105	35.99	29.74	4.03	1.29	22.77
120	31.10	29.18	3.96	1.27	22.34
135	26.87	28.18	3.83	1.23	21.57
150	23.22	26.18	3.66	1.17	20.58
165	20.07	25.38	3.45	1.11	19.43
180	17.34	23.77	3.24	1.04	18.20
210	12.95	20.44	2.79	0.89	15.65
240	9.67	17.22	2.35	0.75	13.19
270	7.22	14.29	1.95	0.62	10.94
300	5.39	11.70	1.60	0.51	8.96
330	4.03	9.49	1.30	0.42	7.27
360	3.01	7.64	1.04	0.33	5.85
480	0.94	3.01	0.41	0.13	2.31
720	0.09	0.40	0.05	0.02	0.30
1440	0.00	0.00	0.00	0.00	0.00
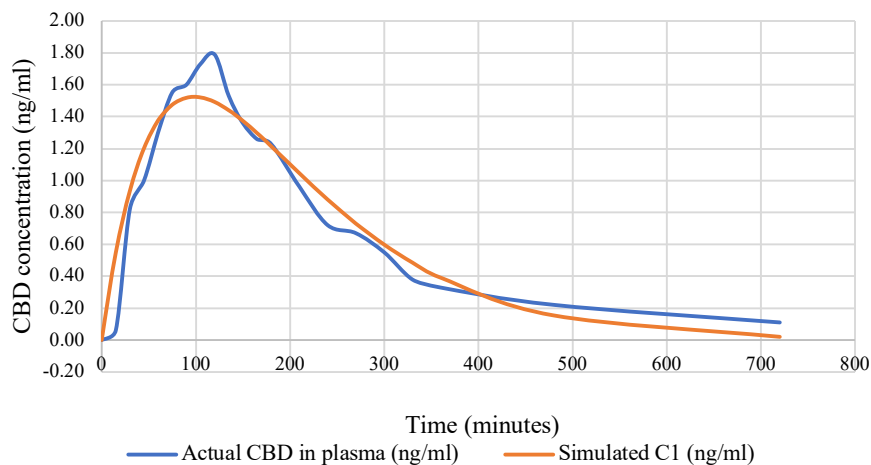
Fig. 2. Graph of actual and simulated CBD concentration in plasma after 10 mg CBD sublingual administration
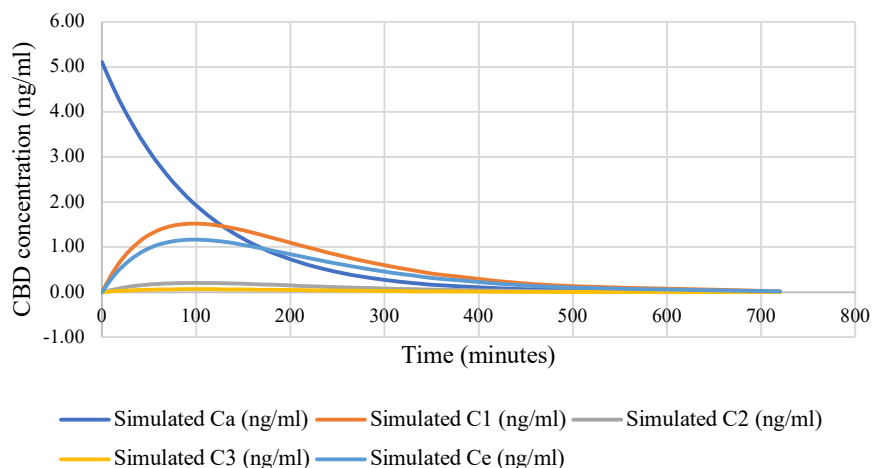
Fig. 3. Graph of simulated CBD concentration after 10 mg CBD sublingual administration in each compartment administration
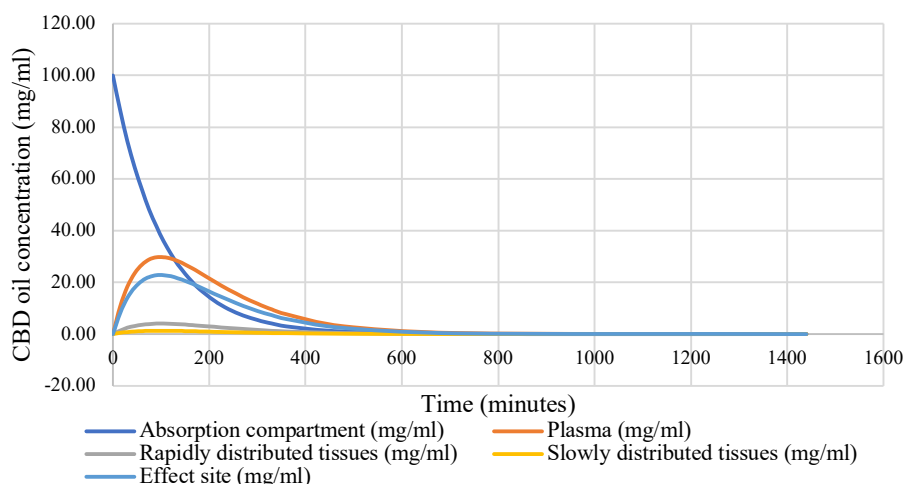
Fig. 4. Graph of simulated CBD concentration after 1 drop CBD oil (2.94 mg CBD) sublingual administration in each compartment

4. Discussion and Conclusion

In this study, we present a compartment model for investigating the pharmacokinetics of medical cannabis containing cannabidiol (CBD) after sublingual administration. The compartment model is converted into a set of ordinary differential equations (ODEs) using the law of mass action, referred to as Eqs. (1) - (5). These equations describe the rate of change in CBD concentration within each compartment. Subsequently, we obtain the exact solutions for this system, denoted as Eqs. (6) - (10), which were derived using mathematical methods.

For the simulation results, the solution curve of the model was plotted in Microsoft Excel using the exact solutions along with the optimal rate constants, which were refined through curve fitting between the exact solutions and the actual data. This process aimed to obtain a graph that closely represented the actual CBD concentration as shown in Fig. 2. It was observed that the CBD concentration in the absorption compartment ($C_a(t)$) rapidly decreased from its initial concentration (5.11 ng/ml), while the CBD concentration in plasma ($C_1(t)$) exhibited a rapid increase within the interval of 0 to 90 minutes (0 to 1.5 hours). The maximum concentration of 1.52 ng/ml was reached at 90 to 105 minutes, followed by

a rapid decrease within the interval of 105 to 360 minutes and a gradual decrease within the interval of 360 to 720 minutes, as depicted in Fig. 2. By comparing the simulation results with the actual data, it was found that the model achieved a good fit with an RMSE of 0.1571 and an R-squared value of 0.9240. Furthermore, the optimal parameters obtained from the curve fitting of $C_1(t)$ (as shown in Table 2) were used to calculate the concentrations of CBD in $C_2(t), C_3(t)$ and $C_e(t)$, the simulation results are presented in Fig. 3.

Moreover, we applied the optimal rate constants to investigate the CBD concentration in cannabis sublingual oil formulations from Chaophraya Abhaibhubejhr hospital. We considered the administration of 1 drop of CBD oil, equivalent to 2.94 mg CBD, with an initial concentration of 100 mg/ml, over a period of 1440 minutes (24 hours). The study revealed that the CBD concentration in plasma increases rapidly within the interval of 0 to 105 minutes, reaching a maximum concentration of 29.74 mg/ml at 105 minutes. Subsequently, it decreases quickly in the interval of 105 to 480 minutes, and gradually decreases after 480 minutes onwards. Concurrently, in the absorption compartment, the CBD concentration decreases rapidly from the initial concentration (100 mg/ml) in the interval of 0 to 480 minutes, and gradually decreases until it reaches 0 mg/ml after 480 minutes onwards. The CBD concentration in other tissues gradually increases with the distribution rate constants. The maximum concentration of CBD in rapidly distributed tissues is 4.03 mg/ml at 105 minutes, in slowly distributed tissues is 1.29 mg/ml at 90 to 105 minutes, and in the effect site is 22.77 mg/ml at 90 to 105 minutes. After that, the concentration in all compartments gradually decreases from the maximum concentration point and reaches 0 mg/ml in the interval of 105 to 1440 minutes. Based on these findings, we concluded that CBD oil should be administered again at least 480 minutes (8 hours) after the initial dose. This ensures that CBD concentrations in plasma and other tissues are sufficiently lowered, thereby minimizing the potential for harm or side effects to the body.

The study demonstrated that our model accurately described the concentration of CBD in plasma compared to the actual data, and it also successfully predicted the CBD concentration in other tissues. Therefore, our model can serve as a valuable tool for guiding the determination of dosage and duration for CBD sublingual products, thereby minimizing potential harm to the body.

Acknowledgements

References

[1] Narcotics Act, No. 7, B.E. 2562, Government Gazette, Volume 136, Special Section 19 (Kor) (Announced 18 Feb. 2019)
[2] Martin W. A Primer to Medicinal Cannabis: An introductory text to the therapeutic use of cannabis, Bedrocan International 2018, p. 1-57.
[3] William JS, William EW, Joseph TD, Robert AB and Jane MP. Concepts in Clinical Pharmacokinetics, Sixth edition, ASHP, 2014, p.2.
[4] Khanday MA et al. Mathematical models for drug diffusion through the compartments of blood and tissue medium, *Alexandria Journal of Medicine* 2017; 53: 245-249.
[5] Guy GW and Robson PJ. A Phase I, Open Label, Four-Way Crossover Study to Compare the Pharmacokinetic Profiles of a Single Dose of 20 mg of a Cannabis Based Medicine Extract (CBME) Administered on 3 Different Areas of the Buccal Mucosa and to Investigate the Pharmacokinetics of CBME per Oral in Healthy Male and Female Volunteers, *Journal of Cannabis* 2003; 3(4): 79-120.
[6] Currie GM. Pharmacology, Part 2: Introduction to Pharmacokinetics, *J Nucl Med Technol* 2018; 46(3): 221-230.
[7] Sophie AM, Nicole LS, Andrew SY and Saoirse EO. A Systematic Review on the Pharmacokinetics of Cannabidiol in Humans, *Front. Pharmacol* 2018; 9(1365): 1-13.

[8] Food and Drug Administration (FDA), Cannabis Products: Cannabis oil drops under the tongue 5 ml CBD (Translated) [Online], Available: https://cannabis.fda.moph.go.th/wp-content/uploads/PDF/Accept/ อภัยภูเบศSAS6N.pdf [2023, June 14].

MERCURY REACTIONS IN NATURAL GAS PROCESSING WITH SOLVATION EFFECTS: A DFT STUDY

Pheerawich Laokulwanich[1], Potchara Sukrason[1], Tat Tangpirul[1], Adis Khetubol[1]*, and Stephen John Turner[2]*

[1]Department of Physics and Astronomy, Kamnoetvidya Science Academy (KVIS), 999 Moo 1 Payupnai, Wangchan, Rayong 21210 Thailand

[2]School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), 555 Moo 1, Payupnai, Wangchan, Rayong 21210 Thailand

*Corresponding authors. Tel.: +66-911136984, +66-883819789
E-mail address: Adis.k@kvis.ac.th, Steve.t@vistec.ac.th

Abstract

Mercury, a toxic element usually found in the earth's crust, contaminates oil and natural gas. As it has a serious impact on health and ecosystem, more details about mercury reactions in the gas industry are very essential in order to regulate the mercury emissions. However, there are only a few works on the contaminated mercury in natural gas production which always faces these harmful substances. In this article, reactions of some mercury compounds such as $(CH_3)_2Hg$, CH_3HgCl, and CH_3HgI, were investigated by Density Functional Theory (DFT) at M062X/DEF2-TZVP and B3LYP/DEF2-TZVP levels of theory. The transition state and activation energy were calculated via Nudged Elastic Band method (NEB). The calculations were integrated with Solvation Model based on Density (SMD) on some solvents of different polarity that are present during the oil-gas transportation and refining processes such as water, toluene, xylene, 1,2-Ethanediol. The simulations were performed under the natural gas environments at controlled temperature 400K. The varied method of product varies reaction pathways. The impact from solvents lowered overall energy. The polar solvents tend to have lower energy than non-polar solvents.

Keywords: Mercury reaction; NEB; Natural gas; SMD; DFT.

1. Introduction

Mercury is a highly toxic element that poses significant risks to living organisms. It naturally occurs in the volcanic rock and can be released into the environment during industrial processes, including oil and gas extraction. [1,2]. The natural gas industry, as a crucial component of fossil fuel mining, contributes to elevated mercury emissions, with detrimental effects on the environment and human health. Over the years, these emissions have reached alarming levels, leading to widespread environmental contamination and associated adverse consequences. Mercury exists in different forms, including elemental mercury (Hg_0), mercury ions (Hg^+, Hg^{2+}), and mercury compounds such as CH_3Hg and $(CH_3)_2Hg$. Some of these compounds possess high toxicity which can affects to human and cause many health problems. They can enter human body through ingestion, inhalation or dermal exposure [2,3].

While much research has been done on mercury reactions in other environment, such as aquatic systems and coal combustion, little emphasis has been paid to understanding the reactions that occur in the natural gas industry. Given that the gas industry contributes significantly to mercury emissions, it is critical to investigate the reactions that occur in this setting in order to develop appropriate restrictions and mitigation techniques.

$$(CH_3Hg^+)_2 + S^{2-} \rightarrow HgS + (CH_3)_2Hg \qquad\qquad (1)$$

$$Hg + CH_3I \rightarrow CH_3HgI \qquad\qquad (2)$$

$$(CH_3)_2Hg + HgCl_2 \rightarrow 2C\ _3HgCl \qquad\qquad (3)$$
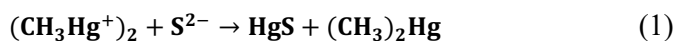
This study focuses on simulating three reactions (eq1.-eq3.) involving mercury compounds, (CH3)2Hg, CH3HgCl, and CH3HgI. These reactions were chosen based on their possible occurrence in the natural gas business, considering the compounds found in gas pipelines [4]. Additionally, iodine compounds are considered because of the presence of brine solution in some offshore processes [5]. To investigate the reaction, a part of Density Functional Theory (DFT), Nudged Elastic Band (NEB) was used to determine the transition state of reactions and calculate minimum energy path (MEP) of the reactions [6]. The simulations also included the Solvation Model based on Density (SMD) module [7], which accounts for the influence of solvents and other chemicals found in gas pipes, allowing for a more accurate portrayal of the reaction environment.

2. Computational details

All density functional theory (DFT) calculations were performed by ORCA5, a free open-source quantum chemistry software widely recognized in the field of computational chemistry [8]. The simulations were running on the M062X /DEF2-TZVP level of theory and running under four solvents from Solvation Model based on Density (SMD) consisting of water, toluene, xylene, and 1,2-Ethanediol. The computational process started with generating reactant and product of each reactions using atomic visualization software, Avogadro. After obtaining the most stable structure of the molecules by geometry optimization, Nudged Elastic Band (NEB). In this study, NEB-TS was employed to create reaction pathways and predict transition state of each reaction [9]. The activation energy, enthalpy, and vibrational frequency at the transition state were determined.

3. Result and discussion

3.1 Potentials of mean force.

In order to compare the corresponding barrier height and reaction energy in terms of free energy with those in the gas phase, The free energy on stationary points along the reaction path were calculated through NEB-TS in ORCA5 package. The calculations were varied by the two methods, M062X and B3LYP, and four solvents from SMD. Comparing the two methods, both methods overall show similar trends, except the first reaction (eq1.) There are two types of graphs of B3LYP method, the one that grows smoothly then becomes constant from water and xylene, and the fluctuated one from toluene and nonsolvent effect, representing two possibilities of rection pathways as shown in fig.2. and fig.3. respectively. Both pathways start with the same reactant, but in the first path, the top mercury atom releases a sulfur atom to bind with

another mercury, then shift to the carbon to form dimethylmercury. Meanwhile in the second path, a carbon atom goes straight forward to the further mercury atoms, forms dimethylmercury directly.
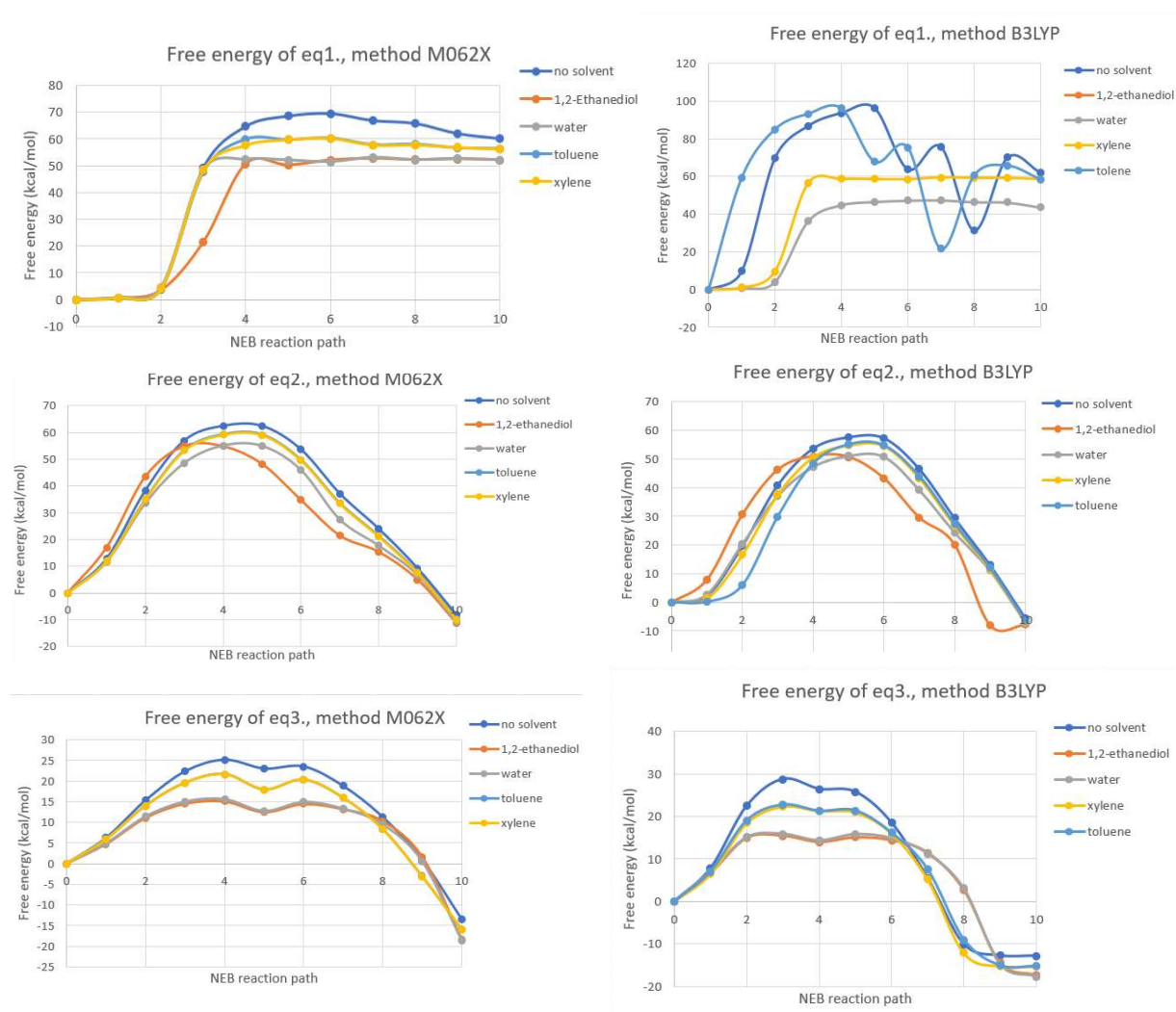

Fig. 1. Potentials of mean force of reaction path under difference solvation effects, calculated at M062X/DEF2-TZVP (left) and B3LYP/DEF2-TZVP (right).
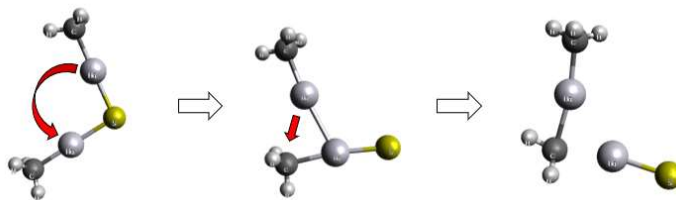

Fig. 2. Reaction pathway of eq.1. using B3LYP methods under water solvent.
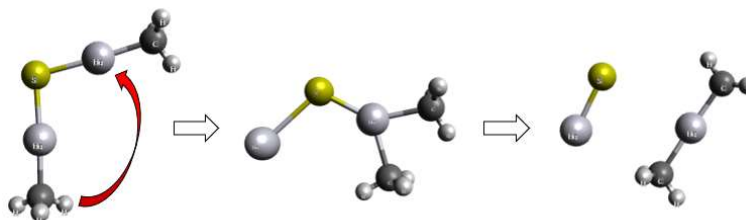
Fig. 3. Reaction pathway of eq.1. using B3LYP methods without solvent effects.

3.2 effects from solvents

From fig.1, the simulations without solvation effects made the highest energy, followed by toluene and xylene having close results, and 1,2-ethanediol and water the least. Polarity may be an important factor as it shows contrast between polar solvents, 1,2-ethanediol and water, and non-polar solvents, xylene and
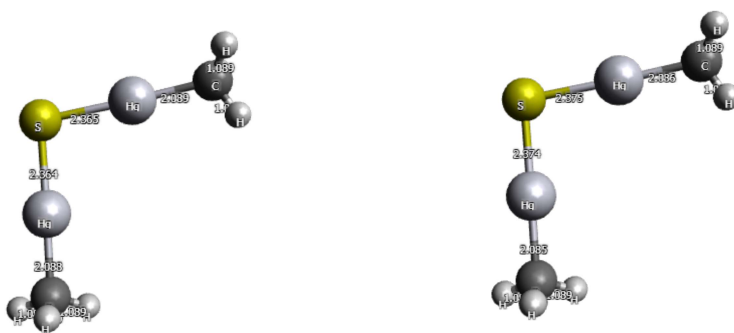
Fig. 4. Bond length of reactant complex comparing between 1,2-ethanediol (left) and toluene (right).

toluene. The bond length in reaction complex also indicates the difference between polar and non-polars. As shown in fig.4., the Hg-C and Hg-S bond length of reaction complex in toluene are 2.336A and 2.375A which is shorter than the those in 1,2-ethanediol which are 2.889A and 2.365A, supporting the higher energy barrier result from fig.1.

4. Conclusions

In this study, three reactions consists of $(CH_3Hg^+)_2 + S^{2-} \rightarrow HgS + (CH_3)_2Hg$, $Hg + CH_3I \rightarrow CH_3HgI$ and $(CH_3)_2Hg + HgCl_2 \rightarrow 2CH_3HgCl$ were investigated by DFT at M062X/DEF2-TZVP and B3LYP/DEF2-TZVP levels of theory, together with solvent effects from 4 different solvents. We find that B3LYP method can express the different reaction pathway. The solvent reduces the energy of the whole reaction. Polarity of solvents create a noticeable difference; polar solvents always creates slightly lower energy than non-polar, corresponding to the longer bound length found in reactant complex under impact from polar solvent.

Acknowledgements

References

[1] Wilhelm, S. M., & Bloom, N. (2000). Mercury in petroleum. Fuel Processing Technology, 63(1), 1-27. https://doi.org/10.1016/S0378-3820(99)00068-5

[2] Vasileios D. Koulocheris, Thermodynamic modelling and simulation of mercury distribution in natural gas, Ph.D. Thesis 2021, page 5 – 15. http://dx.doi.org/10.26240/heal.ntua.20774

[3] Bernhoft, R. A. (2012). Mercury toxicity and treatment: a review of the literature. Journal of environmental and public health, 2012. https://doi.org/10.1155/2012/460508

[4] Faramawy, S., Zaki, T., & Sakr, A. E. (2016). Natural gas origin, composition, and processing: A review. Journal of Natural Gas Science and Engineering, 34, 34-54. https://doi.org/10.1016/j.jngse.2016.06.030

[5] Obot, I. B., Onyeachu, I. B., & Umoren, S. A. (2019). Alternative corrosion inhibitor formulation for carbon steel in CO2-saturated brine solution under high turbulent flow condition for use in oil and gas transportation pipelines. Corrosion Science, 159, 108140.

[6] Henkelman, G., Uberuaga, B. P., & Jónsson, H. (2000). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. The Journal of chemical physics, 113(22), 9901-9904. https://doi.org/10.1063/1.1329672

[7] Marenich, A. V., Cramer, C. J., & Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. The Journal of Physical Chemistry B, 113(18), 6378-6396. https://doi.org/10.1021/jp810292n

[8] Neese, F. (2022). Software update: The ORCA program system—Version 5.0. Wiley Interdisciplinary Reviews: Computational Molecular Science, 12(5), e1606 https://doi.org/10.1002/wcms.1606

[9] Ásgeirsson, V., Birgisson, B. O., Bjornsson, R., Becker, U., Neese, F., Riplinger, C., & Jónsson, H. (2021). Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following. Journal of chemical theory and computation, 17(8), 4929-4945. https://doi.org/10.1021/acs.jctc.1c00462

THE DIFFUSION COEFFICIENT OF MERCURY AND MERCURY COMPOUNDS IN CRUDE OIL AND WATER AT DIFFERENT POSSIBLE TEMPERATURES

Tat Tangpirul[1], Potchara Sukrason[1], Pheerawich Laokulwanich[1], Adis Khetubol[1]*, and Stephen John Turner[2]*

[1]*Department of Physics and Astronomy, Kamnoetvidya Science Academy (KVIS), 999 Moo 1 Payupnai, Wangchan, Rayong 21210 Thailand*

[2]*School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), 555 Moo 1, Payupnai, Wangchan, Rayong 21210 Thailand*

*E-mail: Steve.t@vistec.ac.th, +66-883819789 (Prof. Dr. Stephen John Turner),
*E-mail: Adis.k@kvis.ac.th, +66-911136984 (Dr. Adis Khetubol)

Abstract

Although a very small amount of mercury is found in crude oil and natural gas, it causes significant harm to their transportation system in both the short and long terms. In the short term, the presence of mercury leads to toxic and contaminated crude oil and natural gas, which requires additional treatment for its removal and therefore increases the production cost. In the long term, its leftovers in the transportation pipeline can cause steel embrittlement, which is one of the major concerns for the pipeline's corrosion. The diffusion of the mercury species in the particular media, such as water and crude oil, determines their possibility of reaching the pipeline's surface and causing the harmful problem. In this work, we mainly focus on the investigation of the transport behaviour of mercury species through the self-diffusion coefficient using all-atom molecular dynamics (MDs) simulation. The self-diffusion of mercury species as Hg^0 and $(CH_3)_2Hg$ in a system of TIP4P water and the crude oil simulated as a mixture of hexane, heptane, octane, nonane, cyclohexane, cycloheptane, toluene, and benzene is determined through the mean-squared-displacement method (MSD) at different temperatures, 300 K, 350 K, and 400 K. The simulations have been repeated and analysed statistically to obtain a reasonable fit for each particular condition. The results reveal that elemental mercury can diffuse better than dimethyl mercury in water environments, but dimethyl mercury can diffuse better than elemental mercury in crude oil environments.

Keywords: Molecular dynamics simulation; diffusion coefficient; MSD; mercury; crude oil

1. Introduction

Crude oil is a petroleum product composed of hydrocarbons and other substances formed by geological formations under high pressure and high-temperature conditions. Mercury is a volatile element that is found naturally in crude oil. Despite its low traceable concentration, due to its high toxicity, it can provoke catastrophic structural fracture when diffused into the pipe and therefore lead to pipeline cracking. In the past, several industrial accidents have been recorded which were caused by the failure of equipment, as a result of mercury corrosion. For instance, the accident at the Skikda LNG plant at Skikda, Algeria, at which the corrosion from mercury triggered the explosion of LNG plant, caused 27 deaths and 74 injuries [1]. Not only does it have a toxic nature, but it also spoils catalysts that are used during processing; it shortens catalysts' life and requires more catalyst change-out, which leads to rising production costs. In addition to its impact on pipeline corrosion, mercury is known to contribute to severe diseases. Direct contact with mercury can lead to the development of severe cutaneous disease [2]. Furthermore, when mercury vapor is inhaled, it is absorbed into lungs, and subsequently transported to other tissues through the bloodstream [2].

The diffusion coefficient determines the rate of transfer of the diffusing substance across a unit area of one particular section, and defines the behaviours of mercury's diffusion [3]. Kuss J. et.al.,

determined the diffusion coefficient of elemental mercury in natural waters [4], Kuss J et.al., determined the diffusion coefficient of elemental mercury in water of oceanic salinity in 278K – 303K [5]. However, the behaviours of mercury in a crude oil environment – where it is typically found – are yet to be investigated, as well as the modelling of the alternative form of mercury, such as dimethyl mercury $((CH_3)_2Hg)$ which is also one of the forms of volatile mercury that exists in crude oil as a dissolved species [6].

Therefore, in this study, molecular dynamics (MD) simulation is used to determine the diffusion coefficient of elemental mercury (Hg^0) and dimethyl mercury $((CH_3)_2Hg)$. The simulations were performed in water and crude oil environments at 300 K, 350 K, and 400 K using Large Atomic/Molecular Massively Parallel Simulator (LAMMPS) [7]. The diffusion coefficient is then computed by using mean-squared displacement (MSD) method.

2. Simulation Settings

2.1 Simulation data acquisition

In this work, MD simulation is performed using large atomic/molecular massively parallel simulator (LAMMPS), a package widely used in modelling of materials at a wide range of temperatures [7]. In this work, TIP4P water model is used for modelling of water [8]. The OPLS force field [10-12] was applied for the parameterization of molecular species of the crude oil, which is composed of the structures proposed by [9], containing 144 hexane, 132 heptane, 156 octane, 180 nonane, 96 cyclohexane, 156 cycloheptane, 40 hexacosane, and 40 octatriacontane. For the modelling of elemental mercury (Hg^0), an all-atoms simulation method was employed. Whereas, for modelling dimethyl mercury, coarse-grained simulation techniques are employed instead. The interaction between atoms is defined by the Lennard-Jones potential with the addition of Coulombic pairwise interaction:

$$E_{lj} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right] \tag{1}$$

$$E_{coul} = \frac{Cq_iq_j}{\varepsilon r} \tag{2}$$

The parameters for dissimilar groups of atoms are computed by the arithmetic mixing rule:

$$\varepsilon_{ij} = \sqrt{(\varepsilon_{ii}\varepsilon_{jj})} \tag{3}$$

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \tag{4}$$

The Lennard-Jones parameters of TIP4P water are obtained from [8], and the parameters of mercury species are obtained from [13]:

Table 1. Lennard jones parameters of the components of the simulation system.

System types	Atom types	σ (Å)	$\mathcal{E}(kcal/mol)$
Water	H	0	0
	O	3.1589	8.0417
Crude oil	C	0.066	3.500
	H	0.030	2.500
	C(aromatic)	0.070	3.550
	H(aromatic)	0.030	2.420

Table 2. Atomic force field parameters of mercury species.

Atom types	σ (Å)	ε (kcal /mol)	Charges Value (e)
Hg^0	2.9	1.59	0
Hg (dimethyl)	2.9	1.59	0.208
CH_3 (dimethyl)	3.6	0.24	-0.104

Additionally, the internal interactions among adjacent atoms in the molecules, involving the bonding and angle potentials, are described according to the OPLS force field by harmonic potential model:

$$E_{bond} = K_{bond}(r - r_0)^2 \tag{5}$$

$$E_{angle} = K_{angle}(\theta - \theta_0)^2 \tag{6}$$

Where K_{bond} is a constant of bonding potential, with the unit of $kcal/mol(\text{Å})^2$, K_{angle} is a constant of angle potential, with the unit of $kcal/mol$, r_0 is the equilibrium bond distance, θ_0 is the equilibrium value of the angle.

2.2 Simulation Method

The simulation system comprises two main parts: the mercury species and the environmental species. The mercury species are created at the centre surrounding by the environment as shown in Fig.1 and Fig.2. The number of mercury species in each simulation is 5 for the systems of elemental mercury and 3 for the systems of dimethyl mercury, and the number of water molecules is 5786. The information about the crude oil's composition in each simulation box has been given in the previous section.
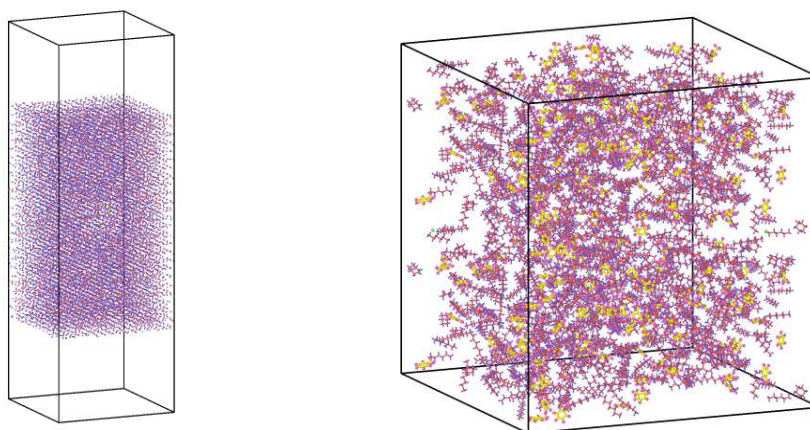
Fig. 1. Initial configuration of water system (left); Initial configuration of crude oil system (right).

MD simulations are performed with the use of periodic boundary conditions in all the direction (x, y, and z) of the simulation box to eliminate the surface effect. The real unit is used in the all the simulations. Firstly, the simulation is started from an equilibration performed by isothermal-isobaric Nose-Hoover thermostat (NPT) for each targeted temperature for 1 ns. Secondly, a canonical NVT ensemble is used to further equilibrate the system at the corresponding temperature via Nose-Hoover

thermostat for 1 ns. Finally, the data collection step – where the MSD is collected for the diffusion coefficient calculation – is performed by a microcanonical NVE ensemble. The timestep used in each simulation is 1 fs and then shortened to 0.5 fs in the data collection step in order to collect the data accurately, the simulation lasts for 1 ns.

2.2 Diffusion coefficient calculation

The mean-squared displacement (MSD) is used for the calculation of the diffusion coefficient based on the Einstein relation:

$$6D = \frac{\partial \langle |r(t) - r(0)|^2 \rangle}{\partial t}$$

where D is the diffusion coefficient and r(t) are the position of the molecule at time t.

2.3 Viscosity determination

The viscosity is determined by the cosine-shaped velocity profile:

$$V = \frac{A\rho}{\eta} \left(\frac{l_z}{2\pi} \right)^2 \tag{7}$$

where V is the velocity amplitude, A is the acceleration amplitude, l_z is the z-length of the simulation box, ρ is the density of the fluid, and η is the shear viscosity.

The cosine-shaped velocity profile is computed from the compute viscosity/cos command in LAMMPS. The simulations are performed with five different values of A. The dynamic viscosity is determined by computing the y-interception of the graph between shear viscosity and A – the acceleration amplitude.

3. Result & Discussion

3.1 Viscosity of water and crude oil

The viscosity of TIP4P water, as proposed by [14], and the viscosity of crude oil, decreases exponentially over the increase of temperature. According to Strokes-Einstein equation, the diffusion coefficient is inversely proportional to the viscosity [15]. Hence, the reduction of viscosity will result in the increase of diffusion coefficient.
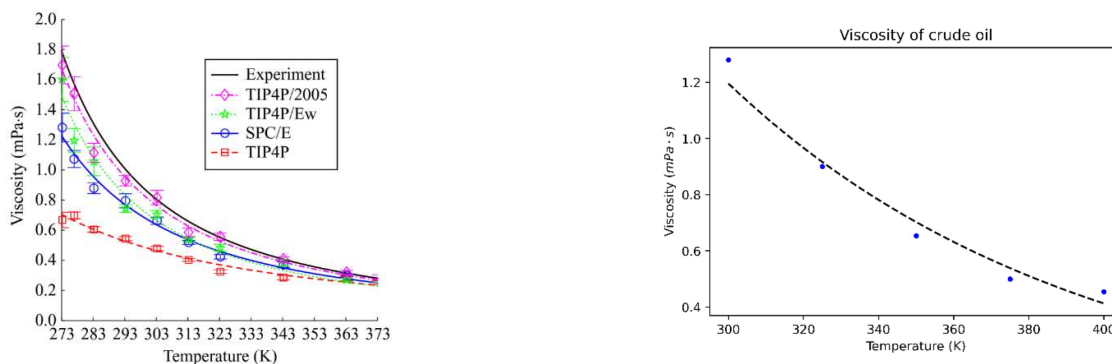
Fig. 2. Viscosity of TIP4P water (left) from [14]; Viscosity of crude oil (right).

3.2 Diffusion coefficients

The results reveal that the mean square displacement (MSD) of mercury species demonstrates a notable increase when diffused within a crude oil environment compared to water, indicating enhanced diffusion capabilities of mercury in crude oil. Furthermore, the results also demonstrate a positive correlation with increasing temperature as MSD increases when the temperature is increased. The simulations are repeated five times each in order to obtain the linear correlation.
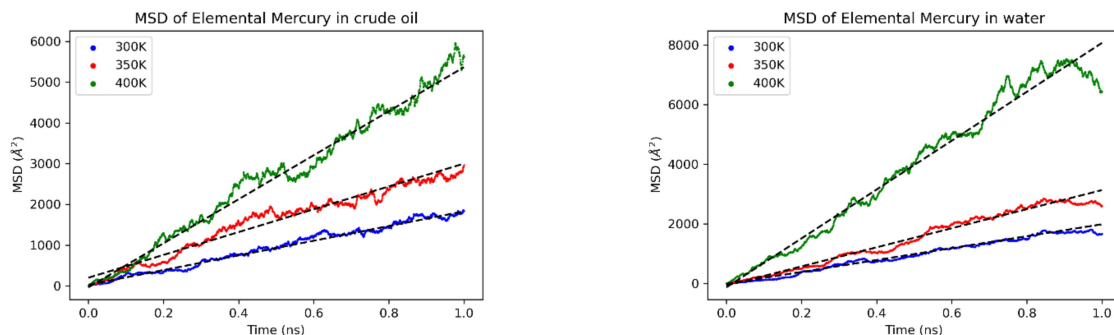

Fig. 3. MSD of elemental mercury in crude oil (left) and water (right).
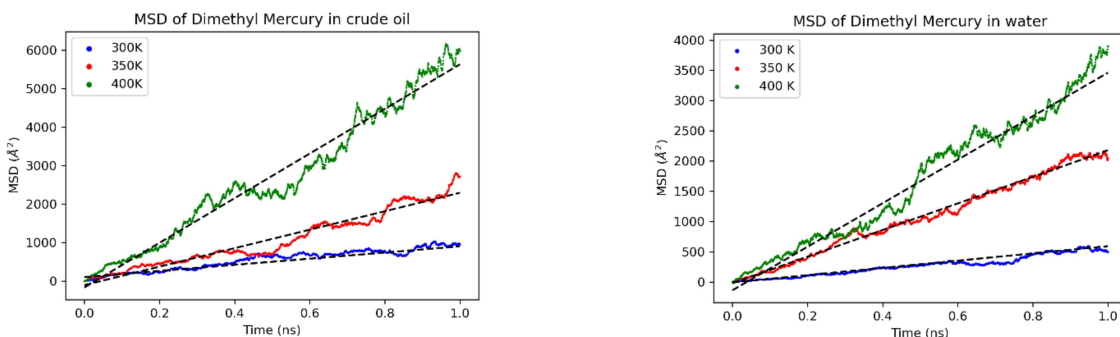
Fig. 4. MSD of dimethyl mercury in crude oil (left) and water (right).

As the temperature goes up, the size of the simulation box increases and the atoms move faster, so the MSD curve at higher temperatures is less smooth compared to lower temperatures. However, the result shows the linear trend with high r-squared value, as shown in Fig 3,4 and Table 3. The diffusion coefficients are then determined by calculating the slope of MSD and time, as also shown in table 3.

Table 3. Diffusion coefficients and root mean square error of mercury and dimethyl mercury in water and crude oil environments.

System types	Temperature (K)	D_{Hg} ($Å^2/fs$)	R^2	$D_{(CH3)2Hg}$ ($Å^2/fs$)	R^2
Water	300	0.00033	0.9852	0.00010	0.9593
	350	0.00053	0.9791	0.00037	0.9911
	400	0.00137	0.9782	0.00060	0.9760
Crude oil	300	0.00030	0.9857	0.00013	0.9351
	350	0.00047	0.9650	0.00060	0.9449
	400	0.00090	0.9846	0.00097	0.9672

4. Conclusion

In this work, molecular dynamics simulation was employed to determine the diffusion coefficients of elemental mercury and dimethyl mercury at 300 K, 350 K, and 400 K in two different environments. The first environment consists of 5786 TIP4P water molecules. The second environment contains 144 hexane, 132 heptane, 156 octane, 180 nonane, 96 cyclohexane, 156 cycloheptane, 40 hexacosane, and 40 octatriacontane. The diffusion coefficient was calculated from the slope of mean-squared displacement (MSD) and time. The viscosity of crude oil was determined by computing the cosine-shaped velocity profile.

The results show that viscosity has a negative correlation with temperature. The viscosity exponentially decreased as the temperature increased. In accordance with the viscosity, diffusion coefficients increase as the temperature rises. The diffusion coefficient of elemental mercury in water environment is greater than dimethyl mercury, meaning that elemental mercury can diffuse better than dimethyl mercury in water environment. Whereas in crude oil environment, the diffusion coefficient of dimethyl mercury is greater than elemental mercury in 350K and 400 K.

Acknowledgements

References

[1]   Mashyanov, N. (2021). Mercury in gas and oil deposits: corrosion problem. E3S Web of Conferences, 225,
        01009. https://doi.org/10.1051/e3sconf/202122501009
[2]   Boyd, A. S., Seger, D., Vannucci, S., Langley, M., Abraham, J. L., & King, L. E., Jr. (2000). Mercury   exposure and cutaneous disease. Journal of the American Academy of Dermatology, 43(1 Pt 1), 81–90.   https://doi.org/10.1067/mjd.2000.106360
[3]   Hartley, G. S., & Crank, J. F. (1949). Some fundamental definitions and concepts in diffusion
        processes. Transactions of the Faraday Society, 45, 801-818.
[4]   Kuss, J., Holzmann, J., & Ludwig, R. (2009). An elemental mercury diffusion coefficient for natural waters   determined by molecular dynamics simulation. Environmental Science & Technology, 43(9), 3183–3186.   https://doi.org/10.1021/es8034889
[5]   Kuss, J. (2014). Water-air gas exchange of elemental mercury: An experimentally determined mercury   diffusion coefficient for Hg0 water-air flux calculations. Limnology and Oceanography, 59(5), 1461–1467.   https://doi.org/10.4319/lo.2014.59.5.1461
[6]   Wilhelm, S. M., Liang, L., & Kirchgessner, D. (2006). Identification and properties of mercury species in   crude oil. Energy & Fuels: An American Chemical Society Journal, 20(1), 180–186.
        https://doi.org/10.1021/ef0501391
[7]   Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu, D. S., Brown, W. M., Crozier, P. S., in 't Veld,   P. J., Kohlmeyer, A., Moore, S. G., Nguyen, T. D., Shan, R., Stevens, M. J., Tranchida, J., Trott, C., &   Plimpton, S. J. (2022). LAMMPS - a flexible simulation tool for particle-based materials modeling at the   atomic, meso, and continuum scales. Computer Physics Communications, 271(108171), 108171.   https://doi.org/10.1016/j.cpc.2021.108171
[8]   Abascal, J. L. F., & Vega, C. (2005). A general purpose model for the condensed phases of water:
        TIP4P/2005. The Journal of Chemical Physics, 123(23), 234505.
https://doi.org/10.1063/1.2121687

[9]   Chen, X., Hou, L., Li, W., Li, S., & Chen, Y. (2018). Molecular dynamics simulation of magnetic field   influence on waxy crude oil. Journal of Molecular Liquids, 249, 1052–1059. https://doi.org/10.1016/j.molliq.2017.11.101

[10]   Jorgensen, W. L., & Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proceedings of the National Academy of Sciences of the   United States of America, 102(19), 6665–6670. https://doi.org/10.1073/pnas.0408037102

[11]   Dodda, L. S., Vilseck, J. Z., Tirado-Rives, J., & Jorgensen, W. L. (2017). 1.14*CM1A-LBCC: Localized   bond-charge corrected CM1A charges for condensed-phase simulations. The Journal of Physical   Chemistry. B, 121(15), 3864–3870. https://doi.org/10.1021/acs.jpcb.7b00272

[12]   Dodda, L. S., Cabeza de Vaca, I., Tirado-Rives, J., & Jorgensen, W. L. (2017). LigParGen web server: an   automatic OPLS-AA parameter generator for organic ligands. Nucleic Acids Research, 45(W1), W331–   W336. https://doi.org/10.1093/nar/gkx312

[13]   Lagache, M. H., Ridard, J., Ungerer, P., & Boutin, A. (2004). Force field optimization for organic mercury   compounds. The Journal of Physical Chemistry. B, 108(24), 8419–8426. https://doi.org/10.1021/jp049676x

[14]   Markesteijn, A. P., Hartkamp, R., Luding, S., & Westerweel, J. (2012). A comparison of the value of   viscosity for several water models using Poiseuille flow in a nano-channel. The Journal of Chemical   Physics, 136(13), 134104. https://doi.org/10.1063/1.3697977

[15]   Kruk, D., Masiewicz, E., Budny, J., Stankiewicz, A., Lotarska, S., Oztop, M., & Wieczorek, Z. (2022).   Diffusion in oils versus their viscosity – Insight from Nuclear Magnetic Resonance relaxometry. Journal of   Food Engineering, 317(110848), 110848. https://doi.org/10.1016/j.jfooodeng.2021.110848

Characterization and optimization of three-dimensional electrospun polymer fibrous scaffolds with carbon-based nanomaterials additives

Tiprat Vitayakitpipat[a], Wiwat Nuansing[b]

[a] School of Physics, Institute of Science, Suranaree University of Technology, 111 University Avenue, Muang, Nakhon Ratchasima, 30000, Thailand.

[b] Center of Excellence on Advanced Functional Materials, Suranaree University of Technology, 111 University Avenue, Muang, Nakhon Ratchasima, Thailand, 30000, Thailand.

Abstract

Three-dimensional (3D) electrospinning is a technique that uses electrical charges to draw a scaffold from a polymer solution and allowing scaffolds to be fabricated in a 3D shape. This research aims to investigate and identify a new polymer solution and appropriate parameters that can fabricate 3D polymer fibrous scaffolds that closely resemble the original or 3D designed shape. The 3D electrospinning method employs a modified 3D printing to print scaffolds, utilizing a positive electrode-connected needle (printed into the negative electrode-connected base) and a syringe containing a polymer solution with the inclusion of carbon-based nanomaterial additives. These additives offer the potential to enhance mechanical and biological properties of the 3D scaffolds. To investigate the effectiveness of the additives, five different types were employed. The characterization techniques include scanning electron microscopy (SEM), X-ray diffraction (XRD) and Fourier-transform infrared spectroscopy (FTIR) were used to classify each type of additive within the 3D polymer scaffolds. Additionally, a comparison is made with 3D scaffolds fabricated without any additives. Furthermore, polymer scaffolds fabricated using 3D electrospinning techniques will be characterized and compared using mechanical testing machines and the characterizations mentioned above. Through this investigation, we aim to optimize the fabrication process and understand the impact of carbon-based nanomaterial additives on the resulting 3D polymer scaffolds.

Keywords: 3D electrospining; polymer fibrous scaffold; carbon-based nanomaterials

1. Introduction

The three-dimensional (3D) printing is an additive manufacturing process that creates a physical object from a digital design. The process works by laying down thin layers of material in the form of liquid or powdered plastic, metal or cement and then fusing the layers together. The 3D printing has been used in many industries such as science, medicine and construction etc. The electrospinning technique is a scaffold fabrication method that uses the electrical charge to draw a very fine scaffold form the polymer solutions. It also can be used in many industries as well, such as medicine and science. And the combination of the 3D printing and the electrospinning is called the 3D electrospinning technique. It is a technique that uses the electrical charge to draw a scaffold from the polymer solution and can be fabricated the scaffolds in the 3D shape.

This research work will investigate and find a new polymer solution that can be able to be fabricate the 3D polymer fibrous scaffold. In addition, three-dimensional (3D) electrospinning will be used to fabricate 3D scaffolds using polymer solution and carbon-based nanomaterials additions. The carbon-based nanomaterials will be used as an additive because this carbon-based nanomaterial can enhance mechanical properties of the 3D scaffolds. Several research reports presented that scaffold containing appropriated amounts of carbon-based nanomaterials (graphene) have the higher tensile strength than that do not contained graphene. Furthermore, polymer scaffolds fabricated by the 3D electrospinning techniques will be compared and characterized using SEM, FTIR, XRD, and mechanical testing machine. This would open new opportunities for biomedical applications and development of innovation in healthcare industry.

For the polymer that suitable for this study, the biocompatibility and biodegradability of polylactic acid (PLA) (powder and pill appearance) is extensively studied. Over the past two decades, biodegradable polymers as PLA, PGA and poly-ε- caprolactone (PCL) have emerged as a class of biomaterials of growing interest in surgical applications, the drug delivery and tissue engineering (for example, sutures for wound healing, internal fixation devices for bone structures, carriers for the release of bioactive molecules, scaffold for the regeneration of tissues or organs). The synthetic copolymer PHEA-PLA shows particular properties that make itself suitable for the production of biocompatible scaffolds and does not cause a decrease in cell viability [1]. Polylactic acid (PLA) is a bioabsorbable polymer that is widely used to manufacture bioabsorbable implants like pins, plates, and screws which can degrade in the body mainly by hydrolysis as bone union gradually progresses. PLA can be utilized as a raw material in FDM to fabricate 3D-printed objects. However, there are some drawbacks that may limit the use of 3D-printed PLA in the medical field. For example, PLA is a relatively hydrophilic polymer and shows a low cell affinity and sometimes causes an inflammatory reaction when it comes into direct contact with surrounding tissues [2]. Polylactic acid (PLA) has low hydrophilicity [3]. From [4], polylactic acid was dissolved with acetone in the concentration of 15%w/v and this research also added tea tree oil or manuka oil into the PLA/acetone solution at five different concentrations of 2.5%v/v, 5.0%v/v, 7.5%v/v, 10.0%v/v and 15.0%v/v for antibacterial activity and mechanical properties improvements.

According to [5], Polyvinyl alcohol (PVA) is the synthetic polymers with the ideal characteristics for the fiber fabrications. It is biodegradable, biocompatible, water-soluble, as well as chemically and thermally stable material. Polyvinyl alcohol (PVA) is non-toxic polymer and is allowed by Food and Drug Administration to use in the materials that are in contact with food. Polyvinyl alcohol (PVA) has some disadvantages such as weak mechanical properties and thermal stability but can be combined with graphene to enhance these properties [6]. In [5] research, polyvinyl alcohol (PVA) is dissolved with the distilled water at a temperature of $90°C \pm 5°C$ in the concentration of 8 %w/v, the dispersions were agitated until it is fully dissolved and cooled at $65°C$. And from [2], polyvinyl alcohol (PVA) is dissolved with water into three different concentrations of 5wt%, 8wt% and 10wt% at the temperature of $90°C$.

The solvents that suitable for this study, Acetone is a relatively ecofriendly and less-toxic solvent compared to other organic solvents. But there are only few reports regarding the use of acetone in order to fabricate the scaffold [7]. Acetone mixed with water had been used. Dimethyl Sulfoxide (DMSO) is seen as alternative non-toxic solvents [8]. DMSO is water soluble and it could be diffused into water very quickly [9], it is produced in nature and is known to be a source of carbon and sulfur. It is used in many experiments such as pharmaceutical preparations, biological experiments and of course, as a polymer solvent. and it must be carefully used at high concentration [10]. Acetonitrile is much less toxic due to its stable C-CN bond and has been widely used as a polar aprotic solvent in organic synthesis and purification [11]. Ethanol or ethyl alcohol that is the organic solvent [12]. Ethanol can be produced by fermentation of sugars and utilized as partial gasoline replacement. It is gaining increasing attention due to its major environment benefits. [13]. It also can be produced by reductive carbonylation of methanol [14]. And methanol or methyl alcohol. It is a colorless, mobile and neutral liquid with light alcohol smell. It is useful as a drying fluid and completely miscible with water that is much lighter. Methanol is a good solvent for the organic industries, but it is volatile, highly flammable and very toxic chemical [14]. Dimethylformamide or DMF has been widely used in the industries (synthetic leather, polyurethane resin, Orlon, polyacrylic fibers, and so on) and has been termed the universal organic solvent because of its extensive miscibility with water and most common organic solvents (toluene, ethyl acetate, ketone, and so on). However, overexposure of DMF has been associated with hepatotoxicity, alcohol intolerance, possible embryotoxicity and teratogenicity in humans and animals. The current permissible exposure limit for DMF concentration in the air of a workplace is 10ppm in the United State and in Taiwan [15]. The industry that uses DMF as a material, requires the airborne levels in the workplace should not exceed a 10ppm. 8h of skin contact be avoided. [16] and also used chloroform as a solvent of PLA. Chloroform or trichloromethane ($CHCl_3$) is a clear, colorless, very volatile chlorinated solvent with characteristic odor and sweet taste. It is sensitive to light therefore it usually contains 0.75% of ethanol as a stabilizer to avoid photochemical transformation to phosgene and hydrogen chloride. It is soluble in most organic solvents, but has limited solubility in water [17]. Chloroform is quite a dangerous solvent, chloroform can affect

reproductive function and congenital anomalies, damage specific organs such as kidneys and liver, damage the nervous system, and exert devastating effects on the circulatory system. [18].

2. Materials and methods

2.1. Solution preparations

Polymers that were used in the preparations are Polycaprolactone (PCL), Polyvinyl alcohol (PVA) and Polylactic acid (PLA) (powder and pill appearance). All polymers that were mentioned are the biocompatible polymer which are non-toxic to the human skin and good for the environment. And the solvents that were used to dissolve the polymer are Acetone, Dimethyl Sulfoxide (DMSO), methanol, ethanol, acetonitrile, Dimethylformamide (DMF) and Chloroform. The additives that were added into some solutions are 5 different types of carbon-based nanomaterials. The first step, the additive carbon-based nanomaterials and the solvent were sonicated together in the sonicator bath for 30 minutes by using normal waves for the additive dispersion, then the polymer were added in the solutions and the solutions were stirred at the temperature that suitable for each solvent and each polymer (the temperature is used around 55°C -120°C) until all of the polymer is dissolved in the solvent. But there are some polymers that were not dissolved in the solvent. All solutions that were prepared had 7.3 and 7.5%W/V of concentrations. And the concentration of the carbon-based nanomaterial additions was 0.35%W/V of all solutions.

2.2. Fiber fabrication (3D electrospinning technique)

After the solutions were prepared, the 3D electrospinning technique is used only the completely dissolved polymer solutions to fabricate the fiber. The first step, the 3D prototype file was created by saving it from the Thingiverse.com website and using the Cura program as the slicer program. The printing parameters were set into 0% of the infill, 0°C of nozzle temperature and bed temperature and 0mm of top/bottom thickness to make the hollow cylinder (the easiest shape to be fabricated in 3D electropinning technique) with the size 5cm of diameter and 5cm of height. After that the file was saved in the SD card and put the card in the 3D electrospinning machine. The next step, the z-offset setting was set in the printer to make the distance between the needle tip to the collector. In this study, the parameters were set into 2.5cm and 5cm of the needle tip to the collector distance. Then, a needle was used with the size of (18G, 0.838mm of inner diameter and 20G, 0.603mm of inner diameter.) and 10mL syringe to pump the solution up and the long plastic tube is also connected with the syringe and the needle as in Fig. 1. (a), to make it moves easily when the 3D electrospinning jet is moving. After that, the syringe was put in the syringe pump to adjust the flow rate (in this research, 3, 5, 7, 8, 9, 10mL/h of flow rate were used) and to pump the solution out. Then, the needle was connected to the 3D electrospinning machine as in Fig. 1. (b), the wire is also connected to the needle tip and the collector to the power supply as in Fig. 1. (c) to create the electric field for drawing the very fine scaffold from the solution. Then the 3D electrospinning machine were started to print the 3D scaffold, it starts to pump the solution from the syringe pump and the applied voltage was slowly increased from the power supply to fabricate the scaffold and to find the most appropriate parameters (applied voltage, flow rate and needle tip to collector distance) that are suitable to these experiments. And as in Fig. 1. (d), some of the experiments had the guiding electrode to accelerate the 3D build-up process and build 3D structures with higher shape fidelity [19]. The guiding electrode is made from the aluminum foil. It had 6cm and 7cm diameter with 3cm of height cylindrical shape. The guiding electrode was placed on the middle of the collector when the fabrication started as in Fig. 1. (d).
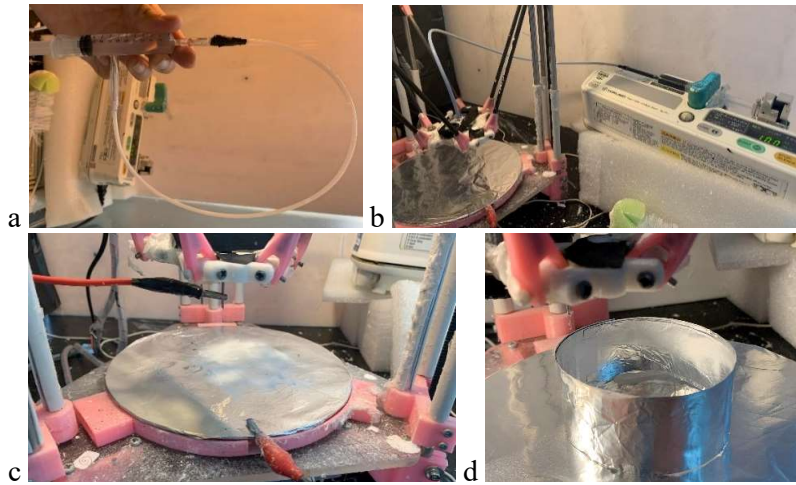
Fig. 1. (a) The syringe was connected to the long tube and the needle; (b) the syringe pump was pumped the solution out; (c) the wire was connected to the needle tip and the collector and to DC power supply; (d) the cylindrical aluminum foil guiding electrode.

2.3. Characterizations

In this research, there are 4 characterizations which are tensile test, morphology test by Scanning Electron Microscope, XRD and FTIR. For the tensile test of the 3D electrospun scaffolds were measured by a universal testing machine as in Fig. 2. (a).

The first step of the tensile test is cutting all samples into 15x45.00mm of rectangular shape. After that, measured the thickness of each sample and put the magnitude in Merlin program including the velocity of load cells (the 5kN of load cell was used in this research). The load cell velocity was set at 5.00mm/min for all samples. Finally, the sample was fixed in the two grip units of the testing machine and run the machine.

For the morphological test, the 3D electrospun scaffolds were studied their morphology, such as the scaffold diameter and the surface of the scaffold with a scanning electron microscope (SEM) and also the 5 unknown carbon-based nanomaterials were studied the morphology by the scanning electron microscope (SEM) as well.

For the other characterizations that are XRD was used to characterize the crystallinity of the scaffold results by putting the result or graphene in the glass holder and put it in the x-ray diffractometer as in Fig. 2. (b) by set the 2θ at 5-90° and 2.4 of the resolutions. FTIR was used to check the chemical interactions between the polymer, solvent and additive, if there were major differences cases by the insertion of the additive, also to characterize the chemical changes in the solutions by putting the result or graphene in the FTIR spectroscope as in Fig. 2. (c) and the FTIR characterizations were recorded in the wavenumber range of 4000-400 cm^{-1} in transmittance mode at room conditions.
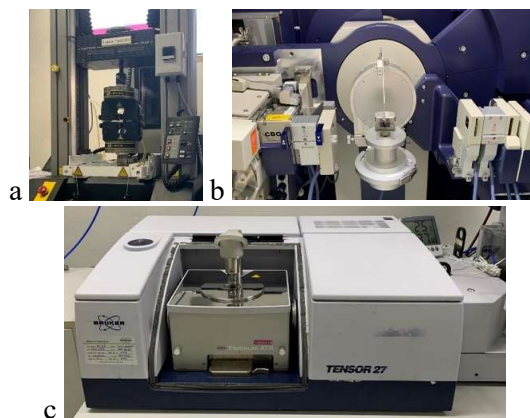
Fig. 2. (a) Instron tensile tester model 5565; (b) Rigaku smarLab x-ray diffractometer (XRD); (c) Bruker Tensor27 FTIR spectroscope.

3. Results

3.1. Solution preparations results

From the 2.1. solutions preparation methods above, the results are shown in the table 1 below, the solutions that can be fabricated are the sample PC1 and PL6-PL12 due to the polymer solubility. Some solvents could not dissolve all the polymer and if the polymers were not completely dissolved, it will be caused the needle clogging in the fiber fabrications in the next step.

Table 1. Solutions preparation results. *G0-G5 are the unknown carbon- based nanomaterials.

Sample	Polymer	Solvent							Addi-tives	Ratio	Concen-tration ($\%w/v$)	Stirring temp. (°C)	Solu-bility
		aceton	acetonitril	chloroform	dmf	dmso	ethanol	methanol					
PC1	PCL	✓	-	-	-	✓	-	-	G0*	5:1	7.5	55	✓
PV1	PVA	✓	-	-	-	-	-	-	-	1	7.5	55	X
PV2	PVA	-	✓	-	-	-	-	-	-	1	7.5	55	X
PV3	PVA	-	-	-	-	-	✓	-	-	1	7.5	55	X
PV4	PVA	-	-	-	-	-	-	✓	-	1	7.5	55	X
PL1	PLA (powder)	✓	-	-	-	-	-	-	-	1	7.5	55	X
PL2	PLA (powder)	✓	-	-	✓	-	-	-	-	1:1	7.5	55	X
PL3	PLA (powder)	-	-	✓	-	-	-	-	-	1	7.5	55	X
PL4	PLA (powder)	✓	-	-	✓	-	-	-	-	1:1	7.5	120	X
PL5	PLA (pill)	✓	-	-	-	-	-	-	-	1	7.5	55	X
PL6	PLA (pill)	-	-	✓	-	-	-	-	-	1	7.5	55	✓
PL7	PLA (pill)	✓	-	-	-	-	-	-	G1*	1	7.3	85	✓
PL8	PLA (pill)	✓	-	-	-	-	-	-	G2*	1	7.3	85	✓
PL9	PLA (pill)	✓	-	-	-	-	-	-	G3*	1	7.3	85	✓
PL10	PLA (pill)	✓	-	-	-	-	-	-	G4*	1	7.3	85	✓
PL11	PLA (pill)	✓	-	-	-	-	-	-	G5*	1	7.3	85	✓
PL12	PLA (pill)	✓	-	-	-	-	-	-	-	1	7.5	85	✓

3.2. Fiber fabrication (3D electrospinning technique) results

After the solutions were well prepared and were fabricated by using the 3D electrospinning technique into the 5cm of height with 5cm diameter of cylindrical shape. These are the results.

Sample PL11; that used PLA as the polymer and dissolved by acetone as solvent with 7.3%wt of concentration. This sample was fabricated by 3D electrospinning technique with 18G (0.838mm of diameter) of needle at 2.5cm distance between needle tip to collector. The result as in Fig. 3. (a) shows that it was not performed in the 3D cylindrical shape and had marks caused by sparks. This sample were fabricated with 9mL/h of flow rate and 10kV of applied voltage.

In Fig. 3. (b) is the sample PL11 result; that used PLA as the polymer and dissolved by acetone as solvent with 7.3%W/V of concentration. It was fabricated with the 18G needle size and the 2.5cm distance from the collector to needle tip. It was begun to fabricate at 3.44kV of applied voltage with 5mL/h of flow rate. But the applied voltage that gave the best result was 7kV and the flow rate was 5mL/h. The result of these parameters is better than Fig. 3. (a) result. And the result had a ring shape scaffold, but not quite in the 3D cylindrical shape.

In Fig. 3. (c) is the sample PL9 result that used PLA as the polymer and dissolved by acetone as solvent with 7.3%W/V of concentration. It was fabricated with the 18G needle size and the 5cm distance from the collector to needle tip. It was begun to fabricate at 10kV of applied voltage with 5mL/h of flow rate. But the flow rate that gave the best result was 10mL/h and the applied voltage was 12kV. The result of these parameters started to fabricate in the 3D cylindrical shape but the height of the scaffold was not stable as in Fig. 3. (c). And the fabrication had an untidy direction when increased the flow rate and applied voltage.

In Fig. 3. (d) is the sample PL9 result that used PLA as the polymer and dissolved by acetone as solvent with 7.3%W/V of concentration. It was fabricated with the 20G needle size and the 5cm distance from the collector to needle tip. It was not performed like the 3D cylindrical shape when using 0-20kV of applied voltage and 5mL/h of flow rate. The result had untidy direction of the fabrication. And in Fig. 3. (e) is the result of the same sample with the same needle size and distance between needle tip to collector, but the flow rate was 10mL/h. It started to fabricate in the 3D shape with 9kV of the applied voltage and gave the best result at 13kV of the applied voltage with 10mL/h of flow rate. It had an untidy direction fabrication when the applied voltage was higher than 14kV.

In Fig. 3. (f) is the sample PL9 result that used PLA as the polymer and dissolved by acetone as solvent with 7.3%W/V of concentration. It was fabricated with the 20G needle size and the 5cm distance from the collector to needle tip. And this result had the 6cm of diameter guiding electrode to accelerate the 3d build-up process. The flow rate that used was 10mL/h and it started to fabricate at 5kV of applied voltage. When the applied voltage was slowly increased to 8kV, it started to fabricate into the outer side of the guiding electrode and when the flow rate was increased to 13kV, it started to spark in the outer side of the guiding electrode.

In Fig. 3. (g) is the sample PL9 result that used PLA as the polymer and dissolved by acetone as solvent with 7.3%W/V of concentration. It was fabricated with the 20G needle size and the 5cm distance from the collector to needle tip. And this result had the 7cm of diameter guiding electrode to accelerate the 3d build-up process. The flow rate that used was 10mL/h and it started to fabricate at 5kV of applied voltage. When the applied voltage was slowly increased to 7- 8kV, it started to fabricate into the edge of the guiding electrode and when the flow rate was increased to 10kV, it started to fabricate in the outer side of the guiding electrode. And when the flow rate was increased to 12-13kV, the scaffold was fabricated the connected scaffold into the inner side of the guiding electrode as in Fig. 3. (g).

In Fig. 3. (h) is the sample PL12 that used PLA as the polymer and dissolved by acetone as solvent with 7.5%W/V of concentration without the carbon-based nanomaterials added. It was fabricated with the 18G needle size and the 5cm distance from the collector to needle tip. With 5mL/h of flow rate, it started to fabricate into the 3D shape at 7kV of applied voltage and when the applied voltage was increased into 8-13kV, the scaffolds had a better fabrication with the fine fibrous scaffolds. With 10mL/h of flow rate, it started to fabricate into the 3D cylindrical shape as in Fig. 3. (h) at 8kV of applied voltage. The scaffold had fine appearance with higher height compared to all results and when the applied voltage increased

into 9-10kV, the result had started to fabricate into the bigger cylinder shape with untidy direction. And when the applied voltage was increased more than 11kV, the fabrication had more untidy direction.
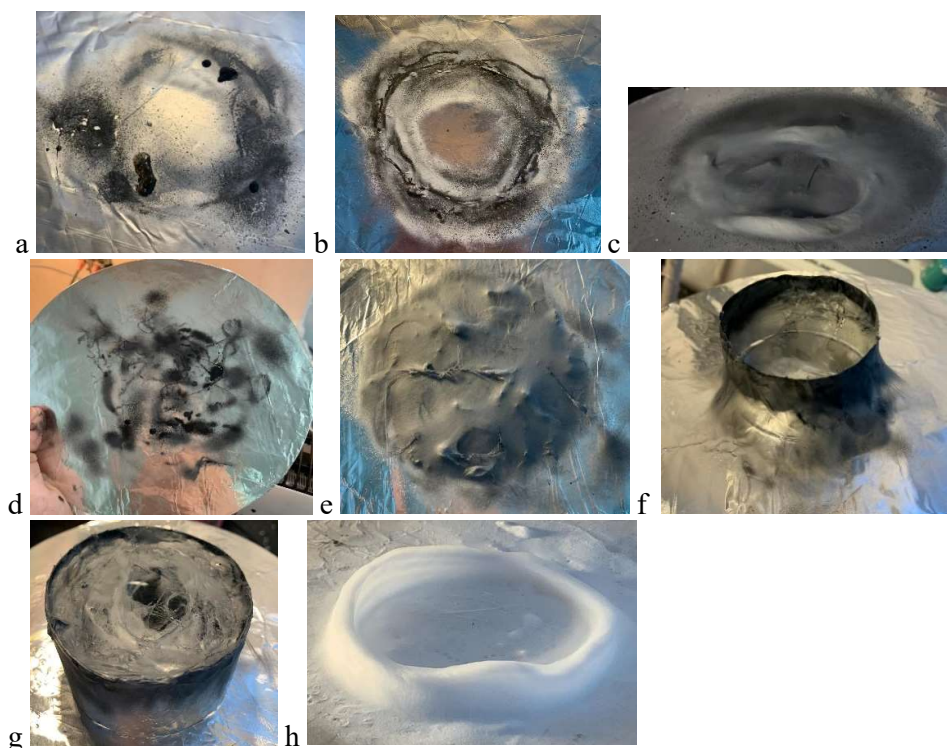

Fig. 3. (a) Sample PL11 result with 10mL/h flow rate; (b) Sample PL11result with 5mL/h flow rate; (c) Sample PL9 result with 18G needle and 10mL/h flow rate; (d) Sample PL9 result with 20G needle and 5mL/h flow rate; (e) Sample PL9 result with 20G needle and 10mL/h flow rate; (f) Sample PL9 result with 6cm diameter guiding electrode; (g) Sample PL9 result with 7cm diameter guiding electrode; (h) Sample PL12 result without carbon-based nanomaterial added.

3.3. Characterizations results

3.3.1. Tensile test

The tensile results of the scaffolds were analyzed by the stress-strain chart. The scaffold that has the carbon-based nanomaterials added compare to the scaffold that has no carbon-based nanomaterial added as in Fig. 4. The carbon-based nanomaterial added scaffold with the average thickness of 1.731mm has 60.8197N of the maximum load, 2.34 MPa of the maximum stress and 56.88% of the elongation at break. And the scaffold that has no carbon-based nanomaterial added chart with the average thickness of 2.494mm has 59.9455N of the maximum load, 1.60MPa of the maximum stress and 57.16% of the elongation at break.
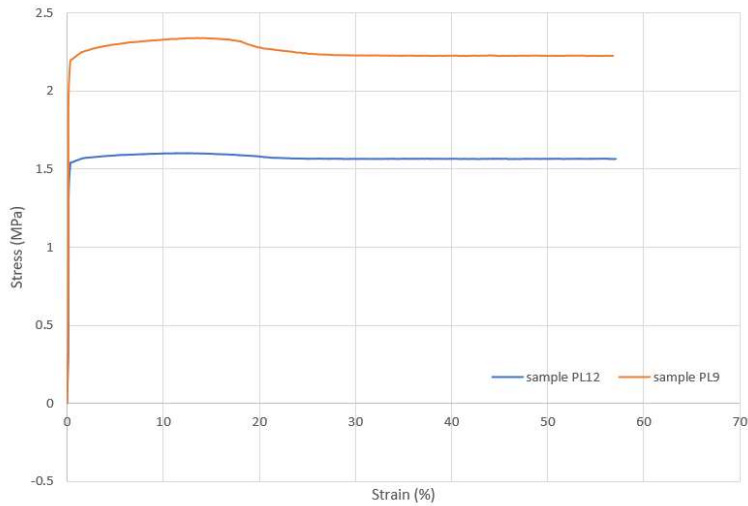
Fig. 4. The stress-strain chart between the carbon-based nanomaterials added scaffold (sample PL9) and the scaffold that has no carbon-based nanomaterial added (sample PL12).

3.3.2. Morphology tests

The morphology of the 3D electrospinning scaffold from sample PL12 without the carbon-based nanomaterials added was analyzed by the scanning electron microscope (SEM) as in Fig.5 with 500times of magnification. It shows a very fine fiber overlapping together with the different sizes of fibers around 1-5µm of width.
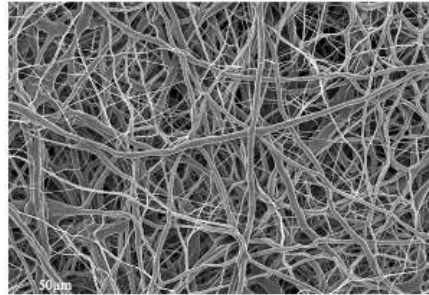
Fig. 5. The SEM image of 3D electrospinning scaffold

The morphology of 5 unknown carbon-based nanomaterials were analyzed by SEM as in Fig. 6. with 10K times magnification. The G1, G2 and G3 carbon-based nanomaterials in Fig. 6. (a), (b), (c) respectively have a tiny flat flake structures which have around 1µm of length and the carbon-based nanomaterials agglomerate together. For the G4 and G5 carbon-based nanomaterials in Fig. 6. (d), (e) respectively have the structures like the plane sheets overlapping together and the length size of the G4 and G5 carbon-based nanomaterials are over than 2µm.
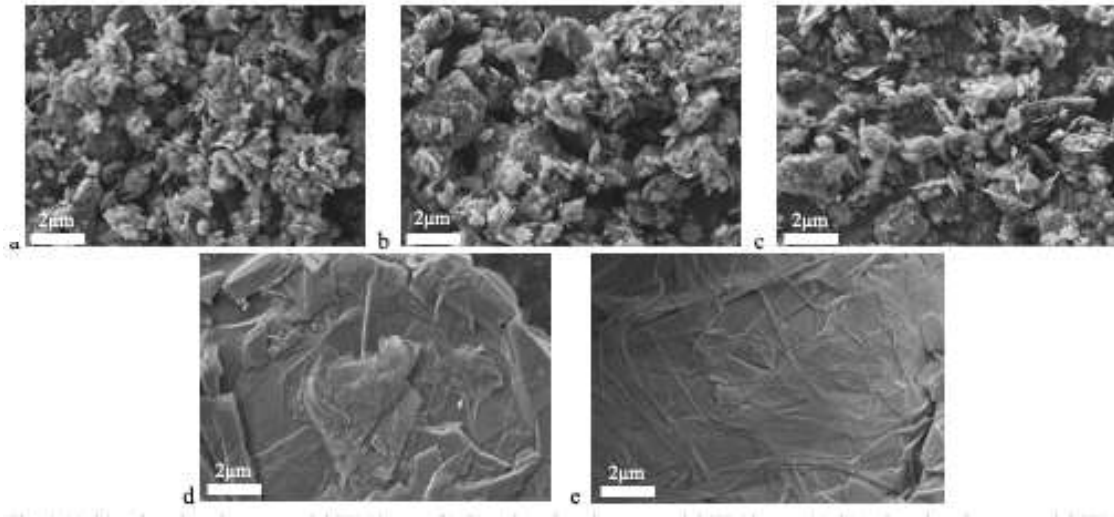
Fig. 6. (a) G1 carbon- based nanomaterial SEM image; (b) G2 carbon- based nanomaterial SEM image; (c) G3 carbon- based nanomaterial SEM image; (d) G4 carbon- based nanomaterial SEM image; (e) G5 carbon- based nanomaterial SEM image

3.3.3. XRD results

In this study, the XRD pattern of 5 unknown carbon-based nanomaterials have been analyzed the crystallinity. In Fig. 7. Shows a shape and tight peak (2θ= 26.45° and 54.48°) which corresponds to the diffraction line C (002) with the intercellular spacing in the crystal (d) respectively is 3.36 and 3.72Å. The XRD patterns show the data of the typical crystal structure of graphite.[20] And there are no significant differences between the 5 carbon-based nanomaterials additives. And the weak peaks in Graphene01-Graphene03, which are indicative of low crystallinity level. Compared to the strong peaks in Graphene04 and Graphene05 that indicated the high crystallinity level [19].
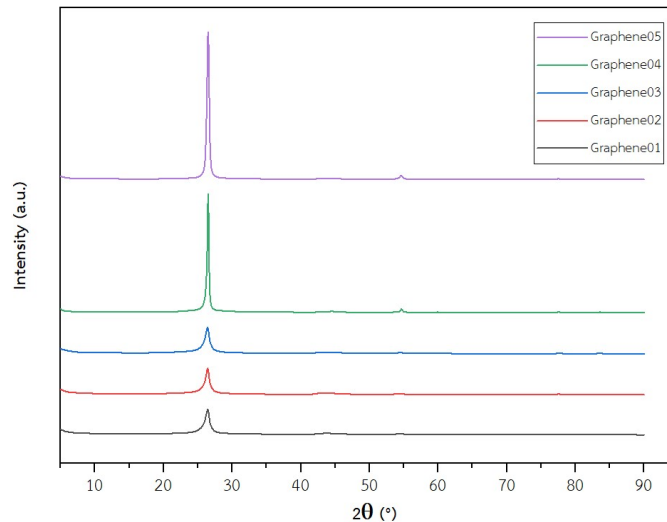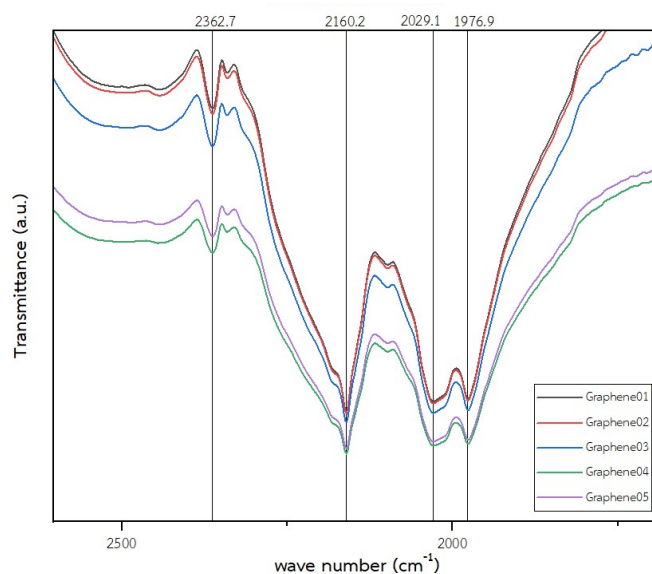

Fig. 7. XRD patterns of carbon-based nanomaterials G1-G5 that added in the solutions. (In Fig. Graphene01 refers to G1 carbon-based nanomaterial, to Graphene05 refers to G5 carbon-based nanomaterial)

3.3.4. FTIR results

In this study, the FTIR spectra of 5 unknown carbon-based nanomaterials have been analyzed the chemical interactions in the additives. In Fig. 8., all spectra show the few absorption signals due to the difference in the state of charges between carbon atoms [21]. This weak difference leads to a very small induced electric dipole, which provides a very clean spectrum [22]. It is clear that when an oxidation treatment is used on the material, bands corresponding to oxygenated functional groups appear, which indicates that the oxidative process was successful [21]. And there are no significant differences between the 5 carbon-based nanomaterials additives.

Fig. 8. FTIR spectra of carbon-based nanomaterials G1-G5 that added in the solutions. (In Fig. Graphene01 refers to G1 carbon-based nanomaterial, to Graphene05 refers to G5 carbon-based nanomaterial)

4. Discussion and conclusion

In this study, the solution preparation results in table 1. The solutions that can be used to fabricate are sample PC1 that is 7.5%W/V of PCL dissolved with acetone and 0.35%W/V of G0 additive, PL6 that is 7.5%W/V of PLA (pill appearance) dissolved with chloroform without the additive, PL7-PL11 that is 7.3%W/V of PLA (pill appearance) dissolved with acetone and 0.35%W/V of G1-G5 additive(respectively). And sample PL12 that is 7.5%W/V of PLA (pill appearance) dissolved with acetone without the additive. And all the solutions that have PVA and PLA (powder appearance) as polymer cannot be fabricated because the polymer did not be dissolved in the solutions.

The 3D electrospinning scaffolds results as in Fig. 3. (a) and (b) that are the sample PL11 results with has G5 carbon-based nanomaterial added. The shapes of the results are flat and not in 3D cylindrical shape because the distance between the collector is close by the needle tip and from the G1-G5 carbon-based nanomaterials morphology results in Fig. 6., G5 carbon-based nanomaterials in Fig. 6. (e) has a big flat sheet flakes that were overlapping together. And in Fig. 3. (d), (e) that are the results from PL9 that has G3 carbon-based nanomaterials added. Fig. 3. (e) that used 10mL/h flow rate has a better scaffold fabrication than Fig. 3. (d) that used 5mL/h flow rate with the same 5cm of the distance between collector to the needle tip and the same 20G of needle size. And in Fig. 3. (c) that used 18G needle size has a better scaffold fabrication than Fig. 3. (e) that used 20G needle size with the same 5cm of the distance between

collector to the needle tip and 10mL/h of flow rate. In Fig. 3. (f) and (g), that used the same PL9 sample, the same 20G needle size and the same 5cm of the distance between collector to the needle tip. But had a different diameter of the guiding electrode that were 6cm and 7cm. The results of these samples are not in the 3D cylindrical shape because the size of the guiding electrode is not suitable for the experiment that has 5cm of diameter cylindrical movement from the 3D electrospinning machine. From the results, the guiding electrode should have the size of diameter larger than 7cm. And in Fig. 3. (h), that used the PL9 sample without the additives, 18G needle size and the 5cm of the distance between collector to the needle tip with 10mL/h flow rate and 8kV of applied voltage. The scaffold could form into the fine appearance scaffolds with the higher height compared to all the results. Because the carbon-based nanomaterials additives weighted so that the results that has the carbon-based nanomaterial added could not fabricated into the higher height of the 3D cylindrical shape compared to the result that did not have the additives in Fig. 3. (h).

The tensile characterization results, the carbon-based nanomaterial additives scaffold has the higher maximum stress and higher maximum load, even the scaffolds thickness is thinner than the no-additives scaffold. The results could conclude that the carbon-based nanomaterial additives scaffold has the higher tensile strength than the scaffold that did not have the carbon-based nanomaterial additives. The morphology results from SEM, shows that the scaffolds that were fabricated from the 3D electrospinning technique had a very fine fiber overlapping together with the different sizes of fibers as in Fig. 5. And the morphology results of the G1-G5 carbon-based nanomaterial with the same magnification in Fig. 6., G1, G2 and G3 carbon-based nanomaterial has smaller size of materials than G4 and G5.

The XRD patterns shows that the 5 unknown G1-G5 carbon-based nanomaterial are graphite due to the 2θ peaks that are 26.45° and 54.48° of the results in Fig. 7. that show the data of the typical crystal structure of graphite. And the weaker peaks in Graphene01-Graphene03, which indicated to the low crystallinity level. Compared to the stronger peaks in Graphene04 and Graphene05 that indicated the high crystallinity level.

The all FTIR spectra of 5 unknown carbon-based nanomaterials shows the few absorption signals due to the difference in the state of charges between carbon atoms. This weak difference leads to a very small induced electric dipole, which provides a very clean spectrum. And all the spectra indicated that 5 unknown carbon-based nanomaterial are graphite.

Acknowledgements

References

[1] Monte, A. I. L., Licciardi, M., Bellavia, M., Damiano, G., Palumbo, V. D., Palumbo, F. S., … Giammona, G. Biocompatibility and biodegradability of electrospun PHEA- PLA scaffolds: our preliminary experience in a murine animal model. *Digest Journal of Nanomaterials and Biostructures*; 2012, 7, 2, 841-851. Retrieved from https://www.res earchgate.net/publication/244280669.
[2] Saniei, H., and Mousavi, S. Surface modification of PLA 3D-printed implants by electrospinning with enhanced bioactivity and cell affinity. Polymer; 2020. doi: https:/doi.org/10.1016/j.polymer.2020.122467.

[3] Liu, Y., Wang, S., Lan, W., and Qin, W. Fabrication of polylactic acid/carbon nanotubes/chitosan composite fibers by electrospinning for strawberry preservation. *International Journal of Biological Macromolecules*; 2018, 121, 1329-1336. doi: 10.1016/j.ijbiomac.2018.09.042.

[4] Zhang, W., Huang, C., Kusmartseva, O., Thomas, N. L., and Mele, E. Electrospinning of polylactic acid fibres containing tea tree and manuka oil. *Reactive and Functional Polymers*; 2017. doi: 10.1016/j.reactfunctpolym. 2017.06.013.

[5] Saavedra, J. P., Ricaurte, L., Perez, N. C. P., and Carvajal, M. X. Q. Development and Development and characterization of Sechium edule starch and polyvinyl alcohol nanofibers obtained by electrospinning. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*; 2022, 649, Retrieved from https://doi.org/10.1016/j.colsurfa .202 2.129456.

[6] Ginestra, P., Riva, L., Fiorentino, A., Zappa, D., Comini, E., and Ceretti, E. Electrospinning of Poly(vinyl alcohol)-Graphene oxide aligned fibers. *Procedia CIRP*; 2019, 89, 110-115. doi: 10.1016/j.procir.2020.05.126.

[7] Bosworth, L. A., and Downes, S. Acetone, a Sustainable Solvent for Electrospinning Poly(e-Caprolactone) Fibres: Effect of Varying Parameters and Solution Concentrations on Fibre Diameter. *Journal of Polymers and the Environment*; 2012, 20, 879-886. doi: 10.1007/s10924-012-0436-3.

[8] Honiggman, P., Sharma, N., Okolo, B., Popp, U., Msallem, B., and Thieringer, F. M. Patient-Specific Surgical Implants Made of 3D Printed PEEK: Material, Technology, and Scope of Surgical Application. *BioMed Research International*; 31, Retrieved from https://doi.org/10.1155/2018/4520636.

[9] Jia, W. J., Gu, Y. C., Gou, M. L., Dai, M., Li, X. Y., Kan, B., … Qian, Z. Y. Preparation of Biodegradable Polycaprolactone/ Poly (ethylene glycol)/ Polycaprolactone (PCEC)Nanoparticles. *Drug Delivery*; 2008, 15, 409-416. doi: 10.1080/10717540802321727.

[10] Kim, K., and Lee, S. E. Combined toxicity of dimethyl sulfoxide (DMSO) and vanadium towards zebrafish embryos (Danio rerio): Unexpected synergistic effect by DMSO. *Chemosphere*; 2020, 270, Retrieved from https://doi.org/10.1016/j.chemosphe re.2020.129405.

[11] Ebrahimlo, A. R. M., and Soreh, P. G. Rapid and sustainable process with low toxicity for cyanation of silver nitrate by DC arc-discharge in presence of acetonitrile. *Arabian Journal of Chemistry*; 2019, retrieved from http://www.sciencedirect.com.

[12] Youn, H. S., Kim, S. J., Kim, G. H., and Um, B. H. Enhancing the characteristics of hyDrochar via hydrothermal carbonization of Korean native kenaf: The effect of ethanol solvent concentration as co-solvent and reaction temperature. *Fuel*; 2022, Retrieved from https://doi.org/10.1016/j.fuel.2022.125738.

[13] Elbakush, A. E., and Guven, D. Evaluation of ethanol tolerance I relation to intracellular storage compound of Saccharomyces cerevisiae using FTIR spectroscopy. *Process Biochemistry*; 2020, retrieved from https://doi.org/10.1016/j.procbio.2020.11.028.

[14] El-Zeftawy, A. M. Focus on the chemical value of methanol. J. King Saud Uni; 1993, 7, 209-256.

[15] Chang, H. Y., Shih, T. S., Guo, Y. L., Tsai, C. Y., and Hsu, P. C. Sperm function in workers exposed to N, N-dimethylformamide in the synthetic leather industry. *Fertility And Sterility*; 2003, 81, 6. doi: 10.1016/j.fertnstert .2003.10.033.

[16] Hurtt, M. E., Placke, M. E., Killinger, J. M., Singer, A. W., and Kennedy, Jr. G. L. 13- Week Inhalation Toxicity Study of Dimethylformamide (DMF) in Cynomolgus Monkeys. *Fundamental and applied toxicology*; 1991, 18, 596-601.

[17] Luttrell, W. E. Toxic tips: Chloroform. *Chemical Health & Safety*; 2005. doi: 10.1016/ j.chs.2005.03.002.

[18] Dehghani, M., Shahsavani, S., Mohammadpour, A., Jafarain, A., Arjmand, S., Rasekhi, M. A., … Conti, G. O. Determination of chloroform concentration and human exposure assessment in the swimming pool. *Environmental Research*; 2021, 203, Retrieved from https://doi.org/10.1016/j.envres.2021.111883.

[19] Vong, M., Sanchez, F. J. D., Keirouz, A., Nuansing, W., and Radacsi, N. Ultrafast fabrication of Nanofiber-based 3D Macrostructures by 3D electrospinning. *Materials & Design*; 2021, 208, Retrieved from https://doi.org/10.1016/j.matdes.2021.109916.

[20] Siburian, R., Sihotang, H., Lumban Raja, S., Supeno, M., and Simanjuntak, C. New Route to Synthesize of Graphene Nano Sheets. *Oriental Journal of Chemistry;* 2017. Doi: http://dx.doi.org/10.13005/ojc/340120.

[21] Sebastian, R., Julian, A. T., Johnanes, D. O., Diana, P. N. P., Mayra, E. V. Z., Jose, H. M. H., … Carlos, D. G. T. Antimicrobial Films Based on Nanocomposites of Chitosan/Poly (vinyl alcohol)/Graphene Oxide for Biomedical Applications. *Biomolecules;*2019, 9, 109. doi: 10.3390/biom9030109.

[22] Lopez mata, C., Aguilar, J., Peña, Y. Caracterización óptica y morfologica de materiales compuestos de P3OT y nanotubos de carbono funcionalizados. *Química Hoy;* 2012. 2, 18–23

DESIGN AND MODELING OF A HAMILTONIAN CONTROL LAW FOR A BIDIRECTIONAL CONVERTER IN DC DISTRIBUTION APPLICATIONS

Methawin Jantra[1*], Uthen Kamnarn[1], Anon Namin[1], Charnyut Karnjanapiboon[1], Suchart Janjornmanit[1], Samart Yachiangkam[1], Pakawadee Wutthiwai[1], Krit Ratchapum[1], Ekkachai Chaidee[1], Surasak Yousawat[1], Teeruch Janjongcam[1], Suparak Srita[1], Pratch Piyawongwi-sal[2], Jedsada Yodwong[3], Noureddine Takorabet[4] and Phatiphat Thounthong[5]

[1]*Department of Electrical Engineering, Rajamangala University of Technology Lanna, Thailand*

[2]*Department of Computer Engineering, Rajamangala University of Technology Lanna, Thailand*

[3]*mu Space and Advanced Technology Company Limited, Thailand*

[4]*Groupe de Recherche en Energie Electrique de Nancy, Université de Lorraine, Nancy, France*

[5]*Renewable Energy Research Centre (RERC), Department of Teacher Training in Electrical Engineering,*
Faculty of Technical Education, King Mongkut's University of Technology North Bangkok, Thailand

*Corresponding address (E-mail: Methawin@narit.or.th, +66-946017811)

Abstract

This article describes the design and modelling of a Hamiltonian control law for a bidirectional converter in dc distribution applications. In comparison to the typical PI compensation technique for the bidirectional converter, the suggested system offers the advantages of a Hamiltonian control law technique that regulates the voltage bus. A Hamiltonian control law does not require complex converter modelling. By simulating MATLAB Simulink, the proposed system confirms the notion. The proposed system consists of three converters, one bidirectional and two buck converters connected on the same distribution bus, all of which are designed for a 24-Volt, 50-Watt DC distribution bus. For the bidirectional converter, the Hamiltonian control law is used. The modelling and practical results show that the Hamiltonian control law technique used in conjunction with a bidirectional converter may significantly regulate the DC distribution bus.

Keywords: Hamiltonian Control Law; Bidirectional Converter; DC Distribution;

1.   Introduction

Due to some areas' inability to access the local power grid from the electricity supplier, there are many areas in many countries today that require the use of electricity. One way used to provide consumers with electricity in remote areas is the microgrid. Microgrids can be divided into two topologies: AC distribution systems and DC distribution systems. Compared to AC distribution systems, the DC distribution system performs better [1, 2]. Many converters are connected to the power distribution bus in a dc distribution system and work in parallel or serial connection together [3]. The power bus has 380 Vdc, 100 Vdc, 28 Vdc, 24 Vdc, and 12 Vdc of voltage potential. The user can receive electricity from the microgrid distribution system [4]. The microgrid can help the system's power flow, reducing the demand on the distribution network [5]. The microgrid is a popular choice for distributing power since it provides higher power quality while being environmentally beneficial. A microgrid system is capable of supplying electricity from various renewable energy sources. It can also draw power from a battery or other form of energy storage [6]. The microgrid system is powered by numerous distribution power sources at low voltage. Using the microgrid with or without a grid connection is possible. One type of power system that

utilizes numerous tiny power sources near to the user is the microgrid. Microgrid systems can eliminate independence, stability, high efficiency, and flexibility [8].

In a dc microgrid, various converter types are employed. The bidirectional converters are one of them. The development of reversible energy flow in both directions, from the energy source to load or load to an energy source, is of importance to bidirectional converters. In systems that distribute electric power for distribution buses and energy sources like EVs, PHEVs, uninterruptible power supplies (UPSs), renewable energy power sources, and smart-grid systems, bidirectional converters are used. Because two converter topologies can be used to perform the same functions, a bidirectional converters architecture can reduce the overall converter volume [15]. Numerous compensation plans exist [16]. Fig. 1. depicts the structure of the bidirectional converter, which can be used as a boost or buck converter.

The buck converter and boost converter's component connections are arranged closely in the schematic. The configuration of the power switch and power diode is different. Create a bidirectional converter by combining these two converters. Fig. 1 depicts the bidirectional converter's configuration design. A compensating mechanism is necessary for this bidirectional converter. There is a lot of study on compensation methods, including fuzzy logic, sliding mode, and basic PID compensation. This article focuses on the application of the Hamiltonian control law to a bidirectional converter in DC distribution systems operating in boost mode with a 24 Vdc power bus and 12 battery power storage sources on the secondary side to transfer 50 W of power. Fig. 1 depicts the bidirectional converter's schematic.
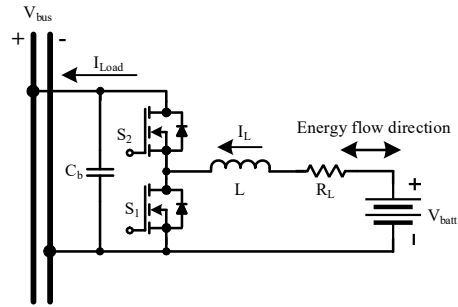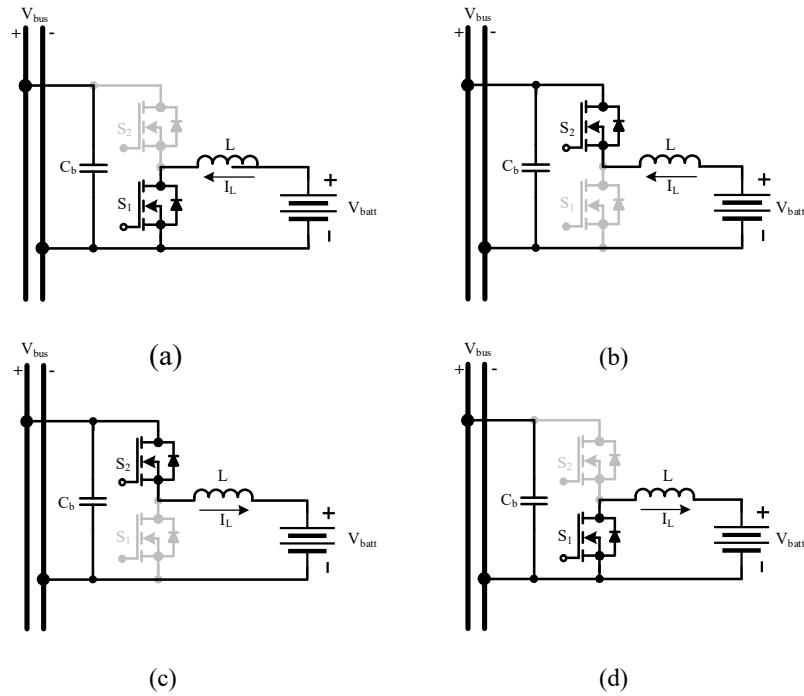

Fig. 1. The schematic of the bidirectional converter

Fig. 2. The operation mode of bidirectional converter: Boost operation mode and Buck operation mode.

The bidirectional converter's primary side is connected to Vbatt. Vbus is connected to the secondary side of the bidirectional converter. The filter capacitors C1 and C2 make up the bidirectional converter. L, s1, and s2 are the major power switches. There are two operating modes for bidirectional converters: boost operation and buck operation. In boost operation, the mode was initiated by switching on S1, moving the battery's energy to be stored in an inductor L, then switching S1 off and switching S2. An inductor L transfers its stored energy to the 24 Vdc bus, and the direction of the inductor current also switches from the right to the left. The total of the inductor voltage and the battery voltage determines the converter's output voltage. Fig. 2(a) and 2(b) depict the operation in detail, respectively.

When switch S2 is turned on, the buck operation begins. The surplus energy from the 24 Vdc bus flows and is stored in an inductor L, with an inductor voltage equal to the difference between the 24 Vdc bus and the battery voltage. The direction of the inductor current is shifting from the left to the right. Next, flip switch S2 to the off position and switch S1 to the on position. The battery receives the energy that is stored in an inductor L. The inductor current is also going from the left to the right in terms of direction. The voltage of the inductor is the same as the voltage of the battery. Figs. 2(c) and 2(d), respectively, depict the operation in detail.

2.   System Configuration Designed

Tables 1 and 2 display the design parameters for the bidirectional converter. The parameters in Table 2 are configured for the design boost operation mode, whereas Table 1 is configured for the design buck operation mode for the bidirectional converter.

Table 1. The parameter applied for the design bidirectional converter in buck operation mode.

Parameter Name	Value	Unit
Bus Voltage (V_{in})	24	Volt
Battery Voltage (V_o)	12	Volt
Output Ripple Voltage (Δvo)	100	mV
Power output (P_o)	50	Watt
Switching Frequency (f_s)	20	kHz

In Table 1, Bus Voltage (Vin) is an input of the buck converter. The output of the buck converter is connected to the battery Voltage (Vo). Let's define the duty cycle for driving the power switch of this mode D_{Buck}. The duty cycle D_{Buck} is for the gate driving signal for power switch S2 from Fig. 2(c), which acts as the main switch of the buck converter.

$$V_o = D_{Buck}(V_{in})$$
(1)

$$D_{Buck} = \frac{V_o}{V_{in}} = 0.5$$
(2)

$$R_L = \frac{Vo^2}{Po} = 2.88\ \Omega$$
(3)

$$L_{min} = (\frac{1 - D_{Buck}}{2f_s})R_L = 36\ \mu H$$
(4)

$$IL_{max} = V_o(\frac{1}{R_L} + \frac{1 - D_{Buck}}{2L_{min}f_s}) = 8.33\ A$$
(5)

$$IL_{\min} = V_o(\frac{1}{R_L} - \frac{1-D_{Buck}}{2L_{\min}fs}) = 0 \text{ A}$$

(6)

$$C_o = \frac{(1-D_{Buck})}{8L_{\min}(\frac{\Delta V_o}{V_o})f_s^2} = 520.83 \ \mu\text{F}$$

(7)

Where:

D_{Buck} is the Duty cycle of Buck converter

R_L is the resistive load of Buck converter

$L_{\min}$ is the minimum inductance that allow converter operated in boundary conduction mode (BDM)

$IL_{\max}$ is the maximum inductor current of the Buck converter

$IL_{\min}$ is the minimum inductor current of the Buck converter

C_o is the output capacitor of the Buck converter

Table 2. The parameter applied for design bidirectional converter in boost operation mode.

Parameter Name	Value	Unit
Battery Voltage (V_{in})	12	Volt
Bus Voltage (V_o)	24	Volt
Output ripple voltage (Δvo)	100	mV
Power output (P_o)	50	Watt
Switching Frequency (f_s)	20	kHz

Table 2 presents the input of the boost converter connected to the battery (Vin). The boost converter's output is connected to Bus Voltage (Vo). Let's define the duty cycle for driving the power switch of this mode as D_{Boost}. The duty cycle D_{Boost} is for the gate driving signal for power switch S1 from Fig. 4(c), which act as a main switch of the boost converter.

$$V_o = (V_{in})\left(\frac{1}{1-D_{Boost}}\right)$$

(8)

$$D_{Boost} = 1 - (V_{in})\left(\frac{1}{V_o}\right) = 0.5$$

(9)

$$R_L = \frac{Vo^2}{Po} = 11.52 \ \Omega$$

(10)

$$L_{\min} = \frac{D_{Boost}(1-D_{Boost})^2 R_L}{2f_s} = 36 \ \mu\text{H}$$

(11)

$$IL_{\max} = \frac{V_{in}}{(1-D_{Boost})^2 R_L} + \frac{1}{2}\frac{V_{in}D_{Boost}}{L_{\min}f_s} = 8.333A$$

(12)

$$IL_{\min} = \frac{V_{in}}{(1-D_{Boost})^2 R_L} - \frac{1}{2}\frac{V_{in}D_{Boost}}{L_{\min}f_s} = 0A$$

(13)

$$C_o = \frac{D_{Boost}}{R_L \frac{\Delta V_o}{V_o} f_s} = 520.83 \ \mu\text{F}$$

(14)

Where:

D_{Boost} is the Duty cycle of Boost converter

R_L is the resistive load of Boost converter

L_{min} is the minimum inductance that allow converter operated in boundary conduction mode (BDM)

IL_{max} is the maximum inductor current of the Boost converter

IL_{min} is the minimum inductor current of the Boost converter

C_o is the output capacitor of the Boost converter

The result of computed parameter from table 1 and table 2 shows that the inductance value and the output capacitance value are also the same. The inductance and capacitance of the buck converter and boost converters can be specific from the calculated parameters.

3.   Hamiltonian control law

The Hamiltonian control law of bidirectional converter operated in Boost mode used only differential equation of the converter.

$$L\frac{di_L}{dt} = v_{batt} - r_L \cdot i_L - v_{bus} + d \cdot v_{bus} \tag{1}$$

$$C_1\frac{dv_{bus}}{dt} = (i_L - d \cdot i_L) - \frac{v_{bus}}{r_P} - \frac{v_{bus}}{R_L} \tag{2}$$

Hamiltonian Standard Form

$$\dot{X} = \left[J_{anti-symmetric} - R_{diagonal} \right] \cdot \frac{\partial H(x)}{\partial x} + g(x) \cdot u_{input} + \xi_{Disturbance}$$

x: the vector state, Hamiltonian variables.
u, y: the port variables.
H(x): the Hamiltonian function (or energy function).
J(x): the skew-symmetric matrix (or anti-symmetric matrix), $J = -J^T$
     : the interconnection matrix.
R(x): the dissipation matrix (positive definite matrix), $R = R^T \geq 0$
g(x): the external port connection matrix.

Define Plant Boost Converter
Boost Converter Differential Equations:

$$L\frac{di_L}{dt} = v_{batt} - r_L \cdot i_L - v_{bus} + d \cdot v_{bus} \tag{3}$$

$$C_1\frac{dv_{bus}}{dt} = (i_L - d \cdot i_L) - \frac{v_{bus}}{r_P} - \frac{v_{bus}}{R_L} \tag{4}$$

$$\frac{d\lambda}{dt} = k_I(v_{batt} - v_{bus}) \tag{5}$$

Then,

$$\frac{di_L}{dt} = \frac{v_{batt}}{L} - \frac{r_L}{L}i_L - \frac{v_{bus}}{L} + \frac{d \cdot v_{bus}}{L} \tag{6}$$

$$\left.\begin{aligned}\frac{dv_{bus}}{dt} = \frac{(1-d)}{C_1}i_L - \frac{v_{bus}}{r_p C_1} - \frac{i_L}{C_1}\end{aligned}\right\} \tag{7}$$

$$\frac{di_L}{dt} = \frac{v_{FC}}{L} - \frac{r_L}{L}i_L - \frac{v_C}{L} + \frac{d \cdot v_C}{L} \tag{8}$$

$$\frac{dv_C}{dt} = \frac{(1-d)}{C_b}i_L - \frac{v_C}{r_p C_b} - \frac{i_{CH}}{C_b} \qquad \dot{X} = [J - R] \cdot \frac{\partial H(X)}{\partial x} + g(X) \cdot u + \xi \tag{9}$$

$$\frac{d\lambda}{dt} = k_I(v_{Cd} - v_C) \tag{10}$$

Step 1: define x, u

$$x = [x_1 \ x_2 \ x_3]^T = [i_L \ v_C \ \lambda_I]^T \tag{11}$$

$$u = d \tag{12}$$

$$\lambda_I = k_I \int (v_{Cd} - v_C) \tag{13}$$

Step 2: Re-Write equation

$$\frac{dx_1}{dt} = \frac{Vbatt}{L} - \frac{rL \cdot x1}{L} - \frac{x2}{L} + \frac{u \cdot x2}{L} \tag{14}$$

$$\frac{dx_2}{dt} = \frac{x1}{Cb} - \frac{u \cdot x1}{Cb} - \frac{iLoad}{Cb} \tag{15}$$

$$\frac{dx_3}{dt} = K_I(x2d - x2); \ KIx2d - KIx2 \tag{16}$$

Step 3: Define $g(x)$ & ξ $[\dot{x}_1 \ \dot{x}_2 \ \dot{x}_3] = [\quad]_{J-R}[\quad]_{\frac{\partial H(X)}{\partial x}} + [\quad]_{g(x)}u_1 + [\quad]_{\xi}$

Step 4: Define $H(x)$;

Tricky $H(x)$ = total stored energy in the system: L, C

$$H(x) = \frac{1}{2}\left(L \cdot i_L{}^2 + C_b \cdot v_C{}^2 + k_I{}^{-1} \cdot x_3^2\right) \tag{17}$$

$$H(x) = \frac{1}{2}\left(L \cdot x_1^2 + C_b \cdot x_2^2 + k_I{}^{-1} \cdot x_3^2\right) \tag{18}$$

Or, define diagonal matrix

$$Q = diag \quad \left. \begin{matrix} C_b & \frac{1}{k_I} \end{matrix} \right]^T \tag{19}$$

$$Q = \begin{bmatrix} L\ 0\ 0\ & 0\ 0\ C_b\ 0\ & 0\ \frac{1}{K_I} \end{bmatrix} \tag{20}$$

$$H(x) = \frac{1}{2}X^T Q X \tag{21}$$

$$H(x) = \frac{1}{2}[x_1\ x_2\ x_3]^T \begin{bmatrix} L\ 0\ & 0\ & 0\ 0\ C_b\ 0 & 0\ \frac{1}{k_I} \end{bmatrix}^{\curvearrowright Matrix} [x_1\ x_2\ x_3]^{\curvearrowright vector} \tag{22}$$

$$H(x) = \frac{1}{2}[x_1\ x_2\ x_3]^T \begin{bmatrix} Lx_1\ 0x_1\ 0x_1\ & 0x_2\ C_b x_2\ 0x_2\ & 0x_3\ 0x_3\ k_1{}^{-1}x_3 \end{bmatrix} \tag{23}$$

$$H(x) = \frac{1}{2}[x_1\ x_2\ x_3]^{T \curvearrowright vector} \begin{bmatrix} Lx_1\ C_b x_2\ k_I{}^{-1}x_3 \end{bmatrix}^{\curvearrowright vector} \tag{24}$$

$$H(x) = \frac{1}{2}\left(X_1 L X_1 + X_2 C_B X_2 + X_3 K_I^{-1} X_3\right)_{\underset{Energy}{}} \tag{25}$$

$$H(x) = \frac{L \cdot X_1^2}{2} + \frac{C_b X_2^2}{2} + \frac{X_3^2}{2k_I} \tag{26}$$

$$H(x) = \frac{1}{2}\left(x_1^2 + C_b \cdot x_2^2 + \frac{x_3^2}{k_I}\right) \tag{27}$$

$$\frac{\partial H(X)}{\partial x} = \left[\frac{d}{dx_1}\left(\frac{1}{2} \cdot L \cdot x_1^2 + \frac{1}{2} \cdot C_b \cdot x_2^2 + \frac{1}{2} \cdot \frac{x_3^2}{k_I}\right)\ \frac{d}{dx_2}\left(\frac{1}{2} \cdot L \cdot x_1^2 + \frac{1}{2} \cdot C_b \cdot x_2^2 + \frac{1}{2} \cdot \frac{x_3^2}{k_I}\right)\ \frac{d}{dx_3}\left(\frac{1}{2} \cdot L \cdot x_1^2 + \frac{1}{2} \cdot C_b \cdot x_2^2 + \frac{1}{2} \cdot \frac{x_3^2}{k_I}\right)\right] = \left[\frac{d}{dX_1}\left(\frac{2}{2} \cdot L \cdot X_1^{2-1} + 0 + 0\right)\ \frac{d}{dX_2}\left(0 + \frac{2}{2} \cdot C_b \cdot X_2^{2-1} + 0\right)\ \frac{d}{dX_3}\left(0 + 0 + \frac{2}{2} \cdot \frac{X_3^{2-1}}{k_I}\right)\right] \tag{28}$$

$$\frac{\partial H(X)}{\partial x} = \left[Lx_1\ C_b x_2\ \frac{x_3}{k_I}\right] \tag{29}$$

$$[\dot{x}_1\ \dot{x}_2\ \dot{x}_3] = [JR_{11}\ JR_{12}\ JR_{13}\ JR_{21}\ JR_{22}\ JR_{23}\ JR_{31}\ JR_{32}\ JR_{33}]_{\underset{J-R}{}} \underbrace{\left[Lx_1\ C_b x_2\ k_I{}^{-1} \cdot x_3\right]}_{\frac{\partial H(X)}{\partial x}} + \underbrace{\left[\frac{x_2}{L}\ -\frac{x_1}{C_b}\ 0\right]}_{g(x)} u_1 + \underbrace{\left[\frac{v_{FC}}{L}\ -\frac{i_{CH}}{C_b}\ k_I x_{2d}\right]}_{\xi} \tag{30}$$

Step 6: Define $J - R$

From (14), (15), (16)

$$\frac{dx_1}{dt} = \frac{v_{batt}}{L} - \frac{r_L}{L}x_1 - \frac{1}{L}x_2 + \frac{u}{L}x_2 \tag{31}$$

$$\frac{dx_2}{dt} = \frac{1}{C_b}x_1 - \frac{u}{C_b}x_1 - \frac{i_{LOAD}}{C_b} \tag{32}$$

$$\frac{dx_3}{dt} = k_I(x_{2d} - x_2) \quad ; k_I x_{2d} - k_I x_2 \tag{33}$$

$$[JR_{11}\, JR_{12}\, JR_{13}\, JR_{21}\, JR_{22}\, JR_{23}\, JR_{31}\, JR_{32}\, JR_{33}\,]_{\underbrace{}_{J-R}} \left[Lx_1\ C_b x_2\ k_I^{-1} \cdot x_3\right]_{\underbrace{}_{\frac{\partial H(x)}{\partial x}}} \tag{34}$$

$$JR_{11}LX_1 = -\frac{r_L}{L}x_1 \tag{35}$$

$$JR_{11} = -\frac{r_L}{L}x_1 \frac{1}{Lx_1} = -\frac{r_L}{L^2} \tag{36}$$

$$JR_{12}C_b x_2 = -\frac{1}{L}X_2 \tag{37}$$

$$JR_{12} = -\frac{1}{L}x_2 \frac{1}{c_b x_2} = -\frac{1}{LC_b} \tag{38}$$

$$JR_{21}Lx_1 = \frac{1}{C_B X_1} \tag{39}$$

$$JR_{21} = \frac{1}{C_b X_1}\frac{1}{LX_1} = \frac{1}{LC_b} \tag{40}$$

$$JR_{32}C_b X_2 = -k_I X_2 \tag{41}$$

$$JR_{32} = -k_I X_2 \frac{1}{C_b X_2} = -\frac{k_I}{C_b} \tag{42}$$

$$[\dot{x}_1\ \dot{x}_2\ \dot{x}_3\,] = \left[-\frac{r_L}{L^2} -\frac{1}{LC_b}\ 0\ \frac{1}{LC_b}\ 0\ 0\ 0\ -\frac{k_I}{C_b}\ 0\right]_{\underbrace{}_{J-R}} \left[Lx_1\ C_b x_2\ k_I^{-1} \cdot x_3\right]_{\underbrace{}_{\frac{\partial H(x)}{\partial x}}} + \left[\frac{x_2}{L} - \frac{x_1}{C_b}\ 0\right]_{\underbrace{}_{g(x)}} u_1 + \left[\frac{v_{batt}}{L} - \frac{i_{load}}{C_b}\ k_I x_{2d}\right]_{\underbrace{}_{\xi}} \tag{43}$$

Step 7: Solve $R, R = R^T \geq 0$

$$R = \left[\frac{r_L}{L^2}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right] \tag{44}$$

Step 8: From $(J - R)$, write $J = -J^T$

$$J = \left[\left[-\frac{r_L}{L^2} -\frac{1}{LC_b} \ 0 \ \frac{1}{LC_b} \ 0 \ 0 \ 0 -\frac{k_I}{C_b} \ 0 \right]\underset{[J-R]}{} + \left[\frac{r_L}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \right]\underset{[R]}{}\right] = \left[0 \ -\frac{1}{LC_b} \ 0 \ \frac{1}{LC_b} \ 0 \ 0 \ 0 \ -\frac{k_I}{C_b} \ 0 \right] \tag{45}$$
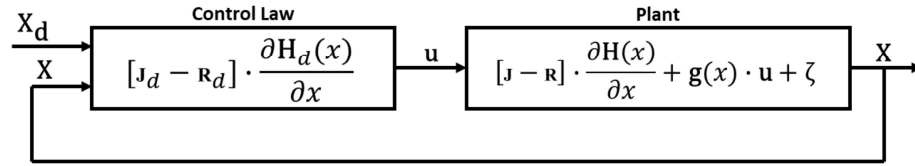
Step 9: Summarize Plant Model

$$\dot{X} = [J - \ R \ ] \cdot \frac{\partial H(x)}{\partial x} + g(x) \cdot u + \xi \overset{\frown}{} Plant \tag{46}$$

$$[\dot{x}_1 \ \dot{x}_2 \ \dot{x}_3] = \left[\left[0 \ -\frac{1}{LC_b} \ 0 \ \frac{1}{LC_b} \ 0 \ 0 \ 0 \ -\frac{k_I}{C_b} \ 0 \right] - \right.$$
$$\left.\left[\frac{r_L}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \right]\right]\underset{J-R}{} \left[Lx_1 \ C_b x_2 \ \frac{x_3}{k_I}\right]\underset{\frac{\partial H(x)}{\partial x}}{} + \left[\frac{x_2}{L} \ -\frac{x_1}{C_b} \ 0\right]\underset{g(x)}{} u_1 + \left[\frac{v_{batt}}{L} \ -\frac{i_{load}}{C_b} \ k_I x_{2d}\right]\underset{\xi}{} \tag{47}$$

$$[J_d - R_d] \cdot \frac{\partial H_d(x)}{\partial x} \overset{\frown}{} Control \ Law \tag{48}$$

$$[J_d - \ R_d] \cdot \frac{\partial H_d(x)}{\partial x} \overset{\frown}{} Control \ Law = [J - \ R \ ] \cdot \frac{\partial H(x)}{\partial x} + g(x) \cdot u + \xi \overset{\frown}{} Plant \tag{49}$$

Define Control law

Step 1: $R_d, R_d = R_d{}^T \geq 0 \ and \ k_\Gamma (\epsilon \ N \geq 0), r_L + k_\Gamma \geq 0$

$$R_{d=} \left[\frac{r_L}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \right]\overset{\frown}{R} + \left[\frac{k_r}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \right]\overset{\frown}{} =^{R_c} \left[\frac{r_L + k_r}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \right] \tag{50}$$

$$k_\Gamma \ is \ constant$$

$$R_{d=} \left[\frac{r_L + k_r}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \right] \tag{51}$$

Step 2: Set: $J_d, J_d = -J_d{}^T$

$$J_d = \left[0 \ -\frac{1}{LC_b} \ 0 \ \frac{1}{LC_b} \ 0 \ 0 \ 0 \ -\frac{k_I}{C_b} \ 0 \right]\overset{\frown}{J} + \left[0 \ -\frac{k_j}{LC_b} \ 0 \ \frac{k_j}{LC_b} \ 0 \ \frac{k_I}{C_b} \ 0 \ 0 \ 0 \right]\overset{\frown}{J_c} \tag{52}$$

$$J_d = \left[0 \ -\left(\frac{1+k_j}{LC_b}\right) \ 0 \ \left(\frac{1+k_j}{LC_b}\right) \ 0 \ \frac{k_I}{C_b} \ 0 \ -\frac{k_I}{C_b} \ 0 \right] \tag{53}$$

$$x_d = [x_{1d} \ x_{2d} \ x_{3d}]^T = [i_{Ld} \ v_{Cd} \ \lambda_I]^T \tag{54}$$

$$e_1 = x_{1d} - x_1, e_2 = x_{2d} - x_2, e_3 = x_{3d} - x_3 \tag{55}$$

$$e = [e_1 \ e_2 \ e_3]^T = [x_{1d} - x_1 \ x_{2d} - x_2 \ x_{3d} - x_3 \ ]^T \tag{56}$$

$$\lambda_I = k_I \int_0^\tau \quad (v_{Cd} - v_C) \tag{57}$$

$$with, x_{3d} = 0$$
$$e = [e_1 \ e_2 \ e_3]^T = [x_{1d} - x_1 \ x_{2d} - x_2 \ -x_3 \ ]^T \tag{58}$$

From (19)

$$H(x) = \frac{1}{2}x^T Q x \tag{60}$$

$$H_d(x) = \frac{1}{2}e^T Q e \tag{61}$$

$$H_d(x) = \frac{1}{2}L \cdot (x_{1d} - x_1)^2 + \frac{1}{2}C_b(x_{2d} - x_2)^2 + \frac{1}{2}\frac{x_3^2}{k_I} \tag{62}$$

$$\frac{\partial H_d(x)}{\partial x} = \left[\frac{d}{dx_1}(\frac{1}{2}L \cdot (x_{1d} - x_1)^2 + \frac{1}{2}C_b(x_{2d} - x_2)^2 + \frac{1}{2}\frac{x_3^2}{k_I}) \ \frac{d}{dx_2}(\frac{1}{2}L \cdot (x_{1d} - x_1)^2 + \frac{1}{2}C_b(x_{2d} - x_2)^2 + \frac{1}{2}\frac{x_3^2}{k_I}) \ \frac{d}{dx_3}(\frac{1}{2}L \cdot (x_{1d} - x_1)^2 + \frac{1}{2}C_b(x_{2d} - x_2)^2 + \frac{1}{2}\frac{x_3^2}{k_I})\right] = \left[L(x_1 - x_{1d}) \ C_b(x_2 - x_{2d}) \ \frac{x_3}{k_I}\right] \tag{63}$$

Step 4: Summarize Control Law

$$[J_d - R_d] \cdot \frac{\partial H_d(x)}{\partial x} \curvearrowright Control \ Law$$

$$\left[\left[0 \ -\left(\frac{1+k_j}{LC_b}\right) \ 0 \ \left(\frac{1+k_j}{LC_b}\right) \ 0 \ \frac{k_I}{C_b} \ 0 \ -\frac{k_I}{C_b} \ 0\right] - \left[\frac{r_L + k_r}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0\right]\right]_{\curvearrowleft Jd-Rd} \left[L(x_1 - x_{1d}) \ C_b(x_2 - x_{2d}) \ \frac{x_3}{k_I}\right]_{\curvearrowleft \frac{\partial H_d(x)}{\partial x}} =$$

$$[\dot{x}_1 \ \dot{x}_2] = \left[\left[0 \ -\frac{1}{LC_b} \ 0 \ \frac{1}{LC_b} \ 0 \ 0 \ 0 \ -\frac{k_I}{C_b} \ 0\right] - \left[\frac{r_L}{L^2} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0\right]\right]_{\curvearrowleft J-R} \left[Lx_1 \ C_bx_2 \ \frac{x_3}{k_I}\right]_{\curvearrowleft \frac{\partial H(x)}{\partial x}} + \left[\frac{x_2}{L} - \frac{x_1}{C_b} \ 0\right]_{\curvearrowleft g(x)}u_1 + \left[\frac{v_{batt}}{L} \ -\frac{i_{load}}{C_b} \ k_Ix_{2d}\right]_{\curvearrowleft \xi} \tag{64}$$

Step 4: Summarize Control Law

$$\left[-\left(\frac{r_L + k_\Gamma}{L^2}\right) -\left(\frac{1 + k_j}{LC_b}\right) 0 \left(\frac{1 + k_j}{LC_b}\right) 0 \frac{k_I}{C_b} 0 -\frac{k_I}{C_b} 0\right]_{Jd-Rd} \left[L(x_1 - x_{1d}) \; C_b(x_2\right.$$
$$\left.- x_{2d}) \frac{x_3}{k_I}\right]_{\frac{\partial H_d(x)}{\partial x} C1 \; C2} =$$
$$\left[-\frac{r_L}{L^2} -\frac{1}{LC_b} 0 \frac{1}{LC_b} 0 0 0 -\frac{k_I}{C_b} 0\right]_{J-R} \left[Lx_1 \; C_b x_2 \frac{x_3}{k_I}\right]_{\frac{\partial H_d(x)}{\partial x}} + \left[\frac{x_2}{L} -\frac{x_1}{C_b} 0\right]_{g(x)} u_1 + \left[\frac{v_{batt}}{L} -\right.$$
$$\left.\frac{i_{load}}{C_b} \; k_I x_{2d}\right]_{\xi \; P1 \; P2} \tag{65}$$

C1, C2 are Control law. P1, P2 are Plant information. Set the C1-P1= 0 then solve variable u. Set the C2-P2=0 then Solve k_J.

$$K = -\frac{i_{load}x_2 - v_{batt}x_1 + x_2 x_3 + x_1 x_{2d} - x_2 x_{1d} - K_r x_1^2 + K_r x_1 x_{1d} + r_L x_1 x_{1d}}{x_1 x_{2d} - x_2 x_{1d}} \tag{66}$$

$$u_1 = \frac{x_{2d} - v_{batt} - K_j x_2 + K_j x_{2d} - K_r x_1 + K_r x_{1d} + r_L x_{1d}}{x_2} \tag{67}$$

4. Simulation Results

Fig. 3 depicts the experimental apparatus for simulation using the suggested Hamiltonian control law method for a bidirectional converter in dc distribution applications. Three converters—one bidirectional converter controlled by Hamiltonians and two conventional PI-controlled buck converters—make up the simulation system. The 24 Vdc distribution bus is shared by all converters. On the 24 Vdc distribution bus buck converter #2 is where the 10 W electrical load is located. MATLAB Simulink is used to run the simulation. When employing a constant power demand with an open-loop bidirectional converter operating in boost mode, oscillation voltage output may occur as depicted in Fig. 4. The output voltage bus may oscillate due to the constant 286 W load. This impact can be reversed by the Hamiltonian control law.
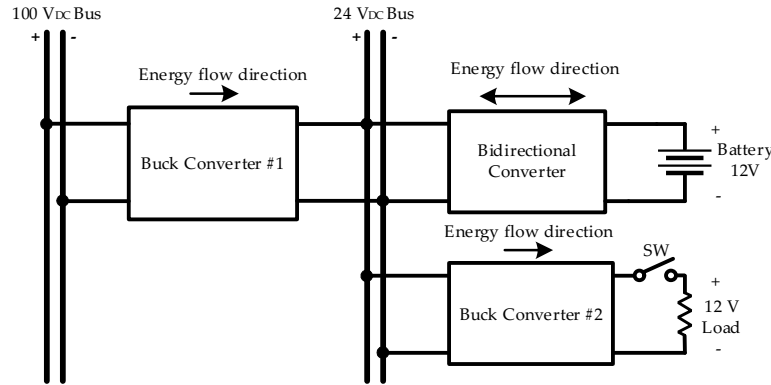
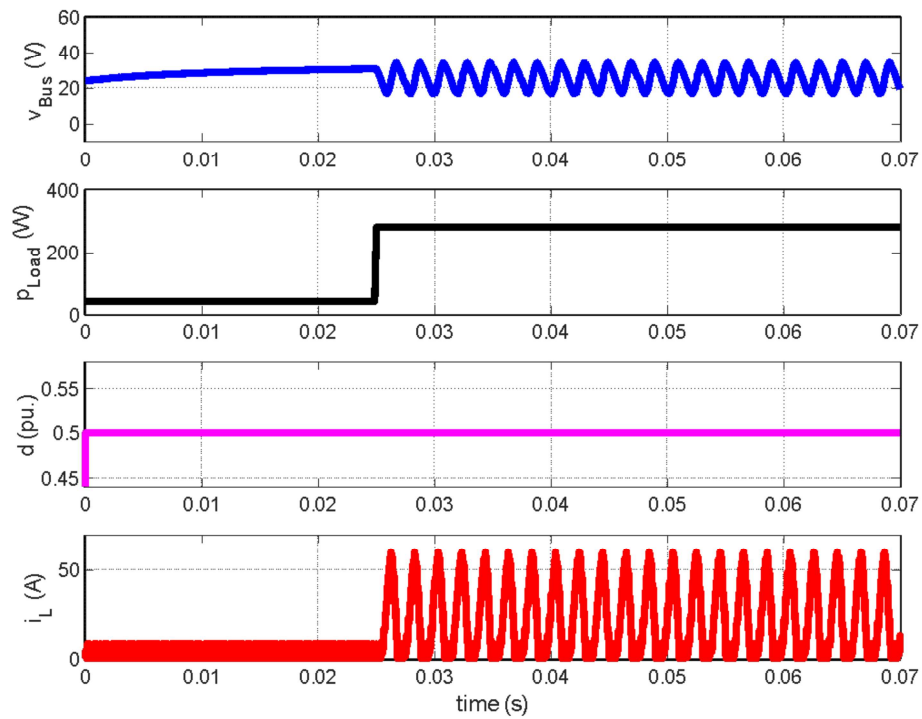**Fig. 3.** The simulation system for Hamiltonian controlled for bidirectional converter.

Fig. 4. The constant power load of 286 W may cause the output voltage bus to oscillate.
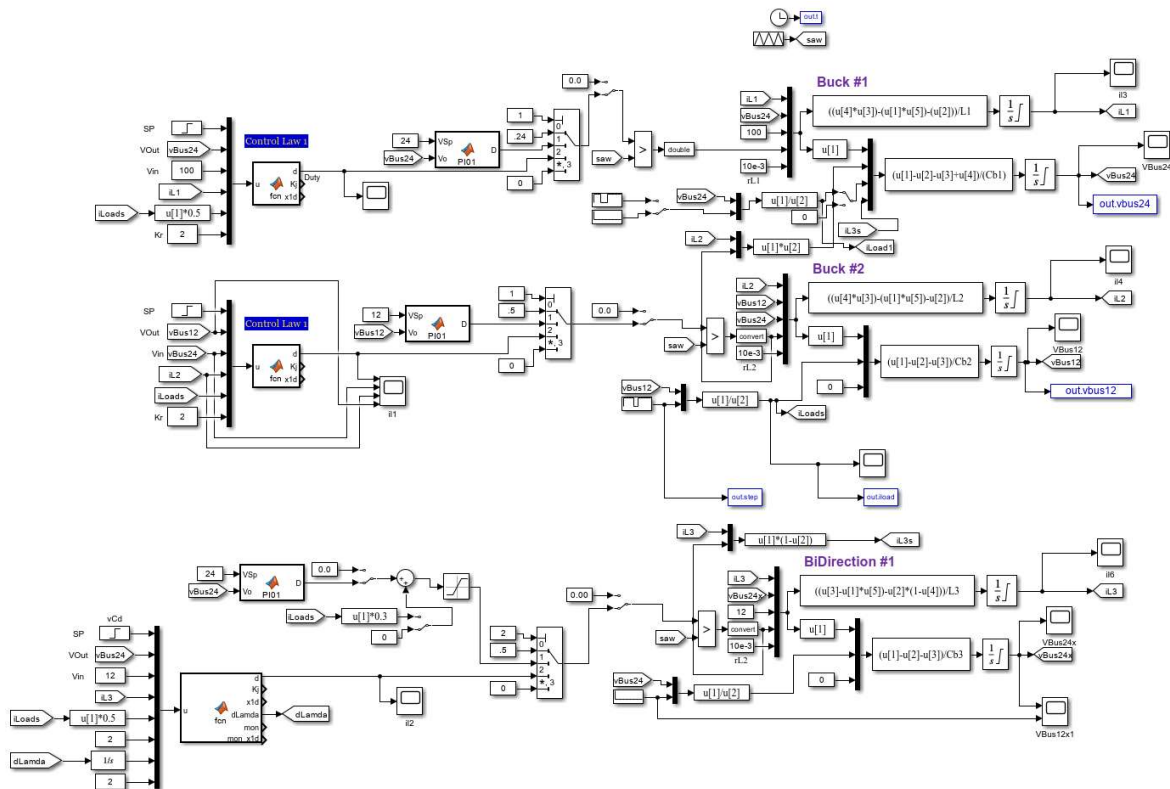

Fig. 5. The MATLAB simulink simulation diagram.

Fig. 5 displays the simulation diagram created by MATLAB Simulink. Fig. 6. displays the outcome of a simulation of a bidirectional converter in dc distribution applications utilizing Hamiltonian control rule.
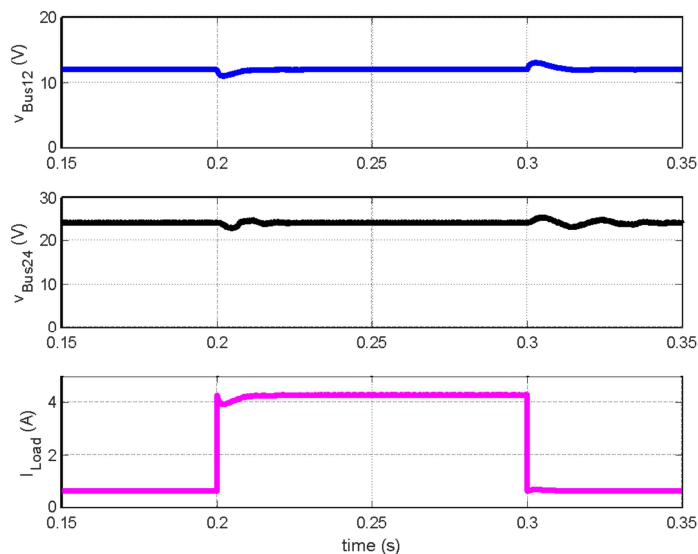

Fig. 6. The simulation result of a bidirectional converter using Hamiltonian control law.

Via Fig. 6. The responses of the 12 V voltage bus of the second buck converter, the 24 V DC distribution bus, and the load current when taking and removing loads are all displayed simultaneously. When a load is applied, the 12 V voltage bus briefly drops in voltage before rising to the voltage setting. After removing the load, the 12 V voltage bus briefly overshoots before falling back to the voltage preset. Similar to the 24 V voltage bus, when a load is applied, the voltage drops somewhat and oscillates before returning to the set voltage. After removing the load, the 24 V voltage bus overshoots briefly and oscillates before returning to the voltage setting. The load step instruction ranges from 1.5 to 50 watts.

5.   Conclusions and Discussion

In this work, a Hamiltonian control law for a bidirectional converter in dc distribution applications is designed and modelled. The proposed system benefits from a Hamiltonian control law method for controlling the voltage bus. The intricate converter modelling is not necessary for a Hamiltonian control law. When used with a steady power load, the Hamiltonian control law can completely remove voltage oscillation. The proposed Hamiltonian control law system's simulation in MATLAB Simulink supports the idea. Then, using a 12 V to 24 V bidirectional converter and the other two buck converters #1 and #2, operate the Hamiltonian control law. The system's ability to operate stably is confirmed by the voltage bus response. When compared to the step response provided in [17], [18], the suggested Hamiltonian control rule for a bidirectional converter has the same step response. The next study will compare the DC-distributed power supply's power balance adjustment technique [19]. The bidirectional converter prototype used in this work will be expanded for our previously proposed wireless EV charging systems [20–24].

References

[1] Xu, C.D.; Cheng, K.W.E. A survey of distributed power system— AC versus DC distributed power system, in Proceedings of the IEEE 4th *International Conference on Power Electronics Systems and Applications (PESA)*. **2011**, Hong Kong, pp. 1-12.

[2] Prabhala, V.A.K.; Baddipadiga, B.P.; Ferdowsi, M. DC distribution systems—An overview. in Proceedings of the IEEE International Conference on *Renewable Energy Research and Application (ICRERA)*, **2014**, Milwaukee, WI, USA.

[3] Wutthiwai, P.; Kamnam, U.; Yodwong, J.; Namin, A.; Thounthong, P.; Takorabet, N. Dynamic and Steady-State Behavior of Distributed Power Supply in DC Architecture with Minimized DC Bus Capacitor. in Proceedings of The 2022 *International Power Electronics Conference* (IPEC-Himeji 2022 -ECCE Asia), **2022**, JAPAN, pp. 1909-1916.

[4] Hartono, B. S.; Budiyanto, Y.; Setiabudy, R. Review of microgrid technology. Paper presented at the 2013 International Conference on *QiR*, **2013**.

[5] Anastasiadis, A. G.; Tsikalakis, A. G.; dan Nikos, D. H. Environmental benefits from dg operation when network losses are taken into account. Paper presented at the Proc. *DISTRES Conf*. **2009**.

[6] Nikos, H. Microgrids: an overview of ongoing research, development, and demonstration projects. *IEEE Power & Energy*. **2007**, 5(4).

[7] Xu, W.; Martel, S.; Mauch, K. An assessment of distributed generation islanding detection methods and issues for Canada. **2004**.

[8] Hatziargyriou, N.; Asano, H.; Iravani, R.; Marnay, C. Microgrids. *IEEE power and energy magazine*. **2007**, 5(4), 78-94.

[9] Onar, O. C.; Kobayashi, J.; Erb, D. C., Khaligh, A. A bidirectional high-power-quality grid interface with a novel bidirectional noninverted buck–boost converter for PHEVs. *IEEE transactions on vehicular technology*. **2012**, 61(5), 2018-2032.

[10] Xuewei, P.; Rathore, A. K.; Prasanna, U. R. Novel soft-switching snubberless naturally clamped current-fed full-bridge front-end-converter-based bidirectional inverter for renewables, microgrid, and UPS applications. *IEEE Transactions on Industry Applications*. **2014**, 50(6), 4132-4141.

[11] Chao, K. H.; Huang, C. H. Bidirectional DC–DC soft-switching converter for stand-alone photovoltaic power generation systems. *IET Power Electronics*. **2014**, 7(6), 1557-1565.

[12] Jin, K.; Yang, M.; Ruan, X.; Xu, M. Three-level bidirectional converter for fuel-cell/battery hybrid power system. *IEEE Transactions on Industrial Electronics*. **2009**, 57(6), 1976-1986.

[13] Mangu, B.; Akshatha, S.; Suryanarayana, D.; Fernandes, B. Grid-connected PV-wind-battery-based multi-input transformer-coupled bidirectional DC-DC converter for household applications. *IEEE journal of emerging and selected topics in power electronics*. **2016**, 4(3), 1086-1095.

[14] Zhang, Z.; Chau, K.-T. Pulse-width-modulation-based electromagnetic interference mitigation of bidirectional grid-connected converters for electric vehicles. *IEEE Transactions on Smart Grid*. **2016**, 8(6), 2803-2812.

[15] Tytelmaier, K.; Husev, O.; Veligorskyi, O.; Yershov, R. A review of non-isolated bidirectional dc-dc converters for energy storage systems. *II International Young Scientists Forum on Applied Physics and Engineering (YSF)*. **2016**.

[16] Gorji, S. A.; Sahebi, H. G.; Ektesabi, M.; Rad, A. B. Topologies and control schemes of bidirectional DC–DC power converters: An overview. *IEEE Access*. **2019**, 7, 117997-118019.

[17] Punna, S.; Banka, S.; Salkuti, S. R. Optimal Energy Management Scheme of Battery Supercapacitor-Based Bidirectional Converter for DC Microgrid Applications. Information. **2022**, 13(7), 350. https://www.mdpi.com/2078-2489/13/7/350

[18] Vijayan, M.; Udumula, R. R.; Mahto, T.; Lokeshgupta, B.; Goud, B. S.; Kalyan, C. N. S.; Balachandran, P. K. C, D.; Padmanaban, S.; Twala, B. Optimal PI-Controller-Based Hybrid Energy Storage System in DC Microgrid. Sustainability. **2022**, 14(22), 14666. https://www.mdpi.com/2071-1050/14/22/14666

[19] Kamnarn, U., Yodwong, J., Piyawongwisal, P., Wutthiwai, P., Namin, A., Thounthong, P., Takorabet, N. Design and simulation of DC distributed power supply with power balance control technique. International Journal of Power Electronics and Drive Systems, **2022**, 13(1), 460.

[20] Namin, A., Chaidee, E., Prachuabroek, T., et al., "Solar Tricycle with Lateral Misalignment Maximum Power Point Tracking Wireless Power Transfer", *2018 15th International Conference on Electrical*

Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 656-659 (2018).

[21] Somsak, T., Namin, A., Tammawam, W., Thongpron, J., Tippachon, W., and Oranpiroj,K. "A Prototype of Block UU Shape Ferrite Cores Inductive Wireless Power Transfer for EV Charger,"*2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2021, pp. 930-935.

[22] Somsak, T., Namin, A., Sriprom, T., Thongpron, J., Kamnarn, U. and Patcharaprakiti, N. "Constant Current - voltage with Maximum Efficiency Inductive Wireless EV Charging Control using Dual - sides DC Converters,"*2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2021, pp. 936-941.

[23] Sriprom, T., Namin, A., Tammawan, W., et al., "Variable Frequency Control for Constant Current Constant Voltage Inductive Wireless EV Charging System", 2022 International Power Electronics Conference (IPEC-Himeji 2022- ECCE Asia), pp. 1481-1488 (2022).

[24] Thongpron, J., Tammawan, W., Somsak, T., et,al. "A 10 kW Inductive Wireless Power Transfer Prototype for EV Charging in Thailand", ECTI-EEC, Vol. 20, No. 1, pp. 83–95, (2022).

Non-isolated onboard EV Charger controller design based on Adaptive Hamiltonian control law

N. Somboonpanya[1], S. Khomfoi[1], and T. Phophongviwat[1]

[1]Department of Electrical Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

*Corresponding address (E-mail: 65016033@kmitl.ac.th, +66-90 724 9776)

Abstract

The growth of the electric vehicles (EVs) market and the inadequacy of EV infrastructure, especially EV charging stations, leads to an increase in charging EVs at home or the workplace. Onboard EV chargers (OBCs) have been installed on electric vehicles to charge batteries which consists of active front-end rectifier (AFE) and bidirectional buck-boost converter (BDC). The trend of increasing power ratings OBCs, the improvement in stability and performance has been interesting. This paper demonstrates non-isolated EV onboard chargers with controller design based on adaptive Hamiltonian control law to approach a robust control system. Finally, the feasibility of the proposed controlled scheme is validated by MATLAB/Simulink with an 11kW non-isolated onboard EV charger. In addition, the performance of the proposed controlled scheme is compared with conventional PI control.

Keywords: Non-isolated onboard EV charger (OBCs), active front-end rectifier (AFE), Hamiltonian control law, The bidirectional buck-boost converter (BDC), Conventional PI control

1. Introduction

Awareness of climate change and competitive prices of electric vehicles (EVs) leads to growth in the electric car market. In 2022, the electric vehicle population increased by approximately 75% from 2021, was contributed by sustained policies and lower costs in the production process. The electric vehicle benefits can be suggested, such as a reduction in gas emissions and high efficiency when compared with internal combustion engines (ICE). With highly extreme demand for clean energy vehicles and a lack of infrastructure. Especially inadequate public EV charging stations leads to higher charging at home or workspaces [1].
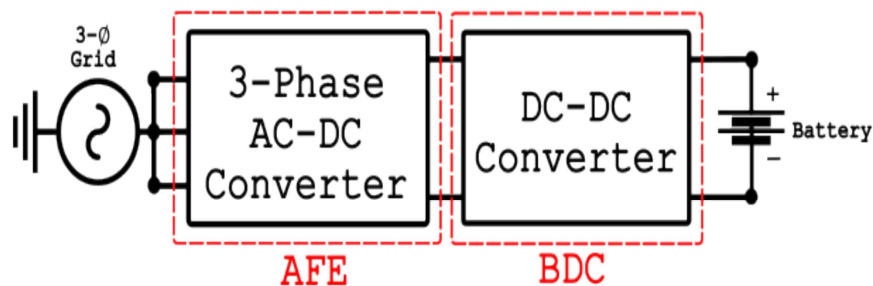

Fig. 1 Diagram of non-isolated EV onboard charger (OBCs)

On-board EV chargers (OBCs) were installed on EVs for charging the battery on electric vehicles. OBCs include the AC/DC and DC/DC stages shown in Fig 1. Active front-end rectifier (AFE) is mainly used for maintaining dc bus voltage and alleviating harmonic current drawbacks. In addition, AFE is applied for power factor correction in AC/DC stage. The bidirectional buck-boost converter (BDC) is integrated with OBCs to regulate charging battery current and voltage during charging operation. With demand in extending driving range, a large pack of batteries is stored, resulting in the development of

---

high-power rating OBCs for a decrease in charging time. Also, OBCs must operates in wide power range deal to power rating of charging stations [2]. Pulse charge technique is suggested to retain battery lifetime and reduce charging duration [3], since charging current is discontinuity leads to high dynamic operation. Therefore, wide operating point, fast response and high performance have become significant factors to design the controller. Currently, OBCs have been design based on convention PI control law which is linear controller, it cannot operates in wide input range since it depend on linearized controlled plant. Therefore, nonlinear control such as energy-based control, model predictive control have been applied to deal with wide operating point [4].

This paper proposes a adaptive Hamiltonian control law which is energy-based control to non-isolated OBCs. The proposed controlled scheme can effectively handle high dynamics and give acceptable performance [5]. The Hamiltonian control law is applied on BDC to regulate charging current and voltage during charging operation.

This paper consists of five sections. Port-Hamiltonian formulation of nonlinear system has been expressed in section II, and section III presents Adaptive Hamiltonian control law formulation on BDC and cascaded PI control on AFE. Section IV details the simulation results between the proposed controlled scheme and conventional PI control law. Finally, the conclusion is illustrated in section V.

2. Port-Hamiltonian Formulation of Nonlinear system

2.1. *Port-Hamiltonian Formulation of Nonlinear system*

Nonlinear differential equation of controlled plant can be formed based on Port-Hamiltonian as [5].

$$\dot{x} = [J\text{-}R]\,\frac{\partial H(x)}{\partial x} + g(x) \cdot u + \xi \tag{1}$$

Where $J = -J^T$ is the interconnection symmetric matrix ($n \times n$), $R = R^T$ is the damping matrix ($n \times n$), $H(x)$ is the Hamiltonian function (energy storage function) of the plant, $g(x)$ is the input matrix ($n \times m$), u is the input control variable ($u \in R^n$), ξ is the disturbance of system. Estimation the control input u is the main objective to eliminate errors between x and x_d (the desired set point). Therefore, the close-loop control can be defined as following equation.

$$\dot{x} = [J_d\text{-}R_d]\,\frac{\partial H_d(x)}{\partial x_d} + \frac{dx_d}{dt} \tag{2}$$

Where $J_d = -J_d{}^T$ is the control interconnection symmetric matrix ($n \times n$), $R_d = R_d{}^T$ is the necessary damping matrix ($n \times n$), $H_d(x)$ is the Hamiltonian function with desired set points. After substitution (1) into (2), the Hamiltonian control law can be obtained as (3). Control input u can be estimated by solving equation (3).

$$[J_d\text{-}R_d]\,\frac{\partial H_{d(x)}}{\partial x} + \frac{dx_d}{dt} = [J\text{-}R]\,\frac{\partial H(x)}{\partial x} + g(x) \cdot u + \xi \tag{3}$$

3. Modelling and Controller design

In this section, the bidirectional buck-boost converter which be integrated on non-isolated OBCs and active-front end rectifier are suggested. Moreover, OBCs is formulated by using adaptive Hamiltonian control law to achieve zero steady-state errors and ability to response during dynamic operations.

3.1. The bidirectional buck-boost converter modelling

Fig 2. displays the schematic of the BDC for non-isolated onboard EV chargers. According to [6]. In Charging mode, BDC operates as buck converter. Hence, the differential equations of BDC in buck converter mode can be expressed as follow.
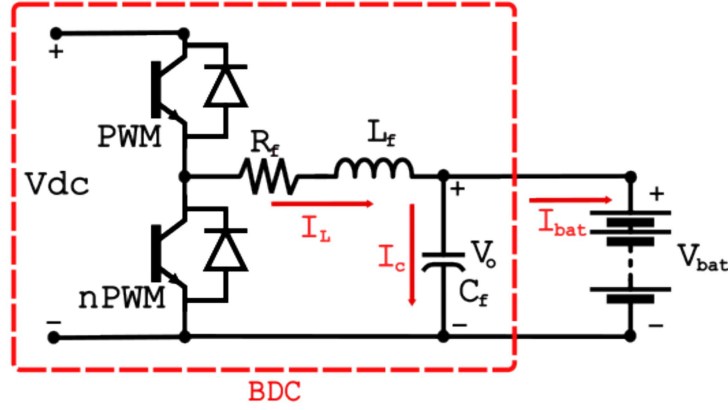
Fig. 2 Bidirectional DC/DC converter (BDC)

$$\frac{dI_L}{dt} = \frac{1}{L_f} (dV_{dc} - R_f I_L - V_o) \tag{4}$$

$$\frac{dV_o}{dt} = \frac{1}{C_f} (I_L - I_{bat}) \tag{5}$$

Where V_{dc}, R_f, L_f, C_f, I_L, V_o, I_{bat} and d are the dc bus voltage, resistor, inductor, capacitor, inductor current, output voltage and battery current and duty ratio, consequently. Then, defines the state variables $x = [x_1\ x_2]^T = [I_L\ V_o]^T$, the desired set values $x_d = [x_{1d}\ x_{2d}]^T = [I_{Lref}\ V_{oref}]^T$ and the control inputs $u = d$. After using storage component in (4) and (5), Hamiltonian energy storage function can be assigned as:

$$H(x) = \frac{1}{2}(L_f x_1^2 + C_f x_2^2) \tag{6}$$

Therefore, the gradient field of the H(x) can be obtained as:

$$\frac{\partial H(x)}{\partial x} = [L_f x_1\ C_f x_2]^T \tag{7}$$

From (1), (4), (5) and (7), the matrix of [J-R], g and ξ can be obtained as following:

$$[J-R] = \begin{bmatrix} \frac{-R_f}{L_f^2} & \frac{1}{L_f C_f} \\ \frac{-1}{L_f C_f} & 0 \end{bmatrix},\ g = \begin{bmatrix} \frac{V_{dc}}{L_f} \\ 0 \end{bmatrix},\ \xi = \begin{bmatrix} 0 \\ -\frac{I_{bat}}{C_f} \end{bmatrix} \tag{8}$$

To satisfy Hamiltonian control law, the error variables between state variable and desired equilibrium point are assigned as following:

$$e = \begin{bmatrix} x_{1d} - x_1 & x_{2d} - x_2 \end{bmatrix}^T \tag{9}$$

According to (9) The Hamiltonian function of error can be obtained as:

$$H_d(x) = \frac{1}{2} (L_f (x_{1d} - x_1)^2 + C_f (x_{2d} - x_2)^2) \tag{10}$$

Then, the gradient field of the $H_d(x)$ can be obtained as (11):

$$\frac{\partial H_d(x)}{\partial x} = \begin{bmatrix} L_f(x_1 - x_{1d}) \\ C_f(x_2 - x_{2d}) \end{bmatrix} \tag{11}$$

Then, the control interconnection symmetric matrix J_d and damping matrix R_d are demonstrated as follow:

$$J_d = \begin{bmatrix} 0 & \frac{(1+K_j)}{L_fC_f} \\ -\frac{(1+K_j)}{L_fC_f} & 0 \end{bmatrix}, R_d = \begin{bmatrix} \frac{(R_f+K_r)}{L_s^2} & 0 \\ 0 & 0 \end{bmatrix} \tag{12}$$

Where $J_d = -J_d{}^T$, $R_d = R_d{}^T \geq 0$ and K_R is the tuning controller for damping transient oscillation. K_J is the periodical adaptive gain. The Hamiltonian control law (3) can be formulated as (13):

$$\begin{bmatrix} -\frac{(R_f+K_r)}{L_s^2} & \frac{(1+K_J)}{L_fC_f} \\ -\frac{(1+K_J)}{L_fC_f} & 0 \end{bmatrix}\begin{bmatrix} L_f(x_1-x_{1d}) \\ C_f(x_2-x_{2d}) \end{bmatrix} = \begin{bmatrix} \frac{-R_f}{L_f^2} & \frac{1}{L_fC_f} \\ \frac{-1}{L_fC_f} & 0 \end{bmatrix}\begin{bmatrix} L_s(x_1) \\ C_f(x_2) \end{bmatrix} + \begin{bmatrix} \frac{V_{dc}}{L_f} \\ 0 \end{bmatrix}\cdot u + \begin{bmatrix} 0 \\ -\frac{I_{bat}}{C_f} \end{bmatrix} \tag{13}$$

Finally, by solving the equation (13), K_J and the control input can be obtained as (14), (15).

$$K_J = -\frac{(I_{bat}-x_{1d})}{(x_1-x_{1d})} \tag{14}$$

$$d = \frac{(x_{2d}-K_Jx_2+K_Jx_{2d}-K_Rx_1+K_Rx_{1d}+R_f x_{1d})}{V_{dc}} \tag{15}$$

3.2. Active front-end power factor correction with cascaded PI control
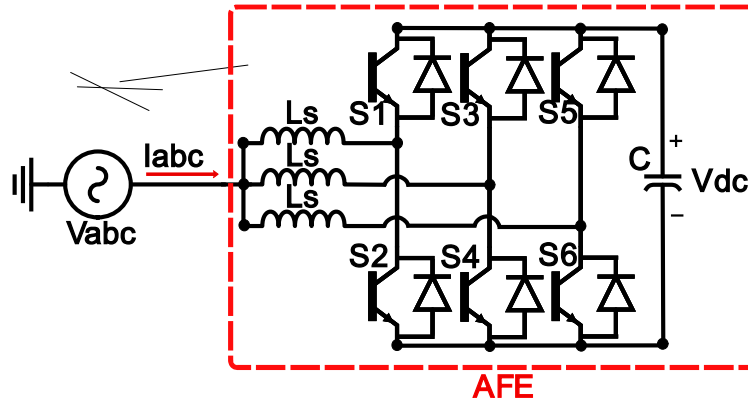

Fig. 3 Active front-end Rectifier (AFE)

Fig 3. illustrates the circuit of three-phase AFE for a non-isolated onboard EV charger. According to [7], the differential equations of AFE are given in DQ-frame as:

$$\frac{dI_d}{dt} = \frac{1}{L_s}\left(V_d - M_d V_{dc} - R_s I_d + \omega_e L_s I_q\right) \tag{16}$$

$$\frac{dI_q}{dt} = \frac{1}{L_s}\left(V_q - M_q V_{dc} - R_s I_q - \omega_e L_s I_q\right) \tag{17}$$

$$\frac{dV_{dc}}{dt} = \frac{1}{C}\left(\frac{3}{2}(M_d I_d + M_q I_q)\right) + I_{dc} \tag{18}$$

Where V_d and V_q are the grid voltage in the DQ-frame, I_d and I_q are the line current in the DQ-frame, ω_e is the rotational speed of the DQ-frame, Ls is line inductance, Rs is the line series resistance,

M_d and M_q are the duty ratio in the DQ-frame , V_{dc} and I_{dc} is the dc bus voltage and dc bus current respectively. From (16), (17) when imposes $\frac{dI_d}{dt}$ and $\frac{dI_q}{dt} = 0$, the M_d and M_q can be expressed as (19), (20)

$$M_d = \frac{1}{V_{dc}}\left(V_d - R_s I_d + \omega_e L_s I_q\right) \tag{19}$$

$$M_q = \frac{1}{V_{dc}}\left(V_q - R_s I_q - \omega_e L_s I_q\right) \tag{20}$$

However, the aforementioned equation cannot ensure zero steady state errors. To eliminate these errors, PI controller is implemented. To maintain dc bus voltage, the sufficient I_d is calculated by following equation.

$$I_{dref} = K_p(V_{dcref} - V_{dc}) + K_I\int(V_{dcref} - V_{dc})dt \tag{21}$$

Where K_p, K_I, V_{dcref}, I_{dref} and I_{qref} is proportional gain, integral gain, desired dc bus voltage, desired current in DQ-frame respectively. Additionally, I_{qref} is set to be 0.
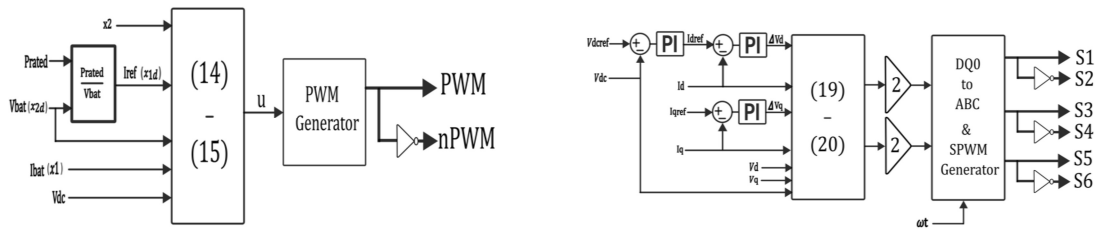
3.3. Onboard-EV charger controller Conclusion

Fig. 4. (a) proposed control scheme; (b) diagram of AFE controller

(a)   (b)

Fig 4(a). shows the proposed adaptive Hamiltonian control on BDC. Equation (14) is used to calculate the adaptive gain (K_J), and then equation (15) calculates control input (u) to control charging current and voltage by using desired and actual state variable and K_J. Fig 4(b). demonstrates diagram of AFE controller based on cascaded PI controller, The outer PI loop provides sufficient direct-axis reference current, inner PI loop eliminates errors between actual current and reference current in DQ-frame and (20), (21) calculates Duty cycle ratio in DQ-frame.

4. Performance Validation

Table 1. An Active front-end rectifier parameters

symbol	Description	Value
Vabc	Nominal grid voltage	380 Vrms
Vdc	Nominal dc-link voltage	600 V
Ls	Line inductance	2 mH
Rs	Line resistance	1 mΩ
C	DC-link capacitance	1500 uF
fs	Switching frequency	20 kHz

Table 2. A Bidirectional buck-boost converter parameters

symbol	Description	value
Vdc	Nominal dc-link voltage	600 V
Vbat	Nominal battery voltage	400 V
Lf	Filter inductance	5 mH
Rf	Series filter resistance	1 mΩ
Cf	Output filters capacitance	500 uF
fs	Switching frequency	20 kHz
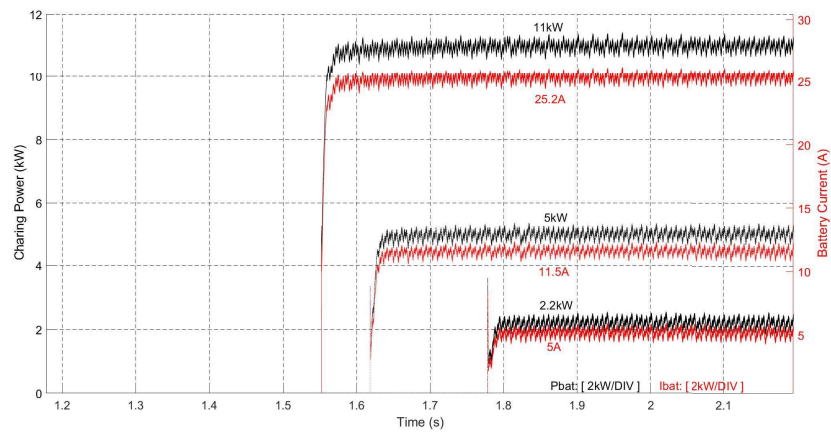
(a)
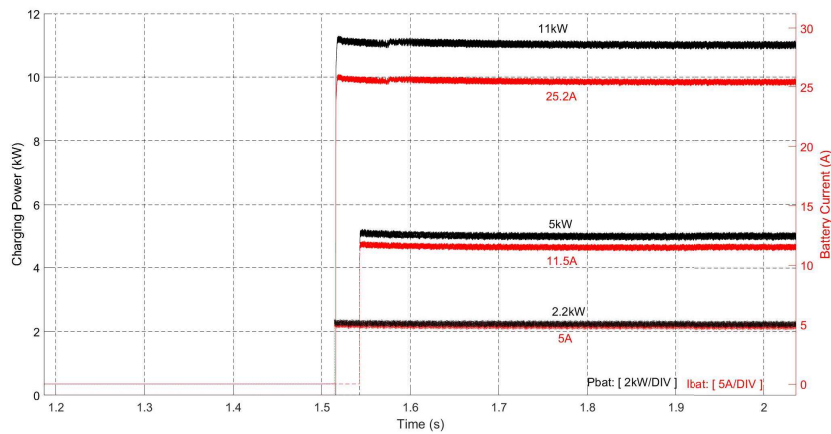
(b)

Fig. 5 Simulation results: (a) proposed control scheme; (b) Conventional PI control law

In this section, MATLAB/Simulink was implemented to simulate non-isolated onboard-EV charger. AFE is designed with cascaded PI control law, while BDC is designed with adaptive Hamiltonian control

law. Additionally, the results of the replacing adaptive Hamiltonian control law in BDC with Conventional PI controller which is designed for 11kW operating point have been demonstrated. Table 1 and Table 2 show the parameters of AFE and BDC respectively. K_P and K_I of AFE controller are 5 and 15 respectively, while K_R is set to 15 and adaptive gain is limited ($-5 \leq K_J \leq 5$). The results include battery charging power (P_{bat}), battery charging current (I_{bat}). There are 3 scenarios, battery was received command at 1.5s and charged with 2.2kW, 5kW and 11 kW respectively.

Fig 5(a) demonstrates the results of adaptive Hamiltonian control law to BDC. As can be seen that P_{bat} was 2.2kW, 5kW and 11kW respectively. Regarding battery current, I_{bat} was 5A, 11.5A and 25.2A respectively. It is clear that the settling time was 20ms, 45ms and 15ms when operating point was 11kW, 5kW and 2.2kW respectively.

Fig 5(b) depicts the results of conventional PI controller to BDC. As can be observed that P_{bat} was 2.2kW, 5kW and 11kW respectively. Charging current was regulated at 5A, 11.5A and 25.2A respectively. The settling time was 84ms, 145ms and 200ms when operating point was 11kW, 5kW and 2.2kW respectively.

5. Conclusion

In this paper, the model of active front-end rectifier (AFE) and bidirectional buck-boost converter (BDC) for a non-isolated onboard EV charger are presented. The adaptive Hamiltonian control law was applied to BDC. Its performance was compared with conventional PI controller. The results of the study indicate that adaptive Hamiltonian control law effectively controls charging voltage and current. Also, it demonstrates the ability to operate in wide operating ranges. While conventional PI controller when operate in lower power, a increase in operating delay can be obviously observed. Non-isolated onboard EV charger was simulated through MATLAB/Simulink for validation the proposed control scheme.

Reference

[1] International Energy Agency, "Global electric vehicle outlook 2022," 2022. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2022
[2] A. Khaligh and M. D'Antonio, "Global Trends in High-Power On-Board Chargers for Electric Vehicles," in IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 3306-3324, April 2019, doi: 10.1109/TVT.2019.2897050.
[3] P. Jampeethong and S. Khomfoi, "An EV quick charging station using a pulse frequency current control technique," 2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Hua Hin, Thailand, 2015, pp. 1-5.
[4] A. Safaeinasab, H. Soltani Gohari and K. Abbaszadeh, "Hamiltonian Energy-Based Sliding Mode Control Approach for a Multi-port Bidirectional EV Charger via Zero Dynamic," 2021 12th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC), Tabriz, Iran, 2021, pp. 1-6.
[5] P. Thounthong, "Port-Hamiltonian Formulation of Adaptive Hamiltonian PID controller to Solve Constant Power Load Stability Issue in DC Microgrid: Control of a Fuel Cell Converter," 2021 IEEE 12th Energy Conversion Congress & Exposition - Asia (ECCE-Asia), Singapore, Singapore, 2021, pp. 1864-1869.
[6]. S. Nayak, S. Mohanty and H. J. Saikia, "An Improved Control Method for the DC-DC Converter in Vehicle to Grid Charging System," 2017 14th IEEE India Council International Conference (INDICON), Roorkee, India, 2017, pp. 1-6.
[7] C. -C. Hou, P. -T. Cheng, S. Bhattacharya and J. Lin, "Modeling and Control of Three-Phase Active front-end Converters," IECON 2007 - 33rd Annual Conference of the IEEE Industrial Electronics Society, Taipei, Taiwan, 2007, pp. 1449-1454.