Universiti Teknologi MARA
Cawangan Melaka

PCMJ

Progress in Computing and Mathematics Journal

volume 1

https://fskmjebat.uitm.edu.my/pcmj/

Progress in Computing and Mathematics Journal
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin
77300, Merlimau, Melaka Bandaraya Bersejarah

PCMJ

Progress in Computing and Mathematics Journal

volume 1

Progress in Computing and Mathematics Journal (PCMJ)
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin
77300, Merlimau, Melaka Bandaraya Bersejarah

EDITORS

PCMJ
Progress in Computing and Mathematics Journal
volume 1

PREFACE

Welcome to the inaugural volume of the **Progress in Computing and Mathematics Journal (PCMJ)**, a publication proudly presented by the College of Computing, Informatics, and Mathematics at UiTM Cawangan Melaka.

This journal represents a significant step in our commitment to fostering a vibrant research culture, initially providing a crucial platform for our undergraduate students to showcase their intellectual curiosity, dedication to scholarly pursuit, and potential to contribute to the broader academic discourse in the fields of computing and mathematics. However, we envision PCMJ evolving into a beacon for researchers both nationally and internationally. We aspire to cultivate a space where groundbreaking research and innovative ideas converge, fostering collaboration and intellectual exchange among established scholars and emerging talents alike.

The manuscripts featured in this first volume, predominantly authored by our undergraduate students, are a testament to the hard work and dedication of these budding researchers, as well as the guidance and support provided by their faculty mentors. They cover a diverse range of topics, reflecting the breadth and depth of research interests within our college, and set the stage for the high-quality scholarship we aim to attract in future volumes.

As editors, we are honored to have played a role in bringing this journal to fruition. We extend our sincere gratitude to all the authors, reviewers, and members of the editorial board for their invaluable contributions. We also acknowledge the unwavering support of the college administration in making this initiative possible.

We hope that PCMJ will inspire future generations of students and researchers to embrace research and innovation, to push the boundaries of knowledge, and to make their mark on the world of computing and mathematics.

Editors
Progress in Computing and Mathematics Journal (PCMJ)
College of Computing, Informatics, and Mathematics
UiTM Cawangan Melaka

TABLE OF CONTENTS

PCMJ

Progress in Computer and Mathematics Journal (PCMJ)
volume 1 [October, 2024]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

DETECTION SYSTEM OF DISEASE FROM TOMATO LEAF USING CONVOLUTIONAL NEURAL NETWORK

Nurkhairunnisa' Binti Azhan
College of Computing, Informatics and Mathematics

Hajar Izzati Binti Mohd Ghazalli
College of Computing, Informatics and Mathematics

Article Info

Received: February 2024
Accepted: August 2024
Available Online: October 2024

Abstract

Tomato farming is crucial for global food production, but diseases affecting tomato plants can harm crop quality and lead to economic losses for farmers. Many farmers lack expertise and education in identifying these diseases, highlighting the need for accessible tools. This study focused on creating a user-friendly web-based system using Convolutional Neural Networks (CNN) to detect tomato leaf diseases. The goal was to empower farmers with a convenient and efficient platform to identify and address diseases, automating the detection process and reducing reliance on manual analysis. The system, achieving over 92.5% accuracy, aimed to enhance productivity by providing timely and accurate identification of tomato leaf diseases. Farmers could easily monitor and assess plant infections through the web-based platform. The research outcomes are expected to benefit the agricultural community by offering a valuable tool for informed decision- making, leading to improved crop quality and increased productivity. Future improvements could include additional functions and information for users, as well as expanding the system to detect more types of tomato leaf diseases.

Keywords: CNN, Detection, Disease

INTRODUCTION

In Malaysia, the highlands produce the majority of tomatoes such as at Kundasang and the Cameron Highlands. Cameron Highlands, Pahang, Sabah, and Kelantan Lodging are where tomatoes are primarily produced. Sarawak, Selangor, Johor, and Melaka all have additional tomato-planting areas (Rahim et al., 2017). More than 20 tomato leaf diseases, have a direct impact on tomato crop productivity and quality, which directly represents economic loss (Gadade & Kirange, 2020). Each year, pathogens cause harm to the tomato harvests. From leaf to fruit, diseases are transmitted. It is essential to spot illnesses on plant leaves in their early stages. Therefore, the fruits have enough nutrition, both in terms of quantity and quality. The procedure for detecting plant leaf illnesses using standard methods is lengthy, and the team's specialists are constantly monitoring it (Kaur & Bhatia, 2019).

The tomato (Solanum Lycopersicon L.), which grows best in shaded areas and blooms at night, is a member of the Solanaceae plant family. The second-most valuable vegetable crop in the world is the tomato. With a total cultivable area of 4.85 million square meters, it now produces 182.3 million tons of fruits annually. It is one of the crops that is cultivated and eaten most often worldwide (Quinet et al., 2019). The tomato is subject to both abiotic and biotic agronomic restrictions, much like any other plant. In order to produce a plentiful and high-quality crop, tomato cultivation does indeed require an environment that is comparatively chilly and dry. Tomatillos are warm-season plants. The availability of water has a significant impact on both yield and quality (Oscar Cyrille Adantchédé Akotowanou et al., 2022).

The tomato is a food that is high in vitamins and minerals that promote excellent health. It also has a small bit of fat and a lot of protein. It is a great source of minerals that serve as cofactors in enzyme processes, vitamins B6, ascorbic acid, and niacin, as well as vitamin B6 (Garuba et al., 2018). These bioactive substances offer several physiological benefits, such as anti-inflammatory, anti-allergenic, antimicrobial, vasodilatory, antithrombotic, cardio-protective, and antioxidant activities. The largest source of lycopene in the human diet comes from tomatoes, which are abundant in carotenoids (Quinet et al., 2019). In terms of production and consumption, tomatoes are the second-largest vegetable and an essential part of diet. According to several studies, tomatoes are a good source of vitamins, pro-vitamins, minerals, and a large number of secondary metabolites (Oscar Cyrille Adantchédé Akotowanou et al., 2022).

Consumers have grown more knowledgeable of the role that food plays in preventing various chronic illnesses and dysfunctions over the past ten years as well as their health advantages (Quinet et al., 2019). Tomatoes may contribute significantly to a healthy diet and can be eaten raw or cooked without losing any of their nutritional content. Over 80% of all tomatoes grown for commercial purposes are used to make processed foods like ketchup, soup, and juice. Numerous health advantages are associated with a diet high in tomatoes and tomato-based products. Many of these advantages are attributed to the antioxidants found in tomatoes (Collins et al., 2022).

There are 3 diseases that researchers choose which are late blight disease, leaf mold disease and mosaic virus disease.

PCMJ
Progress in Computer and Mathematics Journal (PCMJ)
volume 1 [October, 2024]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

LITERATURE REVIEW

This chapter discussed the relevant literature that related to the implemented study. This discussion stressed on the type of diseases tomato leaf that may be used. This chapter may be discussed about the image processing technique and the classification technique that were used in the research. There may be comparisons between the technique, the method and existing application were also stated in this research.

The tomato (Solanum Lycopersicon L.) is a valuable vegetable crop, ranking second globally. It thrives in shaded areas, blooms at night, and belongs to the Solanaceae plant family. Tomatoes are cultivated in a total area of 4.85 million square meters, yielding 182.3 million tons of fruits annually, making it a widely grown and consumed crop worldwide (Quinet et al., 2019). The plant faces both abiotic and biotic agronomic challenges, requiring a relatively cool and dry environment for optimal cultivation. Water availability significantly influences yield and quality (Oscar Cyrille Adantchédé Akotowanou et al., 2022).

In the past decade, consumers have become more aware of the role of food in preventing chronic illnesses. Tomatoes, whether consumed raw or cooked, retain their nutritional content and offer various health benefits. Over 80% of commercially grown tomatoes are used in processed foods like ketchup, soup, and juice, providing numerous health advantages attributed to their antioxidants (Collins et al., 2022).

Image processing

Image processing, also known as autonomous machine perception or image enhancement processing (Bhargav & Chakrapani, 2018), involves altering a picture's characteristics to improve its graphical information for human understanding. It can be defined as the addition of features, enhancement, or extraction of information from an image, with the input being a picture and the output either another image or a set of attributes (Lende et al., 2022).

Deep learning algorithms, such as convolutional neural networks (CNN), facilitate automatic feature extraction, minimizing the need for human effort and

domain knowledge. These algorithms utilize a multi-layer data representation architecture, where initial layers focus on low-level features, and subsequent layers extract high-level features (Alzubaidi et al., 2021).

Classification algorithms play a vital role in image processing, categorizing extracted features into multiple classes based on their distinct characteristics (Rani & Kaur, 2019). Deep learning, an advanced form of neural networks, has demonstrated superior performance in various applications, including audio and speech processing, visual data processing, and natural language processing (NLP) (Alzubaidi et al., 2021).

Machine learning, a key component of artificial intelligence, promotes the idea that robots should learn from data independently to simplify human work (DeGregory et al., 2018). CNN, specifically designed for data with a grid pattern like images, draws inspiration from the organization of the visual cortex in animals. Its popularity stems from its ability to automatically extract features from images through the convolution operation (Patil & Rane, 2021; Altaf & Wahid, 2023). Figure 1 shows the illustration of architecture CNN.

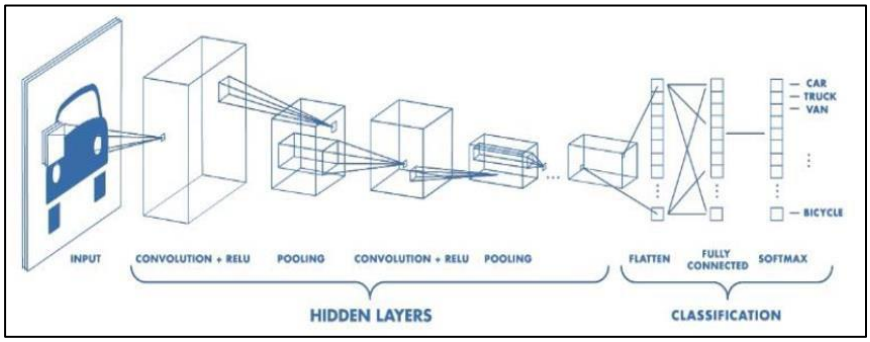

Figure 1

Method VGG-16

A convolutional neural network model called VGG uses extremely deep convolutional networks to recognise images on a huge scale. Because the multi-layer nonlinear layer may enhance the network depth to assure learning more complicated patterns, the tiny convolution kernel is superior than the big convolution kernel and has a relatively modest cost. Nevertheless, VGG employs more parameters and needs greater processing power, which increases memory use. VGG-16 contains three

completely linked layers, the first of which provides the majority of the parameters. The pictures in the experimental data set are ultra-low pixel remote sensing images, hence multiple feature extraction and additional parameters are not required (Ye et al., 2021)

Related work

There were some existing and similar applications to detect an image such automated image capturing system for deep learning-based tomato plant leaf disease detection and recognition, tomato plant leaf disease detection using CNN and identification of tomato plant disease by leaf image using Squeezenet model.

In conclusion based on Table 2.3, three different system with different algorithm had been compared. Those system can detect tomato leaf disease but the dataset that have been used is different. The highest accuracy is system An intelligent VGG-16 framework for brain tumor detection using MRI scans which is 96%.

The researcher chooses three types of disease on tomato leaf which is Phoma Rot, Leaf Miner, and Target Spot. Those disease are from specific breed of tomato Diamante Max. It can detect those disease on leaf and healthy leaf. This innovative technology helps to improve the quantity and quality production of tomato. It can capture the tomato and leaf disease from four side by motor-controlled image capturing box. The researcher collects 4923 datasets of image. By identify those three diseases, the researcher trains a deep convolutional neural network (De Luna et al., 2019)

Detection of Begomovirus disease for identification of disease severity level in tomato leaves using CNN paper was written by (Mubarokah et al., 2022). The researcher built CNN model to identify the image. This tomato leaf disease detection can detect 2 types of leaves which is Begomovirus disease and healthy leaves. In this instance, the dataset comprises 1600 images of tomato leaves captured from different perspectives to enhance its variety.

In An intelligent VGG-16 framework for brain tumor detection using MRI-scans study, the researcher aim to detect brain tumor using MRI scans. This system using CNN method which is VGG-16 to train the data. The dataset was obtained from Kaggle, an online repository, consisting of a total of 253 original images. These images have been organized into two folders, 'yes' and 'no,' with 155 images in the former and 98 in the latter.

System	Algorithm	Data set	Accuracy
Automated image capturing system for deep learning-based tomato plant leaf disease detection and recognition	Convolutional neural network (CNN). F-RCNN	4923 images. Type of disease, Phoma Rot, Leaf Miner, and Target Spot	91.67 % accuracy
Detection of Begomovirus disease for identification of disease severity level in tomato leaves using Convolutional Neural network (CNN)	SSD Mobilenet V1 FPN, EfficientDet D1 and SSD Resnet 50 V1 FPN.	1600 images of tomato leaves captured. This dataset is comprised of two categories of tomato leaves, which include leaves afflicted with Begomovirus disease and leaves that are healthy.	Average accuracy value of SSD Mobilenet V1 FPN, SSD Resnet 50 V1 FPN and EfficientDet D1 model are 56.17%,61.58% and 66.03%
An intelligent VGG-16 framework for brain tumor detection using MRI scans	Convolutional neural network (CNN). VGG-16	There were 253 original photos has been separated into two folders, 155 and 98 in each case.	96% accuracy
Implemented system	Convolutional neural network (CNN).	There are 8020 images representing 3 types of diseases image collected from Kaggle.	-

PCMJ

Progress in Computer and Mathematics Journal (PCMJ)
volume 1 [October, 2024]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

METHODOLOGY

The methodology of this project development is System Development Life Cycle (SDLC). In order to develop software, researcher need to adhere to a specific procedure known as the software life cycle, which is alternatively referred to as the SDLC model (Saravanan, 2020). There are several phases in SDLC which is requirement analysis, design, development, implementation, testing and maintenance. The common models that have been used in SDLC are waterfall model, iterative model, spiral mode, and agile model. Therefore, one of the models may be used in the development of this project which is the modified waterfall model. This model meets the requirements that are involved in the project

Data Collection

The data used for this project is a secondary data that are taken from Kaggle.com. Kaggle is a renowned platform in the field of data science, known for its global prominence and extensive offerings. As the largest data science platform in the world, Kaggle provides a robust set of innovative tools and services aimed at aiding individuals and teams in accomplishing their data science goals. There are 3 type of disease which is late blight, leaf mold and mosaic virus as shown in figure 2. All images for tomato leaf disease is 8020. Those images need to be separate for training so that machine learning can detect the image before classification later.

Figure 2

Implementation

Implementation is the third phase in SDLC where the system is implemented in the production environment. The system may be developed using specific tools and programming

language. The implemented system is tested for its accuracy in detecting the disease leaves. In this phase, the implemented system may be used Phyton and Java programming languages

Testing

Testing is the last phase in SDLC. During this phase, tasks such as identifying system operations, managing data, identifying issues, and updating documentation that underwent further analysis may be performed. Additionally, a User-Satisfaction report might be generated in order to identify any potential user-related problems that need to be addressed in future development iterations. The type of testing that used in designing this project is functionality testing. Accuracy also needs to be tested to make sure the percentage of accuracy is in the best range

Training

The process of data training produced a VGG16 model for tomato leaf disease detection by using a pre-trained model. The initial phase of the process involved loading essential libraries for tasks such as evaluation, data manipulation, visualization, and deep learning. Key libraries encompassed scikit-learn for machine learning, NumPy and Pandas for data management, Matplotlib/Seaborn for visualization, and TensorFlow and Keras for deep learning applications. Batches of picture data were generated for training, validation, and testing using TensorFlow's ImageDataGenerator. The target size could be specified and pixel values could be rescaled when the generator was configured. Various generators were developed for every split dataset. Figure 3 shows the code in splitting the dataset.

```
In [6]: train_split=.8
        test_split=.1
        dummy_split=test_split/(1-train_split)
        train_df, dummy_df=train_test_split(df, train_size=train_split, shuffle=True, random_state=123)
        test_df, valid_df=train_test_split(dummy_df, train_size=dummy_split, shuffle=True, random_state=123)
        print ('train_df length: ', len(train_df), '  test_df length: ', len(test_df), '  valid_df length: ', l

        train_df length:  6416   test_df length:  802   valid_df length:  802
```

Figure 3

RESULT AND DISCUSSION

During accuracy phase, the system undergo testing and determined the accuracy results. Testing is crucial for every system as it helps identify and rectify any bugs, errors or issues before its deployment. Accuracy can be calculate using formula in equation 1

PCMJ
Progress in Computer and Mathematics Journal (PCMJ)
volume 1 [October, 2024]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

$$Accuracy = \frac{Total\ number\ of\ image\ recognized}{Total\ number\ of\ input\ image} \times 100 \quad (1)$$

Precision is the ratio of true positive predictions to the total number of positive predictions made by the model. It is calculated using the formula in equation below. Precision focuses on the accuracy of positive predictions, indicating how many of the predicted positive instances are actually positive. A high precision value means that the model is making fewer false positive predictions. Figure 4 shows the curve graph about precision.

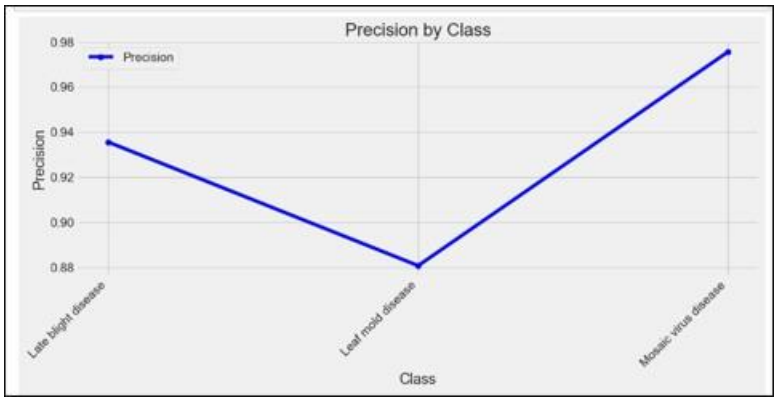

Figure 4

Recall is the ratio of true positive predictions to the total number of actual positive instances. It is calculated using equation below. Recall measures the ability of the model to correctly identify all positive instances. A high recall value indicates that the model is capturing a large portion of the actual positive instances. Figure 5 shows the curve graph about recall or can be known as sensitivity.

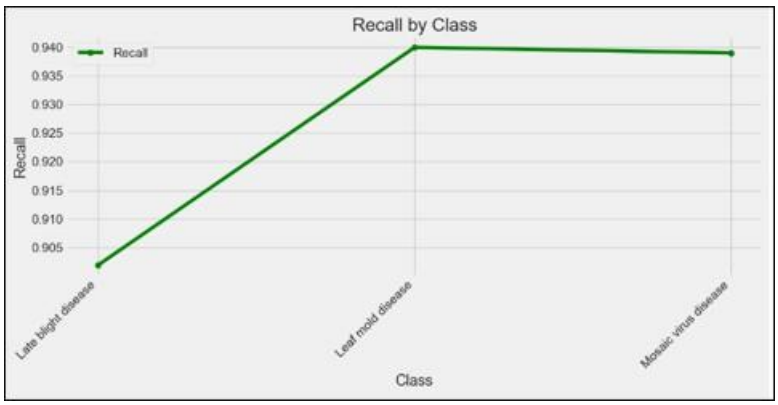

Figure 5

The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall, considering both false positives and false negatives. Figure 6 shows the F1-score.

PCMJ
Progress in Computer and Mathematics Journal (PCMJ)
volume 1 [October, 2024]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

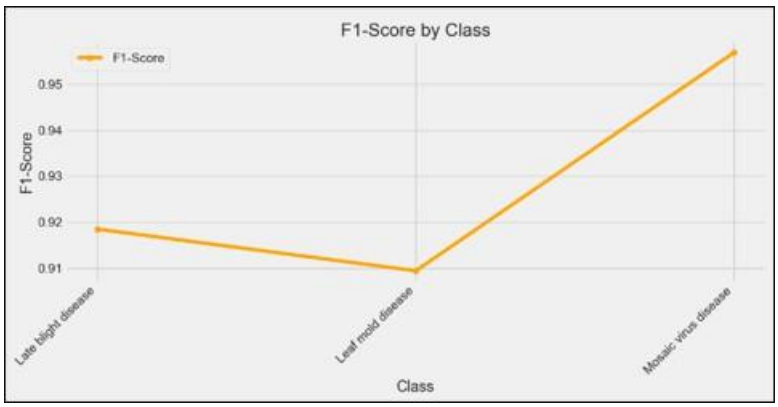

Figure 6

In conclusion, the first chapter in of the project, it focusses on the objective of the project development. The first objective is to design a tomato leaf disease detection system that can help user differentiate between three disease which is late blight disease, leaf mold disease and mosaic virus disease. This objective achieved since the system give the result of name disease after user input the image. Next objective is to develop detection system of tomato leaf disease on web platform using CNN. This objective success with using VGG-16 algorithm under Convolutional Neural Network method. Lastly, to test the accuracy of the application to be developed is achieved by the result that shows in the system.

REFERENCES

Collins, E. J., Bowyer, C., Tsouza, A., & Chopra, M. (2022). Tomatoes: An Extensive Review of the Associated Health Impacts of Tomatoes and Factors That Can Affect Their Cultivation. *Biology*, *11*(2). https://doi.org/10.3390/biology11020239

De Luna, R. G., Dadios, E. P., & Bandala, A. A. (2019). Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, *2018-Octob*(October), 1414–1419. https://doi.org/10.1109/TENCON.2018.8650088

Gadade, H. D., & Kirange, D. K. (2020). Tomato leaf disease diagnosis and severity measurement. *Proceedings of the World Conference on Smart Trends in Systems, Security and Sustainability, WS4 2020*, 318–323. https://doi.org/10.1109/WorldS450073.2020.9210294

Garuba, T., Mustapha, O. T., & Oyeyiola, G. P. (2018). Shelf life and proximate composition of tomato (*Solanum lycopersicum* L.) fruits as influenced by storage methods. *Ceylon Journal of Science*, *47*(4), 387. https://doi.org/10.4038/cjs.v47i4.7557

Kaur, M., & Bhatia, R. (2019). Development of an improved tomato leaf disease detection and classification method. *2019 IEEE Conference on Information and Communication Technology, CICT 2019*, 1–5. https://doi.org/10.1109/CICT48419.2019.9066230

Mubarokah, I., Laksono, P., Cepy, Safitri, R., & Idris, I. (2022). Detection of Begomovirus Disease for Identification of Disease Severity Level in Tomato Leaves Using Convolutional Neural Network (CNN). *ISESD 2022 - 2022 International Symposium on Electronics and Smart Devices, Proceeding*. https://doi.org/10.1109/ISESD56103.2022.9980675

Oscar Cyrille Adantchédé Akotowanou, Euloge Sènan Adjou, Adéyèmi Berane Olubi, Sylvain Daton Kougblenou, Edwige Dahouenon Ahoussi, & Dominique CK Sohounhloué. (2022). The tomato (Solanum lycopersicum L.) in community development: An overview focused on nutritional properties, agronomic constraints, reThe tomato (Solanum lycopersicum L.) in community development: An overview focuscent achievements and future prospective. *International Journal of Frontiers in Biology and Pharmacy Research*, *3*(2), 008–016. https://doi.org/10.53294/ijfbpr.2022.3.2.0061

Quinet, M., Angosto, T., Yuste-Lisbona, F. J., Blanchard-Gros, R., Bigot, S., Martinez, J. P., & Lutts, S. (2019). Tomato Fruit Development and Metabolism. *Frontiers in Plant Science*, *10*(November), 1–23. https://doi.org/10.3389/fpls.2019.01554

Rahim, H., Mukmin, A., Wahab, A., Zaffrie, M., Amin, M., Harun, A., & Haimid, M. T. (2017). Technological adoption evaluation of agricultural and food sectors towards modern agriculture: Tomato (Penilaian penggunaan teknologi bagi sektor pertanian dan makanan ke arah pertanian moden: Tomato). *Economic and Technology Management Review*, *12*, 41–53.

Saravanan, T. (2020). *Comparative Analysis of Software Life Cycle Models*. 906–909. https://doi.org/10.1109/ICACCCN51052.2020.9362931

Ye, M., Ruiwen, N., Chang, Z., He, G., Tianli, H., Shijun, L., Yu, S., Tong, Z., & Ying, G. (2021). A Lightweight Model of VGG-16 for Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 6916–6922. https://doi.org/10.1109/JSTARS.2021.3090085

PCMJ

Progress in Computing and Mathematics Journal

UNIVERSITI TEKNOLOGI MARA
Cawangan Melaka

eISSN 3030-6728