Durian Disease Classification using Vision Transformer for Cutting-Edge Disease Control

Marizuana Mat Daud[1], Abdelrahman Abualqumssan[2], Fadilla ‘Atyka Nor Rashid[3],
Mohamad Hanif Md Saad[4], Wan Mimi Diyana Wan Zaki[5], Nurhizam Safie Mohd Satar[6]
Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[1]
Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[2, 4, 5]
Centre for Artificial Intelligence Technology, Faculty of Information Science & Technology,
Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[3]
Centre for Software Technology & Management, Faculty of Information Science & Technology,
Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia[6]

***Abstract*—The durian fruit holds a prominent position as a beloved fruit not only in ASEAN countries but also in European nations. Its significant potential for contributing to economic growth in the agricultural sector is undeniable. However, the prevalence of durian leaf diseases in various ASEAN countries, including Malaysia, Indonesia, the Philippines, and Thailand, presents formidable challenges. Traditionally, the identification of these leaf diseases has relied on manual visual inspection, a laborious and time-consuming process. In response to this challenge, an innovative approach is presented for the classification and recognition of durian leaf diseases, delves into cutting-edge disease control strategies using vision transformer. The diseases include the classes of leaf spot, blight sport, algal leaf spot and healthy class. Our methodology incorporates the utilization of well-established deep learning models, specifically vision transformer model, with meticulous fine-tuning of hyperparameters such as epochs, optimizers, and maximum learning rates. Notably, our research demonstrates an outstanding achievement: vision transformer attains an impressive accuracy rate of 94.12% through the hyperparameter of the Adam optimizer with a maximum learning rate of 0.001. This work not only provides a robust solution for durian disease control but also showcases the potential of advanced deep learning techniques in agricultural practices. Our work contributes to the broader field of precision agriculture and underscores the critical role of technology in securing the future of durian farming**.

Keywords—Vision transformer; durian disease; deep learning; disease control

I. Introduction

The durian fruit's popularity has surged in recent years, primarily driven by increased consumer demand, notably from China [1]. Moreover, it has found a substantial export market in Southeast Asian countries, Hong Kong, Australia, and Western nations such as United States. This upswing in the durian market can be attributed in part to the cultivation of premium varieties renowned for their exceptional flavor and consistent pulp quality. Notably, varieties like D24, D197 (Musang King), and D200 (Black Thorn) from Malaysia, as well as traditional Thai cultivars such as Monthong, Chanee, and Kanyau, have garnered significant attention and are in high demand, painting a promising future for the fruit.

Thailand maintains its position as the primary producer and exporter of durians, with other countries like Malaysia, Indonesia, Vietnam, Cambodia, and the Philippines also cultivating this unique fruit [2]. The global durian fruit trade is characterized by a dominant duopoly, with China taking the lead in imports, while Thailand leads in exports. In 2021, Thailand's durian exports reached an impressive value of 3,920 million USD, making up a significant 82.7% of the total global trade. In contrast, Malaysia's contribution ranked fourth, comprising about 0.67% of the trade volume, with a total value of 31.8 million USD. Simultaneously, China asserted its dominance in global durian imports in the same year, with an astonishing 4,240 million USD, constituting a substantial 89.4% of the overall trade. Additionally, notable participants in the market, following China's lead, included Hong Kong, Vietnam, Chinese Taipei, and Singapore, accounting for 89.4%, 5.37%, 2.43%, 0.72%, and 0.36% of the trade, respectively [3]. Fig. 1 and Fig. 2 depict the top importer and exporter of durians, respectively.

The adoption of modern agricultural practices, including drip irrigation, enhanced fertilizer formulations and application techniques, and improved cultural and postharvest methods, has significantly contributed to the increased productivity of durian farms [4]. Nevertheless, growers remain vigilant due to the persistent threat of diseases in the industry. Durian trees are susceptible to a range of diseases, such as spot cancer, base rot, base disease, seedling disease, dead tip disease, fungal infections, leaf spots, leaf blight, and fruit rot. Among these, stem rot disease, primarily caused by P. Palmivora, stands out as a particularly perilous ailment. This disease severely impairs the tree's nutrient transport system within the stem.

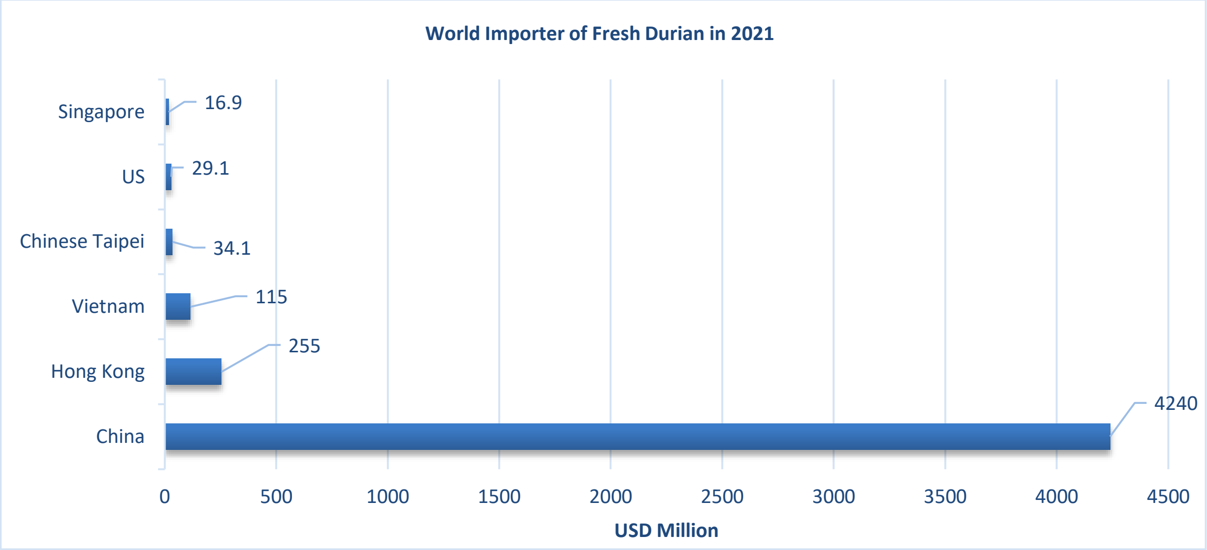


Fig. 1. World importer of fresh durian in 2021.

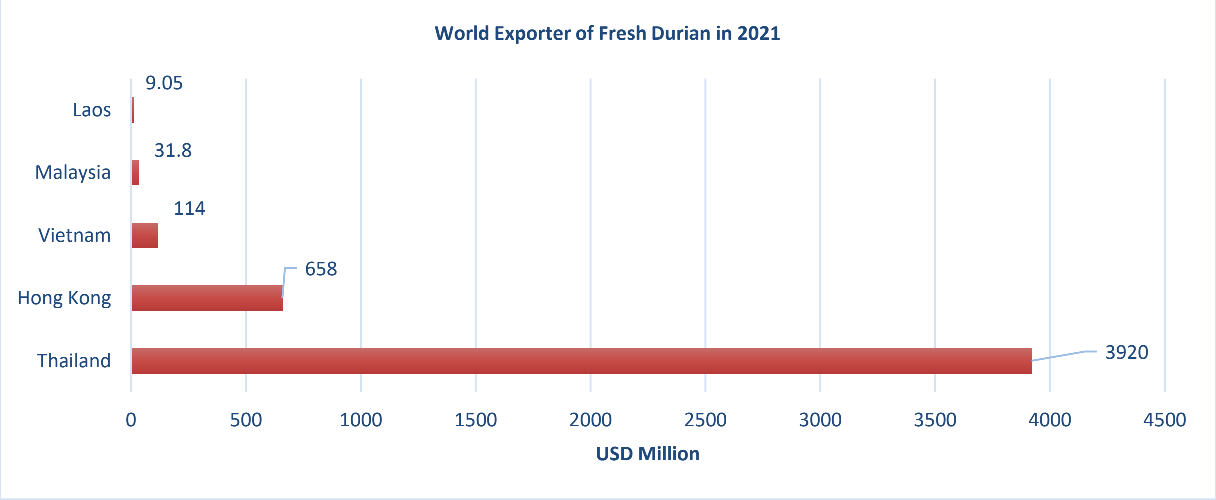


Fig. 2. World exporter of fresh durian in 2021.

Durian trees are susceptible to various diseases, such as algal diseases caused by cephaleuros virescens, characterized by the appearance of orange, rust-colored velutinous spots on the upper surfaces of leaves, twigs, and branches. Another concern is anthracnose, resulting from Colletotrichum gloeosporioides, which manifests as dark lesions on fruit and premature fruit drop. Moreover, Phomopsis leaf spot, induced by diplodia heobromae and C. Gloeosporioides, presents as necrotic, brown circular spots, approximately 1 mm in diameter, featuring dark margins and yellow halos on leaves. The sinister pink disease, attributed to erythricium salmonicolor, is marked by pinkish-white mycelial threads that envelop branches and shoots.

Additionally, postharvest fruit rots caused by Phyllosticta sp. and curvularia eragrostidis result in irregular necrotic patches in varying shades of brown. Rhizoctonia leaf blight, originating from Rhizoctonia solani, leads to water-soaked spots on leaves that coalesce to form larger, irregular, water-soaked patches, eventually drying into light brown necrotic lesions. Lastly, sooty mold and black mildew, caused by Black Mildew fungi, form a hard, lumpy crust on twigs and leaf petioles, and on fruit, they create a spongy crust on the surface. However, a particularly dangerous ailment is stem rot disease, resulting from P. Palmivora, which damages the tree's nutrient transport system in the stem.

A significant aspect of the challenges that arise in agricultural areas can be addressed by computer vision [5]. Traditionally, the detection of plant diseases heavily relied on manual inspections conducted by farmers or laborers, typically with the naked eye (Singh et al., 2017 & Petrellis, 2015). Table I presented traditional disease monitoring procedure for disease management with limitations, such as visual inspection, scouting, weather-based disease forecasting and etcetera. This method can be both laborious and repetitive, especially when dealing with tall Durian trees. However, the advent of artificial intelligence (AI) has revolutionized disease detection in various tree types, including Durian.

TABLE I. TRADITIONAL DISEASE MONITORING PROCEDURE FOR DISEASE MANAGEMENT

Disease Monitoring Procedure	Description	Limitations
Visual Inspection	Regular visual assessment of crops for symptoms of disease.	- Subject to human error and bias. - May miss early or subtle symptoms. - Time-consuming for large fields.
Scouting by Field Observers	Trained personnel systematically inspecting fields for signs of disease.	- Labor-intensive and costly. - Limited coverage and potential variations in observer expertise.
Weather-Based Disease Forecasting	Using weather data to predict disease outbreaks based on favourable conditions for pathogen development.	- Accuracy depends on the quality and availability of weather data. - Doesn't account for all factors affecting disease.
Sampling and Lab Testing	Collecting plant or soil samples for laboratory analysis to identify and confirm disease presence.	- Requires specialized equipment and expertise. - Results may not be available quickly enough for immediate action.
Disease Severity Rating Scales	Assigning numerical scores to rate disease severity, helping quantify disease progression.	- Subjective and dependent on the assessor's judgment. - Can be time-consuming, especially for large areas.
Trap Crops and Indicator Plants	Planting susceptible species near valuable crops to serve as early warning indicators of disease presence.	- May not always provide timely detection. - May require additional land and resources.
Neighbouring Farm Communication	Exchange of information among neighbouring farms about disease outbreaks or observations.	- Relies on the willingness of nearby farmers to share information. - Limited to local awareness.

In agriculture, diseases are a common occurrence across different fruit varieties. When it comes to monitoring fruit diseases, researchers and practitioners often grapple with the challenge of finding a balance between the accuracy of deep learning models and the computational resources necessary for efficient monitoring. To tackle this challenge and enhance both precision and efficiency, various deep learning architectures and techniques have been explored.

Considering there are more pixels in an image than there are words in NLP applications, the use of the attention mechanism in vision applications has been considerably more constrained due to the high computing cost [6]. This means that typical attention models cannot be applied to visuals.

II. RELATED WORKS

Transformer network applications in computer vision were recently reviewed in [7] and vision transformer (ViT) is a major step towards adopting transformer-attention models for computer vision tasks [8]. Compared to CNN-based models that consider picture pixels, using image patches as information units for training is groundbreaking. ViT uses self-attention modules to analyze the relationship between image patches included in a shared region. ViT was demonstrated to outperform CNNs in image classification accuracy given vast quantities of training data and processing resources [8].

State-of-the-art deep learning models can achieve impressive results and are well-suited for drone applications, but they come with a hefty need for computational resources during the training process. In contrast, Vision Transformer (ViT) offers a promising alternative [26]. ViT avoids using Convolutional Neural Networks (CNNs) and performs at a level similar to top-tier CNN models. ViT, a relative of the Transformer model, utilizes a smart technique called self-attention to establish a global reference for each pixel in an image during training. It breaks the image into smaller patches, assigns a position to each patch, and learns from them. In the final layers of the ViT model, the similarity between these patch representations significantly improves. Interestingly, adding more layers to the model doesn't enhance its performance [27]. Nevertheless, ViT does pose a challenge when dealing with high-resolution images due to its four-fold increase in memory requirements, making them more difficult to handle.

Numerous research studies have focused on mitigating the shortcomings of Transformer-based models which tend to fall into two main categories: hybrid models and pure Transformer enhancements. Table II illustrates both hybrid and pure transformer enhancements. Furthermore, by employing the Vision Transformer, researchers achieved a minimum of 1% higher accuracy in classifying cassava leaf diseases than well-known CNN models. They also effectively implemented this model on the Raspberry Pi 4, an edge device, showcasing the substantial potential for its application in the realm of smart agriculture [17]. To the best of our knowledge, only [19] has performed durian disease detection using deep learning approach. The durian disease classification was performed using Resnet-9 and VGG-19 where Resnet-9 was outperformed VGG-19 with accuracy of 100% and 99.11%, respectively. Recent advancements in plant disease detection have seen substantial enhancements using CNN-based models. Nevertheless, these models have limitations such as translation invariance, locality sensitivity, and a lack of comprehensive global image understanding.

TABLE II. HYBRID AND ORIGINAL VISION TRANSFORMER METHOD

Paper	ViT Techniques	Results	Limitations
[9]	Ghost-Enlightened Transformer (GeT)	98.14%	-Relies on large labelled data
[10]	PlantXViT	98.33%	-unable to maintain a lower count of Gega floating operation points.
[11]	Convolution vision Trasnformer (CvT)	87.7%	-higher accuracy will increase training and inference times and memory used.
[12]	Convolution-enhanced image Transformer (CeiT)	99.1%	
[13]	LocalVit	94.2%	
[14]	Swin Transformer	81.3%	-larger resolution needed to increase the accuracy
[15]	k-NN attention (KvT)	73.0%	-need to be paired up as the boosting agent for the vision transformer.
[16]	RegionViT	83.8%	Not stated

To overcome these challenges, this study introduces a novel approach that employs a Vision Transformer-based model for more effective plant disease classification. ViT results will be compared with ResNet-9 and VGG-19 [19] in results and discussion part. This approach combines computer vision and deep learning technologies to revolutionize agricultural production management, utilizing large-scale datasets to address current agricultural issues and improve the overall performance of agricultural automation systems, especially in Durian disease classification, thereby propelling agricultural automation equipment and systems toward a more intelligent future [18].

The paper is organized as follows: Section I presents a brief introduction to the type of durian diseases and current method used to detect the diseases. Section II delves into related works. Section III covers the methodology of ViT and how the experiment conducted. Then, Section IV presents the results and discussion of durian disease detected using ViT and Section V gives the conclusion and future work of the research.

III. Methodology

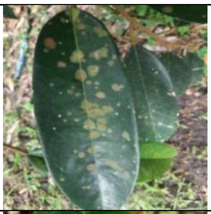
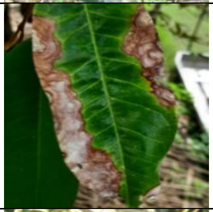
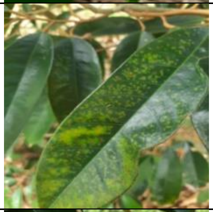
A. Dataset Preparation

In this experimental study, our primary objective was to develop and train a robust deep learning model specifically Vision Transformer capable of accurately classifying diseases that affect durian plants. Our dataset included a total of 1,344 images, which were distributed across four distinct classes. The diseases aimed to precisely classify were ‘durian_leaf_spot’, ‘durian_leaf_blight’, ‘durian_algal_leaf_spot’, and ‘durian_ healthy’, as presented in Table III [20].

To effectively manage the dataset, the original dataset is divided into two sets: training and validation. This was achieved by applying a validation split ratio of 20%, meaning that 80% of the data was designated for training purposes, while the remaining 20% was reserved for validation.

Additionally, to enhance the model's generalization and diversify the training data, data augmentation techniques was implemented. The 'ImageDataGenerator' class is provided by Keras, which facilitated various augmentations of the training data. These augmentations included random rotations of up to 40 degrees, horizontal and vertical shifts of up to 20% of the image dimensions, shearing transformations up to 20%, random zoom adjustments that could expand or contract by up to 20%, and horizontal flipping to create mirror images. When generating new pixels, the 'nearest' method was utilized. This process resulted in the creation of four augmented versions of each original image, significantly expanding the size of the training dataset.

TABLE III. Durian Disease Dataset Extracted From [20]

Dataset	Description	Total number of images	Sample of images
'durian___leaf_spot'	Images of durian leaves with leaf spot disease.	336	
'durian___leaf_blight'	Images of durian leaves with leaf blight disease	336	
'durian___algal_leaf_spot'	Images of durian leaves with algal leaf spot disease	336	
'durian___healthy'	Images of healthy durian leaves	336	

B. Durian Disease Classification using Vision Transformer (ViT)

The ViT model consists of patch creation, patch encoding, multiple Transformer layers, and a final classification head, as shown in Fig. 3.

- Patch creation: Instead of processing entire images at once, the ViT model divides each image into smaller non-overlapping patches or tiles. This patch-based approach allows the model to process large images efficiently. Each patch is treated as a separate input and processed independently by the model.
- Patch Encoding: After splitting the image into patches, each patch is encoded into a numerical representation that the model can work with. This process typically involves linearly projecting the patch's pixel values into a lower-dimensional vector, allowing the model to learn spatial relationships and features within each patch.
- Multiple Transformer Layers: The heart of the ViT model consists of multiple Transformer layers. These layers process the encoded patches and capture contextual information, enabling the model to understand how different patches relate to one another. The self-attention mechanism in the Transformer architecture is particularly crucial for this step, as it helps the model weigh the importance of different patches when making predictions.
- Final Classification Head: At the end of the ViT model, there is a classification head. This part of the model takes the information from the previous Transformer layers and makes predictions based on the features learned during the earlier stages of processing. For tasks like image classification, this is where the model assigns labels or probabilities to different classes.

The ViT model employs two optimizers, Adam and SGD (Stochastic Gradient Descent), which include a regularization technique called weight decay. Weight decay is a regularization method used to prevent overfitting in deep learning models. It works by adding a penalty term to the loss function during training, encouraging the model to have smaller weight values. Smaller weights can make the model more robust and less prone to overfitting.

To improve training efficiency, a learning rate schedule is defined. In this schedule, the learning rate, which determines how much the model's parameters are updated during training, is reduced by 50% every 10 training epochs. This gradual reduction in learning rate is a common strategy to help the model converge to a good solution without making overly large updates to its weights, which can cause instability.

During training, the ViT model periodically saves its current state as checkpoints. These checkpoints capture the model's parameters, allowing you to resume training from where you left off or use the model for inference. The saving of model checkpoints is typically based on validation accuracy, meaning that the model's performance on a separate validation dataset is used as a criterion to determine when to save a checkpoint. This ensures that the saved models are based on their ability to generalize to unseen data.

Furthermore, the training data is divided into two parts: the main training data (90%) and a validation set (10%). The main training data is used to train the ViT model, while the validation set is used to monitor the model's performance during training. This split is important for assessing how well the model is learning and for tuning hyperparameters like the learning rate. After the model has been trained, it is evaluated on a separate test dataset that the model has never seen during training. This evaluation assesses the model's performance on unseen data and provides an indication of how well it can generalize to real-world scenarios. The evaluation reports two metrics: accuracy, which measures the overall correctness of predictions, and top-5 accuracy, which indicates how often the correct label is among the top five predicted labels.

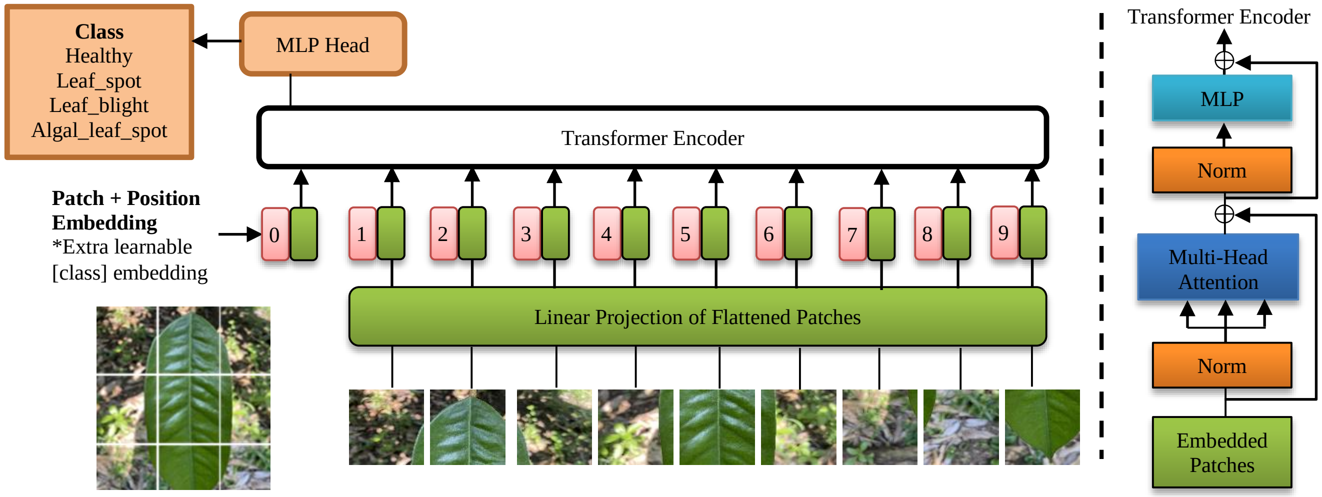


Fig. 3. Vision transformer architecture.

IV. Results and Discussion

In this study, ViT has been deployed and fine-tuned it to perform Durian Disease classification using two well-known optimization algorithms, namely, Stochastic Gradient Descent (SGD) and ADAM optimizer. The experiments encompassed a range of hyperparameters, including different learning rates (0.001, 0.005, and 0.01) and various epoch settings (20, 30, and 40). The outcomes provide valuable insights into how the model's performance varies with alterations in these key hyperparameters, shedding light on the most effective configurations for the task.

As revealed in Table IV, a validation accuracy of 94.12% was attained with the utilization of the ADAM optimizer, a maximum learning rate of 0.01, and an extended training period of 400 epochs. In contrast, Table V displays the outcomes with the SGD optimizer, where a comparable learning rate and epoch setting yielded an accuracy of 85.82%. ADAM's superior performance in this context can be attributed to its adaptability and the fusion of techniques from both momentum and RMSprop optimization. ADAM excels in scenarios involving intricate loss surfaces and fluctuating learning rates. It dynamically tailors the learning rate for each parameter, guided by historical gradient information. This adaptability often results in swifter convergence and enhanced generalization capabilities. In contrast, SGD adheres to a more conventional optimization approach. Achieving parity with ADAM's performance, particularly with complex models like ViT, often necessitates manual fine-tuning of the learning rate and other hyperparameters.

In many cases, the maximum learning rate of 0.01 might lead to faster convergence but might also make the model diverge or not settle into the optimal solution. By reducing the maximum learning rate during training (e.g., from 0.01 to 0.001), the model is allowed to fine-tune and reach a more stable and accurate solution. This phenomenon can be observed where both optimizers are performed well with maximum learning rate of 0.001 compare to 0.01.

Certainly, our training approach involved the incorporation of a learning rate schedule to foster training stability and mitigate overfitting. Simultaneously, data augmentation proved instrumental in enhancing the model's robustness by enabling it to adapt to a broader range of image conditions. Techniques such as resizing, flipping, rotation, and zooming effectively contributed to this augmentation strategy.

However, in Table VI, ResNet-9 is outperformed the other two methods, ViT and VGG-19. When evaluating why ViT might not be as good as the accuracy of ResNet-9 and VGG-19 [19], various factors come into play. First, ResNet-9 and VGG-19, as convolutional neural networks (CNNs), have been explicitly tailored for image classification tasks, boasting a proven track record in this domain. ViT, on the other hand, is a relatively newer architecture that demands substantial fine-tuning for optimal performance.

Additionally, ViT models often require more extended training schedules, specialized initialization methods, and specific architectural considerations, necessitating higher computational resources. Moreover, ViT's capacity to generalize might be challenged when confronted with smaller datasets, as its architecture is not as well-suited to such scenarios. Comparatively, ResNet-9 and VGG-19 [19] models tend to deliver robust performance with limited data, given their established history in this context.

Furthermore, the availability of pretrained weights customized for specific tasks can provide a performance advantage to ResNet-9 and VGG-19 over ViT. It's important to note that ViT is relatively more susceptible to overfitting, especially in cases involving smaller datasets or exceptionally large models. In addition to these considerations, our dataset for durian leaf disease classification is relatively small, posing a challenge for ViT's generalization capabilities. Addressing this issue would necessitate the utilization of larger models, fine-tuning with different hyperparameters, and more extensive tuning.

TABLE IV. VIT WITH ADAM OPTIMIZER

Maximum Learning Rate	Epochs	Train Loss	Validation Loss	Validation Accuracy
0.001	100	0.0647	0.2388	0.8447
0.001	200	0.0167	0.1010	0.9118
0.001	400	0.0400	0.0873	0.9412
0.005	100	0.7316	0.7451	0.6471
0.005	200	0.7406	0.7898	0.6765
0.005	400	0.7406	0.7898	0.7353
0.01	100	1.0367	1.1862	0.4706
0.01	200	1.0376	1.1377	0.4982
0.01	400	1.0270	1.0367	0.5010

TABLE V. VIT WITH SGD OPTIMIZER

Maximum Learning Rate	Epochs	Train Loss	Validation Loss	Validation Accuracy
0.001	100	0.1200	0.2689	0.8009
0.001	200	0.0951	0.2564	0.8511
0.001	400	0.0894	0.2416	0.8582
0.005	100	0.8766	0.9394	0.5843
0.005	200	0.8105	0.8324	0.6056
0.005	400	0.7791	0.8289	0.6082
0.01	100	1.1245	1.3457	0.3943
0.01	200	1.1369	1.3363	0.3973
0.01	400	1.2046	1.2298	0.4085

TABLE VI. COMPARISON OF VIT, RESNET-9 AND VGG-19 [19] USING ADAM AND SGD OPTIMIZER

Maximum	Adam Optimizer			SGD optimizer		
Learning Rate	VGG-19	ResNet-9	VIT	VGG-19	ResNet-9	VIT
0.001	0.8271	0.9521	0.8447	0.8542	0.8113	0.8009
0.001	0.8500	0.9797	0.9118	0.8542	0.8447	0.8511
0.001	**1.0000**	**0.9911**	**0.9412**	0.8875	0.8896	0.8582
0.005	0.8438	0.9667	0.6471	0.8708	0.8081	0.5843
0.005	0.8708	0.9792	0.6765	0.8771	0.8447	0.6056
0.005	0.8812	0.9792	0.7353	0.8708	0.8792	0.6082
0.01	0.8500	0.9729	0.4706	0.8542	0.7934	0.3943
0.01	0.8604	0.9896	0.4982	0.8542	0.8073	0.3973
0.01	0.8438	0.9896	0.5010	0.8812	0.8358	0.4085

V. CONCLUSION AND FUTURE WORKS

This research represents a significant advancement in the field of durian leaf disease detection and recognition, addressing a critical issue faced by durian farmers in ASEAN countries and beyond. The traditional manual identification of leaf diseases has been a labor-intensive and time-consuming process, posing substantial challenges to the agricultural sector's sustainability. Through the application of cutting-edge deep learning techniques and the utilization of well-established models like ViT, an automated system has been successfully developed and capable of accurately classifying and recognizing durian leaf diseases. Notably, our results demonstrate the remarkable performance, achieving an impressive accuracy rate of 94.12% when utilizing the Adam optimizer. Moreover, this research underscores the broader implications of utilizing cutting-edge machine learning techniques in agriculture. It opens the door to the development of precision agriculture systems that can revolutionize crop management practices. The implementation of ViT-based disease control not only safeguards the economic stability and food security of the Southeast Asian region but also paves the way for further advancements in the field of agriculture. With technology as a key ally, durian farmers and the agricultural sector as a whole are better equipped to overcome the challenges posed by disease and secure a more prosperous future.

ACKNOWLEDGMENT

This work was carried out with the financial support from the Ministry of Higher Education of Malaysia under the research grant LRGS/1/2019/UKM-UKM/5/2.

REFERENCES

[1] Y. Ahmad, "TECHNICAL OP-ED: Durian cultivation faces serious threat from disease caused by Phytophtora sp.," https://www.itfnet.org/v1/2018/12/management-of-phytophtora-in-durian/. (accessed on 20 Oktober 2023).

[2] T. M. UN Comtrade, "Durian Global Market Report 2018," http://www.plantationsinternational.com/ docs/durian-market.pdf. (accessed on 20 Oktober 2023).

[3] OEC, "Fruit, edible: durians, fresh," https://oec.world/en/profile/hs/fruit-edible-durians-fresh#:~:text=Exporters%20and%20Importers&text=In%202021%2C%20the%20top%20exporters,and%20Laos%20(%249.05M) (accessed on 20 Oktober 2023).

[4] 27Group, "Technology Advancement in the Durian Industry. ," https://27.group/technology-advancement-in-the-durian-industry/ (accessed on 22 Oktober 2023).

[5] M. K. Tripathi and D. D. Maktedar, "A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey," Information Processing in Agriculture, vol. 7, no. 2, pp. 183–203, 2020, doi: https://doi.org/10.1016/j.inpa.2019.07.003.

[6] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, "Vision Transformers For Weeds and Crops Classification Of High Resolution UAV Images," CoRR, vol. abs/2109.02716, 2021, [Online]. Available: https://arxiv.org/abs/2109.02716.

[7] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," Jan. 2021, doi: 10.1145/3505244.

[8] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations, 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy.

[9] X. Lu et al., "A hybrid model of ghost-convolution enlightened transformer for effective diagnosis of grape leaf disease and pest," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 5, pp. 1755–1767, 2022, doi: https://doi.org/10.1016/j.jksuci.2022.03.006.

[10] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT." 2022.

[11] H. Wu et al., "CvT: Introducing Convolutions to Vision Transformers." 2021.

[12] K. Yuan, S. Guo, Z. Liu, A. Zhou, F. Yu, and W. Wu, "Incorporating Convolution Designs into Visual Transformers." 2021.

[13] Y. Li, K. Zhang, J. Cao, R. Timofte, and L. Van Gool, "LocalViT: Bringing Locality to Vision Transformers." 2021.

[14] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." [Online]. Available: https://github.

[15] X. and W. F. and L. M. and C. S. and L. H. and J. R. Wang Pichao and Wang, "KVT: k-NN Attention for Boosting Vision Transformers," in Computer Vision – ECCV 2022, G. and C. M. and F. G. M. and H. T. Avidan Shai and Brostow, Ed., Cham: Springer Nature Switzerland, 2022, pp. 285–302.

[16] C.-F. Chen, R. Panda, and Q. Fan, "RegionViT: Regional-to-Local Attention for Vision Transformers." Oct. 2021.

[17] H.-T. Thai, N.-Y. Tran-Van, and K.-H. Le, "Artificial Cognition for Early Leaf Disease Detection using Vision Transformers," in 2021 International Conference on Advanced Technologies for Communications (ATC), 2021, pp. 33–38. doi: 10.1109/ATC52653.2021.9598303.

[18] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation —A review," Information Processing in Agriculture, vol. 7, no. 1, pp. 1–19, 2020, doi: https://doi.org/10.1016/j.inpa.2019.09.006.

[19] M. M. Daud, A. Abualqussan, F. A. N. Rashid and M. H. M. Saad, "Durian disease classification using transfer learning for disease management system," Journal of Information System and Technology management (JISTM), 2023.

[20] Roboflow. (2022). Durian Diseases Image Dataset. Retrieved from https://universe.roboflow.com/new-workspace-7ly0p/durian-diseases/dataset/1.