BIS INFORMATION TECHNOLOGY AND COMPUTER SCIENCE

Volume 3 2026 | DOI: 10.31603/bistycs.523
7th Borobudur International Symposium on Science and Technology (BIS-STE) 2025

UNIMMA PRESS
UNIVERSITAS MUHAMMADIYAH MAGELANG

Performance optimization VGG16 architecture a lightweight hybrid transformer integrated with convolutional neural network for disease classification of porang plant leaves

Fauzan Masykur[1*], Angga Prasetyo[1], Arief Rahman Yusuf[1], Ellisia Kumalasari[1], Sugianti[1], Indah Puji Astuti[1], and Rifqi Rahmatika Az-Zahra[1]

[1] Muhammadiyah University of Ponorogo, Ponorogo, Indonesia
[*]Corresponding author's email: fauzan@umpo.ac.id

Abstract

The porang plant (*Amorphophallus Muelleri Blume*) has high economic value with export demand reaching 6,064,947 kg worth IDR 297 billion in 2019. However, the productivity of porang plants in Indonesia is seriously threatened by disease attacks that can reduce harvest yields by 23-48%. Conventional disease detection methods require a long time and specialized expertise, while economic losses due to plant diseases reach trillions of rupiah annually. The development of an artificial intelligence-based automatic detection system is a critical solution to support national food security and increase the export competitiveness of Indonesian porang commodities. This study aims to develop a classification system for porang plant leaf diseases using the VGG16 architecture on a convolutional neural network to achieve detection accuracy above 95%. The developed system will integrate transfer learning technology to optimize model performance with limited datasets. This research also aims to design an algorithm capable of identifying multiple types of porang leaf diseases in real-time with a high level of precision to support farmer decision-making. The main output is a Based on the empirical results presented, the hybrid transformer model (98%) clearly outperforms the VGG16 model (95%) in the classification accuracy metric. Performance improvement, the 3% advantage of the hybrid transformer model indicates that combining the local feature extraction capabilities of CNNs. This research will produce a standardized dataset of porang leaf diseases for the development of this system, which is expected to reduce crop losses by up to 30% and increase the efficiency of plant disease diagnosis from days to minutes.

Keywords

VGG16, CNN, Hybrid transformer, Plant disease, Amorphophallus mueller

Published:
May 04, 2026

Selection and Peer-review under the responsibility of the 7th BIS-STE 2025 Committee

Introduction

The Porang plant (*Amorphophallus muelleri Blume*) is a commodity classified into the Araceae family. This native Indonesian shrub (*herbaceous*) plant is characterized by a single tuber located in the soil [1]. Initially, Porang tubers were sold without processing, which resulted in low economic value. However, currently Porang has become a high-value commodity along with the discovery of its various benefits, including its role in preventing and reducing disease, as well as its use as an important raw material in the beauty industry and as an alternative food ingredient to replace rice [2].

Porang cultivation is often faced with the problem of attacks by various diseases that attack the leaves of the plant. The main factors that influence the emergence of this disease include weather conditions, environmental conditions around the plant, and the quality of the seeds used. Some examples of common leaf diseases are leaf spots or rust (*Puccinia Arachidis*) and white mottled leaves caused by Konjac Mosaic virus infection [3].

Diseases in Porang plants, if not adequately treated, have the potential to significantly reduce tuber production, resulting in economic losses for farmers. This condition requires the need for accurate and rapid disease diagnosis to ensure effective treatment. The use of technology in this field is carried out by building a deep learning system to detect Porang leaf diseases. previous research initial system developed using a Convolutional Neural Network (CNN) with the VGG16 model has been trained and is able to detect disease types based on leaf images [4]. Although it has achieved 81% accuracy, this accuracy still needs to be improved to 90%. This improvement effort is planned through the implementation of the Lightweight Hybrid Transformer System by adding 50 hidden layers [5].

The objective of this research is to propose the development of a system that facilitates Porang farmers in identifying and categorizing leaves into five classes: healthy (normal) leaves, leaf blight, leaf spots, and striped leaves (konjac mosaic). The recognition process utilizes a CNN algorithm combined with the Lightweight Hybrid Transformer, a deep learning technique that is very suitable for digital image processing applications [6], [7]. The main contributions offered by this research are as follows:

1. Providing an image dataset of diseased Porang leaves, obtained from a limited amount of primary data.
2. Application of image augmentation techniques to overcome data limitations. Augmentation is carried out to ensure the availability of adequate image data for the Hybrid Transformer model, thereby increasing the model's generalization capabilities and preventing overfitting.

Method

The initial stage in this research was to explore datasets in the form of images of Porang plant leaves attacked by pests and diseases, which were collected directly from the cultivation locations (gardens and fields) of Porang farmers. The data samples were divided into five types of classes, namely: leaf rot (late blight), brown spots (early blight), white spots (konjac mosaic), insect pests, and healthy leaves. A total of 1,500 Porang leaf images were collected and classified into these five categories, with each class including between 100 and 200 images for the training process (as documented in the Figure 1). Considering that the method used is supervised learning, images of various sizes and resolutions must be labeled according to their class. In addition, augmentation techniques were applied to increase the amount and variety of data without changing the existing dataset classification [8].

Figure 1. Leaves attacked diseases

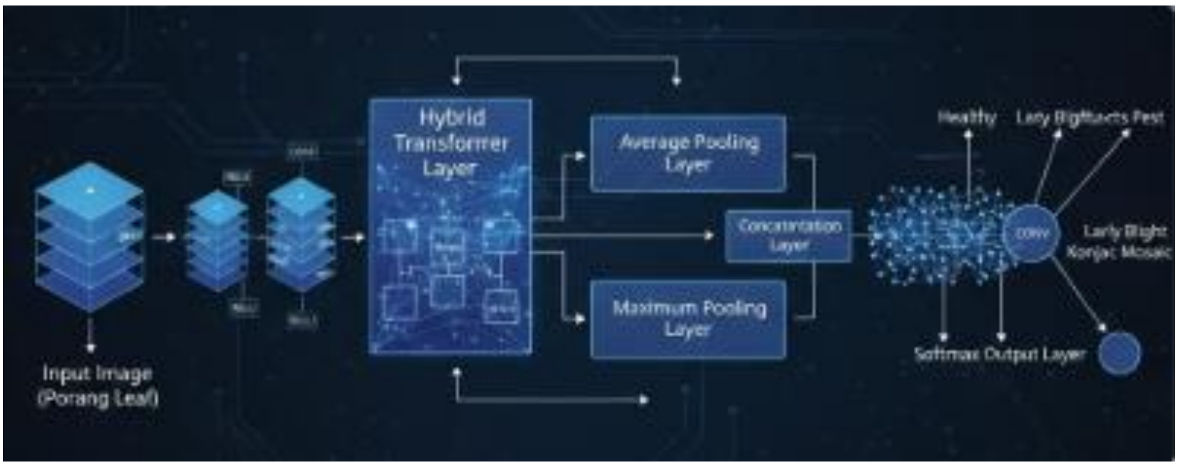

Figure 2. Hybrid transformer architecture

To overcome this (as documented Figure 2), the hybrid transformer model is proposed. This model excels due to its ability to model global dependencies and long-distance relationships between data elements using the Self-Attention mechanism. In this architecture, the features extracted by CNN are split (patched) and fed into the transformer encoder layer. The Transformer then processes these feature patches to understand the correlations between image parts globally, resulting in a much more contextual representation [9].

The training process begins with the initialization of the main parameters, namely: Number of epochs Set in the range of 20 to 75 epochs, Step per epoch: Adjusted based on the amount of training and testing data. Batch size: Set at 32. For optimization, the Adam optimizer is used with the Sparse Categorical Cross Entropy (SCCE) loss function, and the metric used is accuracy. In addition, validation data is adjusted to the specific needs of ongoing research [2], [10].

$$Vaug = \frac{1}{Wf - Hf} \sum_{w=1}^{Wf} \sum_{h=1}^{Hf} Fw,h$$

Global Average Pooling CNN (VAug); N = (W F / p) x (H F / p) where p is the patch size); d = dept; w = width; R = resolution. The steps for multi-class classification in deep learning include adding commonly used loss functions, Sparse Categorical Cross Entropy (SCCE) and Categorical Cross Entropy (CCE). These two functions calculate the loss based on the probability distribution generated by the model for each class [11], [12], [13].

The next step is to implement Adaptive Moment Estimation (Adam) as the optimization algorithm. Adam is a combination of RMSprop and SGD with Momentum; it leverages RMSprop for an adaptive learning rate per parameter, and Momentum for adjusting weights based on gradients. This combination significantly accelerates convergence and optimizes model performance [6], [7], [12], [14], [15].

Results and Discussion

Results this study evaluates and compares the testing accuracy between custom CNN, VGG16, and hybrid transformer models trained on image datasets for the classification of pests and diseases of porang plants.

The dataset used for this research is the image of porang plant leaves, which includes original data and augmented data. During the training process, early stopping and callback mechanisms were implemented. Performance results, including accuracy, loss, and training time per epoch, are calculated and presented in separate tables to compare the performance between the original dataset and the augmented dataset. The best performance values achieved during training, based on relevant performance criteria, are detailed in Table 1.

Model calculation Table 1, this study focuses on training initial data using two model approaches: custom CNN and VGG16 model, hybrid transformer. In this phase, key parameters, such as learning rate, number of epochs, and number of fully connected layers, are adjusted to find the optimal configuration.

On the original dataset, the average accuracies of all models are very close (as documented in the Figure 3 and Figure 4). Nevertheless, the VGG16 model demonstrates superior performance, providing the best average accuracy across all classes. Furthermore:

1. The average predictive ability of the models for classes other than the primary target class is relatively similar across all models.
2. The precision (correct positive classification rate) of the Vgg16 model is the highest.
3. The training time per epoch of Vgg16 is shorter than that of the other models, at only milliseconds (ms).

Based on Table 1, the test results for disease and pest detection and classification in Porang plants indicate that the Hybrid Transformer Model achieved the best accuracy.

Table 1. Comparative calculation model

Model Basic CNN						
Epoch	**VGG 16**			**Hybrid Transformer**		
	Loss	**Acc**	**Time(ms)**	**Loss**	**Acc**	**Time(ms)**
24	40	75	126	61	80	22
28	27	50	133	45	93	24
33	56	66	144	33	96	23
42	47	72	132	33	83	24
46	43	58	128	172	96	24

Table 2. VGG16 classification performance on dataset

	Precision	***Recall***	***F1-Score***	***Support***
(0)	1.00	1.00	1.00	24
(1)	1.00	0.92	0.96	23
(2)	0.95	1.00	0.97	14
(3)	0.95	1.00	0.97	19
(4)	0.96	0.96	0.94	21
accuracy			0.95	95
macro avg	0.95	0.95	0.95	95
weighted avg	0.95	0.95	0.95	95

Table 3. Hybrid transformer performance on dataset

	Precision	***Recall***	***F1-Score***	***Support***
(0)	1.00	1.00	1.00	24
(1)	1.00	0.96	0.96	23
(2)	0.98	1.00	0.97	14
(3)	0.97	1.00	0.97	19
(4)	0.96	0.97	0.98	21
accuracy			0.98	96
macro avg	0.98	0.97	0.98	97
weighted avg	0.98	0.97	0.97	97

Table 2 and Table 3 show the precision, recall, and f1-score values for each class on two datasets, namely the original dataset and the dataset that has been augmented using the VGG16 model. It can be seen that this model has the best performance on the augmented dataset, with precision values between 95% and 100%, while the hybrid transformer model has the best performance on the original dataset, with precision values between 98% and 100%. This model is considered successful in classifying leaf diseases of porang plants.

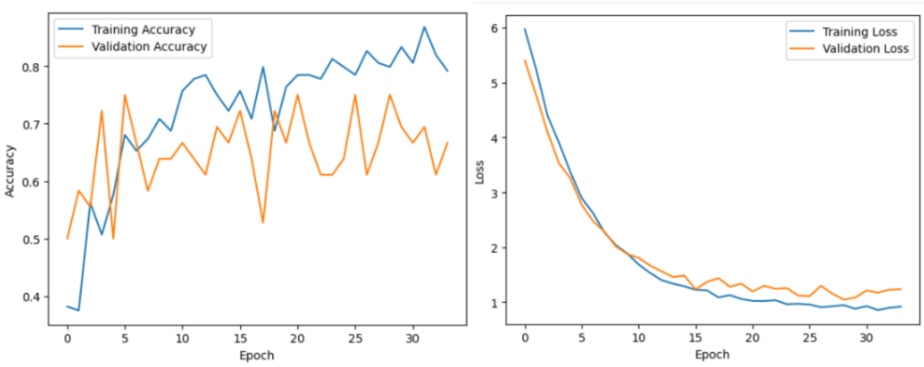

Figure 3. VGG16 training accuracy and training loss

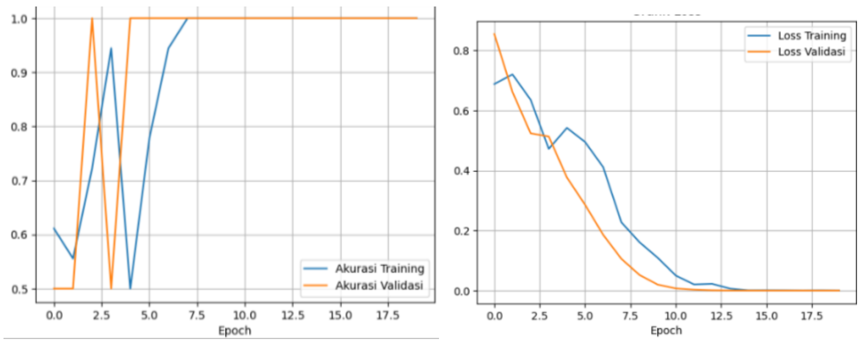

Figure 4. Hybrid transformer training accuracy and training loss

The empirical results in Table 4 show a significant performance difference between the VGG16 (Conventional CNN) and the hybrid transformer models in image classification tasks. The VGG16 model, which focuses on local features and spatial hierarchy through convolution operations, achieves a strong accuracy of 95%. This performance demonstrates the effectiveness of classical CNN architecture in extracting basic visual patterns. However, the hybrid transformer model, which integrates the local feature extraction capabilities of CNN with the self-attention mechanism of transformer to process long-range dependencies and global context, managed to achieve a superior accuracy of 98%. This 3% performance improvement empirically demonstrates that the Transformer's ability to effectively model relationships between distant image regions effectively compensates for the limitations of CNNs that tend to focus on locality. Although the hybrid transformer exhibits significantly superior performance, it should be noted that this model generally has higher computational complexity (mainly due to the intensive self-attention mechanism), an important consideration in terms of deployment efficiency and resources. In conclusion, for maximizing accuracy, the Hybrid Transformer architecture proves superior, while the VGG16 offers a better balance between solid performance and lower computational complexity.

Table 4. Significant performance

Model	accuracy	Precision	Recall	F1-Score	Support
VGG16	95%	95%	95%	95%	95%
Hybrid transformer	98%	97%	97%	98%	97%

Conclusion

This study shows that the integration of CNN with the hybrid transformer model is a successful approach for the classification of porang plant leaf diseases. Based on the empirical results presented, the Hybrid Transformer model (98%) clearly outperforms the VGG16 model (95%) in the classification accuracy metric.

1. Performance improvement: The 3% advantage of the hybrid transformer model indicates that combining the local feature extraction capabilities of CNNs (such as VGG16) with the transformer self-attention mechanism (which excels at capturing global context) is very effective in improving image classification performance.
2. Goal implication: If the main goal is to maximize classification accuracy, then the Hybrid Transformer architecture is a better choice. However, this advantage may come at the cost of higher computational and memory costs compared to VGG16.

These results confirm the superior efficacy of the hybrid transformer architecture for disease detection and classification tasks in an agricultural context.

Acknowledgement

The research team would like to express its deepest gratitude to the Ministry of Research, Technology, and Higher Education (Kemenristekdikti) of the Republic of Indonesia for the support and trust given through the Fundamental Research Grant Scheme for the 2025 Fiscal Year.

References

1. D. Rustandi, S. H. Wijaya, Mushthofa, and R. Damayanti, "Anatomy Identification of Bamboo Stems with The Convolutional Neural Networks (CNN) Method," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 8, no. 1, Art. no. 1, Jan. 2024, doi: 10.29207/resti.v8i1.5370.
2. M. Zuhan and Y. Kristian, "Detection of Porang Plant Diseases and Pests (Amorphophallus Muelleri) Based on Leaf Imagery Utilizing DCNN Transfer Learning," *J. Sist. Telekomun. Elektron. Sist. Kontrol Power Sist. Dan Komput.*, vol. 3, no. 2, pp. 129–140, July 2023, doi: 10.32503/jtecs.v3i2.3709.
3. H. O. Velesaca, R. Mira, P. L. Suárez, C. X. Larrea, and A. D. Sappa, "Deep Learning based Corn Kernel Classification," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2020, pp. 294–302. doi: 10.1109/CVPRW50498.2020.00041.
4. "Analysis for Detecting Banana Leaf Disease Using the CNN Method | JUITA: Jurnal Informatika." Accessed: June 02, 2025. [Online].
5. A. Prasetyo, F. Masykur, and A. R. Yusuf, "Deteksi Gula Aren dengan CNN Berbasis AlexNet sebagai Solusi Keamanan Pangan untuk Penderita Diabetes," *SinarFe7*, vol. 7, no. 1, pp. 302–309, Aug. 2025.
6. F. Masykur, A. Prasetyo, I. A. Zulkarnain, E. Kumalasari, and P. Utomo, "Yolo-Drone: Detection Paddy Crop Infected Using Object Detection Algorithm Yolo and Drone Image," *JOIV Int. J. Inform. Vis.*, vol. 9, no. 5, pp. 2104–2111, Sept. 2025, doi: 10.62527/joiv.9.5.3472.
7. A. Prasetyo *et al.*, "Analisis Deteksi Citra Mata Ikan Nila dengan Metode Convolutional Neural Network Arsitektur Alexnet," *J. Pustaka Data Pus. Akses Kaji. Database Anal. Teknol. Dan Arsit. Komput.*, vol. 5, no. 1, pp. 48–53, June 2025, doi: 10.55382/jurnalpustakadata.v5i1.995.
8. S. Dwianto, F. N. Mubarok, D. Satriatama, and T. Agustin, "Penerapan Convolutional Neural Network (CNN) dalam Deteksi Penyakit pada Tanaman Terong," *Pros. Semin. Nas. Amikom Surak.*, vol. 2, pp. 270–280, Dec. 2024.
9. L. Fitriani, D. Tresnawati, and M. B. Sukriyansah, "Image Classification On Garutan Batik Using Convolutional Neural Network with Data Augmentation," *JUITA J. Inform.*, pp. 107–115, May 2023, doi: 10.30595/juita.v11i1.16166.

10. N. Helmawati and E. Utami, “Analysis for Detecting Banana Leaf Disease Using the CNN Method,” *JUITA J. Inform.*, pp. 29–36, Mar. 2025, doi: 10.30595/juita.v13i1.24514.
11. V. D. T. Wijaya, S. Novianto, and U. Rosyidah, “DETEKSI HURUF ARAB MENGGUNAKAN METODE FREEMAN CHAIN CODE,” *Techno.Com*, vol. 13, no. 4, Art. no. 4, 2014, doi: 10.33633/tc.v13i4.606.
12. W. N. Waluyo, R. R. Isnanto, and A. F. Rochim, “Comparison of Mycobacterium Tuberculosis Image Detection Accuracy Using CNN and Combination CNN-KNN,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 7, no. 1, Art. no. 1, Feb. 2023, doi: 10.29207/resti.v7i1.4626.
13. A. TiaraSari and E. Haryatmi, “Penerapan Convolutional Neural Network Deep Learning dalam Pendeteksian Citra Biji Jagung Kering,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 2, Art. no. 2, Apr. 2021, doi: 10.29207/resti.v5i2.3040.
14. Y. Aufar, M. H. Abdillah, and J. Romadoni, “Web-based CNN Application for Arabica Coffee Leaf Disease Prediction in Smart Agriculture,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 7, no. 1, Art. no. 1, Feb. 2023, doi: 10.29207/resti.v7i1.4622.
15. R. Shen, T. Zhen, and Z. Li, “Segmentation of Unsound Wheat Kernels Based on Improved Mask RCNN,” *Sensors*, vol. 23, no. 7, Art. no. 7, Jan. 2023, doi: 10.3390/s23073379.