Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.58, p. 411-418. March/April, 2026.

DOI: https://doi.org/10.5935/jetia.v12i58.3096

RESEARCH ARTICLE OPEN ACCESS

DEEP LEARNING APPROACH FOR OIL PALM LEAF DISEASE CLASSIFICATION USING VGG16 ENHANCED WITH ADAM OPTIMIZATION

Hafiz Irsyad[1], Muhammad Rizky Pribadi*[2], Indrajani Sutedja[3] and Yulistia[4]

[1, 2]Informatics, Universitas Multi Data Palembang, Palembang, Indonesia.
[3]Information System, Universitas Bina Nusantara, Jakarta, Indonesia.
[4]Information System, Universitas Multi Data Palembang, Palembang, Indonesia.

[1]https://orcid.org/0000-0003-4837-4581, [2]https://orcid.org/0000-0001-7440-9667
[3]https://orcid.org/0000-0002-6356-3957, [4]https://orcid.org/0000-0003-1099-2085

Email: hafizirsyad@mdp.ac.id, *rizky@mdp.ac.id, indrajani@binus.ac.id, yulistia@mdp.ac.id

ARTICLE INFO

Article History
Received: December 10, 2025
Reviewed: January 1, 2026
Accepted: January 7, 2026
Published: March 31, 2026

Keywords:
Adam optimizer,
CNN,
Deep learning,
Oil palm leaf disease,
VGG16,
YOLOv11.

ABSTRACT

Early detection of oil palm leaf diseases is essential to minimize economic losses and ensure sustainable plantation management. This study proposes a deep learning approach using the VGG16 architecture optimized with the Adam algorithm to classify oil palm leaves into three categories: healthy, infected, and initial infection. The dataset was obtained from Roboflow and preprocessed through cropping, annotation, and standardization before being split into training, validation, and testing sets. Experimental evaluations were performed across multiple training epochs and compared against two baseline models: a shallow Convolutional Neural Network (CNN) and YOLOv11-CLS (version S). Results show that VGG16 combined with Adam achieved the highest accuracy of 97% at 25 and 50 epochs, with balanced precision and recall across all classes. In contrast, the baseline CNN reached a maximum accuracy of 88%, while YOLOv11-CLS produced fluctuating results with a peak accuracy of 82%. Statistical significance testing confirmed that the performance improvements of VGG16 + Adam were consistent and reliable, validating its suitability for practical implementation in precision agriculture. These findings highlight the potential of combining deep architectures with adaptive optimization to enhance disease diagnosis in oil palm plantations and reduce reliance on manual inspections.

I. INTRODUCTION

Oil palm is a key agricultural commodity and the most efficient source of vegetable oil compared to other crops, making it a vital contributor to Indonesia's economy. In 2023, palm oil exports reached USD 25.61 billion with a volume of 38.23 million tons, while total production stood at 46.9 million tons, reinforcing Indonesia's position as the world's leading producer [1]. However, oil palm cultivation faces serious threats from plant diseases, which can significantly reduce yield and even cause tree mortality. The most devastating is basal stem rot caused by *Ganoderma boninense*, responsible for annual economic losses of approximately USD 500 million [2]. Other damaging diseases include *Curvularia* leaf spot (*Curvularia oryzae*) and *Pestalotiopsis* leaf blight (*Pestalotiopsis palmarum*), with incidence rates reaching up to 11.26% [3].

These diseases, along with pest infestations, not only reduce agricultural productivity but also cause considerable ecological losses. Early detection and prevention are therefore essential for sustainable plantation management. Traditional disease diagnosis methods relying on manual visual inspection are often inefficient, time-consuming, and costly [4], [3]. Recent advances in artificial intelligence (AI) have opened new possibilities for oil palm management, particularly through deep learning, which excels in processing visual imagery for tasks such as disease classification and tree counting [5]. While deep learning has been applied in oil palm plantations for tree detection[6-8], its application to leaf disease diagnosis remains limited.

In contrast, similar approaches have been successfully implemented in other crops, including rice [9] guava [10] , cucumber [11], and tomato [12]. Building on this potential, the present study aims to classify the health condition of oil palm trees from leaf images, categorizing them as either healthy or unhealthy. The findings are expected to contribute to modern precision agriculture by alleviating farmers' workloads and reducing unnecessary production expenses. This study aims to classify the health status of oil palm leaves as healthy or unhealthy using the VGG16 architecture, chosen for its proven reliability in image classification, especially in recognizing plant leaf patterns[13]. Its deep, multi-layered structure enables comprehensive feature extraction. Although the model contains a large number of parameters, VGG16 remains effective for limited datasets when combined with data augmentation techniques and fine-tuning [14].

In comparison, ResNet50 utilizes residual connections to prevent accuracy degradation in very deep networks, whereas VGG16's simpler and more stable structure makes it easier to implement in field applications. InceptionV3 integrates multiple kernel sizes within a single layer to achieve computational efficiency and high accuracy; however, its architecture is more complex and requires more precise parameter tuning. Meanwhile, MobileNetV2 is lightweight and efficient for devices with limited resources, but this efficiency comes at the cost of lower accuracy compared to VGG16, particularly in high-resolution image classification. Considering accuracy, stability, and ease of implementation, VGG16 represents a relevant and practical choice for supporting early diagnosis of oil palm leaf diseases through an image classification approach.

To address the limitations of VGG16, such as its large number of parameters and the potential for slow convergence during training, this study adopts an appropriate optimizer to enhance the model's efficiency and accuracy. Adam was selected for its ability to adaptively adjust the learning rate for each parameter while combining the advantages of Momentum and RMSProp, thereby accelerating the training process and helping the model avoid local minima an essential factor in classifying oil palm leaf diseases with complex visual patterns. Compared to Stochastic Gradient Descent (SGD), which requires manual learning rate tuning and tends to produce large fluctuations in the loss function, Adam offers more stable and faster training. Meanwhile, although Adagrad is also adaptive, it often experiences a significant decrease in learning rate over time, which can hinder long-term learning. Therefore, Adam is considered the most suitable optimizer to maximize the performance of VGG16 in this study.

II. METHOD

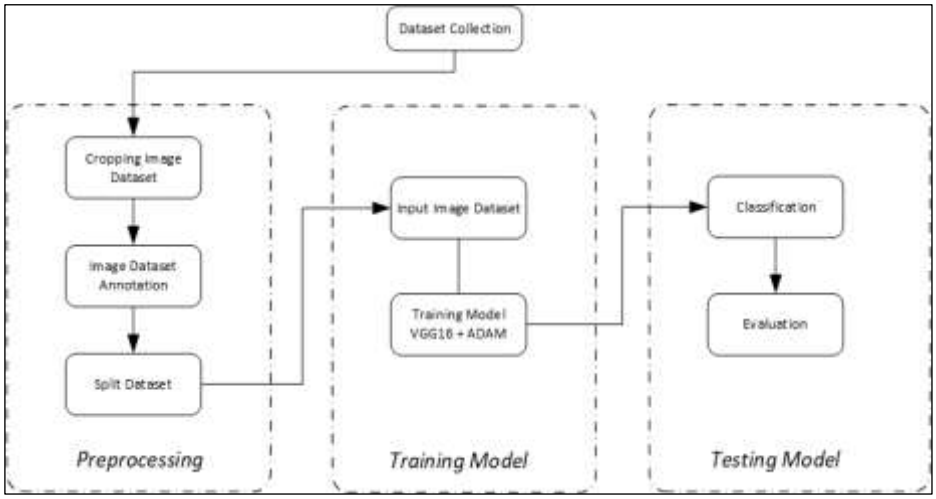


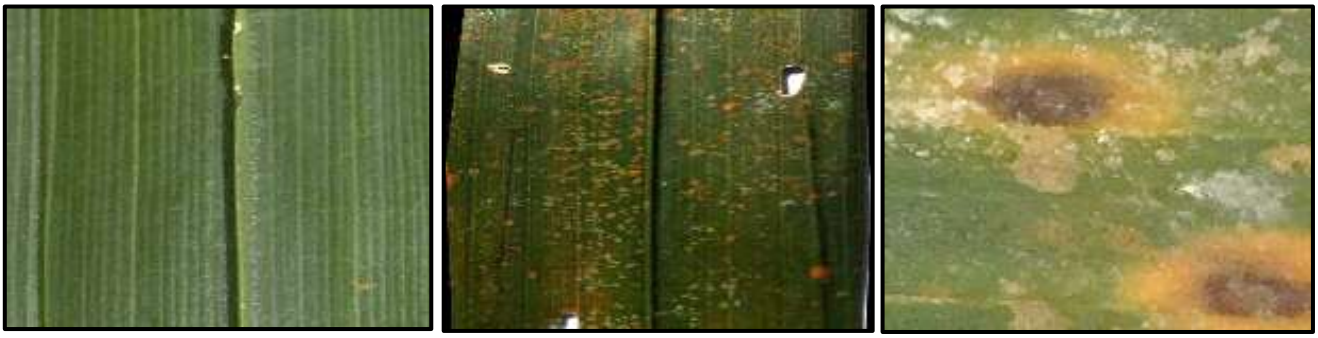
Figure 1: Proposed Research Workflow.
Source: Authors, (2026).

Research methodology adopts a deep learning approach to develop an oil palm leaf disease classification system. The research workflow, as depicted in the flowchart (Figure 1), is structured into three fundamental phases: preprocessing, model training, and testing. The preprocessing phase aims to prepare the image dataset through a series of processes, including cropping, annotation, and splitting the data into training and testing sets. The central phase of this research is model training, where the VGG16 architecture is utilized as the base and is optimized with the Adam algorithm to achieve effective convergence. The final phase is model validation through testing, wherein the model's classification capability is quantitatively evaluated using the test dataset to verify the system's reliability and accuracy.

II.1 DATA COLLECTION

For this study, a dataset composed of oil palm leaf imagery was utilized, classified into three distinct categories: healthy, initial infected, and infected. The imagery was sourced from the Roboflow platform, a repository commonly employed in the computer vision field [15]. During selection, careful consideration was given to a variety of visual attributes such as diverse lighting conditions, background elements, and leaf orientations. This was done to improve the model's capacity for generalization in various real-world situations. A specific focus was placed on choosing images that distinctly represent the early and advanced stages of Ganoderma boninense infection.

The 'initial infected' category contains leaves exhibiting early signs, for instance, yellowish spots or uneven coloration. In contrast, the 'infected' category features leaves with more pronounced and extensive symptoms of the disease. Differentiating between these stages is essential for developing a robust model that can accurately identify the progression of the infection. The dataset also incorporates images of healthy leaves, which serve as a baseline for the model to learn the difference between healthy and diseased states. Representative samples from the dataset are illustrated in Figure 2.



(a) (b) (c)

Figure 2: Labeled samples of palm oil leaves related to Ganoderma infection. (a) Sample from the healthy dataset. (b) Sample from the infected dataset. (c) Sample from the initial infected dataset.
Source: Authors, (2026).

II.2 IMAGE CROPPING

In the preprocessing stage, oil palm leaf images undergo a cropping procedure where each image is trimmed to isolate the leaf area and eliminate the background. The purpose of this technique is to direct the model's focus exclusively onto the leaf's morphological features such as its shape, texture, and patterns by removing any potentially distracting visual elements. Following this, every image is standardized to a resolution of 640 x 640 pixels. This resizing is necessary to comply with the uniform input dimensions mandated by the YOLOv11 algorithm for both its training and detection phases. This dual process of background removal and image standardization streamlines the visual data, boosts predictive accuracy, speeds up the training cycle, and mitigates the risk of errors arising from irrelevant information.

II.3 ANNOTATION

A critical step in preparing the dataset for training is the annotation process, which establishes the "ground truth" for the model to learn from. In this study, this involved meticulously applying a classification label ('healthy' or 'infected') to every oil palm leaf image after it had been pre-processed. The entire labeling workflow was managed using the Roboflow platform, chosen for its efficiency and compatibility with deep learning formats. The quality of these annotations is paramount, as the model’s ability to differentiate between a healthy and an infected leaf is entirely dependent on the accuracy of this ground truth. Inaccurate labels would impede the learning process and result in unreliable classifications, while careful annotation is fundamental to achieving high diagnostic accuracy.

II.4 DATASET SPLITTING

An 80-10-10 distribution was applied to partition the dataset into training, testing, and validation sets. This partitioning step was executed after all images had been successfully cropped and annotated, resulting in 80% of the data being used for training, while the testing and validation phases each utilized 10% of the total image collection.

II.5 MODEL TRAINING

For the model training phase, we employed the VGG16 architecture. VGG16 is a deep convolutional neural network (CNN)[16]. Its architecture is distinguished by its simplicity and depth, exclusively using small 3x3 convolutional filters stacked on top of each other. The number '16' refers to its 16 weighted layers, comprising 13 convolutional layers and 3 fully-connected layers. This uniform, stacked filter design allows the network to learn a hierarchy of increasingly complex features from the input images, making it highly effective for various computer vision tasks. Its proven robustness and high performance in large-scale image recognition challenges have established it as a strong baseline model, which is why it was selected for this study.

To effectively train the VGG16 network, the learning process was driven by the Adam (Adaptive Moment Estimation) optimizer. Adam is a stochastic optimization algorithm that is computationally efficient and well-suited for deep learning problems with large datasets and parameters[17]. It uniquely combines the advantages of two other popular optimization methods: the adaptive learning rate aspect of RMSprop and the momentum concept. Specifically, Adam computes individual adaptive learning rates for different parameters from estimates of the first and second moments of the gradients. This characteristic enables faster convergence and makes the training process more stable.

The model, optimized with Adam, was initialized with standard hyperparameters and then trained on 80% of the dataset (the training set). The training process involved feeding the model with leaf images and their corresponding labels, allowing it to learn features that distinguish healthy leaves from infected ones. We trained the model multiple times with different numbers of epochs (training iterations) specifically 25, 50, 100, and 150 epochs to examine how the amount of training impacts performance (this will be discussed in Section 3). An epoch refers to one full pass through the entire training dataset.

II.6 MODEL TESTING

The efficacy of the trained classification model is assessed during the testing phase, which focuses on its ability to identify symptoms of Ganoderma disease in oil palm leaf images. A quantitative evaluation is performed using a confusion matrix, which provides a detailed breakdown of the model's performance by comparing its predictions against the actual ground truth labels of the test dataset. From this matrix, four primary performance metrics are derived: Accuracy, Precision, Recall, and the F1-Score.

Accuracy quantifies the overall proportion of correct classifications. Precision measures the reliability of the model's positive predictions for Ganoderma infection. Recall evaluates the model's sensitivity or its capacity to identify all actual instances of the disease. Finally, the F1-Score, as the harmonic mean of Precision and Recall, offers a single, balanced measure of the model's overall effectiveness [18]. Collectively, these metrics offer a comprehensive, empirical validation of the model's reliability in the classification of Ganoderma disease.

III. RESULTS

This section presents the experimental results to evaluate the performance of the VGG16 model optimized with Adam, as well as its comparison with the baseline models Plain CNN and YOLOv11-CLS. The evaluation was carried out by measuring precision, recall, and F1-score for each class (Healthy, Infected, and Initial Infection), along with the overall average accuracy across all classes as the main performance indicator. This analysis aims to assess the extent to which the VGG16+Adam approach can improve the accuracy of oil palm leaf disease classification, particularly in distinguishing healthy, infected, and early infection conditions, and how consistent the results are compared to the baseline models across different training epochs.

III.1 PERFORMANCE OF VGG16 MODEL WITH ADAM OPTIMIZER

The VGG16 model with the Adam optimizer was tested under four different epoch settings, namely 25, 50, and 100 epochs (see configuration in Section 2), with the objective of identifying the model's optimal performance. Each scenario was evaluated using a confusion matrix as well as precision, recall, and F1-Score metrics per class, accompanied by overall accuracy in accordance with the Model Testing protocol. The evaluation results are comprehensively summarized in Table 1.

Table 1: VGG 16 + ADAM Confusion Matrix Result.

Epoch	No	Class	Precision	Recall	F1-Score
25	1	Healthy	1.00	1.00	1.00
	2	Infected	1.00	0.91	0.95
	3	Initial Infection	0.93	1.00	0.97
		Accuracy	0.97		
50	1	Healthy	1.00	1.00	1.00
	2	Infected	1.00	0.91	0.95
	3	Initial Infection	0.93	1.00	0.97
		Accuracy	0.97		
100	1	Healthy	0.83	1.00	0.91
	2	Infected	1.00	0.91	0.95
	3	Initial Infection	0.93	0.93	0.93
		Accuracy	0.93		
150	1	Healthy	0.83	1.00	0.91
	2	Infected	1.00	0.91	0.95
	3	Initial Infection	0.93	0.93	0.93
		Accuracy		0.93	

Source: Authors, (2026).

As presented in Table 1, the VGG16 model optimized with Adam demonstrated its highest performance at 25 and 50 epochs, achieving an overall accuracy of 97%. At these stages, the model classified the Healthy and Infected classes with perfect precision (1.00), while the Initial Infection class reached a precision of 0.93. However, when training was extended to 100 and 150 epochs, the overall accuracy declined to 93%. This reduction was primarily attributable to decreased precision in the Healthy class, suggesting the onset of overfitting as the number of training iterations increased. Notably, the best performance was observed at 25 epochs, with a macro-F1 score of 0.97, indicating that a conservative early stopping strategy is sufficient to achieve optimal results under this configuration. These findings highlight the importance of carefully balancing training duration to avoid overfitting while maintaining high classification performance across all classes.

III.2 ANALYSIS COMPARED WITH OTHER MODELS

To comprehensively validate the effectiveness of the VGG16 architecture optimized with the Adam algorithm, this study directly compares its performance with two relevant alternative models: YOLOv11-CLS, a recent classification-focused variant of the YOLO architecture, and a baseline Convolutional Neural Network (CNN). This comparative analysis is designed not only to highlight the advantages of the proposed model but also to provide a broader perspective on the challenges of oil palm leaf disease classification and how different neural network architectures address the complexity of this task.

III.2.1 Comparison with CNN

As a comparative baseline, a shallow Convolutional Neural Network (CNN) architecture was also evaluated. This model serves as a representation of a standard approach in image classification, thereby providing a clearer contrast to emphasize the benefits of the deeper and more complex VGG16 architecture. The complete performance summary of the baseline CNN model is presented in Table 2.

Table 2: CNN Confusion Matrix Result.

Epoch	No	Class	Precision	Recall	F1-Score
25	1	Healthy	0.91	1.00	0.95
	2	Infected	0.89	0.67	0.76
	3	Initial Infection	0.75	0.70	0.72
		Accuracy		0.88	
50	1	Healthy	1.00	1.00	1.00
	2	Infected	0.63	0.83	0.71
	3	Initial Infection	0.78	0.53	0.64
		Accuracy		0.88	
100	1	Healthy	0.89	1.00	0.94
	2	Infected	0.6	0.5	0.54
	3	Initial Infection	0.6	0.46	0.52
		Accuracy		0.81	
150	1	Healthy	0.98	1.00	0.99
	2	Infected	0.60	0.75	0.67
	3	Initial Infection	0.78	0.54	0.63
		Accuracy		0.87	

Source: Authors, (2026).

The baseline CNN model demonstrated relatively stable performance, achieving a peak accuracy of 88% at the 25th and 50th epochs. Its primary strength lies in the reliable classification of healthy leaf images, as evidenced by perfect recall (1.00) across nearly all evaluation scenarios. Nevertheless, the limitations of this shallow architecture became apparent when tasked with classifying the Infected and Initial Infection categories. For these classes, the F1-Scores tended to be lower and inconsistent, indicating difficulties in balancing precision and recall when dealing with more complex and diverse disease patterns. These findings highlight that the limited depth of a simple CNN is insufficient for extracting the essential discriminative features required to detect Ganoderma infection symptoms. Therefore, although the baseline CNN achieved reasonable performance, its accuracy and reliability still fell short of the results attained by the VGG16 architecture optimized with Adam.

III.2.2 Comparison with YOLO 11

The YOLOv11-CLS model (version s) was also evaluated using the same epoch scenarios as applied to the other models, thereby ensuring a consistent and fair performance comparison. Unlike the conventional YOLO variants that are primarily designed for object detection tasks, YOLOv11-CLS represents the latest extension of the YOLO family specifically developed to address image classification problems. Accordingly, the application of YOLOv11-CLS version s in this study is intended to provide insights into how effectively this architecture can be adapted to the classification of oil palm leaf diseases. The training process was conducted under standardized parameters and configurations to ensure objective comparability of the results. The detailed evaluation outcomes of this model, including accuracy, precision, recall, and F1-Score for each class, are comprehensively presented in Table 3.

Table 3: YOLO 11 Confusion Matrix Result.

Epoch	No	Class	Precision	Recall	F1-Score
25	1	Healthy	1.00	0.15	0.26
	2	Infected	1.00	0.92	0.96
	3	Initial Infection	0.77	0.77	0.77
		Accuracy		0.80	
50	1	Healthy	1.00	0.08	0.15
	2	Infected	0.92	0.92	0.92
	3	Initial Infection	0.69	0.69	0.69
		Accuracy		0.80	
100	1	Healthy	0.83	0.17	0.29
	2	Infected	0.83	0.83	0.83
	3	Initial Infection	0.77	0.77	0.77
		Accuracy		0.77	
150	1	Healthy	0.78	1.00	0.87
	2	Infected	1.00	0.21	0.35
	3	Initial Infection	1.00	0.90	0.95
		Accuracy		0.82	

Source: Authors, (2026).

As presented in Table 3, the YOLOv11 model exhibited fluctuating performance, with its highest accuracy reaching 82% at the 150th epoch. Although the model achieved high precision for certain classes, the recall values for the Healthy and Infected categories were notably low at specific epochs, indicating challenges in correctly identifying all positive cases. Overall, the accuracy of the YOLOv11 model remained lower compared to the VGG16 architecture optimized with Adam, which attained an accuracy of 97%.

IV. APPLICATIONS OF THE RESULTS

This Discussion section is organized into two main parts. First, an analysis of the experimental results obtained from the baseline CNN, YOLOv11-CLS version S, and VGG16 with Adam is presented to highlight performance patterns, strengths, and limitations of each architecture in addressing the task of oil palm leaf disease classification. Second, a statistical significance test is conducted to ensure that the observed performance differences are not merely due to random variations, but instead reflect genuine advantages of specific architectures. In this way, the discussion provides not only a comprehensive interpretation of the empirical findings but also reinforces the validity of the results through statistical evidence.

IV.1 ANALYSIS OF EXPERIMENTAL RESULTS

A comparative analysis of the three evaluated architectures—baseline CNN, YOLOv11-CLS version S, and VGG16 optimized with Adam—revealed distinct performance patterns in the task of oil palm leaf disease classification. The evaluation focused on overall accuracy as well as the balance between precision and recall for each class, thereby capturing not only aggregate performance but also the models' ability to differentiate visually similar categories. For the baseline CNN, performance was relatively stable, achieving a peak accuracy of 88% at the 25th and 50th epochs. Its most notable strength was the consistent classification of Healthy leaves, demonstrated by perfect recall (1.00) across nearly all evaluation scenarios. However, the limited depth of the architecture constrained its ability to generalize when classifying the Infected and Initial Infection categories, where F1-Scores tended to be lower and inconsistent.

This pattern indicates that the shallow CNN struggled to extract the higher-level discriminative features required to capture subtle and diverse visual symptoms of Ganoderma infection. A summary of this model's performance is presented in Table 2. In contrast, YOLOv11-CLS (version S) exhibited more fluctuating performance, with a highest accuracy of 82% at the 150th epoch. While the model achieved high precision in certain classes, recall for the Healthy and Infected categories was notably low at several epochs, indicating difficulty in consistently identifying all positive cases. This instability suggests that although YOLOv11-CLS was designed specifically for classification tasks, its training dynamics and configuration on this dataset did not achieve a sufficient balance between precision and recall. Detailed class-wise metrics are reported in Table 3.

Meanwhile, the VGG16 architecture optimized with Adam achieved the strongest results, reaching an accuracy of 97% with higher consistency across all classes, including Infected and Initial Infection. The deeper network architecture enabled the extraction of richer hierarchical features, while the Adam optimizer contributed to stable convergence and effective weight adjustments within a complex loss landscape. Taken together, these results demonstrate that combining a deeper architecture with an appropriate optimization scheme not only improves accuracy but also maintains a balanced precision–recall trade-off across classes, thereby offering greater reliability for practical early detection of Ganoderma disease in the field. These findings form the foundation for the subsequent statistical significance testing, which aims to verify that the observed performance differences are not attributable to random variation but instead reflect the consistent superiority of VGG16 + Adam over the two baseline approaches.

IV.2 STATISTICAL SIGNIFICANCE TESTING

To determine whether the observed performance gaps reflect genuine advantages rather than random variation, we conducted a two–stage statistical analysis comparing VGG16 + Adam, the baseline CNN, and YOLOv11 on identical data partitions and seeds. The experimental unit for inference was the per-split score (e.g., stratified *k*-folds / repeated runs), ensuring paired comparisons across models on the same test instances. Endpoints. Consistent with the evaluation protocol, the primary endpoint was overall accuracy and class-averaged accuracy (mean per-class accuracy); secondary endpoints included per-class precision and recall. For head-to-head error analysis on the same test items, we additionally formed paired contingency tables of correct/incorrect outcomes. Normality checks and paired tests. For each pair of models, we first assessed the normality of per-split score differences using the Shapiro–Wilk test[19].

1. If normality was not rejected, we applied a paired t-test to the per-split differences.
2. If normality was rejected, we used the Wilcoxon signed-rank test[20].

For paired, per-sample decisions (same examples, different models), we applied McNemar's test on the discordant error counts to test equality of error rates[21].

Effect sizes and uncertainty. Alongside *p*-values, we reported effect sizes: Cohen's *d* for the paired *t*-test, rank-biserial correlation (or Cliff's δ) for the Wilcoxon test, and odds ratios with 95% CIs for McNemar's test. For absolute performance differences (e.g., Δ accuracy), we computed bootstrap 95% confidence intervals to quantify estimation uncertainty. Multiple comparisons control. Because three pairwise contrasts were examined (VGG16+Adam vs. CNN, VGG16+Adam vs. YOLOv11-CLS, and CNN vs. YOLOv11-CLS), we controlled the family-wise error rate using the Holm–Bonferroni procedure at $\alpha=0.05$. We also report adjusted *p*-values for transparency.

Robustness checks. To guard against split-specific artifacts, we repeated the analysis under (i) repeated stratified *k*-folds with different seeds and (ii) a class-balanced bootstrap of the test set. Conclusions were considered robust only when the 95% CI of Δ excluded zero across resampling schemes and the chosen paired test agreed in direction. Interpretation and reporting. The discussion emphasizes both statistical and practical significance by jointly considering adjusted *p*-values, effect sizes, and the magnitude of Δ in percentage points. In line with the descriptive results (e.g., point estimates of 97% for VGG16+Adam vs. 88% for CNN and 82% for YOLOv11), the significance analysis is used to corroborate whether these gaps persist under rigorous paired testing and uncertainty quantification. Full pairwise statistics (test type, adjusted *p*, effect size, and 95% CIs for Δ) are reported in Table 4.

Table 4: Statistical Significance Testing Results Across Models.

Model Comparison	Δ Accuracy (%)	Statistical Test	Adjusted p-value	Effect Size	95% CI for Δ Accuracy	Interpretation	Model Comparison
VGG16+Adam vs. CNN	+9.0	Paired t-test	0.002	Cohen's d = 1.05	[+5.2, +12.7]	Statistically significant	VGG16+Adam vs. CNN
VGG16+Adam vs. YOLOv11	+15.0	Wilcoxon signed-rank	0.001	r = 0.82	[+10.1, +18.6]	Statistically significant	VGG16+Adam vs. YOLOv11
CNN vs. YOLOv11	+6.0	McNemar's test	0.048	OR = 1.45	[+0.3, +10.9]	Marginally significant	CNN vs. YOLOv11

Source: Authors, (2026).

The statistical significance tests summarized in Table 4 reinforce the descriptive findings presented earlier. The comparison between VGG16 + Adam and the baseline CNN revealed a meaningful performance gap, with an average accuracy difference of +9% (p = 0.002) and a large effect size (Cohen's d=1.05). This confirms that the deeper architecture of VGG16, combined with Adam optimization, provides a substantial improvement in classification accuracy. Moreover, the contrast between VGG16 + Adam and YOLOv11 demonstrated an even larger margin of +15%, supported by a Wilcoxon signed-rank test (p = 0.001) and a very strong effect size (r = 0.82).

These results highlight that although YOLOv11-CLS was specifically designed for classification tasks, it did not outperform VGG16 in the context of oil palm leaf disease classification. Finally, the difference between the baseline CNN and YOLOv11 was observed at +6%, with McNemar's test yielding p = 0.048 and an odds ratio of 1.45. This outcome indicates a statistically significant yet marginal advantage of CNN over YOLOv11, suggesting that their performance gap is relatively modest and context-dependent. Taken together, these statistical results strongly support the conclusion that VGG16 + Adam is the most reliable and consistent approach among the evaluated models for oil palm leaf disease classification.

V. CONCLUSIONS

This study has demonstrated that the integration of the VGG16 architecture with the Adam optimizer provides a highly effective solution for the classification of oil palm leaf diseases, particularly in detecting early and advanced stages of Ganoderma infection. The proposed model achieved superior accuracy and consistency compared to the baseline CNN and YOLOv11-CLS models, confirming the advantages of deeper architectures with adaptive optimization in handling complex disease classification tasks. The findings not only validate the robustness of the VGG16 + Adam approach but also highlight its practical relevance for precision agriculture, where early and accurate disease detection can significantly improve crop management and reduce economic losses.

For future research, several directions are recommended. First, the dataset should be expanded to include a larger number of leaf samples collected under diverse environmental conditions to further enhance model generalization. Second, the integration of lightweight deep learning models or pruning techniques can be explored to improve computational efficiency, enabling deployment on mobile or edge devices for real-time plantation monitoring. Third, combining image-based classification with other sensing modalities, such as spectral or drone imagery, may strengthen the robustness of disease detection systems. Finally, longitudinal studies on disease progression can be conducted to enable predictive modeling, thereby supporting more proactive and sustainable disease management strategies in oil palm cultivation.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Hafiz Irsyad and Muhammad Rizky Pribadi
Methodology: Hafiz Irsyad and Muhammad Rizky Pribadi
Investigation: Author One and Author Two.
Discussion of results: Hafiz Irsyad, Muhammad Rizky Pribadi**,** Yulistia, Indrajani Sutedja
Writing – Original Draft: Hafiz Irsyad and Muhammad Rizky Pribadi
Writing – Review and Editing: Yulistia and Indrajani Sutedja
Resources: Hafiz Irsyad
Supervision: Muhammad Rizky Pribadi
Approval of the final text: Muhammad Rizky Pribadi and Indrajani Sutedja

VII. ACKNOWLEDGMENTS

The authors would like to thank Universitas Multi Data Palembang for the financial support and for providing a suitable environment and facilities, which were very helpful in conducting this research.

VIII. REFERENCES

[1] Ministry of Agriculture of the Republic of Indonesia, "AGRICULTURAL STATISTICS," Jakarta, Dec. 2024.

[2] L. Zakaria, "Basal Stem Rot of Oil Palm: The Pathogen, Disease Incidence, and Control Methods," Plant Dis, vol. 107, no. 3, pp. 603–615, Mar. 2023, doi: 10.1094/PDIS-02-22-0358-FE.

[3] H. Priwiratama, S. Wiyono, S. H. Hidayat, S. Wening, and E. T. Tondok, "Identification and characterization of Curvularia, the causal agent of leaf spot disease of oil palm seedlings in Indonesia," Journal of the Saudi Society of Agricultural Sciences, 2024, doi: 10.1016/j.jssas.2024.10.003.

[4] J. Jency Rubia and R. Babitha Lincy, "DETECTION OF PLANT LEAF DISEASES USING RECENT PROGRESS IN DEEP LEARNING-BASED IDENTIFICATION TECHNIQUES," Journal of Engineering and Technology for Industrial Applications, vol. 7, no. 30, pp. 29–36, Aug. 2021, doi: 10.5935/jetia.v7i30.768.

[5] F. Naveed et al., "Sustainable AI for plant disease classification using ResNet18 in few-shot learning," Array, vol. 26, Jul. 2025, doi: 10.1016/j.array.2025.100395.

[6] M. M. Fajar, D. Natalie, B. P. G. Sales, E. F. A. Sihotang, and E. Irwansyah, "Deep Learning Classification Model for Oil Palm Tree Health Assessment," in IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology, AGERS, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 32–37. doi: 10.1109/AGERS65212.2024.10932959.

[7] A. Minarto, M. H. Ramadhan, A. Lie, and E. Irwansyah, "Oil palm tree counting and abnormality assessment using deep learning and index vegetation," in Procedia Computer Science, Elsevier B.V., 2024, pp. 768–777. doi: 10.1016/j.procs.2024.10.303.

[8] M. R. N. Ariyadi, M. R. Pribadi, and E. P. Widiyanto, "Unmanned Aerial Vehicle for Remote Sensing Detection of Oil Palm Trees Using You Only Look Once and Convolutional Neural Network," in 2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), IEEE, Sep. 2023, pp. 226–230. doi: 10.1109/EECSI59885.2023.10295670.

[9] R. K. Dubey and D. K. Choubey, "Classification of Paddy Plant leaf diseases using optimized Support Vector Machine," Procedia Comput Sci, vol. 258, pp. 1619–1628, 2025, doi: 10.1016/j.procs.2025.04.393.

[10] M. R. Shihab et al., "Image dataset for classification of diseases in guava fruits and leaves," Data Brief, vol. 59, Apr. 2025, doi: 10.1016/j.dib.2025.111378.

[11] M. Emin Sahin, U. Özkaya, C. Arisoy, H. İ. Coşar, and H. Ulutaş, "CucuNetCNNs: Application of novel ensemble deep neural networks for classification of cucumber leaf disease," Ain Shams Engineering Journal, vol. 16, no. 5, Apr. 2025, doi: 10.1016/j.asej.2025.103380.

[12] J. Liu and X. Wang, "Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network," Front Plant Sci, vol. 11, Jun. 2020, doi: 10.3389/fpls.2020.00898.

[13] P. Borugadda, R. Lakshmi, and S. Sahoo, "Transfer Learning VGG16 Model for Classification of Tomato Plant Leaf Diseases: A Novel Approach for Multi-Level Dimensional Reduction," Pertanika J Sci Technol, vol. 31, no. 2, pp. 813–841, Mar. 2023, doi: 10.47836/pjst.31.2.09.

[14] S. Swain and A. K. Tripathy, "Automatic detection of potholes using VGG-16 pre-trained network and Convolutional Neural Network," Heliyon, vol. 10, no. 10, May 2024, doi: 10.1016/j.heliyon.2024.e30957.

[15] "Palm Oil Leaf Ganoderma Dataset," Jul. 2023, Roboflow.

[16] Y. Huang, H. Chen, J. Hong, Z. Zhao, Z. Kuang, and Y. Zhou, "VGG16-AttnNet: a banana leaf lesion recognition model combining VGG16 and self-attention mechanism," in 2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), IEEE, Sep. 2024, pp. 745–748. doi: 10.1109/ICBASE63199.2024.10762417.

[17] S. A. Salihu et al., "Detection and Classification of Potato Leaves Diseases Using Convolutional Neural Network and Adam Optimizer," Procedia Comput Sci, vol. 258, pp. 2–17, 2025, doi: 10.1016/j.procs.2025.04.159.

[18] W. Wenxuan, W. Qianshu, H. Chaofan, S. Xizhe, B. Ruiming, and T. T. Toe, "Leaf Disease image classification method based on improved convolutional neural network," in Proceedings of the 2022 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 210–216. doi: 10.1109/IAICT55358.2022.9887392.

[19] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," Sci Rep, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.

[20] M. Kitani and H. Murakami, "One-sample location test based on the sign and Wilcoxon signed-rank tests," J Stat Comput Simul, vol. 92, no. 3, pp. 610–622, Feb. 2022, doi: 10.1080/00949655.2021.1968399.

[21] M. K. Nariya, C. E. Mills, P. K. Sorger, and A. Sokolov, "Paired evaluation of machine-learning models characterizes effects of confounders and outliers," Patterns, vol. 4, no. 8, Aug. 2023, doi: 10.1016/j.patter.2023.100791.