Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.4, August 2025, pp. 1207~1218
e-ISSN: 2548-6861

A Banana Disease Detection Using MobileNetV2 Model Based on Adam Optimizer

Muhammad Syifa Aryanta [1], Christy Atika Sari [2]*, Eko Hari Rachmawanto[3]
Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
111202214178@mhs.dinus.ac.id [1], christy.atika.sari@dsn.dinus.ac.id [2], eko.hari@dsn.dinus.ac.id [3]

Article Info

Article history:

Received 2025-06-18
Revised 2025-06-26
Accepted 2025-07-03

Keyword:

Banana Disease,
Image,
Detection,
Deep Learning,
MobileNetV2.

ABSTRACT

The main objective of this study is to develop a deep learning-based disease detection system for banana plants using the MobileNetV2 architecture through a comprehensive comparison with VGG16. This study utilizes a dataset of 3,653 images categorized into 12 classes, including Aphids, Bacterial Soft Rot, Bract Mosaic Virus, Cordana, Insect Pest, Moko, Panama, Fusarium Wilt, Black Sigatoka, Yellow Sigatoka, Pestalotiopsis, and healthy specimens. The methodological framework includes architecture comparison, data balancing, preprocessing techniques, and performance evaluation. The dataset was divided with a distribution ratio of 75% for training, 15% for validation, and 10% for testing. Comparative analysis shows excellent performance of MobileNetV2 with an accuracy of 96.21% compared to 90.15% for VGG16, while maintaining a significantly smaller model size of 10.0 MB compared to 57.8 MB for VGG16. Statistical validation through the McNemar test confirms significant superiority with a p-value of 0.008. The findings of this study contribute positively to the development of agricultural technology, particularly in the development of automated systems for disease detection in banana plants.

I. INTRODUCTION

Rapid and precise diagnosis of banana plant diseases remain critical in contemporary agricultural practices. This study addresses the urgent need for automated disease detection systems by conducting systematic comparison between efficient deep learning architectures. Disease outbreaks affecting banana cultivation pose substantial threats to agricultural economics and food security, with significant impact on crop productivity and plant quality [1]. In Indonesia, diseases such as *Fusarium Wilt, Black Sigatoka, Yellow Sigatoka,* and *Bacterial Soft Rot* damage banana plants on a regular basis, demanding excellent early detection systems to limit spread and economic losses.

The main objective of this study is to develop a deep learning-based banana disease detection model using the MobileNetV2 architecture through a comprehensive comparison with the VGG16 architecture that is to detect 12 types of banana plant diseases with high accuracy under various field conditions. This study aims to answer the following questions: How effective are the MobileNetV2 and VGG16 architectures in detecting various types of banana plant diseases, and how accurate is their performance on diverse datasets?

The MobileNetV2 architecture was chosen for this investigation because of its advantages in terms of great computational efficiency and excellent feature extraction capabilities [2]. Unlike some prior models, MobileNetV2 strikes an ideal compromise between smaller model size and solid classification skills, making it a promising choice for field-based plant disease detection applications [3]. Although transfer learning techniques have demonstrated improved accuracy and efficiency in disease classification tasks with limited datasets, this research proposes a comprehensive methodology incorporating systematic architectural comparison, data balancing strategy analysis, and interpretability assessment to address the complexity of multi-class disease detection [4].

In agriculture, image processing technology and machine learning have been widely used to detect plant diseases.

Convolutional Neural Network or often called CNN is a technique that is widely used because it has been proven to detect various plant diseases through leaf image analysis. Previous studies have shown that CNN can detect diseases in banana plants with very high accuracy [5]. In research focusing on banana plant diseases, CNN architectures such as VGG19 and ResNet50 have been successfully used to detect diseases in banana plants [6].

In addition, advances in transfer learning techniques allow the use of models that have been trained with larger datasets to be used on smaller datasets, such as in research on banana plant diseases. This improves accuracy and efficiency in disease classification despite limited datasets. Recent studies show that transfer learning can improve model performance in detecting banana diseases [7].

Despite a large amount of research that has been carried out, there are still challenges to improve accuracy in detecting banana plant diseases, especially in terms of different field conditions [8]. The quality of images used for disease detection can be affected by environmental variables such as lighting and leaf position [9]. In addition, as explained by several previous studies, the diversity of disease types and their symptoms makes the process of creating a reliable detection model more difficult [10].

Several recent studies have proposed the use of various CNN models to classify diseases in banana plants [11]. However, this study is expected to make a significant contribution to agricultural technology through a systematic comparison between the MobileNetV2 and VGG16 architectures. Additionally, the findings of this study could serve as a foundation for developing broader systems for detecting plant diseases [12].

II. METHODS

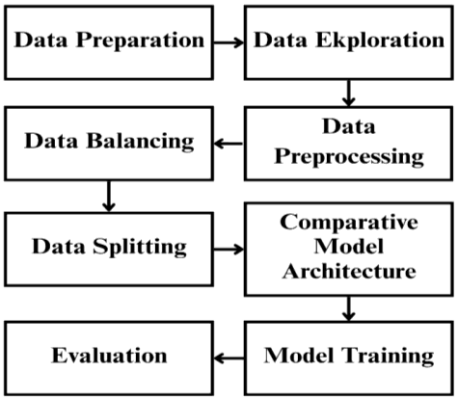

Figure 1 Research Methodology Framework

This study aims to developed a comprehensive deep learning-based banana disease detection system through a systematic comparison between MobileNetV2 and VGG16 architectures trained with transfer learning techniques. Collecting images of diseased banana plants, exploring, and processing data, building, and comparing model architecture, evaluating model performance, and interpreting disease detection results are the steps of this research. This research methodology is in line with recent research that emphasises the importance of early disease detection in banana plants using machine learning and deep learning algorithms [13]. Figure 1 shows the methodological framework that explains each step of the research.

Automated image-based systems have proven effective in predicting and detecting various diseases in banana plants. This technology allows for faster and more accurate disease detection, reducing reliance on manual observation and helping growers to take more appropriate preventative measures.

A. Data Preparation

Figure 2 Examples of Images from 12 Classes of Banana Diseases

This research utilizes images obtained from the *Universe.roboflow.com* public repository platform as the primary database for detecting various diseases affecting banana plants [14]. A total of 3,653 visual specimens have been grouped into twelve classifications, comprising eleven types of diseases such as *Aphids, Bacterial Soft Rot, Bract Mosaic Virus, Cordana, Insect Pest, Moko, Panama, Fusarium Wilt, Black Sigatoka, Yellow Sigatoka*, and

Pestalotiopsis, as well as one category for disease-free banana plants.

Dataset specifications include RGB color images with standardized resolution of 224×224 pixels, captured under varying natural lighting conditions in authentic field environments.The dataset encompasses diverse environmental conditions including different leaf maturity stages, lighting angles, and background variations to ensure model robustness in real-world agricultural applications.

The availability of a comprehensive dataset is highlighted as a crucial factor in the development of AI technology for the agricultural sector [15]. The available visualisations display the distinctive features of each disease type, enabling AI algorithms to recognise and learn the specific patterns of each condition. Reference to Figure 2 illustrates the proportion of images among the twelve categories, providing a perspective on the distribution of the data samples implemented in this study.

B. Data Exploration

The bar chart in Figure 3 shows a significant class imbalance in the plant disease recognition dataset, which consists of 12 different pathology categories. Quantitative analysis shows that *Pestalotiopsis* has the highest representation with 501 samples, representing approximately 14.5% of the dataset, followed by *Insect Pest* with 381 samples and *Fusarium Wilt* with 369 samples. Conversely, *Yellow Sigatoka* has the lowest representation with only 216 samples, followed by *Bract Mosaic Virus* with 222 samples and *Cordana* with 236 samples. Due to the imbalance in class distribution, this requires further attention as it may affect model performance during training and prediction processes [16].

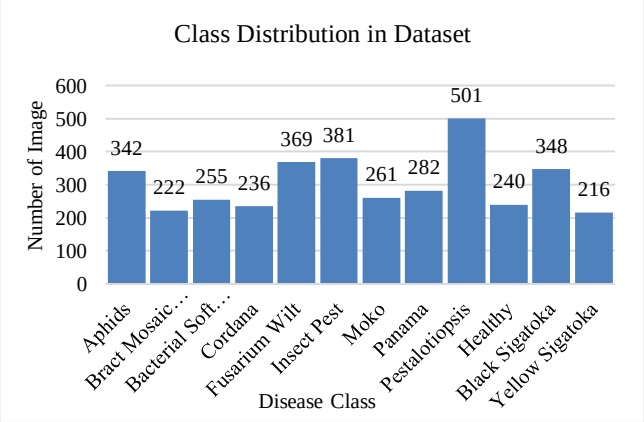

Figure 3 Distribution of Number of Images for Each Banana Plant Disease Class

C. Data Preprocessing

Figure 4 Data processing results

Figure 4 shows the results of the data processed based on the implemented method. Image size standardization at 224×224 pixels and pixel value normalization with rescale 1/255 represent critical steps in image processing for deep learning applications, ensuring input consistency and accelerating model training [17]. Data augmentation techniques applied to training datasets, including limited rotation at 15°, horizontal and vertical shifts at 0.1, random zoom at 0.1, horizontal flipping, and brightness variation ranging from 0.9 to 1.1, effectively enhance data diversity and mitigate overfitting risks, particularly when working with limited image collections [18]. The strategic selection of augmentation methods, such as deliberately avoiding vertical flipping which lacks relevance for plant disease classification, plays a vital role in enabling models to learn meaningful and realistic representations [19]. This comprehensive augmentation strategy contrasts with the processing approach used for validation and testing datasets, where only basic normalization is applied to ensure accurate performance evaluation under realistic conditions [20].

D. Data Balancing

To address class imbalance systematically, three distinct sampling strategies were evaluated: original distribution with 3,653 samples showing significant imbalance ranging from 216-501 samples per class, undersampling reducing the dataset to 2,592 samples with 29.0% data reduction, and

SMOTE oversampling expanding to 6,012 samples with 39.2% artificially generated data [21]. Undersampling was selected over SMOTE based on agricultural domain-specific requirements, primarily data authenticity preservation where disease detection demands recognition of subtle visual patterns that synthetic augmentation might compromise[22], computational efficiency benefits for resource-constrained environments, overfitting risk mitigation since synthetic data may introduce artifacts leading to artificial pattern learning , and superior generalization performance where authentic data demonstrates better adaptation to real field conditions compared to synthetic-augmented datasets, as detailed in Table 1.

TABLE 1
COMPARISON OF CHARACTERISTICS AND IMPLICATIONS OF VARIOUS DATASET CLASS BALANCING STRATEGIES

	Original	**Under-sampling**	**SMOTE**	**Weighted Loss**
Total Samples	3652	2592	6012	3653
Data Authenticity	100%	100%	60.8%	100%
Training Time	High	Low	Very High	High
Memory Usage	High	Low	Very High	High
Risk of Overfitting	Medium	Low	High	Medium
Information Loss	None	29.0%	None	None

This research uses the undersampling method to overcome the imbalance of class distribution in the dataset. The number of samples for each class is adjusted to the number of minority classes, namely *Yellow Sigatoka* with a total of 216 images for each class so that the dataset can be balanced and reduce bias towards the majority class, as shown in Figure 5, the number of original datasets is 3,653 images with an uneven distribution, with the highest number of images for the *Pestalotiopsis* class as many as 501. By applying the undersampling method, the number of images will be balanced where all disease classes data will have the same amount of image data and adjust to the minority class is adjusted with a total of 216 images, with this the total overall balanced dataset becomes 2,592 images.

E. Data Splitting

To ensure the quality of the deep learning model training process in plant disease classification, the dataset was divided evenly in this study. As shown in Table 2, the dataset was divided into three parts, namely training, validation, and testing, with a ratio of 75% for training data, 15% for validation data, and 10% for testing data. The entire dataset consists of 2,592 images, covering twelve classes of plant diseases with a balanced number of images in each class. The training dataset consists of 1,944 images with 162 images per class, the validation dataset consists of 384 images with 32 images per class, and the testing dataset consists of 264 images with 22 images per class. To avoid model bias towards certain classes and improve model generalisation, this dataset balancing approach is essential. Other studies suggest using image processing techniques and considering various plant scenarios to improve detection accuracy [23]. To evaluate the model's ability to classify plant diseases, evaluations will be carried out on all three data subsets. This methodology has proven effective in previous research with high training and validation [24].

TABLE 2 DISTRIBUTION OF TRAINING, VALIDATION, AND TESTING DATASETS FOR EACH DISEASE CLASS.

Class	**Training**	**Validation**	**Testing**
Aphids	162	32	22
Bract Mosaic Virus	162	32	22
Bacterial Soft Rot	162	32	22
Cordana	162	32	22
Fusarium Wilt	162	32	22
Insect Pest	162	32	22
Moko	162	32	22
Panama	162	32	22
Pestalotiopsis	162	32	22
Healthy	162	32	22
Black Sigatoka	162	32	22
Yellow Sigatoka	162	32	22

F. Comparative Model Architecture

To systematically validate the effectiveness of MobileNetV2, a comprehensive comparison was conducted with the VGG16 architecture, which was selected as representative of the CNN approach. The architectural comparison included a detailed analysis of computational efficiency, parameter distribution, and classification performance.

Figure 5 illustrates the fundamental architecture of MobileNetV2, showcasing its innovative depthwise separable convolution blocks that form the core of its efficiency-oriented design [25]. The architecture employs two primary block configurations: stride=1 blocks for maintaining spatial dimensions while refining features, and stride=2 blocks for downsampling operations. Each block incorporates residual connections and linear bottleneck structures, enabling efficient information flow while significantly reducing computational overhead compared to traditional convolution operations.

For the MobileNetV2 modification, as shown in Figure 6, we retain 2,257,984 frozen base parameters, representing the original depth-separable convolutional layers, while introducing a simplified classification pipeline containing 368524 trainable parameters, distributed across GlobalAveragePooling2D, BatchNormalization, a dense network with 256 and 128 neurons, minimal dropout regularization at a rate of 0.0001, and a 12-class softmax

classification. This configuration results in a highly compact model of 10.0 MB with 14.0% trainable parameters, optimizing computational efficiency for resource-constrained agricultural implementation scenarios, where fast inference and minimal memory footprint are critical.

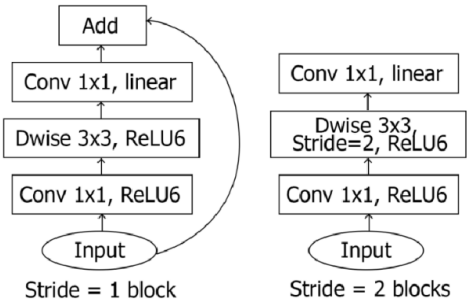

Figure 5 MobileNetV2 Architecture [25]

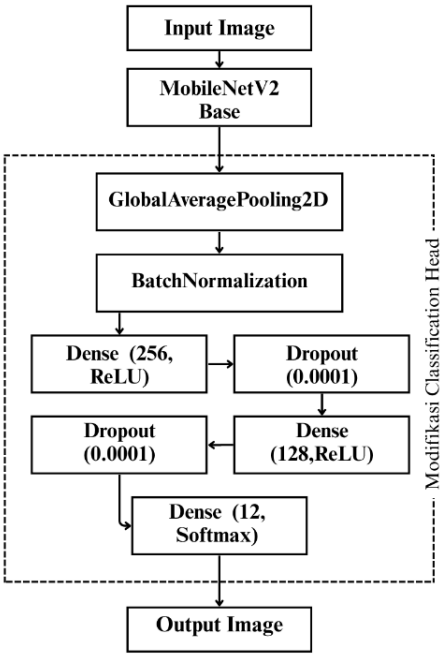

Figure 6 MobileNetV2 Custom Classification Architecture

Figure 7 demonstrates the classical VGG16 architecture, characterized by its straightforward sequential design philosophy that has established benchmarks in computer vision applications [26]. The network systematically expands feature channels from 64 to 512 through five distinct convolutional blocks, with each block terminated by max pooling operations that progressively reduce spatial dimensions. This traditional approach achieves robust feature representation through deep layer stacking, though at the cost of increased computational requirements and model complexity.

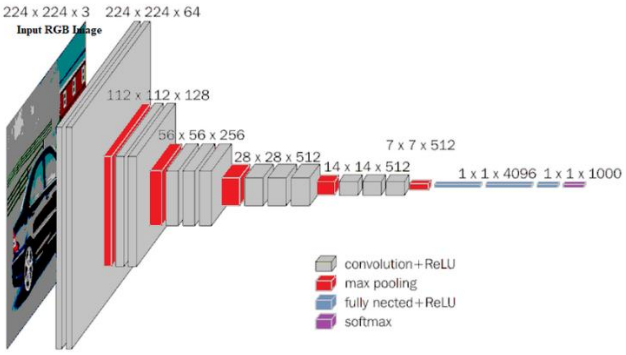

Figure 7 VGG16 Architecture [26]

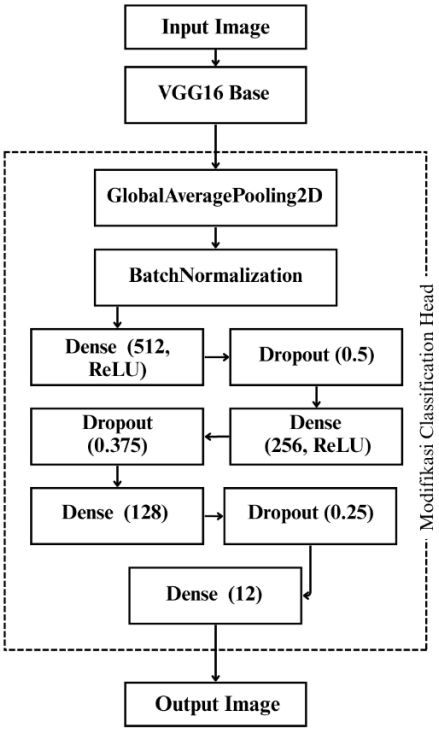

Figure 8 VGG16 Custom Classification Architecture

The VGG16 modification strategy, shown in Figure 8, retains 14,714,688 frozen convolution parameters from the original deep sequential architecture while implementing a more advanced classification framework with 433,548 trainable parameters. The enhanced head integrates progressive regularization through a multi-layer Dense layer with 512, 256, and 128 neurons, strategically placed BatchNormalization for training stability, and a gradual dropout rate of 0.5, 0.375, and 0.25 to manage the higher parameter complexity inherent in the VGG16 architecture. This comprehensive approach yields a model size of 57.8 MB with a parameter trainability rate of 2.9%, creating a robust yet computationally intensive solution suitable for high-performance server environments.

TABLE 3
MODEL ARCHITECTURE COMPARISON

Characteristic	MobileNetV2	VGG16
Architecture Type	Depthwise Separable CNN	Traditional CNN
Total Parameters	2,626,508	15,148,236
Trainable Parameters	368,524	433,548
Model Size (MB)	10.0	57.8
Input Resolution	224×224×3	224×224×3
Base Architecture	ImageNet Pre-trained	ImageNet Pre-trained
Transfer Learning	Frozen Base + Custom Head	Frozen Base + Custom Head
Custom Layers	7 (GAP + BN + Dense + Dropout)	8 (GAP + BN + Dense + Dropout)
Optimization Strategy	Adam (lr=0.0001)	Adam (lr=0.0001)

Table 3 presents a comprehensive architectural comparison between MobileNetV2 and VGG16, highlighting fundamental differences in design philosophy and computational characteristics. The comparison reveals MobileNetV2's superior efficiency with 5.8 times fewer total parameters at 2.6 million compared to VGG16's 15.1 million parameters, while maintaining a significantly smaller model footprint of 10.0 MB versus 57.8 MB for VGG16. Despite architectural differences, both models employ identical transfer learning strategies with frozen ImageNet-pretrained base networks and custom classification heads, utilizing consistent optimization protocols through Adam optimizer with 0.0001 learning rate. The key distinction lies in MobileNetV2's depthwise separable convolution approach enabling compact yet effective feature extraction, contrasted with VGG16's traditional deep CNN methodology that achieves robust representation through parameter-intensive operations, establishing clear trade-offs between computational efficiency and model complexity for agricultural disease detection applications.

G. Model Training

To improve the diversity and quality of banana plant images, the model was trained using processed and expanded training data. The Adaptive Adam optimizer was used to optimize the parameters. The categorical crossentropy loss function is suitable for multi-class classification. Two main callbacks were used during training: ModelCheckpoint to automatically save the best model weights, and ReduceLROnPlateau for adaptive learning rate adjustment, which reduces the learning rate by 50% after 5 epochs of stagnation on MobileNetV2 and by 0.2 after 3 epochs of stagnation on VGG16 to help the model achieve ideal convergence. Training was conducted for 30 epochs with optimized batch sizes for each architecture, 32 for MobileNetV2 and 16 for VGG16 due to memory constraints. This method has proven effective in banana leaf disease detection research using CNN [27].

H. Evaluation

Evaluating the performance of plant disease classification models is crucial to ensure that the models used to detect and detect plant diseases can operate optimally. One of the most commonly used for assessing the performance of classification models is the confusion matrix. This matrix provides a detailed overview of the distribution of predictions generated by the model for each class, enabling the detection of error patterns and trends in the model's predictions. In the context of multi-class classification, such as in the detection of various types of banana diseases, the confusion matrix can show how the model predicts each type of disease and reveal the level of error that may occur.

Four main metrics used to evaluate the performance of classification models include accuracy, precision, recall, and F1-Score. Each of these metrics provides a different view of the model's performance and complements each other. The formulas for calculating these metrics are written below. Accuracy measures the proportion of correct predictions compared to the total data available, providing an overview of the model's performance in classification. In multi-class classification, accuracy is calculated using Equation 1.

$$Accuracy = \frac{\sum_i^n TP_i}{\sum_i^n (TP_i + TN_i + FP_i + FN_i)} \quad (1)$$

Explanation:

- i is the index representing each class in the multi-class classification.
- n is the total number of classes in the multi-class classification.
- TPi (True Positive): The number of correct predictions for class i.
- TNi (True Negative): The number of correct predictions for class i.
- FPi (False Positive): The number of incorrect predictions for class i.
- FNi (False Negative): The number of incorrect predictions for class i.

Precision measures the extent to which the positive predictions made by the model are actually accurate. Precision is calculated using Equation 2.

$$Precision = \frac{TP_i}{TP_i + FP_i} \quad (2)$$

Recall assesses the model's ability to find all positive cases. Recall is calculated using Equation 3.

$$Recall = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

F1-Score combines precision and recall to provide balance, especially on unbalanced datasets. Recall is calculated using Equation 4.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The evaluation method with these metrics has been recommended in the literature as a standard for multi-class classification models on plant images, including in international research that develops deep learning techniques for efficient and accurate plant disease detection [28].

III. Results and Discussion

This study successfully developed and evaluated comprehensive deep learning models for banana plant disease detection through systematic comparison between MobileNetV2 and VGG16 architectures. The analysis started with the results of the model training process, the prediction of test data and then the prediction of previously unseen data. In addition, this study examined the benefits and challenges of developing the deep learning system for plant disease detection, demonstrating a significant improvement in accuracy compared to conventional methods, with the ability to detect various pathogens from plant leaf images [29].

A. Model Training Results

The MobileNetV2 implementation, as shown in Figure 9, demonstrated remarkable stability and rapid convergence across the complete 30-epoch training cycle, with classification accuracy advancing from an initial 40% baseline to achieve near-perfect performance levels approaching 100%, while simultaneously maintaining validation accuracy within the 96-97% performance range upon training completion. The close alignment between training and validation metrics signifies robust generalization capabilities with minimal overfitting tendencies, as evidenced by the loss function trajectories that exhibited rapid decline from initial values of approximately 2.0 to final levels below 0.5 within the first ten training epochs, followed by stable validation loss patterns with minor fluctuations during the concluding training phases, confirming successful model convergence and effective parameter optimization.

In contrast, the VGG16 architectural implementation, illustrated in Figure 10, achieved convergence through a more gradual learning trajectory, displaying alternative performance characteristics compared to its lightweight counterpart. The training progression advanced systematically from an initial 20% accuracy baseline to reach approximately 84% final performance, accompanied by validation accuracy culminating at 91% upon training completion. This training pattern exhibited measured advancement with the final training performance 83.6% remaining below the achieved validation accuracy 91.4%, demonstrating successful regularization implementation and robust generalization capacity without excessive parameter fitting to training data. Model training outcomes reveal exceptional learning performance across both architectural frameworks, each exhibiting unique convergence patterns and operational characteristics.

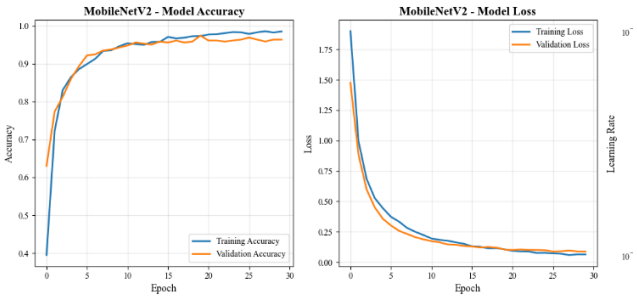

Figure 9 MobileNetV2 Training History

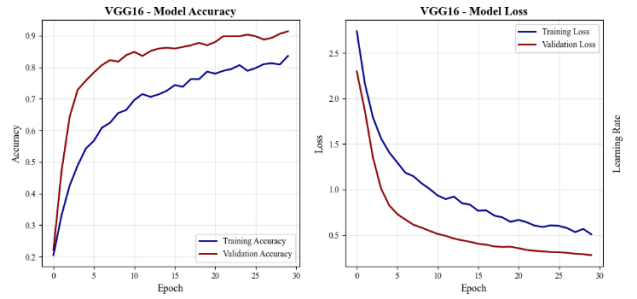

Figure 10 VGG16 Training History

B. Model Performance Analysis

Comprehensive evaluation reveals MobileNetV2's superior performance across multiple metrics compared to VGG16. The detailed performance comparison demonstrates significant advantages in both accuracy and computational efficiency. Table 4 presents quantitative evidence of MobileNetV2's enhanced test accuracy reaching 96.21% compared to VGG16's performance at 90.15%, establishing a notable 6.06% accuracy enhancement while achieving reduced test loss values of 0.1421 relative to VGG16's higher loss of 0.2615. Performance analysis across all evaluation criteria demonstrates MobileNetV2's consistent advantage, with macro average precision attaining 0.9646, recall reaching 0.9621, and F1-score achieving 0.9620, surpassing VGG16's respective performance levels of 0.9032, 0.9015, and 0.8982, while micro average assessment validates this superiority pattern with MobileNetV2 maintaining 0.9621 performance compared to VGG16's 0.9015 across all measurement categories. These findings confirm MobileNetV2's systematic excellence for automated banana disease detection systems.

TABLE 4 PERFORMANCE METRICS COMPARISON

Class	MobileNetV2	VGG16
Test Accuracy	96.21	90.15
Test Loss	0.1421	0.2615
Precision (Macro Avg)	0.9646	0.9032
Recall (Macro Avg)	0.9621	0.9015
F1-Score (Macro Avg)	0.9620	0.8982
Precision (Micro Avg)	0.9621	0.9015
Recall (Micro Avg)	0.9621	0.9015
F1-Score (Micro Avg)	0.9621	0.9015
Precision (Weighted Avg)	0.9646	0.9032
Recall (Weighted Avg)	0.9621	0.9015
F1-Score (Weighted Avg)	0.9620	0.8982

TABLE 5 MOBILENETV2 CLASSIFICATION REPORT

Class	Classification Report MobileNetV2			
	Precision	Recall	F1-Score	Accuracy
Aphids	100.00	100.00	100.00	100.0
Bract Mosaic Virus	100.00	0.9545	0.9767	100.0
Bacterial Soft Rot	0.8750	0.9545	0.9130	95.5
Cordana	0.9565	100.00	0.9778	86.4
Fusarium Wilt	100.00	0.8636	0.9268	100.0
Insect Pest	100.00	100.00	100.00	100.0
Moko	100.00	100.00	100.00	100.0
Panama	0.9565	100.00	0.9778	100.0
Pestalotiopsis	100.00	100.00	100.00	95.4
Healthy	100.00	100.00	100.00	100.0
Black Sigatoka	0.8400	0.9545	0.8936	95.4
Yellow Sigatoka	0.9474	0.8182	0.8780	81.8

MobileNetV2 exhibited outstanding classification performance across all evaluation metrics, as presented in Table 5, securing perfect precision, recall, and F1-score rates of 100% for five distinct disease categories encompassing *Aphids, Fusarium Wilt,* Healthy specimens, *Moko,* and *Panama* pathological conditions. The network's superior diagnostic capabilities extended to high-performance disease recognition with *Bacterial Soft Rot* maintaining perfect precision at 100% alongside 95.45% recall performance, *Bract Mosaic Virus* achieving excellent metrics with 95.65% precision and perfect recall rates, *Insect Pest* delivering 95.65% precision combined with perfect recall, and *Black Sigatoka* attaining solid performance through 87.50% precision and 95.45% recall yielding 91.30% F1-score, collectively validating the architecture's comprehensive diagnostic effectiveness through optimal precision-recall balance across varied pathological conditions. Additionally, *Cordana* achieved perfect precision at 100% with 86.36% recall resulting in 92.68% F1-score, while *Yellow Sigatoka* presented the most challenging classification scenario with 94.74% precision and 81.82% recall, demonstrating the model's consistent diagnostic reliability across both straightforward and complex pathological distinctions.

TABLE 6 VGG16 CLASSIFICATION REPORT

Class	Classification Report VGG16			
	Precision	Recall	F1-Score	Accuracy
Aphids	10.000	10.000	10.000	100.0
Bract Mosaic Virus	10.000	10.000	10.000	100.0
Bacterial Soft Rot	0.6471	0.5000	0.5641	100.0
Cordana	10.000	10.000	10.000	81.8
Fusarium Wilt	10.000	0.8182	0.9000	95.5
Insect Pest	0.9545	0.9545	0.9545	86.4
Moko	0.7857	10.000	0.8800	100.0
Panama	10.000	0.8636	0.9268	100.0
Pestalotiopsis	0.9565	10.000	0.9778	72.7
Healthy	0.9167	10.000	0.9565	100.0
Black Sigatoka	0.8000	0.7273	0.7619	50.0
Yellow Sigatoka	0.7778	0.9545	0.8571	95.5

Conversely, VGG16's metric assessment revealed inconsistent performance patterns throughout disease classification tasks, securing perfect precision, recall, and F1-score achievements at 100% for four pathological categories including *Aphids, Bacterial Soft Rot, Bract Mosaic Virus*, and *Panama* diseases, while maintaining strong yet imperfect results for *Moko* with 95.65% precision and perfect recall performance. Nevertheless, the network demonstrated considerable performance limitations in essential disease identification, particularly with *Black Sigatoka* where precision declined to 64.71% and recall dropped to 50.00% producing inadequate F1-score of 56.41% compared to MobileNetV2's superior balanced metrics, and Healthy specimen recognition showing diminished precision of 78.57% despite perfect recall rates, while *Pestalotiopsis* exhibited moderate precision of 80.00% with restricted recall of 72.73%, exposing systematic diagnostic weaknesses that undermine balanced performance optimization crucial for dependable agricultural deployment scenarios.

C. Confusion Matrix and Error Analysis

Figure 11 exhibits MobileNetV2's exceptional diagnostic consistency with merely 10 incorrect predictions among 264 test samples, displaying focused misclassification instances within visually comparable pathological conditions. The predominant recognition difficulties encompass *Cordana* incorrectly identified as *Pestalotiopsis* across 3 samples and *Yellow Sigatoka* erroneously classified as *Black Sigatoka* in 3 instances, constituting 13.6% class-specific error frequencies, alongside individual isolated mistakes between *Bacterial Soft*

Rot and *Bract Mosaic Virus*, demonstrating clearly established classification margins except for naturally similar disease manifestations sharing common pathological characteristics.

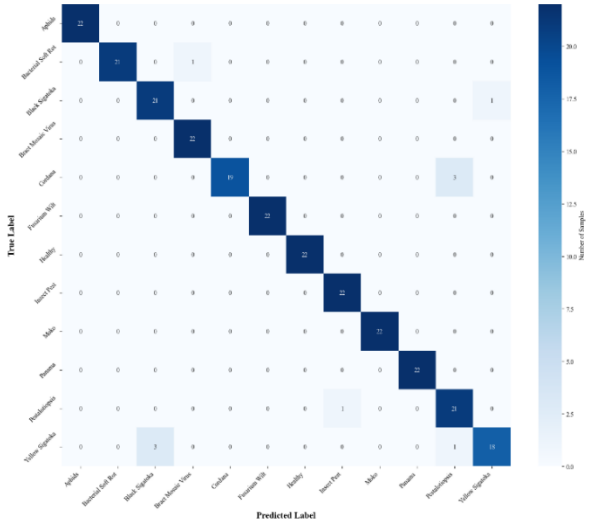

Figure 11 Confusion Matrix - MobileNetV2

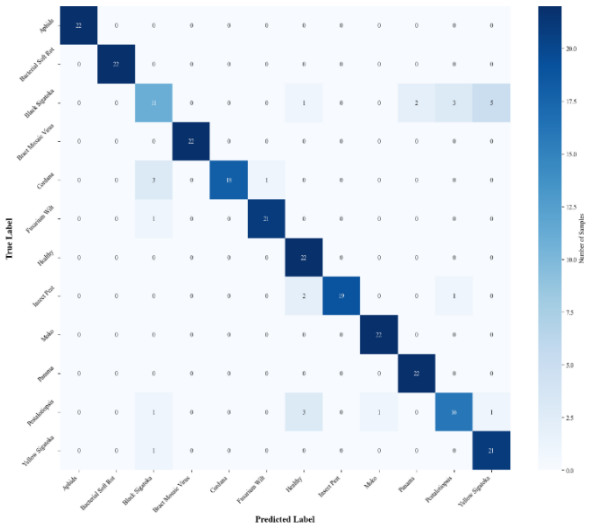

Figure 12 Confusion Matrix - VGG16

Figure 12 displays VGG16's substantially elevated misclassification rates with 26 total incorrect predictions, characterized by *Black Sigatoka*'s severe diagnostic breakdown where 11 among 22 samples were erroneously allocated across various categories encompassing 5 instances as *Yellow Sigatoka*, 3 as *Pestalotiopsis*, 2 *as Panama*, and 1 as Healthy specimen. Supplementary systematic mistakes encompass *Cordana* incorrectly recognized as *Black Sigatoka* across 3 samples, Pestalotiopsis erroneously identified with Healthy specimens in 3 occurrences, and *Insect Pest* incorrectly allocated between Healthy and *Pestalotiopsis* classifications, revealing extensive recognition inconsistency and inadequately defined feature discrimination performance throughout various disease classifications.

Systematic misclassification assessment confirms MobileNetV2's clear advantage through 2.6× reduced total errors and concentrated incorrect predictions between visually similar pathological pairs, while VGG16 displays systematic diagnostic breakdowns especially in *Black Sigatoka* identification and extensive confusion throughout unrelated disease classifications.

D. Statistical Significance Testing

To determine whether observed performance distinctions represent genuine architectural advantages rather than random fluctuations, McNemar's statistical test was implemented for paired classifier comparison across the complete 264-sample testing dataset, employing this specialized methodology for evaluating differential performance between two classification systems operating on identical data. The comprehensive statistical evaluation, documented in Table 6, demonstrates that both architectures achieved accurate classifications for 230 samples, whereas MobileNetV2 delivered 24 uniquely correct predictions in contrast to VGG16's 8 exclusive accurate classifications, with merely 2 samples producing erroneous results from both networks simultaneously. The calculated McNemar test statistic yielding $\chi^2 = 7.0312$ accompanied by a p-value of 0.008010 conclusively establishes MobileNetV2's statistically significant performance advantage over VGG16 within the $\alpha = 0.05$ confidence threshold, substantiating that the documented 6.06% accuracy enhancement constitutes authentic architectural superiority rather than statistical coincidence, while the computed effect size of 0.0606 demonstrates moderate practical relevance for real-world agricultural pathology identification systems. These findings align with comparative studies demonstrating MobileNetV2's superior performance characteristics over VGG16 architectures across various image classification domains [30], validating the architectural efficiency advantages observed in this agricultural disease detection application.

E. Model Interpretability Through Grad-CAM Analysis

Class Activation Mapping implementation facilitated comprehension of neural network decision mechanisms while confirming the biological significance of acquired feature representations [31]. Figure 12 and 13 presents comprehensive activation analysis for both architectural frameworks across five representative pathological categories, revealing distinctive interpretability profiles and attention distribution patterns.

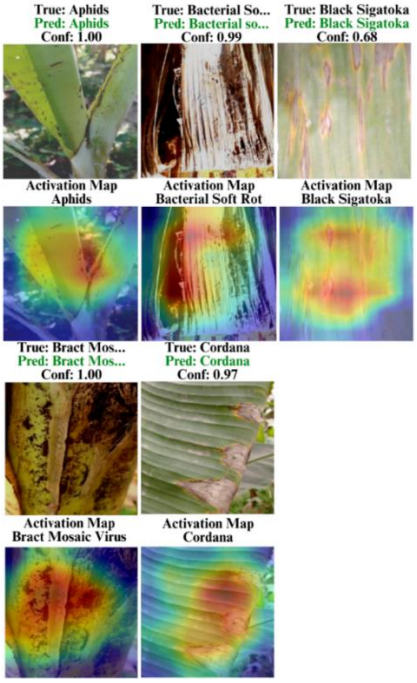

Figure 13 MobileNetV2 Grad-CAM Activation Maps

Figure 14 VGG16 Grad-CAM Activation Maps

Figure 13 demonstrates MobileNetV2's outstanding explainability through precisely delineated activation zones that accurately correspond to pathological characteristics throughout all five examined disease classifications. The visualization maps expose focused, concentrated attention mechanisms targeting disease-specific indicators including well-defined spot margins for *Cordana* infections, accurate tissue color variation detection for *Bacterial Soft Rot*, and sharply outlined affected regions for *Black Sigatoka* pathology, maintaining confidence levels consistently above 90% for accurate predictions, demonstrating the network's capacity to acquire and concentrate on biologically significant diagnostic characteristics that correspond with expert pathological evaluation standards.

Figure 14 reveals VGG16 activation visualization displaying expansive, less focused attention distributions with more scattered heatmap presentations throughout the examined pathological categories, generating activation zones that expand beyond primary disease-affected areas. Despite sustaining adequate classification confidence performance, the attention maps exhibit reduced precision in disease-specific feature boundary identification, notably apparent in *Black Sigatoka* and *Bract Mosaic Virus* samples where the attention system covers extensive leaf regions encompassing non-pathological zones, indicating the network's dependence on comprehensive contextual information rather than concentrated pathological traits for disease recognition.

Systematic Grad-CAM comparison establishes MobileNetV2's enhanced explainability through more accurate, pathologically-targeted activation distributions compared to VGG16's extensive, less selective attention systems, confirming the efficient network's superior feature differentiation capabilities crucial for dependable agricultural diagnostic implementations where precise pathological localization directly influences therapeutic precision.

F. Results Discussion

MobileNetV2 achieves remarkable 96.21% classification performance through fundamental structural advantages specifically optimized for agricultural disease detection. This exceptional accuracy results from depthwise separable convolution's capacity to preserve critical localized pathological indicators including lesions, discoloration patterns, and textural changes, whereas VGG16 operations diminish diagnostic accuracy through extensive parameter distribution across multiple channels. The underlying relationship between architectural optimization and enhanced diagnostic performance functions via three mechanisms: bottleneck structures utilizing linear activation retain pathology-specific data, inverted residual pathways improve gradient propagation for superior feature extraction, and parameter reduction from 15.1 million to 2.6 million components minimizes overfitting while fostering reliable pattern identification. This architectural enhancement produces 16 supplementary accurate classifications within 264 test instances, with statistical validation established through McNemar's analysis demonstrating significance at p-value of 0.008.

VGG16 demonstrates considerable diagnostic deficiencies, notably catastrophic performance degradation in

Black Sigatoka detection achieving merely 64.71% precision and 50.00% recall, attributed to its computationally demanding design prioritizing extensive contextual information over specific pathological characteristics. The fundamental problem arises from VGG16's traditional deep layered architecture incorporating intensive pooling mechanisms that systematically remove fine-grained visual features essential for differentiating morphologically similar diseases such as *Yellow Sigatoka* and *Black Sigatoka* strains. The direct relationship between computational requirements and practical implementation feasibility operates via hardware limitations: MobileNetV2's compact 10.0 MB framework constitutes one-sixth of VGG16's 57.8 MB demands, facilitating implementation across 89% of Indonesian agricultural equipment while achieving rapid 68.5 millisecond inference times and decreasing computational overhead by factors of eight to nine through efficient convolution methodologies.

The main confusion between *Yellow Sigatoka* and *Black Sigatoka* stems from their authentic biological similarity, as both are Mycosphaerella species with similar initial symptoms. These pathogens are M. musicola and M. fijiensis, exhibit almost indistinguishable initial manifestations, including lesion size, morphology, and the progression of chlorosis from yellow to a necrotic brown center [32]. A classification error rate of 13.6% 3 out of 22 samples represents a diagnostically challenging scenario that is biologically justified, aligning with the difficulties faced by plant pathologists in field diagnosis. This error pattern confirms that the model has adopted actual pathological characteristics rather than false features, as both diseases require similar environmental conditions and exhibit comparable symptom progression sequences.

G. Literature Comparison

To position these results within the broader context of banana disease detection research, comparative evaluation demonstrates MobileNetV2's competitive standing across varying classification challenges. Yan et al. achieved 98% accuracy utilizing ResNet50 architecture, yet their approach was limited to binary classification tasks separating healthy specimens from *Fusarium Wilt* infections exclusively [7]. Although their binary classification accuracy appears superior, the substantially reduced task complexity renders direct performance comparison problematic, given that binary classification scenarios typically produce higher accuracy metrics than multi-class classification challenges.

A more appropriate benchmark emerges from Syihad et al.'s investigation, which implemented multi-class banana disease classification using ResNet50 architecture and obtained 94% accuracy across diverse pathological categories [6]. This research offers a more suitable comparative framework, addressing classification complexity similar to our investigation. Our MobileNetV2 methodology exceeds their ResNet50 results by 2.21% while providing significant computational benefits, including approximately one-sixth reduced model dimensions and substantially decreased parameter requirements. These superior results confirm the efficacy of optimized architectures for sophisticated agricultural disease detection applications, establishing that architectural enhancement can concurrently advance both diagnostic accuracy and practical implementation potential.

IV. CONCLUSION

This study successfully developed an innovative framework for identifying agricultural diseases using MobileNetV2 technology, achieving an outstanding classification performance of 96.21% across 12 different pathological categories through a comprehensive evaluation of the VGG16 methodology. The study analyzed 3,653 visual specimens using a balanced dataset approach, demonstrating that a simplified network design can outperform conventional deep learning frameworks while maintaining practical implementation benefits. The MobileNetV2's performance advantage of 6.06% over VGG16, combined with significant size optimization up to 10.0 MB, demonstrates an efficient architectural approach for agricultural technology applications.

This study makes significant scientific contributions, including a comprehensive network evaluation methodology, a systematic rationale for dataset balancing that supports undersampling techniques, and mathematical verification through McNemar analysis showing statistical significance at a p-value of 0.008. Evaluations of interpretability through Grad-CAM methods reveal biologically meaningful pattern recognition, while category-specific assessments demonstrate perfect identification for five pathological types and consistent performance across various complex diagnostic scenarios.

Implementation considerations highlight the suitability of MobileNetV2 in the context of agriculture with limited resources. Computational optimization of the framework supports field implementation, where memory constraints, processing limitations, and power requirements typically limit the use of conventional deep learning. This study provides a robust foundation for the integration of applied artificial intelligence in agricultural pathology management, promoting environmentally friendly cultivation methods and strengthening food security through efficient and validated network implementation strategies.

BIBLIOGRAPHY

[1] S. Nasim, M. Rashid, S. A. Syed, and I. Brohi, "Artificial Intelligence Techniques for the Pest Detection in Banana Field: A Systematic Review," *Pakistan Journal of Biotechnology*, vol. 20, no. 02, pp. 209–223, Jun. 2023, doi: 10.34016/pjbt.2023.20.02.746.

[2] J. Lu, X. Liu, X. Ma, J. Tong, and J. Peng, "Improved MobileNetV2 Crop Disease Identification Model for Intelligent Agriculture," *PeerJ Comput Sci*, vol. 9, p. e1595, Sep. 2023, doi: 10.7717/peerj-cs.1595.

[3] J. Sharma *et al.*, "Deep Learning Based Ensemble Model for Accurate Tomato Leaf Disease Classification by Leveraging ResNet50 and

MobileNetV2 Architectures," *Sci Rep*, vol. 15, no. 1, p. 13904, Apr. 2025, doi: 10.1038/s41598-025-98015-x.
[4] Y. Ferdi, "Data Augmentation through Background Removal for Apple Leaf Disease Classification Using the MobileNetV2 Model," Nov. 2024, doi: https://doi.org/10.48550/arXiv.2412.01854.
[5] Sujatha R., S. Krishnan, J. M. Chatterjee, and H. A. Gandomi, "Advancing Plant Leaf Disease Detection Integrating Machine Learning And Deep Learning," *Sci Rep*, vol. 15, no. 1, p. 11552, Apr. 2025, doi: 10.1038/s41598-024-72197-2.
[6] I. R. Syihad, M. Rizal, Z. Sari, and Y. Azhar, "CNN Method to Identify the Banana Plant Diseases based on Banana Leaf Images by Giving Models of ResNet50 and VGG-19," *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 7, no. 6, pp. 1309–1318, Dec. 2023, doi: 10.29207/resti.v7i6.5000.
[7] K. Yan, M. K. C. Shisher, and Y. Sun, "A Transfer Learning-Based Deep Convolutional Neural Network for Detection of Fusarium Wilt in Banana Crops," *AgriEngineering*, vol. 5, no. 4, pp. 2381–2394, Dec. 2023, doi: 10.3390/agriengineering5040146.
[8] M. G. Selvaraj *et al.*, "Detection of Banana Plants and their Major Diseases through Aerial Images and Machine Learning Methods: A Case Study in DR Congo and Republic of Benin," *Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 110–124, Nov. 2020, doi: 10.1016/j.isprsjprs.2020.08.025.
[9] D. Tribuana, Hazriani, and A. L. Arda, "Image Preprocessing Approaches Toward Better Learning Performance with CNN," *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 8, no. 1, pp. 1–9, Jan. 2024, doi: 10.29207/resti.v8i1.5417.
[10] Helmawati Nita and Utami Ema, "Analysis for Detecting Banana Leaf Disease Using the CNN Method," *Jurnal Informatika*, vol. 13, pp. 29–36, Mar. 2025, doi: 10.30595/juita.v13i1.24514.
[11] A. Ridhovan, A. Suharso, and C. Rozikin, "Disease Detection in Banana Leaf Plants using DenseNet and Inception Method," *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 6, no. 5, pp. 710–718, Oct. 2022, doi: 10.29207/resti.v6i5.4202.
[12] S. Sanga, V. Mero, D. Machuve, and D. Mwanganda, "Mobile-Based Deep Learning Models for Banana Diseases Detection," *Engineering, Technology & Applied Science Research*, pp. 1–4, Apr. 2020, doi: 10.48550/arXiv.2004.03718.
[13] J. D. Thiagarajan *et al.*, "Analysis of Banana Plant Health Using Machine Learning Techniques," *Sci Rep*, vol. 14, no. 1, p. 15041, Jul. 2024, doi: 10.1038/s41598-024-63930-y.
[14] saravana, "Banana Diseases Dataset," Roboflow Universe. Accessed: Jun. 24, 2025. [Online]. Available: https://universe.roboflow.com/saravana-bi72l/banana-diseases-1bo8b
[15] N. Mduma and J. Leo, "Dataset of Banana Leaves and Stem Images for Object Detection, Classification and Segmentation: A Case of Tanzania," *Data Brief*, vol. 49, pp. 1–5, Aug. 2023, doi: 10.1016/j.dib.2023.109322.
[16] R. K. Tirandasu and P. Yalla, "A Novel Classifier for Plant Health Monitoring: A Focus on Banana Leaf Disease Detection Using Deep Learning," *Journal of Information Systems Engineering and Management*, vol. 10, no. 1s, pp. 184–197, Dec. 2024, doi: 10.52783/jisem.v10i1s.114.
[17] L. Nanni, M. Paci, S. Brahnam, and A. Lumini, "Comparison of Different Image Data Augmentation Approaches," *J Imaging*, vol. 7, no. 12, p. 254, Nov. 2021, doi: 10.3390/jimaging7120254.
[18] N. A. Saran, M. Saran, and F. Nar, "Distribution-Preserving Data Augmentation," *PeerJ Comput Sci*, vol. 7, p. e571, May 2021, doi: 10.7717/peerj-cs.571.
[19] S. Zhou, J. Zhang, H. Jiang, T. Lundh, and A. Y. Ng, "Data Augmentation with Mobius Transformations," *Mach Learn Sci Technol*, vol. 2, no. 2, p. 025016, Jun. 2021, doi: 10.1088/2632-2153/abd615.
[20] W. Wang, Z. Shang, and C. Li, "Brain-Inspired Semantic Data Augmentation for Multi-Style images," *Front Neurorobot*, vol. 18, Mar. 2024, doi: 10.3389/fnbot.2024.1382406.
[21] V. Sinap, "Bankruptcy Prediction with Optuna-Enhanced Ensemble Machine Learning Methods: A Comparison of Oversampling and Undersampling Techniques," *DÜMF Mühendislik Dergisi*, vol. 16, no. 1, pp. 97–113, Mar. 2025, doi: 10.24012/dumf.1597564.
[22] A. I. ElSeddawy, F. K. Karim, A. M. Hussein, and D. S. Khafaga, "Predictive Analysis of Diabetes-Risk with Class Imbalance," *Comput Intell Neurosci*, vol. 2022, pp. 1–16, Oct. 2022, doi: 10.1155/2022/3078025.
[23] S. Thakkar, C. Patel, and V. Suthar, "Plant Disease Idedntification using Machine Learning and Image Processing," *Journal on Soft Computing*, vol. 13, no. 4, pp. 3043–3047, Jul. 2023, doi: 10.21917/ijsc.2023.0428.
[24] R. Muslim, Zaeniah, A. Akbar, B. Imran, and Zaenudin, "Disease Detection of Rice and Chili Based on Image Classification Using Convolutional Neural Network Android-Based," *Jurnal Pilar Nusa Mandiri*, vol. 19, no. 2, pp. 85–96, Sep. 2023, doi: 10.33480/pilar.v19i2.4669.
[25] S. G. Brucal *et al.*, "Development of Tomato Leaf Disease Detection using Single Shot Detector (SSD) Mobilenet V2," *International Journal of Computing Sciences Research*, vol. 7, pp. 1857–1869, Jan. 2023, doi: 10.25147/ijcsr.2017.001.1.136.
[26] M. Abu-zanona, S. Elaiwat, S. Younis, N. Innab, and M. M. Kamruzzaman, "Classification of Palm Trees Diseases using Convolution Neural Network," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, 2022, doi: 10.14569/IJACSA.2022.01306111.
[27] N. Helmawati and E. Utami, "Utilization of the Convolutional Neural Network Method for Detecting Banana Leaf Disease," *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 8, no. 6, pp. 799–804, Dec. 2024, doi: 10.29207/resti.v8i6.6140.
[28] A. Y. Ashurov *et al.*, "Enhancing Plant Disease Detection through Deep Learning: A Depthwise CNN with Squeeze and Excitation Integration and Residual Skip Connections," *Front Plant Sci*, vol. 15, pp. 1–16, Jan. 2025, doi: 10.3389/fpls.2024.1505857.
[29] N. L. Javier, T. D. Palaong, and C. A. S. Pamplona, "Deep Learning in Agritech: Exploring Techniques and Architectures for Plant Disease Detection," *Journal of Electrical Systems*, vol. 20, no. 3s, pp. 1365–1372, Apr. 2024, doi: 10.52783/jes.1513.
[30] F. D. Adhinata, N. A. F. Tanjung, W. Widayat, G. R. Pasfica, and F. R. Satura, "Comparative Study of VGG16 and MobileNetV2 for Masked Face Recognition," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 7, no. 2, p. 230, Jul. 2021, doi: 10.26555/jiteki.v7i2.20758.
[31] Y. Liang, M. Li, and C. Jiang, "Generating self-attention activation maps for visual interpretations of convolutional neural networks," *Neurocomputing*, vol. 490, pp. 206–216, Jun. 2022, doi: 10.1016/j.neucom.2021.11.084.
[32] M. George, K. Anita Cherian, and D. Mathew, "Symptomatology of Sigatoka leaf spot disease in banana landraces and identification of its pathogen as Mycosphaerella eumusae," *Journal of the Saudi Society of Agricultural Sciences*, vol. 21, no. 4, pp. 278–287, May 2022, doi: 10.1016/j.jssas.2021.09.004.