Bulletin of Informatics and Data Science
Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

Lightweight CNNs for Eggplant Leaf Disease Classification on a Balanced Dataset: A Comparative Study of MobileNetV3 and EfficientNetV2-B2

Masparudin, Yonky Pernando*, Ihsan Verdian, Fazlul Rahman

Department of Informatics Engineering, Universitas Universal, Batam, Indonesia
Email: [1]masparudin.mahmud@gmail.com, [2,*]yongkyfernando194@gmail.com, [3]ihsanverdian@uvers.ac.id, [4]fazlul.rahman@uvers.ac.id

ARTICLE INFORMATION

ARTICLE HISTORY:
Submited : May 15, 2026
Revised : May 28, 2026
Accept : May 30, 2026
Publish : May 30, 2026

KEYWORD
Eggplant Leaf Disease;
MobileNetV3;
EfficientNetV2;
Data Imbalance;
Background Removal

CORRESPONDENCE AUTHOR
Email: yongkyfernando194@gmail.com

A B S T R A C T

Eggplant (Solanum melongena) is a vital agricultural commodity, but its yield is highly vulnerable to foliar diseases. Early and accurate detection using deep learning is essential for effective crop management. However, deploying automated detection in real-world agricultural settings faces two primary challenges: (1) severe classification bias caused by complex background noise and data imbalance, and (2) extreme computational constraints that hinder the deployment of conventional deep learning models on farmers' edge devices. This study presents a robust methodology for classifying four conditions of eggplant leaves (Healthy, Leaf Spot, Mosaic Virus, and Insect Pest) by implementing an automated background removal technique and targeted data augmentation, resulting in a perfectly balanced dataset of 1,400 images. Furthermore, this research conducts a comparative analysis between two distinct categories of lightweight Convolutional Neural Networks (CNNs): MobileNetV3-Large (representing ultra-lightweight architectures with 224x224 input resolution) and EfficientNetV2-B2 (representing medium-lightweight architectures with 260x260 input resolution). The models were evaluated based on their accuracy, loss convergence, and computational efficiency using an 80:20 data split and early stopping callbacks to prevent overfitting. Experimental results demonstrate that both models achieved exceptional performance. EfficientNetV2-B2 exhibited superior stability and precision, achieving a peak validation accuracy of 97.50% and a validation loss of 0.073. Meanwhile, MobileNetV3-Large reached a validation accuracy of 96.07% with significantly faster training iterations. These findings indicate that while EfficientNetV2-B2 is highly recommended for precision-critical agricultural diagnostics, MobileNetV3-Large remains a formidable alternative for deployment on edge devices with extreme computational constraints

1. INTRODUCTION

Eggplant (Solanum melongena) is a highly cultivated and economically significant agricultural commodity worldwide [1] .However, the overall quality, nutritional value, and agricultural yield of eggplant production are frequently threatened by various foliar diseases and aggressive pest infestations, such as Leaf Spot, Mosaic Virus, and Insect Pest damages [1], [2], [3], [4], [5], [6]. Traditionally, the identification and management of these foliar diseases rely heavily on the direct visual inspection conducted by farmers or agricultural experts. This manual diagnostic approach presents significant challenges; it is not only heavily time-consuming and labor-intensive but also highly prone to subjective misdiagnosis, especially during the early stages of infection when morphological symptoms and chlorotic patches remain visually ambiguous. Establishing a rapid, automated diagnostic mechanism is therefore highly crucial to prevent widespread crop failure and ensure sustainable agricultural output. In recent years, the implementation of Deep Learning, particularly Convolutional Neural Networks (CNNs), has revolutionized the paradigm of automated plant disease detection [1], [7], [8], [9], [10]. Theoretically, CNNs operate by extracting complex spatial features, textural gradients, and hierarchical patterns from input images through interconnected convolutional layers, pooling mechanisms, and non-linear activation functions [10]. Despite their high theoretical accuracy and success in controlled laboratory environments, the performance of conventional CNNs often degrades significantly when deployed in real-world agricultural scenarios [11]. One of the primary causes of this degradation is the presence of complex background noise—such as soil textures, weeds, plant pots, and human hands—which the network may falsely learn as discriminatory features, leading to severe classification bias [11]. Furthermore, publicly available agricultural datasets frequently suffer from severe class imbalance, a condition that forces the mathematical loss optimization function to favor the majority class while effectively ignoring the critical minority diseases [12]. To address the extreme computational constraints of farmers' mobile devices, foundational lightweight architectures such as MobileNetV3 [13] and EfficientNetV2 [14] were introduced, providing highly efficient baselines for edge-based agricultural applications. Building upon these foundations, numerous studies have explored deep learning in precision agriculture. For instance, researchers have utilized squeeze-and-excitation MobileNet models via transfer learning for specific plant disease identification [15]. Subsequent structural modifications, such as the LDL-MobileNetV3S model, have been formulated to enhance the diagnosis of potato leaf diseases through multi-module fusion techniques [16]. Furthermore, advanced object detection frameworks like YOLOv11n-12D have recently been developed as a privacy-protecting framework specifically targeted at real-time eggplant disease detection [17]. Despite the wealth of existing literature, a critical research gap remains largely unaddressed [8], [15], [16]. While individual lightweight CNN architectures have been extensively tested, there is a distinct lack of comparative research that systematically evaluates the combined impact of rigorous automated background removal coupled with strict mathematical data balancing on the performance of these models.

Many previous studies evaluate architectures on datasets containing residual environmental noise or inherent class imbalances, which artificially inflates accuracy metrics for majority classes and masks the true diagnostic precision of the algorithms [11], [12]. Furthermore, a definitive comparative analysis specifically exploring the operational trade-offs between ultra-lightweight and medium-lightweight architectures, evaluated under extreme edge deployment constraints for eggplant foliar diseases, remains remarkably sparse [13], [14]. To facilitate real-time disease detection directly on farmers' smartphones or agricultural edge devices, the deployed AI models must be computationally efficient without sacrificing diagnostic precision. Therefore, this study employs a comparative methodology focusing specifically on two distinct categories of lightweight CNNs. MobileNetV3-Large represents an ultra-lightweight architecture theoretically optimized for mobile CPUs; it operates utilizing a hardware-aware neural architecture search (NAS) and incorporates Squeeze-and-Excitation (SE) modules within its inverted residual blocks to maintain extremely low latency [13]. Conversely, EfficientNetV2-B2 introduces a medium-lightweight theoretical approach, utilizing a compound scaling method alongside Fused-MBConv blocks in its early layers to achieve a mathematically optimal balance between faster training convergence and higher parameter efficiency[14]. The comparative method utilized in this study ensures an unbiased evaluation by subjecting both models to customized identical classification heads, identical training callbacks, and identically preprocessed balanced data. The primary objective of this research is to propose and implement a robust preprocessing pipeline that eliminates background noise and absolutely resolves data imbalance in raw eggplant leaf images. Furthermore, the study aims to conduct a comprehensive comparative analysis between MobileNetV3-Large and EfficientNetV2-B2 to empirically determine the most optimal architecture capable of balancing high diagnostic accuracy, continuous loss stability, and computational efficiency for automated eggplant disease classification [17].

The specific contributions of this study to the field of agricultural informatics are fundamentally threefold. First, it involves the successful implementation of an automated background extraction pipeline that strictly isolates pathological leaf features from complex agricultural backgrounds, thereby eliminating environmental classification bias. Second, this research formulates a perfectly balanced training dataset consisting of exactly 1,400 curated images across four classes, ensuring that no individual disease category dominates the loss optimization function during model training. Finally, it provides a definitive empirical evaluation comparing the inference speed, diagnostic safety, particularly the minimization of False Negatives—and spatial visual interpretability of MobileNetV3-Large against EfficientNetV2-B2. This comprehensive evaluation ultimately provides actionable recommendations for implementing lightweight diagnostic models on severely computationally constrained edge devices.

2. RESEARCH METHODOLOGY

This section outlines the systematic framework designed to execute the automated classification of eggplant leaf diseases. To provide a clear overview of the computational process, the entire experimental workflow is illustrated in Figure 1. This flowchart visually maps the application of our proposed methodological solutions across distinct research stages, encompassing raw data acquisition, automated background extraction, mathematical data balancing, deep learning model execution, and explainability validation.

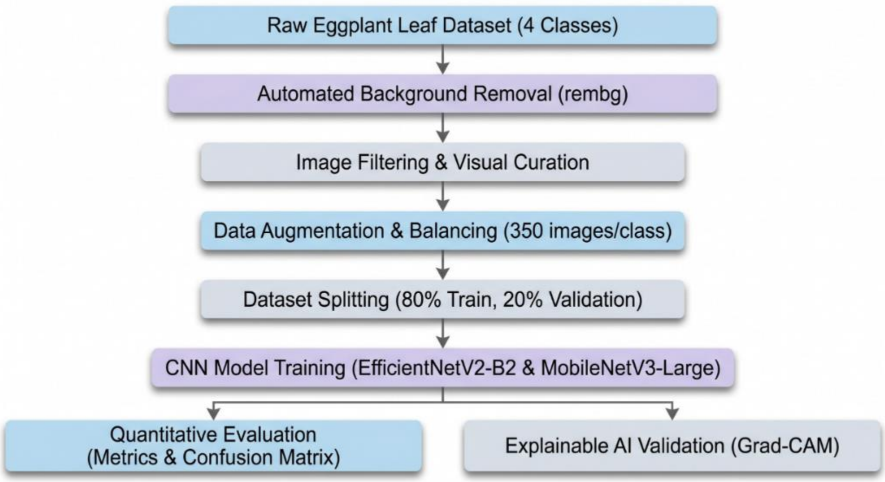


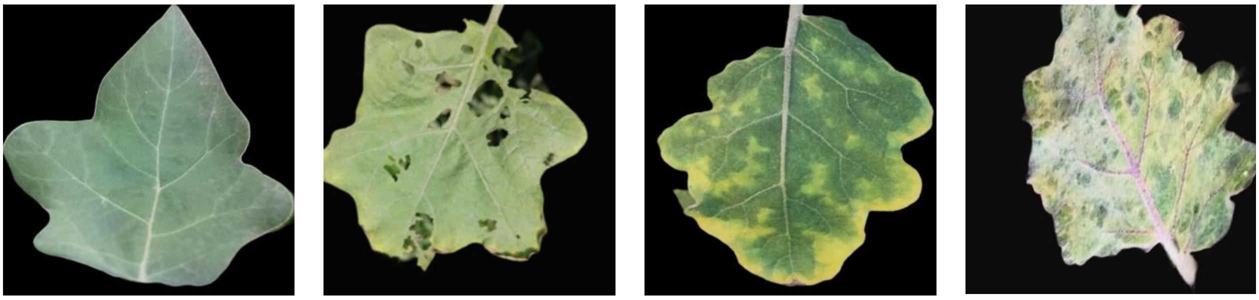
Figure 1. Research Framework

Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

2.1. Dataset Acquisition and Preprocessing

The primary dataset utilized in this research consists of eggplant leaf images categorized into four distinct classes: Healthy Leaf, Insect Pest Disease, Leaf Spot Disease, and Mosaic Virus Disease. The raw eggplant leaf images utilized in this study were originally sourced from a publicly available agricultural repository. To meet the specific requirements of our experimental setup and eliminate environmental classification bias, this raw data was subsequently processed locally through a strict preprocessing pipeline.

Based on existing literature regarding image segmentation, automated background extraction utilizes pre-trained salient object detection networks designed to mathematically distinguish foreground leaf pixels from complex environmental backgrounds [11], [18]. By applying the rembg library, which acts as the applied solution for this stage, the convolutional layers are forced to focus exclusively on the phenotypic characteristics and pathogenic lesions of the leaves rather than learning residual environmental noise [18], Following the background extraction, a pixel-intensity thresholding algorithm and manual visual curation were executed to eliminate corrupted images, heavily shadowed leaves, and samples suffering from severe spatial distortion. This curation step acts as a critical solution to prevent the introduction of label noise that literature confirms could severely degrade the model's decision boundaries [19]. To explicitly demonstrate the result of this preprocessing phase, the visual characteristics of each category, displayed entirely without their original backgrounds, are illustrated in Figure 2.



(a) Healthy Leaf (b) Insect Pest Disease (c) Leaf Spot Disease (d) Mosaic Virus Disease

Figure 2. Representative samples of the eggplant leaf dataset after the automated background removal process: (a) Healthy Leaf, (b) Insect Pest Disease, (c) Leaf Spot Disease, and (d) Mosaic Virus Disease.

As shown in Figure 2, the Healthy Leaf is characterized by a uniform green canvas with structurally intact margins and a complete absence of chlorotic spots. The Insect Pest Disease samples display severe physical perforation, characterized by distinct necrotic holes and jagged edges resulting from chewing insects. The Leaf Spot Disease manifests as localized, circular dark-brown lesions with distinct boundary concentric rings scattered across the leaf blade. Lastly, the Mosaic Virus Disease presents an extensive distribution of irregular yellow-green mottling and chlorotic patches that cause a mottled texture across the remaining leaf structure. These diverse surface patterns represent the complex spatial and color gradients that the convolutional layers must effectively interpret.

The raw eggplant leaf images utilized in this study were originally sourced from a publicly available dataset https://www.kaggle.com/datasets/sujaykapadnis/eggplant-disease-recognition-dataset/code. To meet the specific requirements of our experimental setup, this raw data was subsequently processed locally through our automated background removal and augmentation pipeline to produce the final balanced dataset. The modified dataset is available from the corresponding author upon reasonable request. To ensure the Convolutional Neural Network (CNN) focuses exclusively on the phenotypic characteristics of the leaves rather than the background environment, a rigorous automated background removal process was applied using the rembg library [11], [18].

Following the background extraction, a pixel-intensity thresholding algorithm and manual visual curation were executed to eliminate corrupted images, heavily shadowed leaves, and samples suffering from severe spatial distortion. This step is critical to prevent the introduction of label noise that could degrade the model's decision boundaries [19].

Figure 3. Sample of eggplant leaf image before (left) and after (right) automated background removal.

Bulletin of Informatics and Data Science
Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

2.2. Data Balancing and Augmentation

The initial preprocessing phase resulted in a significant class imbalance, which naturally occurs in agricultural datasets. To prevent the CNN models from developing a classification bias toward the majority class, a targeted data augmentation strategy was implemented. Spatial transformations, including random rotations, horizontal/vertical flipping, and brightness adjustments, were exclusively applied to the minority classes.

Subsequently, a random downsampling technique was applied to the overrepresented classes to unify the sample size. This pipeline successfully yielded a perfectly balanced final dataset consisting of exactly 350 images per class, accumulating to a total of 1,400 clean images.

Table 1. Distribution of the Balanced Eggplant Leaf Dataset

Class Name	Training Set (80%)	Validation Set (20%)	Total Images
Healthy Leaf	280	70	350
Insect Pest Disease	280	70	350
Leaf Spot Disease	280	70	350
Mosaic Virus Disease	280	70	350
Total	1120	280	1400

The uniform distribution depicted in Table 1 establishes a mathematically balanced dataset where no individual class dominates the loss optimization function during training. By partitioning exactly 280 images for the training set and 70 images for the validation set across all categories, the models are forced to develop generalized decision boundaries. This uniform class distribution is a mandatory prerequisite to ensure that the final evaluation metrics reflect true diagnostic capability rather than an artificial bias toward overrepresented classes.

2.3. Proposed CNN Architectures

To evaluate the optimal lightweight model for agricultural edge devices, this study conducted a comparative analysis applying MobileNetV3-Large and EfficientNetV2-B2 as the primary algorithmic solutions.MobileNetV3-Large serves as the baseline for ultra-lightweight architectures. Theoretically, this architecture minimizes computational overhead through depthwise separable convolutions and a hardware-aware network architecture search (NAS) [13]. Literature confirms that integrating Squeeze-and-Excitation (SE) modules within its inverted residual blocks allows the model to recalibrate channel-wise feature responses dynamically, maintaining extremely low latency on mobile CPUs [20], [21]. The input resolution for this model was fixed at 224 × 224 pixels. EfficientNetV2-B2 represents the medium-lightweight category. From a theoretical standpoint, EfficientNetV2 leverages a compound scaling method that uniformly scales network width, depth, and resolution[14]. The inclusion of Fused-MBConv blocks in the early layers replaces standard depthwise convolutions, which existing literature indicates significantly accelerates training speed while maintaining high parameter efficiency [22] To accommodate its structural design, the input resolution was strictly set to 260 × 260 pixels. For a rigorous and fair evaluation, both pre-trained base models were subjected to transfer learning using ImageNet weights, and their top layers were replaced with a customized identical classification head. The classification head consists of a Global Average Pooling 2D layer, a Dense layer with 128 units utilizing a ReLU activation function, a Dropout layer with a 0.3 rate to penalize over-reliance on specific nodes, and a final output Dense layer with 4 units applying a Softmax activation corresponding to the target classes [23].

2.4. Experimental Setup and Evaluation Metrics

The dataset was automatically partitioned into an 80% training set (1,120 images) and a 20% validation set (280 images) using a fixed random seed to ensure reproducibility. Both models were compiled using the Adam optimizer and Categorical Crossentropy loss function. The training process was set for a maximum of 30 epochs with a batch size of 32. An Early Stopping callback with a patience of 6 epochs was integrated to halt the training dynamically and restore the best weights, effectively mitigating overfitting.

The models' performances were quantitatively measured using standard classification metrics derived from the Confusion Matrix: Accuracy, Precision, Recall, and F1-Score [24]. The mathematical formulations for these metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Bulletin of Informatics and Data Science
Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

Where TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) represent the outcomes of the model's predictions against the actual ground truth. Furthermore, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visually interpret the spatial focus of the CNN architectures during the classification process [25].

3. RESULT AND DISCUSSION

This section delivers a comprehensive evaluation of the experimental findings obtained from the trained lightweight Convolutional Neural Network architectures. To fully address the core research problem, the analysis is structured into four progressive evaluation dimensions: a detailed clarification of the applied algorithmic processes, a macro-level investigation of training convergence and continuous loss stability, a micro-level cross-examination of predictive accuracy per individual disease class via confusion matrices, and a qualitative inspection of feature localization using Gradient-weighted Class Activation Mapping (Grad-CAM). Ultimately, a comprehensive discussion is provided to contextualize these findings within the broader spectrum of agricultural informatics and edge device deployment constraints.

3.1. Algorithmic Process and Execution Steps

To fully comprehend the performance metrics generated in this study, it is imperative to first clarify the precise step-by-step algorithmic mechanisms that processed the input data into the final diagnostic predictions. The computational process initiated immediately after the raw dataset was passed through the automated background removal pipeline and balanced via targeted data augmentation. The mathematically balanced dataset of 1,400 curated images was then fed into the customized classification heads of both MobileNetV3-Large and EfficientNetV2-B2. During the forward propagation phase, the input matrices—representing the pixel intensity values of the eggplant leaves—were sequentially processed through the interconnected convolutional layers. In these layers, multiple mathematical filters (kernels) slid across the spatial dimensions of the images, executing dot product operations to extract critical hierarchical features. For EfficientNetV2-B2, the Fused-MBConv blocks actively optimized the extraction of these textural and color gradients, while MobileNetV3-Large utilized Squeeze-and-Excitation modules to dynamically recalibrate the channel-wise feature responses.

Following the extraction of these complex spatial features, the high-dimensional representations were flattened and passed through the fully connected Dense layers. The final output layer, equipped with a Softmax activation function, subsequently calculated the probability distribution across the four mutually exclusive categories: Healthy Leaf, Insect Pest Disease, Leaf Spot Disease, and Mosaic Virus Disease. Once the forward pass generated a diagnostic prediction, the algorithm computed the discrepancy between this predicted probability and the actual ground truth using the Categorical Crossentropy loss function. The resulting loss value mathematically quantified the network's classification error for that specific batch of images.

To minimize this error, the backpropagation algorithm was executed. Utilizing the Adam optimizer, the network calculated the gradients of the loss function with respect to every single weight and bias within the architecture. These computational weights were iteratively updated in the opposite direction of the gradient, forcing the model to continuously refine its feature extraction capabilities. This continuous loop of forward propagation, loss calculation, backpropagation, and weight updating constituted the fundamental learning algorithm of the evaluated models. To strictly prevent the networks from aggressively memorizing the training data—a phenomenon known as overfitting—an Early Stopping callback mechanism was dynamically integrated. This algorithm continuously monitored the validation loss metric at the end of each training epoch. By establishing a patience threshold of exactly six epochs, the algorithmic process was mathematically designed to halt the training iterations the moment the validation loss ceased to improve, subsequently restoring the optimal network weights that yielded the highest generalization capability.

3.2. Training Performance and Convergence Trajectory

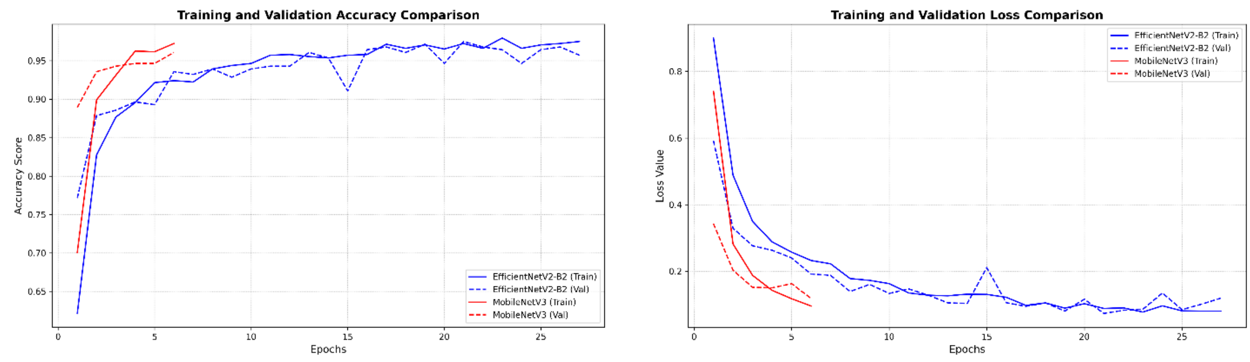


Figure 4. Graphic Accuracy and loss comparison.

Bulletin of Informatics and Data Science
Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

The dynamic training process of both models was rigorously monitored, yielding comprehensive data on learning trajectories, computational efficiency, and overall convergence stability. The graphical representations of the accuracy progression and the continuous loss reduction for both the training and validation datasets are depicted in Figure 4.

As depicted in Figure 4, both algorithmic architectures demonstrated exceptionally rapid learning capabilities during the initial epochs. This accelerated initial convergence can be directly attributed to the robust preprocessing methodology; providing the networks with a completely clean, background-free dataset effectively eliminated the computational burden of processing irrelevant environmental noise, allowing the convolutional layers to immediately isolate pathogenic features.

EfficientNetV2-B2 exhibited a highly stable and mathematically smooth convergence trajectory throughout the training phase. The network systematically optimized its weights without exhibiting erratic fluctuations in the validation loss. The Early Stopping algorithm dynamically halted the EfficientNetV2-B2 training process at epoch 27. Upon halting, the mechanism automatically restored the highly optimized weights generated at epoch 21, at which point the model achieved an exceptional peak validation accuracy of 97.50% alongside a remarkably low validation loss of 0.073. The smoothness of this loss curve fundamentally indicates that the algorithmic integration of Fused-MBConv blocks successfully captured the subtle morphological variations of the foliar diseases without developing a detrimental reliance on the training data noise.

Conversely, MobileNetV3-Large displayed an extremely aggressive and hyper-accelerated learning trajectory. The architecture converged at such an exceptional rate that it triggered the Early Stopping threshold prematurely at epoch 6, forcing the algorithm to restore the best validation weights recorded at the very first epoch. Despite this rapid halting, MobileNetV3-Large successfully achieved a highly competitive validation accuracy of 96.07% and a validation loss of 0.118. While these accuracy metrics are objectively excellent, the severely compressed convergence window suggests that the ultra-lightweight architecture possesses a slight theoretical tendency toward early overfitting when presented with this specific, moderately sized dataset volume.

However, a critical evaluation of computational efficiency fundamentally alters the comparative landscape. MobileNetV3-Large significantly outperformed EfficientNetV2-B2 in raw training speed and processing latency. When executed on a standard computational environment, MobileNetV3-Large required approximately only 26 seconds to complete a full training epoch. In stark contrast, EfficientNetV2-B2 demanded approximately 57 seconds per epoch. This extreme reduction in computational overhead empirically confirms MobileNetV3's architectural superiority regarding raw inference latency, validating its fundamental design philosophy as a model highly optimized for severely constrained mobile processors.

3.2. Classification Metrics and Error Analysis

To transition from a macro-level evaluation of overall accuracy to a micro-level dissection of diagnostic capabilities, comprehensive confusion matrices were generated for the validation dataset. The confusion matrix acts as an essential mathematical tool to cross-examine the true predictive power of the algorithms across individual disease classes, exposing underlying classification biases that are often masked by overall accuracy averages.

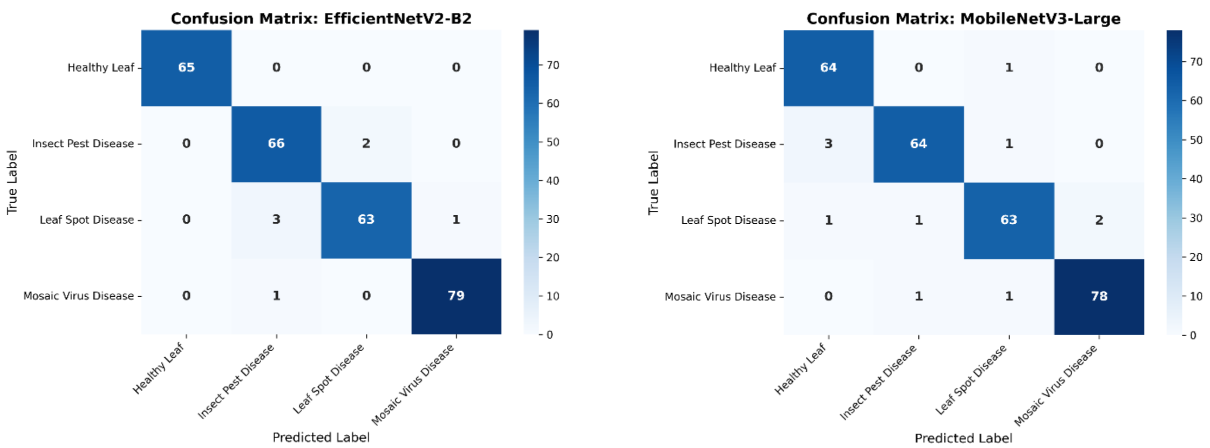


Figure 5. Confusion matrix of EfficientNetV2-B2 and MobileNetV3-Large

Based on the detailed cross-tabulations presented in the confusion matrix, EfficientNetV2-B2 mathematically proved its diagnostic robustness. The most critical finding from this matrix is the model's flawless identification of uninfected plants; EfficientNetV2-B2 correctly classified 100% of the Healthy Leaf samples, securing 65 True Positives out of 65 total healthy instances. In the context of precision agriculture, maintaining a zero False Negative rate for healthy crops is a paramount metric. This absolute precision guarantees that farmers utilizing the diagnostic application will not receive false disease alarms, thereby preventing the unnecessary and ecologically harmful application of expensive chemical treatments. The minor misclassifications generated by the EfficientNetV2-B2 algorithm occurred

Bulletin of Informatics and Data Science
Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

predominantly within the pathogenic categories, specifically between the Leaf Spot Disease and Insect Pest Disease classes. For example, the model incorrectly predicted three Leaf Spot instances as Insect Pest damages. Agronomically, this algorithmic confusion is a highly justifiable error. The physical necrotic tissues and jagged perforations left by chewing insect pests often mimic the dark, localized, and circular lesions caused by severe fungal leaf spots. Despite these minor discrepancies, the model maintained an exceptional classification integrity.

Table 2. Detailed Classification Report of the Evaluated Models

Model Architecture	Class Name	Precision	Recall	F1-Score
EfficientNetV2-B2 (Overall Accuracy: 96.79%)	Healthy Leaf	1.0000	1.0000	1.0000
	Insect Pest Disease	0.9429	0.9706	0.9565
	Leaf Spot Disease	0.9692	0.9403	0.9545
	Mosaic Virus Disease	0.9875	0.9875	0.9875
	Macro Average	0.9749	0.9746	0.9746
MobileNetV3-Large (Overall Accuracy: 96.07%)	Healthy Leaf	0.9412	0.9846	0.9624
	Insect Pest Disease	0.9697	0.9412	0.9552
	Leaf Spot Disease	0.9545	0.9403	0.9474
	Mosaic Virus Disease	0.9750	0.9750	0.9750
	Macro Average	0.9601	0.9603	0.9600

An in-depth review of Table 2 mathematically confirms these observations through standardized metrics. EfficientNetV2-B2 achieved a perfect Precision, Recall, and F1-Score of 1.0000 for the Healthy Leaf class. Furthermore, it secured exceptional F1-Scores for the Mosaic Virus Disease (0.9875), Insect Pest Disease (0.9565), and Leaf Spot Disease (0.9545), culminating in an outstanding Macro Average F1-Score of 0.9746 across all categories.

MobileNetV3-Large, despite its highly competitive overall accuracy of 96.07%, exhibited mathematical vulnerabilities that pose a higher risk regarding diagnostic reliability. An analysis of its confusion matrix reveals that the model generated critical False Negative errors. Specifically, the algorithm misclassified three separate instances of Insect Pest Disease and one instance of Leaf Spot Disease as perfectly Healthy Leaves. In a real-world agricultural deployment, classifying actively infected and damaged crops as completely healthy is a fundamentally detrimental algorithmic failure. These False Negatives actively prevent the timely intervention of critical pesticide or fungicide applications, which can rapidly lead to widespread pathogenic transmission and catastrophic crop failure.

Consequently, the calculated Recall metric for the Insect Pest Disease class in MobileNetV3-Large dropped to 0.9412. Therefore, from a strict agronomic reliability standpoint, the slightly larger architectural parameter size of EfficientNetV2-B2 (which operates with approximately 8.7 million parameters) provides an absolutely necessary and mathematically justified trade-off. The superior safety margin and absolute avoidance of dangerous False Negatives render it significantly safer for diagnostic precision compared to the ultra-lightweight MobileNetV3-Large (which operates with approximately 5.4 million parameters).

3.4. Interpretability through Grad-CAM Explanable AI Analysis

A pervasive and highly valid criticism of implementing deep learning models in critical sectors like agriculture is the inherent "black-box" nature of convolutional networks. While confusion matrices prove that a model is highly accurate, they fail to prove how the model arrived at its conclusions. To mathematically validate that the evaluated CNN architectures were genuinely identifying true pathogenic features rather than learning residual environmental noise or artificial pixel artifacts, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied as a crucial explainability validation step.

The Grad-CAM algorithm works by utilizing the spatial gradient information flowing entirely into the final convolutional layer of the network. By computing the importance of each individual feature map for a specific target class, the algorithm generates a coarse visual heatmap that highlights the exact spatial regions of the input image that predominantly influenced the network's final prediction.

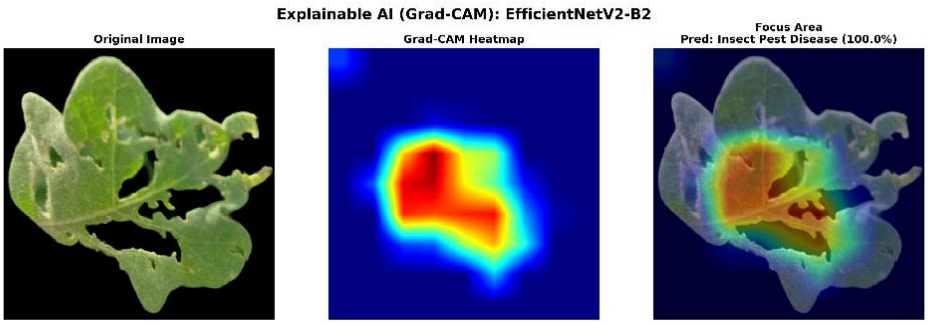


Figure 6. Grad-cam visualization of EfficientNetV2-B2

Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

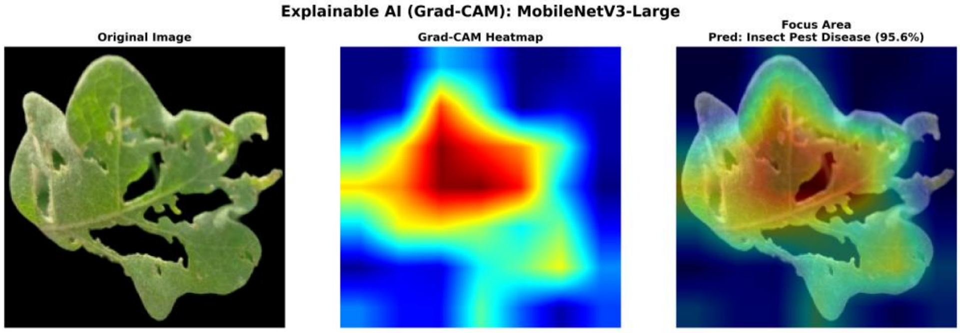


Figure 7. Grad-cam visualization of MobileNetV3-Large

The visual Grad-CAM heatmaps unequivocally confirm the profound efficacy of the automated background removal preprocessing step. In all generated visualizations, the intense activation zones (indicated by deep red and orange spectrums) are strictly confined within the biological perimeters of the leaf structures. The complete absence of background pixels fundamentally forced the convolutional layers of both models to strictly extract structural, textural, and color gradients exclusively from the foliar surfaces.

However, a meticulous qualitative evaluation of the representative validation samples illustrated in Figures 6 and 7 reveals distinctly different spatial attention mechanisms between the two evaluated architectures. EfficientNetV2-B2 exhibits a highly aggressive, deeply localized, and incredibly precise spatial focus. As demonstrated in Figure 6, the network's maximum activation zones directly pinpoint the specific morphological disease features—in this case, anchoring fiercely onto the jagged necrotic holes and severely damaged margins caused by insect pests. The algorithm aggressively ignores the surrounding healthy green tissues, proving a highly sophisticated understanding of pathological lesions.

Conversely, the Grad-CAM visualization for MobileNetV3-Large displayed in Figure 7 illustrates a broader, more diffuse, and significantly smoother attention map. While the ultra-lightweight model still accurately maintains its peak activation over the critical pathogenic lesions, its gradient focus encompasses a much larger overall morphology of the leaf structure. This broader visual attention explains the model's slightly diminished diagnostic precision; because the network dilutes its focus across a wider spatial area rather than isolating the microscopic damage, it occasionally misinterprets minor lesions, leading to the previously discussed False Negative misclassifications. Nonetheless, this vital explainability aspect fundamentally verifies that the high quantitative accuracy scores achieved by both models are phenomenologically valid, mathematically sound, and directly tied to biological reality rather than the result of an artificial dataset bias.

3.5. Comprehensive Discussion and Deployment Implications

The integration of these diverse analytical metrics paints a definitive picture regarding the deployment of lightweight CNNs for precision agriculture. The primary objective of this research was to construct a methodology capable of overcoming severe classification bias caused by complex background noise and data imbalance. The automated preprocessing pipeline successfully achieved this, as validated by the pristine Grad-CAM heatmaps and the perfectly symmetrical confusion matrices where no class disproportionately dominated the predictions.

The comparative evaluation between the two architectural paradigms reveals a critical operational dichotomy. EfficientNetV2-B2 has empirically proven itself as the optimal algorithm for precision-critical agricultural diagnostics. Its medium-lightweight structure possesses enough computational depth to thoroughly differentiate between highly similar pathogenic symptoms, ultimately achieving 100% precision on healthy samples and eliminating the existential risk of False Negatives. The stability of its loss convergence ensures that the model can be deployed safely without the risk of erratic diagnostic behavior.

However, the extreme computational constraints of modern agricultural edge devices—such as older generation smartphones utilized by farmers in remote regions—cannot be entirely ignored. In scenarios where raw inference speed, battery conservation, and minimal memory footprints are the absolute determining factors for deployment, MobileNetV3-Large remains a formidable and highly capable alternative. Despite its slight propensity for False Negatives, its ability to classify highly complex diseases with an accuracy of 96.07% while demanding less than half the processing latency of EfficientNetV2-B2 makes it an incredibly valuable asset for ultra-low-end hardware ecosystems. The decision to deploy either model ultimately depends on the specific hardware limitations of the end-user and the agronomical risk tolerance regarding automated misdiagnoses.

4. CONCLUSION

This research successfully demonstrates that integrating rigorous automated background extraction with mathematical data balancing significantly enhances the diagnostic integrity of lightweight Convolutional Neural Networks in identifying eggplant foliar diseases. By systematically processing raw agricultural images to remove complex

ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

environmental noise and generating a perfectly uniform dataset of 1,400 instances across four categories, the foundational classification bias prevalent in raw datasets was entirely eradicated. The empirical comparative evaluation subsequently established clear operational paradigms for deploying these diagnostic algorithms on computationally constrained edge devices. EfficientNetV2-B2 decisively proved to be the most optimal architecture for precision-critical agricultural applications, characterized by a highly stable learning trajectory and an outstanding peak validation accuracy of 97.50%. Crucially, this medium-lightweight model achieved flawless recognition of uninfected plants, completely eliminating the agronomical risk of False Negative predictions that could otherwise disrupt farm management protocols. Conversely, while MobileNetV3-Large exhibited a slightly diminished diagnostic safety margin with a 96.07% accuracy rate and a propensity for minor misclassifications, its exceptionally rapid inference latency firmly positions it as a highly viable alternative for deployment on ultra-low-end mobile processors. The interpretability validation via Grad-CAM further cemented these findings by proving that both mathematical models correctly anchored their spatial attention exclusively onto actual pathological lesions rather than artificial pixel anomalies. Moving forward, subsequent research initiatives must prioritize the direct compilation of these trained architectural weights into an Android-based mobile framework using TensorFlow Lite to actively measure real-time battery consumption and processing thermal dynamics under actual field conditions, while simultaneously expanding the training corpus to encompass varying chronometric stages of infection severity.

REFERENCES

[1] A. Krishnaswamy Rangarajan and R. Purushothaman, “Disease Classification in Eggplant Using Pre-trained VGG16 and MSVM,” Sci. Rep., vol. 10, no. 1, p. 2322, 2020, doi: 10.1038/s41598-020-59108-x.

[2] S. Lika, Y. Pernando, and A. Kurniawan, “Lightweight Deep Learning for Object Detection on Mobile Device,” Bull. Informatics Data Sci., vol. 2, no. 2, pp. 106–114, 2023, doi: 10.61944/bids.v2i2.82.

[3] R. E. Saragih, A. R. Purnajaya, I. Syafrinal, and Y. Pernando, “Mango and Banana Ripeness Detection based on Lightweight YOLOv8,” J. Buana Inform., vol. 15, no. 2, pp. 79–88, 2024. doi: 10.24002/jbi.v15i2.8895

[4] Y. Pernando, E. Febrianti, I. Syafrinal, Y. Roza, and U. Afifah, “Deep Learning for Faces on Orphanage Children Face Detection,” JURTEKSI (Jurnal Teknol. dan Sist. Informasi), vol. 9, no. 1, pp. 25–32, Dec. 2022, doi: 10.33330/jurteksi.v9i1.1858.

[5] Y. Pernando, Y. Heryadi, I. Sonata, L. A. Wulandhari, and A. S. Girsang, “Use of CNN in Diagnosing Banana Leaf Diseases an Agricultural Technology-Based Solution BT - Advances in Smart Knowledge Computing: Towards Post Artificial Intelligence Era,” in Advances in Smart Knowledge Computing, F. Lumban Gaol, T. Matsuo, and T. Ito, Eds., Cham: Springer Nature Switzerland, 2026, pp. 243–255. doi: 10.1007/978-3-032-01133-6_17.

[6] R. E. Saragih, Y. Pernando, A. J. Santoso, Y. Roza, M. KH, and I. Verdian, “Optimizing The Sorting Of Barangan Bananas Through Hybrid And Transfer Learning Approaches,” ICIC Express Lett, vol. 19, no. 4, pp. 459–466, 2025.

[7] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fredes, and A. Valenzuela, “A Review of Convolutional Neural Network Applied to Fruit Image Processing,” Applied Sciences, vol 10, no 10, 2020. doi: 10.3390/app10103443.

[8] Masparudin, I. Fitri, and Sumijan, “Pengembangan Sistem Klasifikasi Buah Apel Menggunakan Convolutional Neural Network (CNN) Arsitektur MobileNet pada Platform Android,” Sist. J. Sist. Inf., vol. 13, no. 1, pp. 230–243, 2024, doi: https://doi.org/10.32520/stmsi.v13i1.3533.

[9] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, “Review of Image Classification Algorithms Based on Convolutional Neural Networks,” Remote Sensing, vol 13, no 22, 2021. doi: 10.3390/rs13224712.

[10] I. Goodfellow, B. Yoshua, and A. Courville, Deep Learning. The MIT Press, 2016.

[11] F. Zhu, Y. Sun, Y. Zhang, W. Zhang, and J. Qi, “An Improved MobileNetV3 Mushroom Quality Classification Model Using Images with Complex Backgrounds,” Agronomy, vol 13, no 12, 2023. doi: 10.3390/agronomy13122924.

[12] A. R. Purnajaya and M. Masparudin, “Evaluating Resampling Methods for Imbalanced Necrosis Classification on CT Scans,” JUTEKOM (Jurnal Teknol. dan Ilmu Komputer), vol. 2, no. 2, 2026, doi: 10.65258/jutekom.v2.i2.56.

[13] G. Chu et al., “Searching for MobileNetV3,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.

[14] M. Tan and Q. V Le, “EfficientNetV2 : Smaller Models and Faster Training,” in International Conference on Machine Learning (ICML), 2021, pp. 8786–8798.

[15] J. Chen, D. Zhang, M. Suzauddola, Y. A. Nanehkaran, and Y. Sun, “Identification of plant disease images via a squeeze-and-excitation MobileNet model and twice transfer learning,” IET Image Process., vol. 15, no. 5, pp. 1115–1127, Apr. 2021, doi: https://doi.org/10.1049/ipr2.12090.

[16] J. Zhang, X. Yang, X. Fu, B. Wang, and H. Li, “LDL-MobileNetV3S : an enhanced lightweight MobileNetV3-small model for potato leaf disease diagnosis through multi-module fusion,” Front. Plant Sci., vol. 16, no. October, pp. 1–19, 2025, doi: 10.3389/fpls.2025.1656731.

[17] J. Han, Z. Wu, Y. Ding, Y. Guo, and R. Fu, “A privacy-protecting eggplant disease detection framework based on the YOLOv11n-12D model.,” Front. Plant Sci., vol. 16, p. 1634408, 2025, doi: 10.3389/fpls.2025.1634408.

[18] R. Utomo, M. Idhom, and T. Trimono, “Rice Leaf Disease Classification Using EfficientNetV2 with Hyperparameter Tuning,” bit-Tech, vol. 8, no. 2, pp. 2038–2047, Dec. 2025, doi: 10.32877/bt.v8i2.3194.

[19] M. Masparudin, A. Abdullah, and U. Usman, “Sistem prediksi kualitas santan kelapa menggunakan nearest mean classifier (NMC),” Sist. J. Sist. Inf., vol. 9, no. 3, pp. 646–655, 2020, doi: https://doi.org/10.32520/stmsi.v9i3.1015.

[20] W. Wang, Y. Hu, T. Zou, H. Liu, J. Wang, and X. Wang, “A New Image Classification Approach via Improved MobileNet Models with Local Receptive Field Expansion in Shallow Layers.,” Comput. Intell. Neurosci., vol. 2020, p. 8817849, 2020, doi: 10.1155/2020/8817849.

[21] Y. Kaya and E. Gürsoy, “A MobileNet-based CNN model with a novel fine-tuning mechanism for COVID-19 infection detection,” Soft Comput., vol. 27, no. 9, pp. 5521–5535, 2023, doi: 10.1007/s00500-022-07798-y.

Vol. 5 No. 1, May 2026, Page 67−76
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v5i1.153
https://ejurnal.pdsi.or.id/index.php/bids/index

[22] S. Zhu et al., “Screening of Common Retinal Diseases Using Six-Category Models Based on EfficientNet,” Front. Med., vol. 9, p. 808402, 2022, doi: 10.3389/fmed.2022.808402.
[23] F. Chollet, Deep Learning with Python. Manning Publications, 2021.
[24] P. Galdi and R. Tagliaferri, “Data Mining: Accuracy and Error Measures for Classification and Prediction,” Elsevier, 2018. doi: 10.1016/B978-0-12-809633-8.20474-3.
[25] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.