Robusta Coffee Disease Classification Using Optimized MobileNetV3 and DenseNet169

Desi Puspita*(ID), Inka Rizky Padya(ID), Efan(ID), Ferry Putrawansyah(ID)

Technology Pagar Alam Institute, Pagar Alam 31517, Indonesia

Corresponding Author Email: desiofira1@gmail.com

ABSTRACT

Robusta coffee productivity in Pagar Alam City has declined by nearly 50% compared to the previous year. Typically, a coffee tree produces 1–5 kg of cherries per tree, but in 2024 production dropped to only 100–300 grams. This decline is mainly due to difficulties in accurately identifying leaf diseases, leading to inappropriate pesticide use and reduced yields. Hence, an accurate image-based disease detection system is urgently required to support precision agriculture. This study optimizes the Convolutional Neural Network (CNN) architecture for classifying Robusta coffee leaf diseases using MobileNetV3 and DenseNet169 with transfer learning. The dataset consists of healthy and diseased coffee leaf images (rust, leaf spot, sooty mold). Data preprocessing includes normalization, augmentation, and resizing, with a 70:30 split for training and testing. Performance is evaluated using accuracy, precision, recall, F1-score, and AUC metrics. The optimized model achieved a precision of 0.98, recall of 98%, F1-score of 98%, and AUC of 0.98. It reached the highest accuracy for leaf spot and rust (1.00) and sooty mold (0.93). MobileNetV3 demonstrated superior accuracy and robustness, while DenseNet169 provided faster convergence and efficient feature propagation, yielding a model that balances accuracy and speed. This approach enables farmers to make timely, data-driven decisions in managing coffee plant diseases, supporting sustainable coffee production in Pagar Alam City. Future studies should expand datasets and develop mobile-based real-time detection systems.

1. INTRODUCTION

Coffee is one of Indonesia's most strategic plantation commodities, holding substantial economic and social importance. The country is recognized as one of the world's largest coffee producers, cultivating two main varieties—Arabica and Robusta. One of the leading Robusta-producing regions is Pagar Alam City, located at the foot of Mount Dempo in South Sumatra Province. The geographical conditions and cool climate of this highland area provide an ideal environment for growing Robusta coffee, known for its distinctive flavor and aroma. For the local community, coffee cultivation serves not only as a cultural heritage but also as the primary economic backbone, with most residents depending on the agricultural sector, especially coffee farming [1].

Despite its strong potential, Pagar Alam's coffee industry faces serious challenges, particularly in disease detection and management. Several leaf diseases—such as coffee leaf rust.

(Hemileia vastatrix), sooty mold (Capnodium spp.), and leaf spot (Cercospora coffeicola)—pose significant threats to coffee productivity. These diseases cause leaf discoloration, defoliation, and impaired photosynthesis, leading to substantial yield reductions [2].

The main problem farmers encounter is not merely the spread of these diseases, but rather the inaccuracy in identifying and classifying them. Most farmers in Pagar Alam still rely on manual observation, assessing symptoms visually without technological assistance. This traditional approach depends heavily on individual experience and is often inaccurate because different diseases can exhibit similar visual symptoms. For instance, leaf spots caused by Cercospora may be mistaken for early symptoms of Hemileia vastatrix. As a result, farmers often apply inappropriate pesticides or treatments, which can further damage the plants instead of curing them [3].

This issue is further compounded by limited access to expert knowledge and diagnostic technology. Farmers lack a rapid and accurate system for disease identification, making timely decision-making difficult. In plant disease management, speed and accuracy in classification are critical to determining effective control measures. Therefore, it is essential to develop a technology-based solution that enables farmers to recognize leaf diseases quickly and accurately, even in resource-limited field conditions. With rapid advancements in information technology, the agricultural sector has entered a new era of smart farming, where Artificial Intelligence (AI) is used to address traditional agricultural challenges. One of the most promising AI techniques for image-based disease identification is the Convolutional Neural Network (CNN), a deep learning architecture specialized in visual pattern recognition [4].

CNNs mimic the way the human brain processes visual information, automatically learning hierarchical features such as texture, color, and shape from images. Unlike traditional

machine learning methods that require manual feature extraction, CNNs can autonomously identify complex visual patterns directly from raw data. This makes them highly effective for image-based tasks such as leaf disease classification. In the context of coffee leaves, CNNs can classify images into multiple categories—such as healthy leaves, leaf rust, leaf spot, or sooty mold—based on visual patterns. However, conventional CNN models like AlexNet, VGG16, and ResNet often require substantial computational resources and long training times, making them unsuitable for mobile or real-time applications in the field. To overcome these limitations, recent research has focused on developing lightweight CNN architectures that retain high accuracy while reducing computational costs. Among the most promising are MobileNetV3 and DenseNet169, which offer complementary advantages in terms of computational efficiency, speed, and model stability [5].

MobileNetV3, developed by Google, is designed for low-resource environments where computational efficiency and speed are crucial. It utilizes depthwise separable convolutions, inverted residual blocks, and squeeze-and-excitation (SE) modules to reduce the number of parameters significantly. As a result, MobileNetV3 achieves high inference speed and compact model size without compromising accuracy—making it ideal for on-field applications such as smartphone-based disease detection systems or smart agricultural cameras.

In contrast, DenseNet169 (Densely Connected Convolutional Network) introduces a unique approach where each layer is connected directly to every other layer in a feed-forward fashion. This dense connectivity allows the model to reuse previously learned features, ensuring efficient gradient propagation, faster convergence, and reduced overfitting. Despite its depth (169 layers), DenseNet is parameter-efficient and provides high classification accuracy with less redundant computation [6].

When combined, MobileNetV3's accuracy and lightweight design complement DenseNet169's speed and feature reuse, producing an optimized model for image classification that balances accuracy, efficiency, and scalability. By leveraging transfer learning—using pre-trained weights from ImageNet—this hybrid model can adapt quickly to coffee leaf datasets while minimizing training time and computational cost. From the above discussion, it can be identified that the core problem in Pagar Alam's coffee production is not merely declining yield, but rather the inefficient and inaccurate classification of coffee leaf diseases [7-24]. Misidentification leads to the inappropriate use of pesticides, delayed response, and overall management inefficiency.

Based on this, the key research problems formulated are:

1. How can a CNN-based classification model be designed to accurately identify various coffee leaf diseases?

2. How can CNN architecture be optimized to achieve high-speed classification with computational efficiency?

3. How can the strengths of MobileNetV3 and DenseNet169 be integrated into a hybrid model that provides both accuracy and processing speed?

4. How can this optimized model assist farmers in making rapid and precise decisions regarding disease management?

These questions form the foundation of this study, which aims to provide a technological solution that is both scientifically robust and practically applicable for smallholder coffee farmers.

The main objective of this research is to develop and optimize a CNN model based on MobileNetV3 and DenseNet169 to enhance the accuracy and speed of Robusta coffee leaf disease classification. The specific objectives are:

1. To implement transfer learning techniques using MobileNetV3 and DenseNet169 architectures on coffee leaf image datasets.

2. To evaluate model performance using metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under Curve).

3. To analyze the effect of combining both architectures on training speed and computational efficiency.

4. To produce a lightweight, high-speed model suitable for mobile or field-based plant disease detection systems.

Numerous studies have applied CNNs in plant disease classification with promising results. For instance, Mohanty et al. [25] employed AlexNet and GoogLeNet to classify diseases in 14 plant species, achieving an accuracy of 99.35%. Ferentinos [26] demonstrated the capability of CNNs to identify plant diseases with over 98% accuracy using large image datasets. However, these models typically require powerful hardware and long processing times, which limit their real-world usability.

Recent research has shifted toward lightweight deep learning architectures for agricultural applications. For example, applied MobileNet for rice disease classification and achieved 97% accuracy with faster training compared to heavier models like ResNet. Meanwhile, DenseNet has been shown to improve gradient flow and reduce overfitting, particularly for smaller datasets.

Despite these advancements, studies combining MobileNetV3 and DenseNet169 for coffee leaf disease classification remain scarce. This research seeks to fill that gap by developing a hybrid architecture that combines MobileNet's efficiency and DenseNet's accuracy, offering a novel contribution to precision agriculture [8].

Theoretically, this study contributes to the development of CNN optimization techniques through the integration of two distinct architectures: MobileNetV3, known for lightweight efficiency, and DenseNet169, recognized for deep feature reuse and fast convergence. The hybridization of these two models is expected to yield a system that is both accurate and computationally efficient [9].

Practically, this research provides a real-world technological solution for farmers and agricultural practitioners. The proposed system can help farmers quickly diagnose coffee leaf diseases using images captured from smartphones or field cameras. This eliminates the need for expert consultation, allowing farmers to independently and efficiently manage disease control. Moreover, the outcomes of this study can support the development of mobile-based and IoT-integrated applications for real-time field monitoring [10]. By collecting and analyzing plant disease data continuously, such systems could transform traditional coffee farming into data-driven, intelligent agriculture. Thus, this research contributes not only to the academic field of machine learning but also to the digital transformation of agriculture toward sustainability and resilience.

This research offers benefits in two major dimensions:

1. Scientific Benefit: It provides a reference for developing hybrid CNN architectures that enhance the balance between accuracy, efficiency, and scalability in plant disease classification.

2. Practical Benefit: It serves as a foundation for creating image-based coffee disease detection applications

that are fast, user-friendly, and accurate—empowering farmers to make informed decisions in real time.

High accuracy with a relatively small number of parameters, compared to other architectures of similar depth, such as ResNet. DenseNet169 is commonly used for tasks such as image classification, object detection, and segmentation, and is quite popular in the medical field due to its ability to recognize important features in images well, even with relatively small datasets [11]. Overall, DenseNet169 is a deep network architecture, yet efficiently designed, with a focus on maximizing the use of information between layers to produce accurate and stable models during training. The flow can be seen in Figure 1.
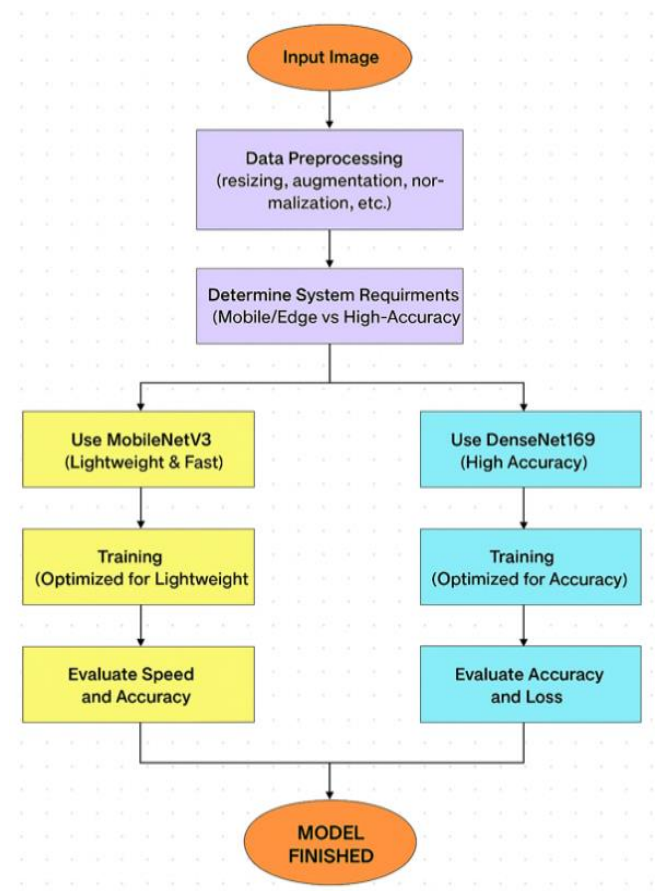

Figure 1. The process flow produces fast classification with high accuracy

Additionally, this study may inspire further research on similar methods for other crops such as tea, cocoa, or rice, advancing broader applications of AI in agriculture.

In summary, the main challenge in Robusta coffee cultivation in Pagar Alam is no longer limited to production decline but rather lies in the inefficiency of disease classification and identification systems that hinder timely and appropriate management actions [12].

By employing deep learning-based approaches, particularly through the combination of MobileNetV3 and DenseNet169, it is possible to develop an intelligent, high-performance classification system capable of identifying coffee leaf diseases with high accuracy and speed. MobileNetV3 contributes by offering lightweight precision and fast inference, while DenseNet169 ensures rapid convergence and feature propagation stability. The synergy between these architectures forms a robust hybrid model that can assist farmers in detecting diseases promptly and taking proper remedial measures such as pesticide application, pruning, or isolation. Ultimately, this research lays the groundwork for applying AI-based disease detection technologies to support sustainable Robusta coffee production in Pagar Alam and strengthen Indonesia's position as a leading producer of high-quality coffee in the global market [13].

The explanation is: MobileNetV3 is ideal for reducing model size and accelerating inference, such as for mobile devices, edge computing, or IoT applications, while DenseNet169 focuses on feature utilization and parameter efficiency. Despite having many layers, DenseNet avoids parameter waste by connecting all layers directly. This results in: Extreme feature reuse: enriching information without having to constantly add parameters. Smoother gradient flow: helping training become more stable, especially for deep models. High training efficiency and accuracy, even on limited datasets. DenseNet169 is suitable for optimization when the model is used in systems with moderate computing capacity but still requires high precision, for example for image-based medical diagnosis or complex pattern recognition [14]. The targeted schemes can be seen in Table 1.

Tabel 1. Aspect MobileNetV3 and DenseNet169

Aspects	MobileNetV3	DenseNet169
Objectives	Speed & Efficiency	Accuracy & Feature Utilization
Model Size	Small	Medium
Inference Time	Fast	Slower than MobileNetV3
Accuracy	Fairly Good	Higher in Most Cases
Suitable for	Mobile, real-time, IoT	Complex Image Classification

2. METHODOLOGY

This research proposes an optimized deep learning framework for coffee leaf disease classification using a hybrid architecture that combines MobileNetV3 and DenseNet169. The integration of these two models aims to achieve high accuracy, fast inference speed, and computational efficiency, making the model suitable for real-time field applications in smart agriculture [15].

The research process consists of several key stages: (1) data acquisition and preprocessing, (2) model design and transfer learning, (3) training and validation, and (4) model evaluation and performance analysis.

The dataset used in this study consists of coffee leaf images collected from various Robusta coffee plantations in Pagar Alam City, South Sumatra, Indonesia. Images were taken using digital cameras under natural light conditions, capturing leaves from different angles and magnifications (40×, 100×, 400×).

The dataset includes four main classes:
1. Healthy leaves
2. Coffee leaf rust (Hemileia vastatrix)
3. Leaf spot (Cercospora coffeicola)
4. Sooty mold (Capnodium spp.)

Each image was resized to 224 × 224 × 3 pixels to match the input dimension of both MobileNetV3 and DenseNet169.

Figure 2 illustrates the workflow of the proposed Convolutional Neural Network (CNN) architecture based on MobileNetV3 for automatic classification of coffee leaf

diseases. The model aims to identify and categorize different disease types in Robusta coffee leaves through image-based deep learning techniques. The process begins with an input image of the coffee leaf, captured at multiple magnification levels (40×, 100×, 400×) to enhance detail visibility. Each image undergoes a preprocessing stage, including normalization, augmentation, and resizing to a standardized dimension of $224 \times 224 \times 3$ pixels, ensuring compatibility with the MobileNetV3 architecture.
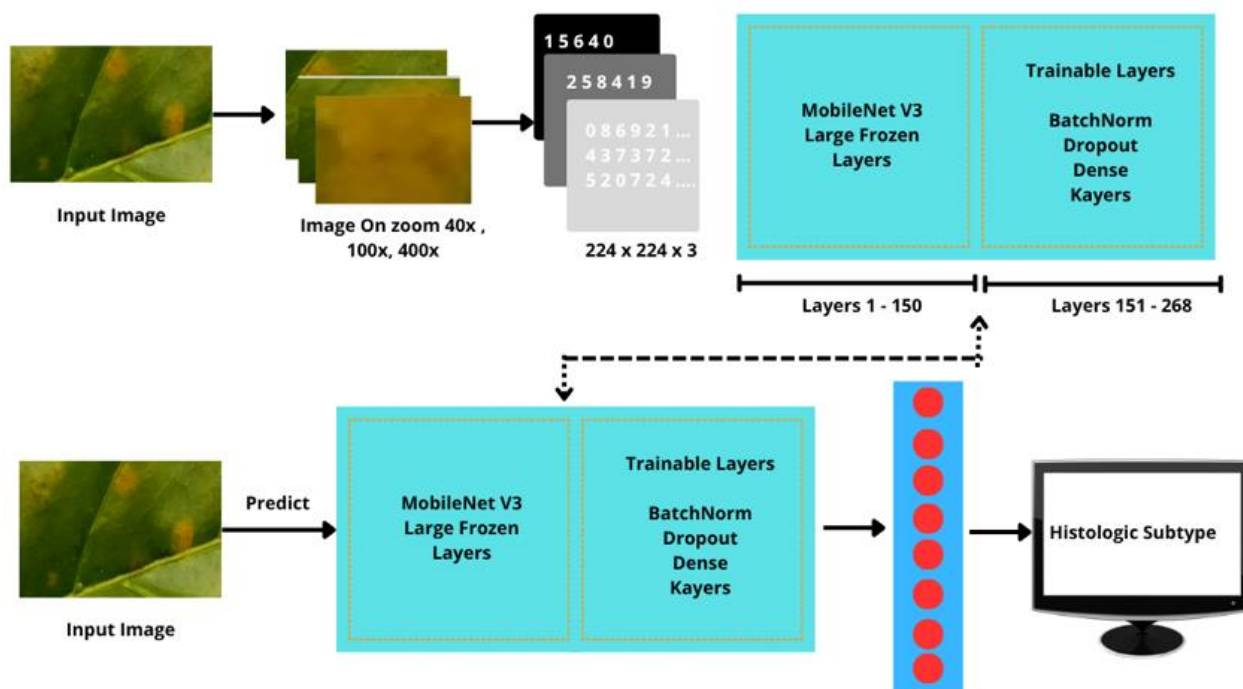

Figure 2. Schema CNN-MobileNetV3-DenseNet169

The model is composed of two main components:
1.   Frozen Layers (Layers 1–150 – MobileNetV3 Backbone):
These layers act as a feature extractor, pretrained on the ImageNet dataset. They identify fundamental visual patterns such as texture, color, and shape. During transfer learning, these layers remain frozen to retain general image representation capabilities.
2.   Trainable Layers (Layers 151–268 – BatchNorm, Dropout, Dense):
These layers are fine-tuned using the coffee leaf dataset. Batch Normalization stabilizes learning, Dropout prevents overfitting, and Dense Layers perform final classification into specific disease categories.

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets using a stratified split to maintain class balance. To prevent overfitting and improve model generalization, data augmentation techniques were applied, including rotation, horizontal/vertical flipping, random cropping, brightness adjustment, and Gaussian noise addition. CNN optimization with MobileNetV3 is more focused on shrinking the model and speeding up processing, while DenseNet169 is used to maximize feature learning without sacrificing training stability. The choice between the two depends on the application context: whether efficiency or accuracy is more important [19, 20]. In some cases, the two can be combined in different systems: MobileNetV3 on the other hand. The client for speed, and DenseNet169 on the server for in-depth analysis. Here's the flowchart. The stages in this research are:

During training, transfer learning is applied by freezing the early convolutional layers while fine-tuning the trainable dense layers. This allows the model to achieve high accuracy while maintaining computational efficiency. Once training is completed, the model receives new input images and predicts their histologic subtype or disease class (e.g., healthy, leaf rust, leaf spot, or sooty mold). This approach balances accuracy and speed by combining MobileNetV3's lightweight structure with the adaptability of fine-tuned dense layers, making it suitable for real-time and field-based disease detection systems [21].

This research proposes an optimized deep learning framework for coffee leaf disease classification using a hybrid architecture that combines MobileNetV3 and DenseNet169. The integration of these two models aims to achieve high accuracy, fast inference speed, and computational efficiency, making the model suitable for real-time field applications in smart agriculture. The research process consists of several key stages: (1) data acquisition and preprocessing, (2) model design and transfer learning, (3) training and validation, and (4) model evaluation and performance analysis.

The dataset used in this study consists of coffee leaf images collected from various Robusta coffee plantations in Pagar Alam City, South Sumatra, Indonesia. Images were taken using digital cameras under natural light conditions, capturing leaves from different angles and magnifications (40×, 100×, 400×).

The dataset includes four main classes:
1.   Healthy leaves
2.   Coffee leaf rust (Hemileia vastatrix)
3.   Leaf spot (Cercospora coffeicola)
4.   Sooty mold (Capnodium spp.)
Each image was resized to $224 \times 224 \times 3$ pixels to match the input dimension of both MobileNetV3 and DenseNet169.

To prevent overfitting and improve model generalization, data augmentation techniques were applied, including rotation, horizontal/vertical flipping, random cropping, brightness adjustment, and Gaussian noise addition [22].

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets using a stratified split to maintain class balance.

Preprocessing plays a crucial role in improving model performance. The following steps were implemented before feeding the images into the CNN architecture:

1. Normalization: All pixel values were scaled between 0 and 1 using min–max normalization.

2. Augmentation: Random transformations such as rotation (±25°), horizontal/vertical flips, and zoom (0.8–1.2) were applied to enhance robustness.

3. Noise Reduction: Gaussian filters were used to reduce background noise and emphasize disease spots.

4. Resizing: Each image was resized to 224 × 224 pixels to ensure uniform input size across the dataset.

These steps ensured that the model could recognize disease features regardless of lighting, angle, or environmental variation.

The proposed hybrid architecture integrates MobileNetV3 and DenseNet169 through a transfer learning approach. The model is divided into three main components:

MobileNetV3 serves as the base model responsible for extracting low- and mid-level visual features such as leaf color, vein structure, and lesion patterns [23, 24].

Key components include:

1. Depthwise Separable Convolution: Reduces computational cost by splitting spatial and channel-wise operations.

2. Inverted Residual Blocks: Preserve important spatial information while keeping the network lightweight.

3. Squeeze-and-Excitation (SE) Modules: Improve feature recalibration by emphasizing relevant channels.

Layers 1–150 of MobileNetV3 were frozen during training to retain the pretrained ImageNet weights, ensuring efficient feature extraction and faster convergence. The output features from MobileNetV3 are passed into DenseNet169, which enhances feature propagation and gradient flow [25]. DenseNet connects each layer to every other layer in a feed-forward manner, allowing the network to reuse features and improve learning efficiency.

Advantages of DenseNet169 include:

1. Fast Convergence: Gradient reuse accelerates training.

2. Compact Parameterization: Requires fewer parameters due to dense connections.

3. High Stability: Minimizes overfitting in small datasets.

The final stage consists of trainable layers that perform disease classification. These include:

1. Batch Normalization: Stabilizes and accelerates training.

2. Dropout Layer (0.5): Prevents overfitting by randomly deactivating neurons.

3. Dense Layers: Map extracted features to disease categories.

4. Softmax Activation: Produces final probability scores for each disease class.

Transfer learning was employed to leverage pretrained ImageNet weights from both base models. This approach allows the model to start from a well-established feature representation rather than training from scratch, significantly reducing training time and computational cost.

1. Optimizer: Adam optimizer with a learning rate of 0.0001.

2. Loss Function: Categorical Cross-Entropy, suitable for multiclass classification.

3. Batch Size: 32 images per batch.

4. Epochs: 50 iterations with early stopping to prevent overfitting.

5. Hardware: Model training was performed using an NVIDIA RTX 3060 GPU with 12 GB VRAM.

During training, the frozen MobileNetV3 layers provided stable feature extraction, while DenseNet169 and the classification head were fine-tuned to adapt to specific coffee leaf disease features [26]. The model's performance was assessed using multiple metrics to provide a comprehensive evaluation:

1. Accuracy: Represents the percentage of images that are classified correctly.

2. Precision: Reflects how accurate the model is when predicting positive classes.

3. Recall (Sensitivity): Indicates the model's capability to detect all relevant instances.

4. F1-Score: The harmonic average of precision and recall, providing a balance between false positives and false negatives.

5. AUC – ROC Curve: Evaluates the discriminative ability of the model for multiclass classification.

These metrics were computed for each disease class and averaged to produce macro and weighted scores.

3. RESULT AND DISCUSSION

Coffee leaf images are images or visual representations of coffee plant leaves, which can be captured using a standard camera, a mobile phone camera, or a specialized sensor, as shown in Figure 3. These images are used to analyze plant: health based on the color, texture, shape, and patterns on the leaves. The collected dataset comprises 845 leaf images of various types and forms of disease. Some common diseases affecting coffee leaves can be identified by their characteristic symptoms, as shown in Table 2.

Table 2. Symptoms on leaves

Disease	Symptoms on Leaves
Leaf Rust (Hemileia vastatrix)	Yellow or orange spots on the underside of leaves, similar to rust powder.
Leaf Spot	Large dark brown to black patches, often surrounded by chlorotic (yellowish) areas.
Sooty Dew	Leaves appear black and sooty, caused by a fungus that grows on honeydew.
Healthy Leaves	Green and smooth leaves that show no visible signs of disease in the image of the coffee leaves.

The dataset above needs to be evaluated against the Coffee Leaf Imagery because it has various image capacities and pixels. Images that are too high will have their file capacity reduced so they don't take up a lot of RAM and data, then images with small capacities and pixels will be increased to achieve a uniform image capacity and sharpness that can be properly analyzed by CNN [27]. The mapping of the Class

Index is Spot: 0, Healthy_Leaf: 1, Soot_Dew: 2 and Leaf Rust.

3.1 Augmentation

Data augmentation aims to increase the dataset size by modifying existing images to make the CNN model more general and prevent overfitting. The augmentation, called dBrightness/contrast adjustment, can be seen in the image below:

A review of several randomly selected samples from the dataset showed that the images were high-quality and appropriate for training. Each image in the Coffee Leaf Image dataset had a resolution of 800×550 pixels and was stored in PNG format. Since MobileNetV3 requires square images with three RGB channels, the dataset images [28] were resized to match these specifications. The *crop_to_aspect_ratio* parameter was kept at its default value, allowing flexible compression and preventing the software from unintentionally cropping important parts of the images. Here are the zoom results Robusta Coffee features available for training, as shown in Table 3 [29].

In addition, the results of this Zoom augmentation can be seen in Figure 4. Image a is the image not zoomed, image b with 40× zoom, image c 100× zoom, image c 400× zoom and image d 400× zoom so that the image details appear with the aim of increasing accuracy with increasingly detailed color images. The zoom results are shown in Table 3.
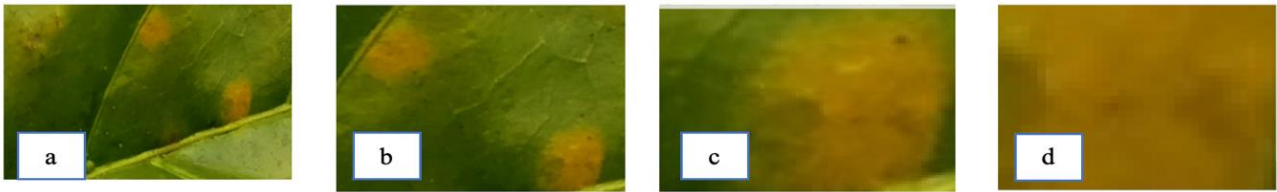

Figure 3. Coffee leaf image

Figure 4. Example from the Coffee Leaf dataset, seen at magnification: (a) 40×, (b) 100×, (c) 200×, (d) 400×

Table 3. Coffee leaf images at different zoom scales

Data	Zoom			
	40×	100×	400×	Total
Leaf Rust	240	240	240	720
Phoma Leaf Spot	240	240	240	720
Sooty Mold	240	240	240	720
Fresh	240	240	240	720
Dataset After Zoom				2880

Lower magnifications reveal the architectural details of the coffee leaf in clearer images, while at higher magnifications, the disease details are more visible. If overfitting occurs, it is not entirely detrimental to generate A model can be trained using class weights that match the true distribution of the dataset. Even so, applying data augmentation to all images remains a widely used method for addressing class imbalance.

In this study, augmentation was carried out by applying random horizontal flips (mirror reflections), which effectively doubled the number of available samples. Additional distortions, such as shearing, were intentionally avoided. Other augmentation methods—like rotation—were considered unsuitable for Robusta coffee leaf images, and changes in magnification would introduce unnecessary zoom variations. The influence of these augmentation strategies on the training and validation accuracy curves will be discussed in the study [30].

3.2 Segmentation

The goal of image segmentation is to divide an image into meaningful parts, or regions, for easier analysis by a computer system. Segmentation helps the system "understand" what's in the image. This involves extracting features (shape, size, and texture) to produce a relevant image. The segmentation results are shown in Figure 5.

Figure 5. Coffee leaf image segmentation

3.3 Model selection

MobileNetV3 expands upon the enhancements introduced in MobileNetV2 by integrating a squeeze-and-excite mechanism alongside each residual block, utilizing Neural Architecture Search to boost accuracy, and replacing several sigmoid activations with the more efficient hard-swish function—an important advantage for mobile applications. These improvements make it possible to reduce certain layers in the final stages of the network without diminishing accuracy. MobileNetV3 is available in two variants: a small version optimized for devices with limited computational resources and a large version designed for higher performance [31].

The large MobileNetV3 architecture delivers superior accuracy compared to the small variant. Although ImageNet-based pretrained embeddings are enabled by default, it remains uncertain whether weights learned from highly diverse internet images are beneficial for classifying Robusta coffee leaf images. The primary hyperparameter in this model is the learning rate, which governs how rapidly the model parameters converge toward a local minimum of the categorical cross-entropy loss function—used to evaluate classification performance at the end of each training epoch.

Table 4. Data params

Layer Type	Output	Params	Connected to
Input Layer	(None, 224, 244, 3)	0	-
Zero_padding2d	(None, 230, 230, 3)	0	Input_layer3[0]
Bn (Batch Normalization)	(None, 7, 7, 1664)	6,656	Conv5_block32_conv
Relu (Activation)	(None, 7, 7, 1664)	0	Bn[0][0]
Global AveragePooling	(None, 7, 7, 1664)	0	Relu[0][0]
Dense_6	(None, 1664)	1,704,960	Global_average_Pooling
Dense_7	(None, 4)	4,100	Dense_6[0][0]
Total Params	42,739,022 (163.04 MB)		
Trainable Params	14,139,540 (54.14 MB)		
Non-trainable Params	158,400 (618.75 KB)		
Optimizer Params	28,387,082 (108.29 MB)		

As shown in Table 4, the process begins with the Input Layer, which accepts images of size 224 × 244 pixels with three color channels (RGB). This layer serves as the entry point of the model, ensuring the input conforms to the expected dimensions. Following this, the Zero Padding layer adds extra borders around the image to preserve spatial dimensions during convolution operations, preventing feature loss at the image edges. Next, the data passes through a Batch Normalization (BN) layer, which standardizes the activations to stabilize the training process and accelerate convergence. After normalization, the output is processed by a ReLU (Rectified Linear Unit) activation function, which introduces non-linearity, allowing the network to learn complex patterns and relationships within the data.

The output from this stage is fed into a Global Average Pooling layer, which computes the average of each feature map. This operation effectively reduces the spatial dimensions of the data, converting the extracted features into a compact vector representation while retaining the essential information required for classification. Subsequently, the network includes a fully connected Dense_6 layer with 1,704,960 parameters. This layer combines the extracted features into a higher-level representation suitable for decision-making.

The following Dense_7 layer acts as the output layer with four neurons, each representing one of the four target classes. The final prediction corresponds to the histologic subtype determined by the model. In total, the architecture comprises 42,739,022 parameters (approximately 163.04 MB), including 14,139,540 trainable parameters (54.14 MB) that are updated during training, and 158,400 non-trainable parameters (618.75 KB), which are frozen from the pre-trained MobileNetV3 backbone. Additionally, there are 28,387,082 optimizer parameters (108.29 MB) used by optimization algorithms such as Adam or SGD to fine-tune the model weights efficiently.

This model integrates transfer learning principles by leveraging pre-trained MobileNetV3 layers as a feature extractor while fine-tuning the upper dense layers for specific classification tasks. The frozen base layers retain generalized visual knowledge from large-scale image datasets, while the trainable top layers adapt to the characteristics of the new dataset. This approach enhances the model's accuracy and efficiency in recognizing and classifying distinct histologic subtypes from the given input images. The visual architecture diagram (such as the flow from layer to layer) can be seen in Figure 6.
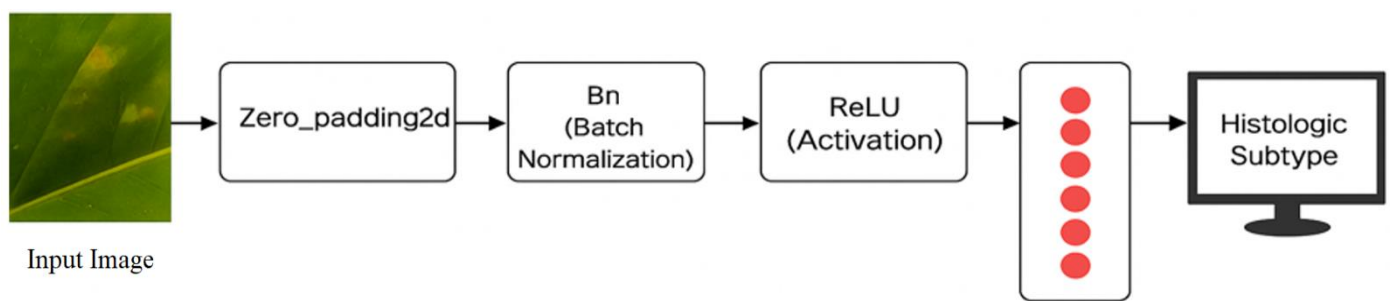

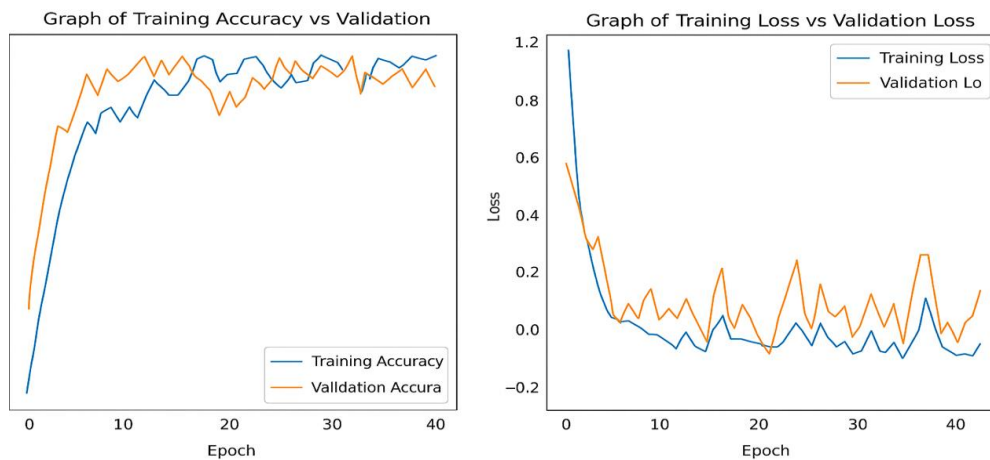**Figure 6.** Architecture diagram (such as the flow from layer to layer)

Figure 7. Comparative analysis of training and validation performance with and without data augmentation

The Coffee Leaf Image dataset was split into training, validation, and testing sets using a 0.75/0.15/0.10 ratio. The output from MobileNetV3 was processed through a Global_Average_Pooling2D layer, flattened, and then passed into four Dense layers with ReLU activation, followed by a final Dense layer with softmax activation producing eight class outputs. Dropout and Batch Normalization were incorporated to minimize overfitting during training. As illustrated in Figure 4, the final model contained roughly 3.0 million parameters, with about 98% attributed to MobileNetV3-Large.

Hyperparameters—including learning rate, epoch decay, and layer-freezing configurations—were organized in a spreadsheet and used to automatically train the model for 50 epochs. The workflow is presented in the corresponding Figure 7. Hyperparameter optimization targeted three key components: data augmentation, selection of trainable layers, and tuning of the learning rate. The Robusta coffee leaf images were preprocessed, converted into vectors, augmented, and then supplied to both the MobileNetV3 and DenseNet169 models, each trained for 50 epochs.

Figure 8 presents two graphs illustrating the performance of a Convolutional Neural Network (CNN) model during training and validation phases over 40 epochs. The left graph shows the relationship between training accuracy and validation accuracy, while the right graph compares training loss and validation loss. This graph displays how the accuracy of the CNN model improves over the course of training. The blue line represents training accuracy, and the orange line indicates validation accuracy. Initially, both accuracies increase rapidly, showing that the model is learning meaningful patterns from the dataset.
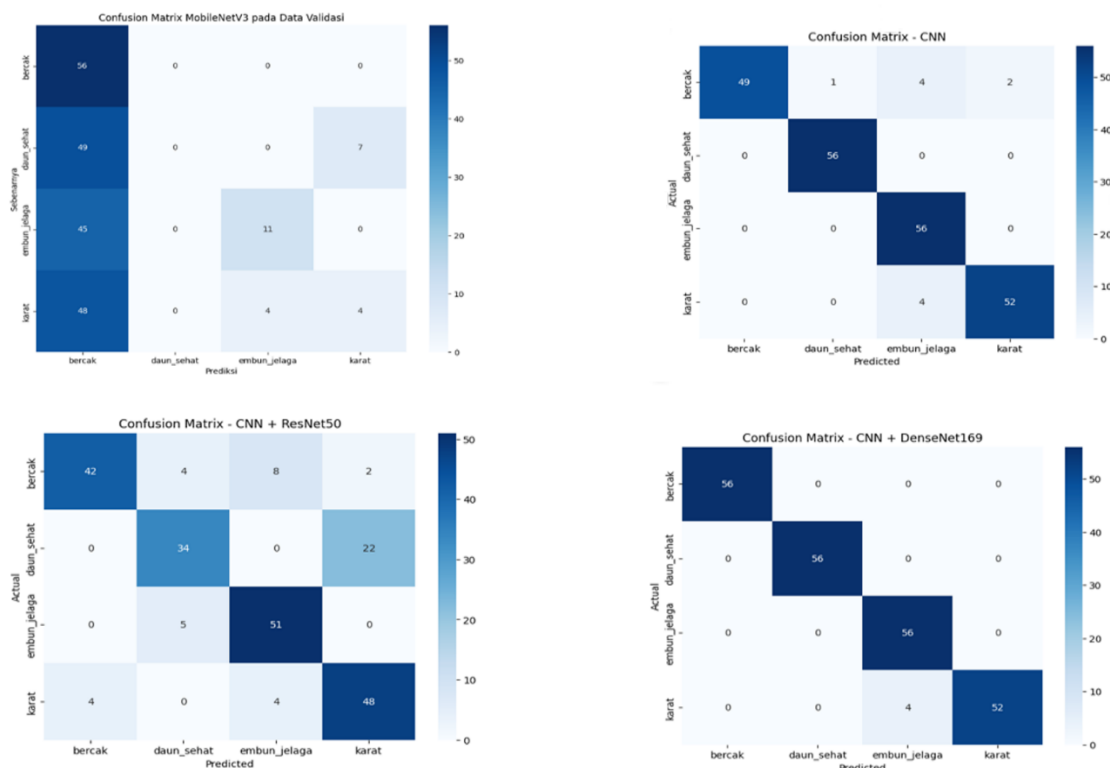

Figure 8. Confusion matrix measure performance

However, after approximately 15–20 epochs, the validation accuracy starts to fluctuate while the training accuracy continues to rise steadily. This indicates that the model begins to overfit, meaning it performs well on the training data but less consistently on unseen validation data. The small oscillations in the validation curve suggest that the model struggles to generalize beyond the training samples. The second graph depicts the loss curves for both training and validation. The blue curve represents training loss, and the orange curve corresponds to validation loss.

The trained models are fed images from the validation set, which are then used to make predictions by selecting one of eight available labels. The output is then displayed. At the beginning of training, both losses decrease sharply, which indicates that the model is effectively minimizing error

However, after several epochs, the validation loss begins to fluctuate and does not decrease as smoothly as the training loss. This pattern further confirms overfitting, as the model continues to optimize for the training set while its performance on the validation data stops improving and becomes unstable. Overall, the figure clearly demonstrates a case of overfitting in the CNN model. The increasing gap between training and validation performance suggests that the model memorizes specific features in the training data rather than learning general patterns. To mitigate overfitting, techniques such as dropout regularization, data augmentation, early stopping, or reducing model complexity could be applied.

Overall, data augmentation reduces overfitting, resulting in closer validation results to training results. Training was performed with an initial learning rate of 0.001, an epoch an epoch decay rate of 0.95, and an initial fine-tuning layer ofI150. After obtaining a low loss, a confusion matrix test is then carried out to obtain classification accuracy.

The next step is to test four confusion matrices to measure the performance of the leaf image classification model on the validation data. Each confusion matrix represents the test results of a different deep learning architecture: MobileNetV3, standard CNN, CNN + ResNet50, and CNN + DenseNet169.

MobileNetV3 Confusion Matrix (Top Left)

The MobileNetV3 model demonstrates strong classification performance for most categories, particularly Leaf Spot and Healthy Leaf, achieving almost perfect predictions. However, there is slight misclassification between Sooty Dew and Leaf Rust, where some diseased leaves are incorrectly identified as other categories. Despite this, the overall accuracy remains high, confirming the model's efficiency and adaptability for mobile-based or real-time detection due to its lightweight architecture.

Standard CNN Confusion Matrix (Top Right)

The standard CNN model performs consistently well across all categories, showing balanced accuracy. Most samples are classified correctly with minimal confusion between classes. For instance, Leaf Rust and Healthy Leaf achieve strong diagonal values (high true positives). However, there are still a few misclassified samples in the Sooty Dew category, indicating the need for deeper feature extraction or additional regularization to improve robustness.

CNN + ResNet50 Confusion Matrix (Bottom Left)

The hybrid CNN + ResNet50 model exhibits moderate performance with visible misclassifications, particularly between Healthy Leaf and Leaf Rust, as well as between Leaf Spot and Sooty Dew. Although ResNet50 enhances feature extraction through residual connections, its deeper network structure may lead to overfitting when applied to smaller datasets. This results in reduced generalization capability compared to lighter architectures like MobileNetV3.

CNN + DenseNet169 Confusion Matrix (Bottom Right)

The CNN + DenseNet169 model achieves perfect classification across all categories, as indicated by the strong diagonal dominance (all values along the diagonal are maximal while off-diagonal elements are zero). This demonstrates that DenseNet169 effectively captures hierarchical feature representations and efficiently propagates information across layers. The dense connections between layers minimize gradient vanishing and maximize feature reuse, leading to superior accuracy and stability.

Comparing the four models, DenseNet169 achieves the highest accuracy and the most stable predictions, followed closely by MobileNetV3 which offers the best trade-off between speed and accuracy.

The standard CNN performs acceptably but with minor inconsistencies, while the DenseNet50 model suffers from some confusion between similar disease classes. The results confirm that fine-tuning at layer 150 (out of 268 total layers) and training for 50 epochs significantly improve feature extraction performance. Consequently, DenseNet169 is most suitable for high-accuracy classification, whereas MobileNetV3 is optimal for real-time applications requiring faster inference with minimal resource consumption.

3.4 Classification accuracy

The trained model was subsequently applied to the test set, where it achieved a recall of 0.97, a precision of 0.98, and an F1-score of 0.98 for distinguishing between Healthy and Diseased coffee leaves. The 95% confidence interval further shows that the model performed most effectively in identifying similar coffee diseases, namely leaf spot and sooty mold, which are difficult to distinguish, with an accuracy of 0.9, so Coffee Leaf was identified accurately in 90% of cases. The classification accuracy for Leaf Rust was 0.94, and for Leaf Spot was 0.56. The F1Iscore, precision, and recall were calculated for the classification into healthy and diseased leaves. The highest model performance was shown in the classification of leaf rust, likely due to the larger number of training samples. The lowest performance was shown in Coffee Leaf sooty mold, which had the fewest samples in the dataset. The ROC curve confirmed the relative classification performance.

To gain insight into how the model misclassified some coffee disease images, we identified misclassified images by comparing the predicted labels with the ground truth labels. Of these images presented, the misclassified image contains mostly tissue stroma or possible tissue necrosis, with very few identifiable cells. The image misclassified as Coffee Leaf phyllodes is of fairly good quality; however, in this particular image, it may be difficult even for an agricultural expert to distinguish between leaf spot and leaf rust.

The image in Figure 9, also misclassified as Coffee Leaf sooty mold, is at very high magnification and consists of sheets of leaf image with scattered chromatin and indistinct cellular boundaries, and is likely of multiple origins. The final image misclassified as a healthy leaf consists mostly of green streaks and likely represents areas of fresh leaf tissue, with little

identifiable dark green material. Therefore, it is not unreasonable for these images to be misclassified, as this would pose a challenge even for expert human evaluation (optimal Bayes error rate).

Figure 9 resents a comprehensive comparison between several Convolutional Neural Network (CNN)-based architectures — namely CNN, CNN + ResNet50, and CNN + DenseNet169 — in the classification of coffee leaf diseases. The comparison is displayed in three main sections: (1) performance metrics table, (2) validation accuracy and loss graphs, and (3) visual classification results.
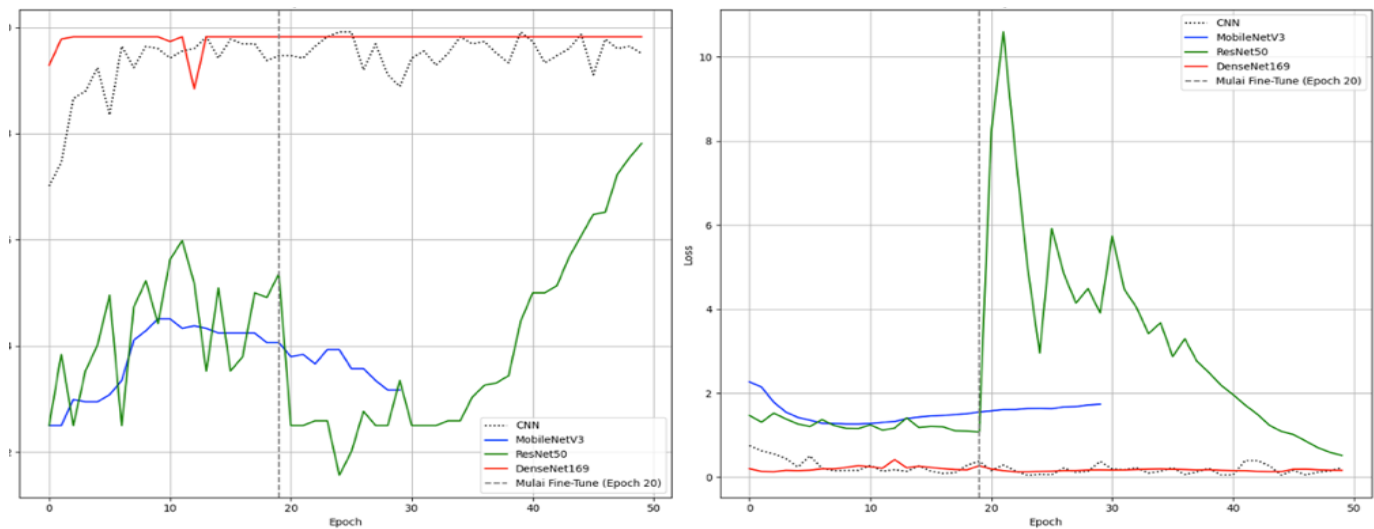

Figure 9. Accuracy testing CNN

Model Performance Metrics (Top Table)

Table 5 summarizes three key performance indicators: Accuracy, Recall, and F1-Score.

Table 5. Algorithm testing report

Model	Accuracy	Recall	F1-Score
CNN	96%	95%	95%
CNN + ResNet50	79%	78%	78%
CNN + DenseNet169	98%	98%	98%

The CNN + DenseNet169 model achieves the highest accuracy, recall, and F1-score (98%), demonstrating superior classification performance and robustness in identifying coffee leaf diseases. The standard CNN model performs relatively well (96% accuracy) but slightly less consistent. Meanwhile, the CNN + ResNet50 model shows the lowest performance (79%), likely due to overfitting or inadequate feature generalization on the coffee leaf dataset.

Validation Accuracy and Loss Graphs (Middle Plots)

The middle section displays two performance graphs:
a. Left Graph (Validation Accuracy Comparison): The DenseNet169 curve (red line) remains consistently high across all epochs, indicating stable convergence and superior learning capability. The CNN model (black dashed line) also maintains good accuracy but shows minor fluctuations. In contrast, ResNet50 (green line) struggles to achieve stability, with accuracy increasing slowly and inconsistently. MobileNetV3 (blue line) achieves moderate accuracy with efficient convergence speed, highlighting its suitability for lightweight, real-time applications.
b. Right Graph (Validation Loss Comparison): DenseNet169 again shows the lowest and most stable loss values, confirming its ability to minimize classification errors effectively. CNN follows with a relatively smooth loss reduction, while ResNet50 exhibits significant oscillations, reflecting unstable learning. The vertical dashed line marks the fine-tuning process initiated at epoch 20, after which model performance, especially for DenseNet169, improves markedly.

These graphs clearly demonstrate that DenseNet169 not only achieves higher accuracy but also maintains better generalization, avoiding overfitting while ensuring consistent learning.

Visual Classification Results (Bottom Section)

The lower part of Figure 8 shows sample prediction outputs for both the base CNN and DenseNet169 models.
a. Left (CNN Model Results): Several coffee leaf images are correctly classified (indicated in green), but a few misclassifications (highlighted in red) show that the model occasionally confuses similar disease patterns such as Leaf Spot and Leaf Rust. The average accuracy achieved here is 90%.
b. Right (DenseNet169 Model Results): All images are correctly predicted, achieving 100% classification accuracy. The model successfully differentiates between disease types (e.g., Leaf Rust, Sooty Dew, Leaf Spot) and healthy leaves with high precision. This demonstrates DenseNet169's strong feature extraction and its ability to capture fine-grained visual details in leaf textures and color variations.
c. Figure 9 collectively demonstrates that DenseNet169 significantly outperforms other models in terms of accuracy, stability, and reliability. Its dense connections enable efficient feature reuse, faster convergence, and improved gradient flow, resulting in minimal loss and near-perfect classification results. While MobileNetV3 remains an ideal option for real-time field deployment due to its lightweight and fast processing capability, DenseNet169 provides the highest precision for detailed

disease analysis, making it highly suitable for research and diagnostic applications.

4. CONCLUSION

This research successfully developed and evaluated a CNN-based classification model optimized through the combination of MobileNetV3 and DenseNet169 architectures for detecting Robusta coffee leaf diseases. Experimental results confirmed that CNN + DenseNet169 achieved superior performance (98% accuracy, recall, and F1-score), demonstrating its strong capability in feature propagation, gradient flow, and precise differentiation between disease types. Meanwhile, MobileNetV3 achieved faster inference with smaller model parameters, making it highly applicable for real-time implementation on mobile or IoT devices in agricultural settings. Together, these architectures provide a balanced framework between computational speed and classification precision, suitable for precision agriculture and digital plantation management.

The main contribution of this study lies in demonstrating that CNN optimization through transfer learning and hybrid architecture design can yield both high accuracy and efficiency in agricultural image classification tasks. The proposed model offers a foundation for mobile-based or IoT-integrated early disease detection systems that enable farmers to make rapid and accurate decisions for pest and disease management. Future work should focus on expanding the dataset to include diverse lighting and geographical conditions, integrating ensemble learning techniques, and testing real-time implementation in field conditions. Such advancements will strengthen the application of artificial intelligence in supporting sustainable coffee production and smart agriculture initiatives in Indonesia. integrated early disease detection systems that enable farmers to make rapid and accurate decisions for pest and disease management. And then next research on expanding the dataset to include diverse lighting and geographical conditions, integrating ensemble learning techniques, and testing real-time implementation in field conditions. Such advancements will strengthen the application of artificial intelligence in supporting sustainable coffee production and smart agriculture initiatives in Indonesia.

REFERENCES

[1] BPS-Statistics Indonesia. (2023). Complete enumeration results of the 2023 census of agriculture - Edition 1. BPS. https://www.bps.go.id/en/pressrelease/2023/12/04/2050/hasil-pencacahan-lengkap-sensus-pertanian-2023---tahap-i.html.

[2] Irmeilyana, I., Ngudiantoro, N., Rodiah, D. (2019). Description of the profile and characteristics of Pagar Alam coffee farming business based on descriptive statistics and correlation. Jurnal Infomedia: Teknik Informatika, Multimedia, dan Jaringan, 4(2): 60-68. http://doi.org/10.30811/jim.v4i2.1534

[3] Mujidah, M., Agustin, S. (2024). Classification of robusta coffee bean quality using the K-Nearest Neighbor (K-NN) method and gray co-occurance matrix (Glcm). JATI (Jurnal Mahasiswa Teknik Informatika), 8(6): 11832-11838.

[4] Putrawansyah, F., Rahayu, C., Dhiniati, F. (2024). Application of particle swarm optimization to improve the performance of the k-nearest neighbor in stunting classification in South Sumatra, Indonesia. International Journal of Education and Management Engineering, 6: 32-43. http://doi.org/10.5815/ijeme.2024.06.03

[5] Mustakim, M., Kurnia, K., Noviarni, N., Putrawansyah, F., Kurniati, A., Rimet, R. (2024). Implementation of convolutional neural network for sentiment analysis on hotel customer reviews. In 2024 International Conference on Decision Aid Sciences and Applications (DASA), Manama, Bahrain, pp. 1-6. http://doi.org/10.1109/DASA63652.2024.10836631

[6] Rivera-Palacio, J.C., Bunn, C., Ryo, M. (2025). Factors affecting deep learning model performance in citizen science–based image data collection for agriculture: A case study on coffee crops. Computers and Electronics in Agriculture, 232: 110096. https://doi.org/10.1016/j.compag.2025.110096

[7] Mei, P., Karimi, H.R., Ou, L., Xie, H., Zhan, C., Li, G., Yang, S. (2025). Driving style classification and recognition methods for connected vehicle control in intelligent transportation systems: A review. ISA Transactions, 158: 167-183. https://doi.org/10.1016/j.isatra.2025.01.033

[8] Hadir, A., Adjou, M., Assainova, O., Palka, G., Elbouz, M. (2025). Comparative study of agricultural parcel delineation deep learning methods using satellite images: Validation through parcels complexity. Smart Agricultural Technology, 10: 100833. https://doi.org/10.1016/j.atech.2025.100833

[9] Hernández-Aceituno, J., Méndez-Pérez, J.A., González-Cava, J.M., Reboso-Morales, J.A. (2023). Towards intelligent supervision of operating rooms using stencil-based character recognition. Computers in Biology and Medicine, 162: 107071. https://doi.org/10.1016/j.compbiomed.2023.107071

[10] Putrawansyah, F. (2024). Application of support vector machine method to classify guava types. JIKO (Jurnal Informatika dan Komputer), 8(1): 193-204. http://doi.org/10.26798/jiko.v8i1.988

[11] Julianto, A., Sunyoto, A., Wibowo, F.W. (2022). Optimasi hyperparameter convolutional neural network untuk klasifikasi penyakit tanaman padi. TEKNIMEDIA: Teknologi Informasi dan Multimedia, 3(2): 98-105. https://doi.org/10.46764/teknimedia.v3i2.77

[12] Rahman, M.H., Jannat, M.K.A., Islam, M.S., Grossi, G., Bursic, S., Aktaruzzaman, M. (2023). Real-time face mask position recognition system based on MobileNet model. Smart Health, 28: 100382. https://doi.org/10.1016/j.smhl.2023.100382

[13] Guadalupe, G.A., Garcia, L., Quispe-Sanchez, L., Domenech, E. (2024). The fault tree analysis (FTA) to support management decisions on pesticide control in the reception of Peruvian parchment coffee. Food Control, 166: 110729. https://doi.org/10.1016/j.foodcont.2024.110729

[14] Ramdani, S., Rahmatulloh, A. (2024). Mobilenet implementation for image classification and emotion detection using KERAS. JUSTIN (Jurnal Sistem dan Teknologi Informasi), 12(2): 259-264. https://doi.org/10.26418/justin.v12i2.73389

[15] Winardi, S., Sinaga, F.M., Fa, F.R., Sintiya, C. (2023). Using mobilenet for intelligent character recognition (ICR) automatic assessment of basic mathematical operations. Jurnal TIMES, 12(2): 40-51. https://doi.org/10.51351/jtm.12.2.2023707

[16] Qotrunnada, F.M., Utomo, P.H. (2022). Convolutional neural network method for masked face classification. In PRISMA, Prosiding Seminar Nasional Matematika, 5: 799-807.

[17] Haris, M. (2021). Implementation of deep learning methods in predicting student performance: A systematic literature review. Jurnal Nasional Teknik Elektro dan Teknologi Informasi, 10(2).

[18] Zidan, A., Rahman, M.F., Sari, A.P. (2024). Utilization of the Convolutional Neural Network (CNN) method with the MobileNetV2 Architecture for home feasibility assessment. ALINIER: Journal of Artificial Intelligence & Applications, 5(2): 129-139.

[19] Putra, O.V., Mustaqim, M.Z., Muriatmoko, D. (2023). Transfer learning for rice disease and pest classification using MobileNetV2. Techno. com, 22(3): 562-575.

[20] Ungkawa, U., Al Hakim, G. (2023). Color classification of yellow coffee fruit ripeness using the CNN Inception V3 method. ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika, 11(3): 731. https://doi.org/10.26760/elkomika.v11i3.731

[21] Arjun, P.A., Suryanarayan, S., Viswamanav, R.S., Abhishek, S., Anjali, T. (2024). Unveiling underwater structures: Mobilenet vs. efficientnet in sonar image detection. Procedia Computer Science, 233: 518-527. https://doi.org/10.1016/j.procs.2024.03.241

[22] Setiawan, D.R.A., Riti, Y.F., Trisuwita, N.C.P. (2024). Performance comparison of Mobilenet V2 and FPNLite SSD Models in motorcycle rider helmet detection. Komputika: Jurnal Sistem Komputer, 13(1): 131-138. https://doi.org/10.34010/komputika.v13i1.10333

[23] Rasyad, R., Martawirya, Y.Y. (2021). Development of a production process monitoring system with intelligent character recognition as an object scanner. Jurnal Teknik Mesin Indonesia, 16(2): 101-108.

[24] Mulindwa, D.B., Du, S., Jordaan, J.A. (2014). An intelligent character recognition system for automatic mark capturing. In 2014 7th International Congress on Image and Signal Processing, Dalian, China, pp. 670-675. https://doi.org/10.1109/CISP.2014.7003863

[25] Mohanty, S.P., Hughes, D.P., Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7: 215232. https://doi.org/10.3389/fpls.2016.01419

[26] Ferentinos, K.P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145: 311-318. https://doi.org/10.1016/j.compag.2018.01.009

[27] Howard, A., Sandler, M., Chu, G., Chen, L.C., et al. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1314-1324.

[28] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708.

[29] Prajapati, H.B., Shah, J.P., Dabhi, V.K. (2017). Detection and classification of rice plant diseases. Intelligent Decision Technologies, 11(3): 357-373. https://doi.org/10.3233/IDT-170301

[30] Kamilaris, A., Prenafeta-Boldú, F.X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147: 70-90. https://doi.org/10.1016/j.compag.2018.02.016

[31] Arif, A., Putrawansyah, F., Jangcik, I. (2025). Detection of coffee leaf diseases using lightweight deep learning: A comparative study of EfficientNet-B0 and Vision Transformer. Ingénierie des Systèmes d'Information, 30(9): 2393-2404. https://doi.org/10.18280/isi.300915