SMATIKA : STIKI Informatika Jurnal
Accredited Sinta 4 Ministry of Higher Education, Science and Technology Republic of Indonesia SK No. 79/E/KPT/2023
E-ISSN: 2580-6939 || P-ISSN: 2087-0256

Comparative Analysis of Transfer Learning-Based Deep Learning Models for Jatropha Leaf Disease Classification

Sarifah Agustiani[1*], Sulistiyah[2], Agus Junaidi[3], Cucu Ika Agustyaningrum[4], Yoseph Tajul Arifin[5]

[1,3] Bina Sarana Informatika University, Faculty of Engineering and Informatics, Informatics, Jl. Kramat Raya No. 98, RT. 2/RW. 9, Kwitang, Senen District, Central Jakarta City, Special Capital Region of Jakarta 10450, Indonesia
[2,4] Bina Sarana Informatika University, Faculty of Engineering and Informatics, Information Systems, Jl. Kramat Raya No. 98, RT. 2/RW. 9, Kwitang, Senen District, Central Jakarta City, Special Capital Region of Jakarta 10450, Indonesia
[5] Bina Sarana Informatika University, Faculty of Engineering and Informatics, Electrical Engineering, Jl. Kramat Raya No. 98, RT. 2/RW. 9, Kwitang, Senen District, Central Jakarta City, Special Capital Region of Jakarta 10450, Indonesia

Keyword

Deep Learning; Image Classification; Jatropha; Plant Disease Classification; Transfer Learning

***Corresponding Author:**
sarifah.sgu@bsi.ac.id

Abstract

Plant disease identification is essential for enhancing agricultural productivity and promoting sustainable crop management practices. Jatropha curcas has considerable potential as a biofuel-producing plant; however, its growth and productivity can be significantly affected by various leaf diseases. Conventional disease diagnosis often requires substantial time and relies heavily on expert knowledge, creating a need for automated solutions based on deep learning techniques. Although deep learning has been widely applied in plant disease recognition, comparative studies focusing on transfer learning models for Jatropha leaf disease classification remain limited, particularly for datasets characterized by distinctive visual features and relatively small sample sizes. This research conducts a comparative assessment of several deep learning architectures to determine the most effective model for classifying Jatropha leaf diseases. The evaluated architectures include MobileNetV2, EfficientNetB0, ResNet50, DenseNet121, and VGG16. All models utilized ImageNet pre-trained weights and were adapted through fine-tuning of the final classification layers to accommodate a dataset containing healthy and diseased Jatropha leaf images. Experimental findings reveal that ResNet50 achieved the highest classification accuracy of 93.81%, followed by VGG16 at 93.58% and EfficientNetB0 at 90.49%. In comparison, DenseNet121 and MobileNetV2 attained accuracies of 85.40% and 74.56%, respectively. Model effectiveness was assessed using accuracy, training duration, confusion matrix analysis, and ROC curve evaluation to examine classification capability across categories. The results demonstrate that ResNet50 offers the most balanced combination of predictive accuracy and performance stability. Overall, the study confirms that transfer learning-based deep learning models are highly effective for Jatropha leaf disease classification, with ResNet50 emerging as the most suitable architecture among those investigated. These findings may serve as a valuable reference for the development of reliable and efficient plant disease detection systems in agricultural environments.

Cite: Agustiani, et al., (2026). Comparative Analysis of Transfer Learning-Based Deep Learning Models for Jatropha Leaf Disease Classification. SMATIKA : STIKI Informatika Jurnal,16(2). doi: https://doi.org/10.32664/smatika.v16i02.2325

1. Introduction

Plant disease detection is an essential component of modern agriculture, contributing significantly to improved productivity and the preservation of crop quality [1], [2]. Plant diseases can lead to significant yield losses if they are not detected and addressed promptly and effectively [3],[4]. Therefore, a detection system is needed that can identify plant diseases accurately and efficiently and can be widely implemented in agricultural practices.

Jatropha is a tropical plant native to the Americas and is now widely distributed across various regions, including Indonesia [5]. This plant has a high adaptability to marginal environmental conditions, such as low-fertility soil and high temperatures [6]. In addition to its use in traditional medicine—for conditions like malaria, fever, as an analgesic, and for blood stabilization [7], Jatropha is also recognized as a potential source of renewable energy, particularly as a feedstock for biodiesel, due to its high vegetable oil content [8].

Plant disease identification is still largely performed using conventional methods, which require considerable time and a high degree of precision. The limited availability of automated systems capable of identifying diseases quickly and accurately poses a challenge to the adoption of modern agricultural technology [9]. Deep learning techniques have been widely adopted in agricultural applications, particularly for the detection and classification of plant diseases CNN-based models are frequently employed due to their capability to automatically extract meaningful features from image data without extensive manual intervention [10]. With the continuous advancement of artificial intelligence technologies, deep learning has emerged as a powerful approach for plant disease identification, offering superior performance in analyzing complex visual information compared with conventional machine learning technique [11].

Advances in artificial intelligence have also significantly contributed to the development of digital image processing applications for agricultural diagnosis [12]. Image classification systems based on deep neural networks have been extensively employed in plant disease recognition due to their effectiveness in capturing complex visual characteristics directly from image inputs [13],[14]. A variety of CNN architectures, including VGG16, ResNet50, MobileNetV2, DenseNet121, GoogLeNet, and Inception-derived networks, have shown reliable performance in disease classification tasks involving numerous crop species such as tea, tomato, cucumber, grape, strawberry, citrus, kiwi, and apple [10], [15],[16] . Despite these advantages, training deep learning models from scratch typically requires a large volume of labeled data and high computational capacity, making the process both resource-intensive and time-consuming [17].

One effective solution to these limitations is the application of transfer learning. Rather than training a model from the beginning, transfer learning utilizes pre-trained networks developed on large and diverse datasets, allowing them to be fine-tuned for a target task, which are then adapted for specific tasks using smaller datasets [18]. Some commonly used architectures in this approach include MobileNetV2, EfficientNetB0, ResNet50, DenseNet, and VGG16 [19]. Various studies have shown that deep learning approaches based on transfer learning are capable of delivering strong performance in plant disease classification [20], [21], [22]. Models such as ResNet and EfficientNet have been reported to perform competitively on various datasets, while lightweight models like MobileNet have been widely developed for implementation on mobile devices and in edge computing [23], [24], [25].

In the biofuel crop domain, several studies have applied computer vision approaches to disease detection in Jatropha curcas L., such as a combination of superpixel-based segmentation, feature extraction, and machine learning algorithms [26]. Another study used a modified InceptionResNetV2 architecture for high-performance classification of Jatropha leaf diseases [27], Moreover, disease detection systems based on image processing techniques, including Mean Shift segmentation, Gaussian smoothing, Canny edge extraction, and contour analysis, have achieved competitive accuracy in identifying leaf disease symptoms [28]. Another approach uses a Fuzzy Neural Network (FNN)-based expert system, which combines neural networks and fuzzy logic for the classification process, although it still has performance limitations [29].

Although various approaches have shown promising results, most still rely on complex pipelines or are not yet fully end-to-end. Furthermore, existing research often prioritizes accuracy enhancement without thoroughly assessing how classification performance relates to computational requirements on Jatropha datasets.

Considering the limitations identified in previous studies, this research examines the performance of various transfer learning-based models for Jatropha leaf disease classification. A comprehensive evaluation is carried out using accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC metrics. In addition to comparing predictive performance, this study emphasizes the assessment of computational resource demands, enabling a more informed selection of models for practical deployment scenarios.

2. Research Method

A deep learning framework for Jatropha leaf disease classification was developed in this study. The research process involved multiple sequential stages, namely data collection, image preprocessing, dataset division, architecture implementation, model training setup, and evaluation of classification performance. A detailed overview of the research workflow is shown in Figure 1.

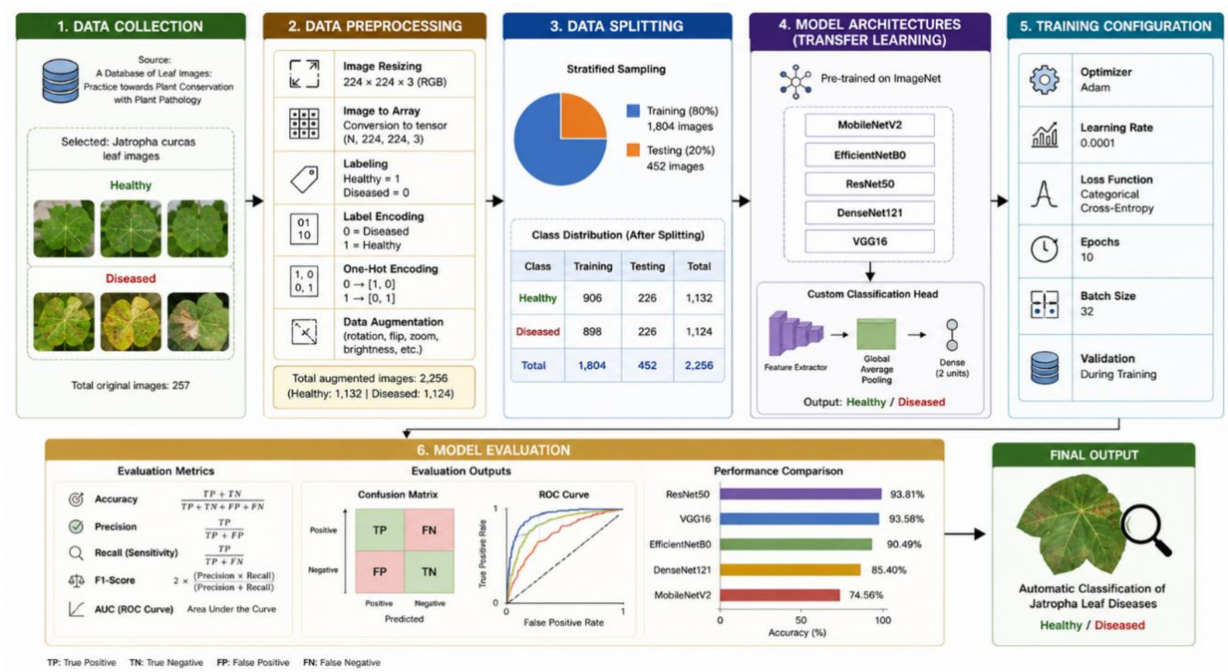

Figure 1. Research Method

2.1 Data Collection

The dataset used in this study was sourced from "A Data of Leaf Images: Practice towards Plant Conservation with Plant Pathology," which includes various plant species [30]. To align with the research objectives, the dataset was filtered to include only Jatropha leaf samples, categorized into healthy and diseased classes. This subset was chosen to focus the study on a single plant species so that the model could learn more specific visual characteristics.

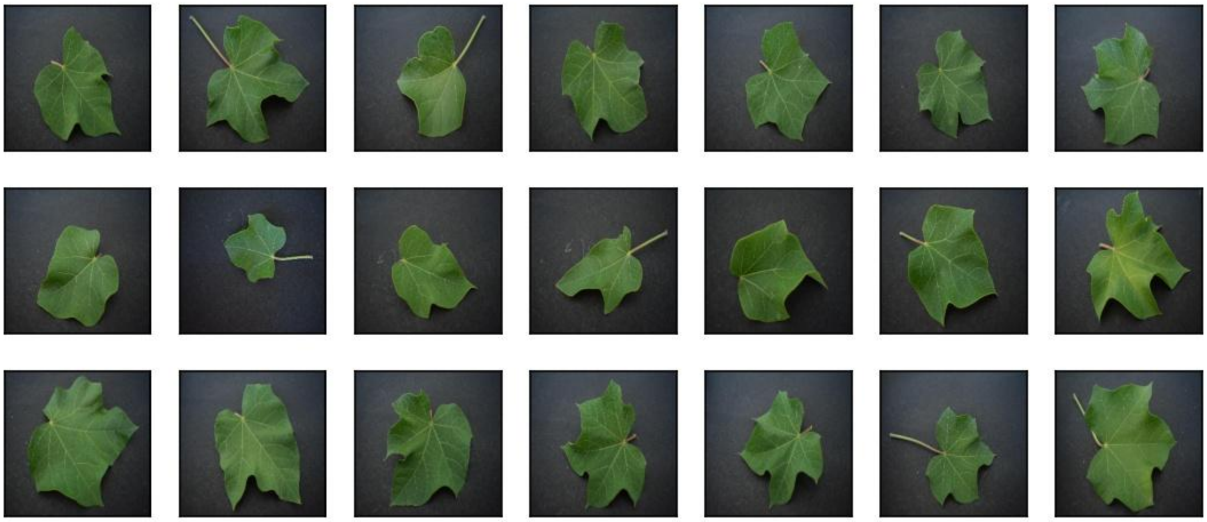

Figure 2. Sample Images of Healthy Jatropha Leaves

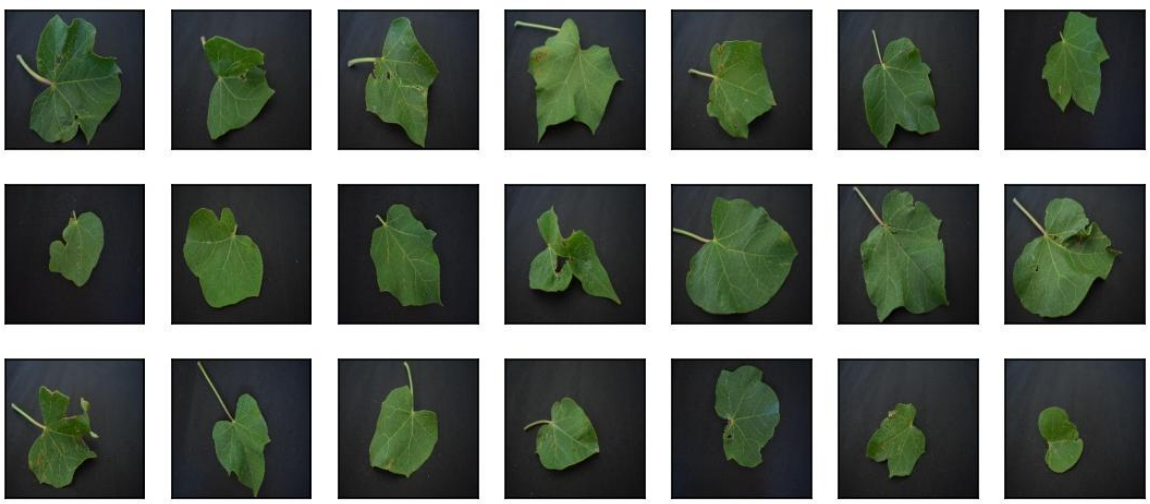

Figure 3. Sample Images of Diseases Jatropha Leaves

The dataset employed in this research consists of two categories, namely healthy and diseased Jatropha leaves, as presented in Figures 2 and 3, with a total of 257 images. Considering the limited number of available samples, data augmentation was applied to enhance dataset variability, improve model generalization, and minimize the likelihood of overfitting during the training phase. Augmentation was implemented using the Keras ImageDataGenerator framework, incorporating several image transformations, including rotation up to 20°, horizontal flipping, zooming with a factor of 0.2, width and height shifts of 0.2, shear transformation of 0.2, and brightness adjustments ranging from 0.8 to 1.2. These transformations were designed to mimic variations in viewing angles, illumination conditions, and leaf orientations that commonly occur in real-world agricultural settings.

Although the augmentation process increased the diversity of training samples, the relatively small number of original images may still limit the robustness and generalization ability of the developed models, especially for architectures with a large number of parameters. A restricted dataset can increase the tendency of a model to memorize training patterns rather than learn representative features, leading to reduced performance on unseen data. To address this challenge, transfer learning was adopted by utilizing feature representations learned from the ImageNet dataset, allowing the models to achieve more reliable learning performance despite the limited training data. A summary of the dataset composition is provided in Table 1.

Table 1. Dataset Distribution

Class	Original Data	Augmented Data	Total Data
Healthy	133	999	1132
Diseases	124	1000	1124
Total	257	1999	2256

2.2 Data Preprocessing

The preprocessing phase consisted of several steps designed to prepare the image dataset for deep learning implementation. These procedures included image resizing, normalization, label assignment, and data encoding. To ensure compatibility with all evaluated architectures, namely MobileNetV2, EfficientNetB0, ResNet50, DenseNet121, and VGG16, each image was converted into an RGB format and resized to a resolution of 224 × 224 pixels before being used as model input.

Each image was then converted into a numerical array using TensorFlow, resulting in a tensor of shape (N, 224, 224, 3). The dataset was labeled into two classes: healthy and diseased, based on directory structure.

Numerical labels were assigned to each class, with healthy leaves labeled as 1 and diseased leaves as 0. The labels were then converted into one-hot encoded vectors, where healthy samples were represented as [0, 1] and diseased samples as [1, 0]. This representation was adopted to support model training using categorical cross-entropy.

2.3 Data Splitting

After the preprocessing stage, the dataset was split into training and testing data in an 80:20 ratio, consisting of 1,804 and 452 images, respectively. The data was split using stratified sampling to ensure that the class distribution remained balanced across both subsets. This approach is crucial for minimizing bias during the training process and ensuring that the model can learn a proportional representation of each class.

2.4 Architecture Models

This study evaluates five deep learning architectures initialized with weights learned from the ImageNet dataset. Each model has distinct architectural characteristics in the feature extraction process. The models used in this study include MobileNetV2, EfficientNetB0, ResNet50, DenseNet121, and VGG16. The selection of these models is based on differences in architectural design strategies, such as parameter efficiency, network depth, and inter-layer connectivity mechanisms, thereby enabling a comparative analysis of each model's performance in classifying Jatropha leaf diseases.

2.4.1 MobileNetV2

MobileNetV2 is a compact convolutional neural network architecture developed to achieve competitive classification accuracy while maintaining low computational requirements. Comprising around 32 layers, the model is particularly suitable for resource-limited platforms, including mobile devices, where low latency and reduced computational overhead are essential requirements [31].

2.4.2 EfficientNetB0

As the baseline variant of the EfficientNet family, EfficientNetB0 utilizes compound scaling to systematically adjust network depth, width, and input resolution, resulting in improved classification performance with efficient resource consumption. The architecture employs a systematic network design that allows it to attain strong classification capability while maintaining a relatively small number of trainable parameters [23].

2.4.3 ResNet50

ResNet50 introduces a residual learning framework in which shortcut pathways bypass certain layers, helping the network learn more effectively despite its considerable depth of 50 layers. These connections facilitate the propagation of information and gradients across layers, reducing optimization difficulties commonly encountered in deep neural networks. The model utilizes knowledge acquired from pre-training on the ImageNet dataset, which contains more than one million labeled images, allowing it to learn rich feature representations that are beneficial for a wide range of image classification applications. The overall structure of the ResNet50 architecture is presented in Figure 4 [32].

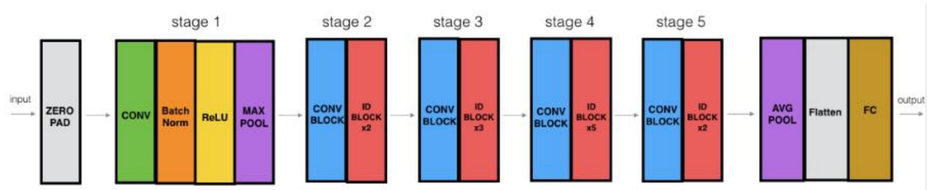

Figure 4. ResNet50 Architecture [33]

2.4.4 DenseNet121

DenseNet121 is a CNN architecture that utilizes dense connectivity, where feature maps generated by each layer are forwarded to all subsequent layers within the network. This connectivity pattern promotes extensive feature reuse, improves information propagation throughout the model, and helps preserve important feature representations during the learning process [34].

2.4.5 VGG16

VGG16 is a widely adopted convolutional neural network architecture that combines architectural simplicity with effective feature extraction capabilities, enabling competitive performance in diverse image processing applications [35].

2.5 Model Training Configuration

During the learning phase, weight updates were conducted using the Adam optimizer configured with a learning rate of 1×10^{-4}. As the classification targets were transformed into one-hot vectors, categorical cross-entropy served as the objective function for minimizing prediction errors. Every evaluated architecture underwent training for ten iterations over the dataset, with 32 samples processed in each mini-batch. The selection of 10 epochs was based on preliminary experiments showing that most transfer learning models achieved stable convergence within a relatively small number of training iterations. Since the models utilized pre-trained weights from ImageNet, the feature extraction process had already been initialized with generalized visual representations, reducing the need for extensive training epochs. Moreover, the restricted amount of available training data necessitated careful control of the training duration, as an excessively large number of epochs may cause the models to memorize training patterns rather than learn generalizable features. Therefore, validation performance was continuously observed during training to evaluate model stability and detect potential overfitting issues.

2.6 Model Evaluation

The predictive quality of the trained networks was quantified using four complementary metrics: accuracy, precision, recall, and AUC. These measures were selected to capture different aspects of classification behavior and provide a balanced evaluation of model effectiveness. Employing various evaluation measures allows a more complete analysis of model behavior, particularly when dealing with potential class distribution imbalances.

2.6.1 Accuracy

To assess the overall predictive performance of a classification model, accuracy is frequently employed as an evaluation criterion. This metric reflects the proportion of correct predictions among all evaluated observations, with higher values indicating stronger classification capability and reduced prediction errors [36].

The accuracy formula can be expressed as follows:

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

2.6.2 *Precision*

The precision metric measures the accuracy of a model when assigning samples to the positive class. It is determined by comparing correctly predicted positive cases against the overall number of positive predictions. Higher precision values suggest that the model makes fewer incorrect positive classifications [37].

The precision formula can be expressed as follows:

$$Precision_{class} = \frac{TP}{TP+FP} \tag{2}$$

2.6.3 *Recall*

The recall metric evaluates the completeness of positive class detection by measuring how many actual positive samples are correctly identified by the model. Models with higher recall are generally more successful in capturing positive instances and minimizing false negative predictions [38].

The recall formula can be expressed as follows:

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

2.6.4 *F1-Score*

To evaluate the trade-off between precision and recall, the F1-Score is employed as a unified performance indicator. This metric summarizes both measures into a single value through harmonic averaging, making it particularly effective for assessing models trained on datasets with unequal class representation [39].

The F1-score formula can be expressed as follows:

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

Higher F1-score values indicate a well-balanced classification performance, reflecting the model's effectiveness in identifying positive samples while maintaining reliable prediction accuracy.

2.6.5 *Area Under Curve (AUC)*

To measure the effectiveness of class discrimination, the Area Under the ROC Curve (AUC) was adopted as a performance indicator. Models with higher AUC values demonstrate superior capability in separating positive and negative classes consistently across varying threshold settings [40].

By using this combination of metrics, the model evaluation in this study focuses not only on accuracy but also takes into account the balance between the model's detection capability and predictive precision.

3. Result and Discussions

The present study compares the classification performance of five deep learning models utilizing transfer learning, including MobileNetV2, EfficientNetB0, ResNet50, DenseNet121, and VGG16, for the identification of diseases in Jatropha leaves. A standardized experimental setup was applied across all architectures to maintain consistency and ensure a fair evaluation of their performance.

The results indicate that each model exhibits different capabilities in extracting and learning visual features from leaf images. These differences are influenced by architectural characteristics such as network depth, feature reuse mechanisms, and parameter efficiency.

3.1 Model Performance Evaluation

3.1.1 Classification Metrics

Table 2 presents the performance evaluation results of all investigated models. The reported precision, recall, and F1-score metrics correspond to weighted-average values generated from the classification reports. This evaluation strategy incorporates the relative frequency of each class, enabling a more comprehensive assessment of model effectiveness and reducing potential bias caused by slight differences in class distribution.

Table 2. Comparative Performance Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (Seconds)	Evaluation Time (seconds)
ResNet50	93.81%	93.93	93.81	93.80	2481.97	46.02
VGG16	93.58%	93.58	93.58	93.58	7967.39	129.42
EfficientNetB0	90.49%	90.58	90.49	90.48	1225.67	25.92
DenseNet121	85.40%	85.58	85.40	85.38	2420.34	46.19
MobileNetV2	74.56%	74.69	74.56	74.52	556.50	13.73

Based on Table 2, Among the evaluated architectures, ResNet50 demonstrates the best overall performance, achieving an accuracy of 93.81%, accompanied by balanced precision (93.93%), recall (93.81%), and F1-score (93.80%). This consistency across all evaluation metrics indicates that the model is not only highly accurate but also stable in distinguishing between healthy and diseased leaf classes without significant bias. The effectiveness of ResNet50 can be attributed to its residual learning mechanism, which enables efficient feature propagation and mitigates the degradation problem in deep networks.

VGG16 demonstrates comparable classification performance, achieving an accuracy of 93.58% with identical values across precision, recall, and F1-score. However, this performance comes at the cost of significantly higher computational requirements, as reflected in its training and evaluation times. This limitation reduces its practicality for real-world deployment, especially in resource-constrained environments.

In contrast, MobileNetV2 provides the fastest computation time, with training and evaluation times of 556.50 seconds and 13.73 seconds, respectively. However, its accuracy is considerably lower at 74.56%, indicating limitations in capturing complex visual patterns. This is primarily due to its lightweight architecture, which prioritizes efficiency over representational capacity. Despite its lower classification performance, MobileNetV2 remains promising for practical deployment on mobile and edge devices due to its low computational complexity and reduced memory requirements. Such characteristics are particularly beneficial for real-time agricultural monitoring systems, smartphone-based disease detection applications, and edge computing environments where computational resources and power consumption are limited.

EfficientNetB0 and DenseNet121 exhibit moderate performance, with accuracies of 90.49% and 85.40%, respectively. Although both architectures are designed to improve feature reuse and scaling efficiency, their performance in this study suggests that increased architectural complexity does not necessarily lead to better results. This is likely influenced by the relatively limited size and variability of the dataset, which may not fully leverage the strengths of these architectures.

The findings of this study can also be compared with several previous studies on Jatropha curcas disease classification. Previous research combining superpixel-based segmentation, feature extraction, and machine learning algorithms achieved classification accuracies exceeding 96% [26]. Another study using a modified InceptionResNetV2 transfer learning architecture reported an accuracy of 97% for Jatropha leaf disease classification [27]. Similarly, image processing-based approaches utilizing Mean Shift, Gaussian Blur, Canny edge detection, and contour analysis achieved an accuracy of 96.93% [28]. In contrast, Fuzzy Neural Network (FNN)-based expert systems demonstrated considerably lower performance, with an accuracy of only 30% [29]. Although some previous studies achieved higher accuracy values than the present work, many relied on customized architectures, specialized preprocessing pipelines, or different dataset characteristics. In comparison, this study emphasizes a comprehensive comparative evaluation of multiple widely used transfer learning architectures under the same experimental setting using a relatively limited dataset. Therefore, the results provide a more balanced analysis of the trade-off between classification performance and computational efficiency for practical implementation. Overall, these results highlight that model performance is strongly influenced not only by architectural design but also by dataset characteristics.

3.1.2 ROC Curve and AUC

The discriminative performance of the evaluated models was analyzed using ROC curves, which illustrate their capability to separate healthy and diseased leaf classes at different classification thresholds. Figure 5 illustrates the ROC curves of all evaluated models. In addition to graphical visualization, the Area Under the Curve (AUC) values were calculated to provide a quantitative evaluation of model discriminative capability. The detailed AUC results are presented in Table 3. The obtained AUC scores were 99% for ResNet50 and VGG16, followed by 97% for EfficientNetB0, 91% for DenseNet121, and 84% for MobileNetV2. These results indicate that ResNet50 and VGG16 demonstrate excellent discriminative capability in distinguishing healthy and diseased Jatropha leaf images across different classification thresholds. The high AUC values are also consistent with the classification performance obtained from the previous evaluation metrics. Comparable discriminative strength for ResNet50 has also been reported in other binary medical image classification settings, although the same study found VGG16 attaining the higher overall classification result between the two architectures [32], suggesting that while both architectures reliably achieve strong class separability across different image domains, their relative ranking in overall accuracy can still vary depending on dataset characteristics.

Table 3. AUC Performance of Evaluated Models

Model	AUC (%)
ResNet50	99%
VGG16	99%
EfficientNetB0	97%
DenseNet121	91%
MobileNetV2	84%

The ROC evaluation reveals that ResNet50 delivers the most favorable classification behavior, as evidenced by a curve situated near the upper-left corner of the plot. Such a pattern indicates effective identification of positive samples with minimal false alarms. Furthermore, the AUC value close to 1 highlights the model's robustness in separating healthy and diseased leaf images.

Other models also demonstrate good classification performance, although their ROC curves are positioned slightly below that of ResNet50. These findings confirm that ResNet50 provides the most stable and reliable classification performance across different threshold settings.

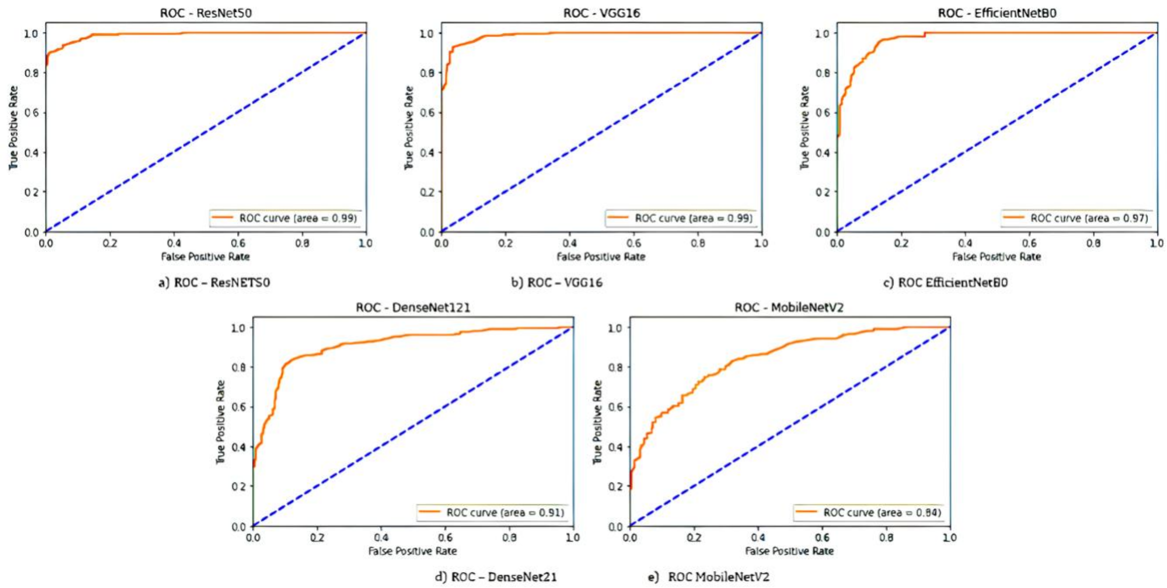

Figure 5. ROC Curve and AUC

3.1.3 Confusion Matrix Analysis

In addition to overall evaluation metrics, the confusion matrix was analyzed to gain deeper insights into class-specific prediction behavior. Figure 6 displays the confusion matrix of ResNet50, highlighting the distribution of true and erroneous classifications produced by the best-performing model.

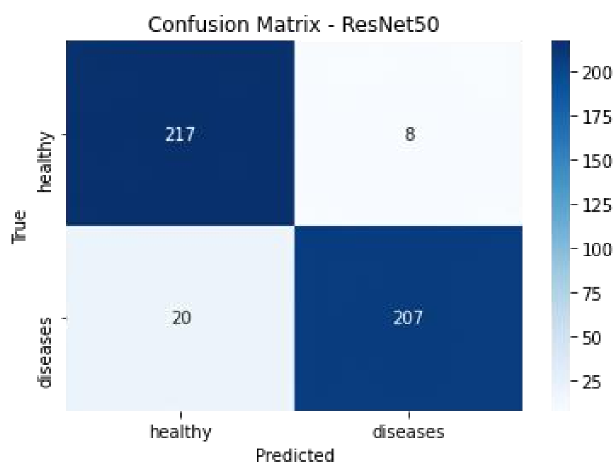

Figure 6. Confusion Matrix of the ResNet50 Model

Based on Figure 6, the values obtained are True Negative (TN) of 217, False Positive (FP) of 8, False Negative (FN) of 20, and True Positive (TP) of 207. Based on these values, manual calculations were performed to validate the model's performance.

$$Accuracy = \frac{207+217}{207+217+8+20} = \frac{424}{452} = 0.9381(93.81\%) \quad (5)$$

$$Precision_{diseases} = \frac{207}{207+8} = \frac{207}{215} = 0.9627(96.27\%) \quad (6)$$

$$Precision_{healthy} = \frac{217}{217+20} = \frac{217}{227} = 0.9119(91.19\%) \quad (7)$$

$$Recall_{diseases} = \frac{207}{207+20} = \frac{207}{215} = 0.9628(96.28\%) \quad (8)$$

$$Recall_{healthy} = \frac{217}{217+8} = \frac{217}{225} = 0.9644(96.44\%) \quad (9)$$

$$F1 - Score_{diseases} = 2\ x\frac{0.9628 \times 0.9119}{0.9628+0.9119} = 0.9367(93.67\%) \quad (10)$$

$$F1 - Score_{healthy} = 2\ x\frac{0.9156 \times 0.9644}{0.9156+0.9644} = 0.9394(93.94\%) \quad (11)$$

To further examine the classification outcomes, performance metrics were manually derived from the confusion matrix shown in Figure 6. The calculated accuracy reached 93.81%, reflecting a high proportion of correctly classified instances among all test samples. This result confirms the effectiveness of the model in recognizing visual patterns associated with both healthy and diseased Jatropha leaves, thereby contributing to its strong overall classification performance.

Additionally, the precision for the "diseases" class of 96.27% indicates that the majority of the model's predictions of diseased leaves are correct. Meanwhile, the precision for the "healthy" class of 91.19% suggests that there are still some errors where leaves that are actually diseased are predicted as healthy.

In terms of recall, the model achieved a value of 96.28% for the "diseases" class, indicating that most infected leaves were successfully detected, although a small number of cases were missed (false negatives). On the other hand, recall for the "healthy" class reached 96.44%, reflecting a high level of accuracy in detecting healthy leaf instances.

Furthermore, the F1 score of 93.67% for the diseased class and 93.94% for the healthy class indicates that the model achieves good balance. This indicates that the model is accurate and consistent in classifying leaf images without significant bias toward either class.

ResNet50 model demonstrates high and stable performance in classifying both classes. The model tends to have higher precision than recall, indicating a strong ability to minimize false positives. However, there are still a small number of false negatives, suggesting that some infected leaves have not been optimally detected. Nevertheless, the model's sensitivity to abnormal features is a key advantage in the context of early plant disease detection, as it helps reduce the risk of missing infected leaves. This balance is plausibly attributable to the residual learning mechanism itself: by allowing gradients and feature information to propagate through shortcut connections across the network's 50 layers, ResNet50 can preserve fine-grained visual cues for both classes during training, rather than disproportionately favoring features associated with the majority or more visually distinct class. This architectural property offers a plausible explanation for the observed balance, although confirming whether it holds consistently across other crops or disease types would require direct empirical comparison in future work.

3.2 Sample Prediction Analysis

Prediction results from the test data, the model demonstrates excellent performance in classifying Jatropha leaf images. Examples of prediction results for the samples are shown in Figure 7.

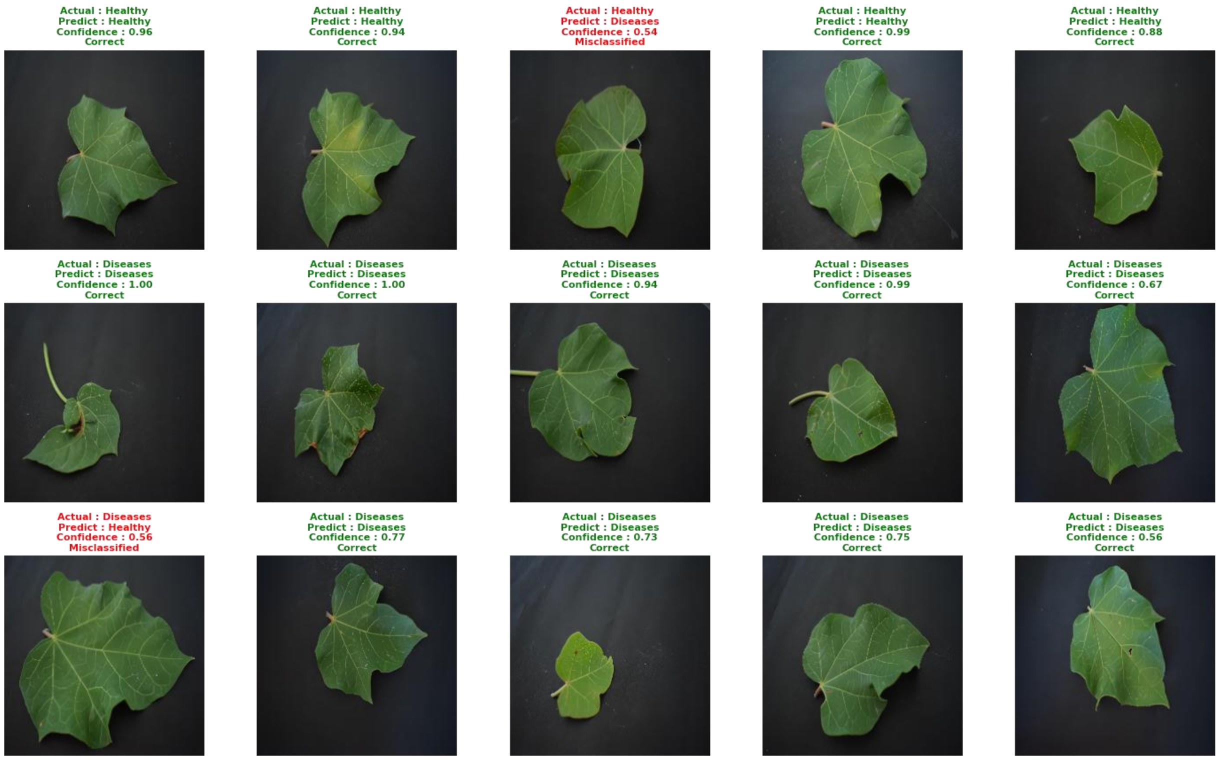

Figure 7. Sample Prediction Results of the ResNet50 Model

Analysis of the prediction visualizations shows that most leaf images were assigned to the correct categories by the ResNet50 model, with confidence values generally remaining at a high level. This result indicates that the model was able to capture representative visual characteristics associated with both healthy and diseased leaves. However, several instances of misclassification still occurred, mainly in cases where disease symptoms appeared subtle or shared visual similarities with healthy leaf features, making class discrimination more challenging.

The misclassified samples are likely caused by several factors. First, some healthy leaves exhibit natural texture variations, uneven color distribution, or small spots that visually resemble early disease symptoms, causing the model to incorrectly classify them as diseased. Second, certain diseased leaves contain only mild infection patterns with limited visible symptoms, making them difficult to distinguish from healthy leaves. Third, variations in illumination, shadows, leaf orientation, and image acquisition conditions may affect the consistency of visual features extracted by the model.

In addition, the relatively limited size and variability of the dataset may reduce the model's ability to generalize subtle feature differences across classes. This limitation is reflected in several misclassified samples that also show lower confidence values compared to correctly classified images. These findings indicate that although the model achieves strong overall performance, additional dataset diversification and real-field image variations are still necessary to improve robustness and reduce classification errors in practical agricultural applications.

These observed sources of misclassification are consistent with limitations that have been broadly documented in the deep learning-based plant disease detection literature. Survey-level evidence on image-based disease detection in dicot plants similarly identifies illumination variability, background complexity, and the visual ambiguity of early-stage symptoms as recurring obstacles that limit the discriminative reliability of convolutional architectures across crop types [15]. A systematic review of deep learning-based disease detection systems likewise reports that subtle or early-stage symptoms are among the most frequently cited causes of model error across studies, regardless of the specific architecture employed [21]. This alignment suggests that the misclassification patterns identified in the present study reflect a known and recurring limitation of vision-based disease detection rather than an issue unique to the ResNet50 model or the Jatropha dataset, reinforcing the relevance of the dataset diversification strategies proposed above.

3.3 Error Analysis and Limitations

The error analysis reveals that incorrect predictions are largely associated with variations in image quality and environmental conditions. Factors such as complex backgrounds, uneven illumination, and naturally occurring leaf defects can generate visual characteristics that resemble disease symptoms. Consequently, the model may encounter difficulties in consistently separating healthy leaves from infected ones when such conditions are present. These observations indicate that classification performance remains influenced by changes in image acquisition settings.

Although the proposed model achieved strong overall performance, the results suggest a slight tendency toward predicting disease in ambiguous cases. This behavior increases the occurrence of false-positive predictions, where healthy leaves are identified as diseased. From an agricultural perspective, however, this outcome may be preferable to false negatives, as overlooking infected plants could delay intervention and potentially increase crop damage. This asymmetric error preference is consistent with the precision-recall pattern reported for the diseased class in the confusion matrix analysis above, where recall slightly exceeded precision, indicating that the model is comparatively more tolerant of misclassifying healthy leaves than of missing genuinely diseased ones. It should be noted, however, that this trade-off is a property of the present model and threshold setting rather than a universally established preference; whether false positives are indeed more acceptable than false negatives is ultimately a context-specific, application-dependent judgment rather than a fixed rule that applies uniformly across all plant disease detection scenarios.

Several limitations of this study also need to be considered. The dataset does not reflect the diversity of environmental conditions encountered in agricultural practices. Consequently, model performance may degrade when applied to data collected outside the training distribution, affecting its generalizability. Variations in illumination, partial occlusion, cluttered backgrounds, and other environmental factors remain underrepresented. Furthermore, differences in class distribution may influence the learning process and contribute to biased predictions.

Additional challenges are expected during real-world deployment, where image quality can be affected by factors such as camera positioning, motion-induced blur, overlapping foliage, and inconsistent lighting. These conditions are considerably more complex than those found in controlled experimental settings and may reduce classification reliability. Future studies should therefore emphasize the collection of larger and more diverse field-based datasets, while also exploring enhanced sampling methods, class-balancing approaches, and advanced image preprocessing techniques. Such improvements are expected to strengthen model robustness and improve generalization performance under practical operating conditions.

3.4 Comparative Model Analysis

A comparative evaluation was performed to investigate the relationship between predictive performance and computational cost among the examined models. This consideration is particularly relevant for agricultural monitoring systems, which are frequently deployed on devices with limited processing power, including mobile platforms and field-operated equipment.

Among the evaluated architectures, ResNet50 achieved the most favorable classification results. The model consistently produced the highest accuracy while maintaining strong precision, recall, and F1-score values. These findings suggest that ResNet50 is highly effective in learning representative features from leaf images, enabling reliable discrimination between disease categories. Nevertheless, the improved predictive capability comes at the expense of increased computational complexity and longer processing time.

In contrast, MobileNetV2 demonstrated the lowest computational burden, requiring less time for both training and inference. Such efficiency makes the architecture attractive for real-time deployment and edge-computing scenarios. However, its comparatively lower classification performance suggests a reduced capacity to capture intricate visual characteristics present in diseased leaf images.EfficientNetB0 and DenseNet121 exhibit moderate performance, suggesting that more advanced architectural designs do not always yield better results. This finding highlights that model effectiveness is highly dependent on dataset characteristics, including size, variability, and feature complexity. Meanwhile, VGG16 achieves competitive accuracy but suffers from extremely high computational costs, making it less suitable for practical deployment despite its strong classification capability.

Overall, the findings confirm that there is no single model that consistently outperforms others across all evaluation aspects. Models with deeper and more complex architectures, such as ResNet50, demonstrate superior capability in extracting discriminative visual features, which contributes to higher classification accuracy. However, these advantages are accompanied by increased computational cost and longer training time. Conversely, lightweight architectures such as MobileNetV2 prioritize computational efficiency and reduced memory consumption, making them more suitable for mobile and edge-device deployment, although their limited representational capacity may reduce performance in distinguishing subtle disease characteristics. In addition, the relatively limited size and variability of the dataset may prevent highly complex architectures from fully utilizing their feature extraction capabilities, thereby affecting model generalization. Therefore, model selection should be aligned not only with accuracy requirements but also with deployment constraints, computational resources, and practical implementation scenarios in agricultural environments.

When positioned against similar comparative studies, the trade-off pattern observed in this research does not consistently replicate across domains, which indicates that the relative ranking of transfer learning architectures is highly dataset-dependent rather than a fixed property of the architectures themselves. A comparison between EfficientNetB0 and ResNet50 on cocoa fruit disease images reported nearly equivalent accuracy between the two models, with EfficientNetB0 even surpassing ResNet50 in precision [23], a much narrower gap than the present study, in which ResNet50 outperformed EfficientNetB0 by more than three percentage points in accuracy. This discrepancy is plausibly related to dataset scale, as the cocoa dataset comprised several thousand images compared with the substantially smaller pool of 257 original Jatropha leaf images used in this study prior to augmentation, suggesting that EfficientNetB0's parameter-efficient design may require a larger and more varied sample to fully realize its competitive advantage. A comparable inconsistency emerges when contrasting ResNet50 and VGG16: whereas the present study found ResNet50 marginally ahead of VGG16, a study comparing the same two architectures on chest X-ray classification reported the opposite ordering, with VGG16 achieving better classification performance than ResNet50 [32]. This reversal may reflect differences in image domain, since the relatively rigid anatomical structure and grayscale intensity patterns of chest radiographs differ considerably from the color and textural variability present in leaf imagery, which could favor different architectural inductive biases. An even more pronounced contrast was found between DenseNet121 and VGG16: the present study recorded DenseNet121 trailing VGG16 by approximately eight percentage points, yet a study on rice plant disease classification using modified VGG-19 and DenseNet-121 architectures reported the opposite pattern, with DenseNet-121 substantially outperforming VGG-19 after architectural modification [34]. Taken together, these contrasting outcomes suggest that the limited size and variability of the Jatropha dataset used in this study may have constrained the feature-reuse mechanisms that typically benefit DenseNet-style architectures, preventing it from reaching the performance levels reported elsewhere. Rather than undermining the present findings, this cross-study divergence reinforces the conclusion that architecture selection for plant disease classification cannot be generalized from a single dataset and must instead account for sample size, class complexity, and domain-specific visual characteristics.

4. Conclusions and Future Works

This research successfully implemented a transfer learning-based deep learning approach for classifying Jatropha leaf diseases using five architectures: MobileNetV2, EfficientNetB0, ResNet50, DenseNet121, and VGG16. The experimental results demonstrate that all models are capable of performing image classification; however, their performance varies significantly depending on architectural characteristics and computational complexity.

Among the evaluated models, ResNet50 produced an accuracy value of 93.81% and balanced recall, precision, and F1-score values. The results reveal that strong classification ability and stable generalization. VGG16 also showed competitive accuracy but required significantly higher computational resources, limiting its practicality. In contrast, MobileNetV2 provided the highest computational efficiency, although with lower accuracy, highlighting a clear trade-off between performance and efficiency. EfficientNetB0 and DenseNet121 demonstrated moderate performance, suggesting that increased architectural complexity does not always guarantee improved results, particularly for datasets with limited variability. While this accuracy remains below the values reported by several prior Jatropha-specific approaches that relied on customized architectures or specialized preprocessing pipelines, the present study offers a more transparent and reproducible benchmark across multiple widely used transfer learning models under a single standardized experimental setting, providing a clearer basis for selecting an architecture according to both accuracy and computational requirements.

Despite these promising results, this study still has limitations in terms of dataset, which may affect the model's generalizability in the real world. Future research should focus on expanding the dataset's variability, including different lighting conditions, backgrounds, and disease severity, to improve robustness. Implementation on mobile or edge-based systems could improve practical applicability, and the incorporation of explainable artificial intelligence methods, such as Grad-CAM, is also recommended to improve model interpretability and support more transparent decision-making. Furthermore, future studies should consider cross-dataset validation to demonstrate model generalization across different image sources and acquisition conditions. Real-time field deployment and testing in real agricultural environments are also needed to assess model confidentiality and practical utility under uncontrolled conditions.

5. References

[1] J. Kotwal, D. R. Kashyap, and D. S. Pathan, "Agricultural Plant Diseases Identification: From Traditional Approach to Deep Learning," *Mater. Today Proc.*, vol. 80, pp. 344–356, 2023, doi: 10.1016/j.matpr.2023.02.370.

[2] T. Raharjo, H. P. Putro, and H. E. Sari, "Early Detection of Tomato Plant Diseases Using the Real-Time Detection Transformer (RT-DETR) Model," *J. Ilm. Sist. Inf.*, vol. 5, no. 1, pp. 584–592, 2026, doi: 10.51903/286bt086.

[3] Y. Gai and H. Wang, "Plant Disease: A Growing Threat to Global Food Security," *Agronomy*, vol. 14, no. 8, 2024, doi: 10.3390/agronomy14081615.

[4] S. Agustiani, R. Aryanti, S. K. Wildah, Y. T. Arifin, S. Marlina, and T. Misriati, "Optimisasi Model Deep Learning untuk Deteksi Penyakit Daun Tebu dengan Fine-Tuning MobileNetV2," *J. Informatics Manag. Inf. Technol.*, vol. 4, no. 4, pp. 150–157, 2024.

[5] Maftuchah and A. Zainudin, *Mengenal Tanaman Jarak Pagar (Jatropha Curcas Linn.)*. Deepublish, 2018.

[6] I. Maulana and I. Darliana, "Efek Mikoriza Arbuskula (MA) dan Zat Perangsang Tumbuh (ZPT) Akar terhadap Pertumbuhan Vegetatif Bibit Tanaman Jarak Pagar (Jatropha curcas L.) Asal Stek Batang," *Compos. J. Ilmu Pertan.*, vol. 7, no. 2, pp. 130–144, 2025, doi: 10.37577/composite.v7i2.888.

[7] D. Wardani, "Isolasi, Identifikasi, dan Aktivitas Antibakteri Fungi Endofit Daun Jarak Pagar (Jatropha Curcas L) terhadap Klebsiella pneumoniae dan Streptococcus pneumoniae," *J. Mhs. Ilmu Farm. dan Kesehat.*, vol. 2, no. 1, 2024, doi: 10.59841/jumkes.v2i1.662.

[8] M. Silalahi, “Jatropha curcas L.: Aspek Botani, Potensi Biofuel, Toksisitas, dan Bioaktivitas Marina,” *J. Pro-Life*, vol. 12, no. 1, pp. 39–50, 2025, doi: 10.33541/pro-life.v12i1.6385.

[9] M. Gusti, T. Ananda, W. Pardosi, A. Maysari, and B. Sembiring, “Literatur Review: Metode Thresholding Untuk Mengidentifikasi Penyakit Pada Daun Bawang Merah,” *FIMERKOM J. Inf. Syst. Technol.*, vol. 1, no. 1, pp. 6–10, 2024.

[10] A. Abbas, S. Jain, M. Gour, and S. Vankudothu, “Tomato Plant Disease Detection Using Transfer Learning with C-GAN Synthetic Images,” *Comput. Electron. Agric.*, vol. 187, p. 106279, 2021, doi: 10.1016/j.compag.2021.106279.

[11] A. T. Adiana, Jumadi, and E. Nurlatifah, “Klasifikasi Penyakit Pada Daun Padi Menggunakan Model Hybrid CNN-SVM,” *SMATIKA STIKI Inform. J.*, vol. 15, no. 2, pp. 258–267, 2025, doi: 10.32664/smatika.v15i02.1548.

[12] K. Anwar, “Sistem Deteksi Wajah Berbasis Deep Learning Menggunakan Convolutional Neural Network (CNN),” *J. Comput. Sci. Inf. Technol.*, vol. 1, no. 2, pp. 46–52, 2025, doi: 10.70716/jocsit.v1i2.258.

[13] S. Agustiani, Y. T. Arifin, A. Junaidi, S. K. Wildah, and A. Mustopa, “Klasifikasi Penyakit Daun Padi Menggunakan Random Forest dan Color Histogram,” vol. 10, no. 1, 2022, doi: 10.23960/komputasi.v10i1.2961.

[14] E. F. Pangestu and B. Irawan, “Klasifikasi Penyakit Tanaman Padi Menggunakan Metode Convolutional Neural Network dengan Arsitektur MobileNetV2,” *JITET*, vol. 14, no. 1, pp. 1003–1013, 2026, doi: 10.23960/jitet.v14i1.8900.

[15] A. Anis, P. Wang, C. Li, and F. Sohel, “A Survey of Deep Learning Techniques for Image-Based Disease Detection in Dicot Plants,” *Inf. Process. Agric.*, 2026, doi: 10.1016/j.inpa.2026.03.014.

[16] A. I. Khan, S. M. K. Quadri, S. Banday, and J. L. Shah, “Deep Diagnosis: A Real-Time Apple Leaf Disease Detection System Based on Deep Learning,” *Comput. Electron. Agric.*, vol. 198, p. 107093, 2022, doi: 10.1016/j.compag.2022.107093.

[17] S. Siswanto, M. U. Dewi, S. Kholifah, G. Widhiati, and W. Aryani, “Penggunaan Model Deep Learning Untuk Meningkatkan Efisiensi Dalam Aplikasi Machine Learning,” *J. Penelit. Sist. Inf.*, vol. 1, no. 4, pp. 215–238, 2023, doi: 10.54066/jpsi.v1i4.1619.

[18] R. Gunawan and others, “Pendekatan Transfer Learning untuk Klasifikasi Penyakit Mata Menggunakan Citra dengan CNN InceptionV3,” *J. Comput. Sci. Inf. Technol.*, vol. 6, no. 1, pp. 60–67, 2025, doi: 10.37859/coscitech.v6i1.8509.

[19] S. Nabila and others, “Analisi Pengaruh Urutan Preprocessing Terhadap Kinerja MobileNetV2 dan VGG16 untuk Klasifikasi Penyakit Daun,” *J. Comput. Sci. Artif. Intell.*, vol. 6, no. 2, pp. 1–6, 2025, doi: 10.32485/jcsai.JOURNAL.

[20] T. Yunan, A. Rajab, and N. Nafiiyah, “Evaluasi Kinerja Model CNN Berbasis Transfer Learning dalam Klasifikasi Penyakit Daun Padi,” *J. Sist. Komput. dan Inform.*, vol. 7, no. 3, pp. 989–995, 2026, doi: 10.30865/json.v7i3.9539.

[21] I. K. P. G. Hartawan and I. M. G. Sunarya, “Tinjauan Sistematis Literatur Tentang Sistem Deteksi Penyakit Berbasis Deep Learning,” *Digit. Transform. Technol.*, vol. 6, no. 1, pp. 112–123, 2026.

[22] F. Putra, “Penerapan Teknologi Machine Learning dalam Deteksi Dini Penyakit Pada Tanaman Pangan,” *J. Kolaborasi Sains dan Ilmu Terap.*, vol. 3, no. 1, pp. 1–5, 2024.

[23] N. P. Maylianti, I. G. Ngurah, L. Wijayakusuma, I. P. Chandra, and A. Wiguna, “Comparison of EfficientNet-B0 and ResNet-50 for Detecting Diseases in Cocoa Fruit,” *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 115–120, 2025, doi: 10.30871/jaic.v9i1.8868.

[24] N. Pratama, A. S. Sunge, and E. Budiarto, "Penerapan Model MobileNetV2 Untuk Prediksi Tingkat Roasting Biji Kopi Berbasis Gambar Pada Bot Telegram," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 4571–4576, 2025, doi: 10.31004/riggs.v4i2.1272.

[25] T. D. A. R. Putra and Y. F. Riti, "Implementasi Arsitektur MobileNetV2 untuk Klasifikasi Rumah Adat Berbasis Android," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 14, no. 2, pp. 217–222, 2026.

[26] S. S. Chouhan, U. P. Singh, U. Sharma, and S. Jain, "Leaf Disease Segmentation and Classification of Jatropha Curcas L. and Pongamia Pinnata L. Biofuel Plants Using Computer Vision Based Approaches," *Measurement*, vol. 171, p. 108796, 2021, doi: 10.1016/j.measurement.2020.108796.

[27] S. S. K. B, A. M, L. G. D, N. R, P. P, and S. R, "Jatropha Leaf Disease Classification Using Improved Inception ResNetV2," in *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, 2023, pp. 1–5. doi: 10.1109/ViTECoN58111.2023.10157396.

[28] E. Saxena, S. S. Chouhan, A. Shukla, R. K. Patel, S. Solanki, and U. P. Singh, "Machine Vision Methodology for Jatropha Plant Leaf Disease Detection," in *Trends in Computer Engineering, Data Science and Artificial Intelligence*, 2026, pp. 159–168. doi: 10.1007/978-3-032-10199-0_10.

[29] T. H. Saragih, D. M. N. Fajri, A. Hamdianah, W. F. Mahmudy, and Y. P. Anggodo, "Jatropha Curcas Disease Identification Using Fuzzy Neural Network," in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, 2017, pp. 305–309. doi: 10.1109/SIET.2017.8304153.

[30] S. S. Chouhan, U. P. Singh, A. Kaul, and S. Jain, "A Data Repository of Leaf Images: Practice Towards Plant Conservation with Plant Pathology," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, 2019, pp. 700–707. doi: 10.1109/ISCON47742.2019.9036158.

[31] S. A. Maulana, S. H. Batubara, Y. Permata, P. Pasaribu, H. Syahputra, and F. Ramadhani, "Deteksi Burung Menggunakan Convolutional Neural Network (CNN) dengan Model Arsitektur MobileNetV2," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 6108–6114, 2024, doi: 10.36040/jati.v8i4.10126.

[32] T. Berliani, E. Rahardja, and L. Septiana, "Perbandingan Kemampuan Klasifikasi Citra X-ray Paru-paru Menggunakan Transfer Learning ResNet-50 dan VGG-16," *J. Media Heal.*, vol. 5, no. 2, pp. 123–135, 2023, doi: 10.28932/jmh.v5i2.6116.

[33] D. A. Agustina, "Klasifikasi Citra Jenis Kulit Wajah dengan Algoritma Convolutional Neural Network (CNN) ResNet-50," *J. Ris. Sist. Inf.*, vol. 1, no. 3, pp. 2–8, 2024, doi: 10.22373/cj.v8i2.25441.

[34] I. Megahaztuti, B. Yuwono, B. M. Akbar, H. C. Rustamaji, and S. P. Nugroho, "Performance Comparison of VGG-19 and DenseNet-121 Architectures for Rice Plant Disease," *Telematika*, vol. 22, no. 2, pp. 74–88, 2025, doi: 10.31315/telematika.v22i2.14437.

[35] D. Setiawan, S. Widodo, T. Ridwan, and R. Ambari, "Perancangan Deteksi Emosi Manusia Berdasarkan Ekspresi Wajah Menggunakan Algoritma VGG16," *Syntax J. Inform.*, vol. 11, no. 1, pp. 1–11, 2022, doi: 10.35706/syji.v11i01.6594.

[36] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 640–651, 2021, doi: 10.30865/mib.v5i2.2937.

[37] D. P. Wijaya, L. D. Murti, and M. R. Rachman, "Recall dan Precision pada Online Public Access Catalog (OPAC) Dinas Arsip dan Perpustakaan Kota Bandung," *Visi Pustaka*, vol. 24, no. 1, 2022, doi: 10.37014/visipustaka.v24i1.2915.

[38] A. Ridhovan and A. Suharso, "Penerapan Metode Residual Network (ResNet) dalam Klasifikasi Penyakit pada Daun Gandum," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 1, pp. 58–65, 2022, doi: 10.29100/jipi.v7i1.2410.

[39] D. Kurniawan, M. Wahyudi, and L. Pujiastuti, "Deteksi dan Prediksi Cerdas Penyakit Paru-Paru dengan

Algoritma Random Forest," *Indones. J. Comput. Sci.*, vol. 3, no. 1, 2024, doi: 10.31294/ijcs.v3i1.6071.

[40] M. C. Rani, F. D. Azkia, R. A. Dewi, M. Wahyudi, Sumanto, and A. S. Budiman, "Perbandingan Algoritma Random Forest, Naive Bayes, dan Neural Network dalam Klasifikasi Penyakit Jantung," *J. Sains Inform. Terap.*, vol. 4, no. 2, pp. 77–84, 2025, doi: 10.62357/jsit.v4i2.609.