Journal of Business, Social and Technology
Volume 7, Issue 3, 1306-1319
e_ISSN: 2807-6362
https://bustechno.polteksci.ac.id/index.php/jbt/index
DOI: doi.org/10.59261/jbt.v7i3.723

Comparative Analysis of YOLOv8 and MobileNetV2 for Digital Image-Based Bird's Eye Chili Leaf Disease Classification

***I Gede Tiar Eka Saputra**
Universitas Pendidikan Ganesha, Indonesia

I Made Gede Sunarya
Universitas Pendidikan Ganesha, Indonesia

***Corresponding author:**
I Gede Tiar Eka Saputra, Universitas Pendidikan Ganesha, Indonesia.
✉tiar@student.undiksha.ac.id

Article Info:
Article history:
Received: June 15, 2026
Revised: July 24, 2026
Accepted: August 12, 2026
Available Online: August 24, 2026

Keywords:
YOLOv8-cls; MobileNetV2; Deep Learning; Leaf Disease Classification; Bird's Eye Chili.

Abstract
Background: Bird's eye chili is an important horticultural commodity in Indonesia but is highly vulnerable to leaf diseases such as mosaic virus, leaf spot, and chlorosis. Visual disease identification is often inaccurate and can delay treatment, while previous deep learning studies have mainly used laboratory-acquired datasets with limited applicability to field conditions.
Objective: This study compares the performance of YOLOv8 classification (YOLOv8-cls) and MobileNetV2 in classifying bird's eye chili leaf diseases using digital images acquired under field conditions.
Methods: A dataset of 800 leaf images representing four classes (healthy, mosaic virus, leaf spot, and chlorosis) was collected under natural field conditions. The images were preprocessed and used to train both models through transfer learning using ImageNet-pretrained weights. Performance was evaluated using stratified 5-fold cross-validation and measured using accuracy, precision, recall, F1-score, and the Matthews correlation coefficient (MCC).
Results: Both models achieved average accuracies above 98%. YOLOv8-cls achieved the best overall performance, with 98.38% accuracy, 98.55% precision, 98.35% recall, and 98.41% F1-score, while requiring only 2.84 MB of storage, making it suitable for deployment on edge devices. MobileNetV2 achieved 98.25% accuracy, 98.36% precision, 98.16% recall, and 98.24% F1-score, while reducing training time by approximately 40%.
Conclusion: YOLOv8-cls is the more suitable architecture for mobile and edge-based plant disease detection because of its superior classification performance and compact model size, whereas MobileNetV2 is advantageous when rapid model retraining is required. These findings provide a practical basis for developing accessible digital agriculture applications for early disease detection.

To cite this article: Saputra, I. G. T. E., & Sunarya, I. M. G. (2026). Comparative Analysis of YOLOv8 and MobileNetV2 for Digital Image-Based Bird's Eye Chili Leaf Disease Classification. *Journal of Business, Social and Technology*, 7(3), 1306-1319. https://doi.org/10.59261/jbt.v7i3.723

INTRODUCTION

Bird's eye chili is one of the horticultural commodities that plays a significant role in the lives of Indonesian people. In addition to being a key ingredient in various culinary dishes, chili peppers also influence national economic stability, as fluctuations in their prices frequently contribute to food inflation. Unstable bird's eye chili production can directly affect market prices and reduce food supply. One of the primary factors affecting chili productivity is the occurrence of plant diseases and pest infestations affecting leaves and fruits. Bird's eye chili leaf diseases can be caused by viruses, bacteria, or fungi, leading to symptoms such as discoloration, necrotic spots, leaf curling, and tissue damage (Sumantara et al., 2024).

Diseases that are not detected at an early stage can significantly reduce both the quality and quantity of crop yields. At the farmer level, disease identification is generally performed manually through visual observation, which carries a relatively high risk of misdiagnosis. Recent advances in computer vision and deep learning technologies have created new opportunities for developing automated plant disease identification systems. Deep learning, particularly Convolutional Neural Networks (CNNs), has the capability to automatically extract visual features from digital images without requiring manual feature engineering. The robustness of CNN-based methods in extracting complex image features and achieving high classification accuracy has been demonstrated across various domains, including medical image classification using periapical radiography (Sumantara et al., 2024)

In the agricultural sector and environmental image processing, CNN-based architectures such as YOLO and MobileNet have demonstrated strong performance. YOLO is a modern deep learning architecture with strong capabilities for feature extraction and real-time visual analysis. The effectiveness of the YOLO family in agriculture has been demonstrated in previous studies, for example, in comparative architectural analyses for crop object estimation, such as coconut fruit detection (Natih et al., 2026). Meanwhile, MobileNetV2 was designed as a lightweight architecture optimized for environments with limited computational resources. This architecture has demonstrated robust performance in processing local image data under dynamic lighting and environmental conditions, as demonstrated by the optimization of MobileNet using camera datasets collected in Bali (Saputra et al., 2026).

Comparing these two models is important for understanding the trade-off between classification accuracy and computational efficiency. Most previous studies have relied on public datasets collected under controlled laboratory conditions, which significantly limits the generalizability of their findings to real-world agricultural environments. Field-collected datasets more accurately represent the visual complexity encountered under actual farming conditions (including varying lighting conditions, natural soil and vegetation backgrounds, and diverse levels of disease severity) thereby providing greater ecological validity and enabling the development of models that perform more reliably when deployed in the field. Therefore, this study utilizes an independently collected dataset obtained directly from cayenne pepper farms in Bali. The primary dataset consists of 800 images covering four major classes: healthy leaves, mosaic virus infection, leaf spot disease, and pest-induced damage. The use of local field data is expected to improve the models' generalization capability when deployed under real agricultural conditions. Based on this background, this study aims to analyze and compare the performance of YOLOv8-cls and MobileNetV2 for cayenne pepper leaf disease classification using digital images.

Deep learning is a subfield of machine learning that employs multilayer artificial neural networks to automatically learn data representations. The core concept of this architecture lies in its ability to perform hierarchical feature extraction, progressing from low-level to high-level representations. According to the seminal work of LeCun et al (2015), the early layers of deep learning networks generally extract basic features such as edges and lines, while deeper layers combine these features into more complex and meaningful visual representations (LeCun et al., 2015; Somvanshi et al., 2026).

This capability for automatic feature representation and learning makes deep learning highly effective for solving digital image processing problems involving complex patterns. Deep learning algorithms have demonstrated excellent performance not only on standard datasets but also on local datasets with unique visual characteristics. One example of a local application is the successful use of deep learning methods to classify traditional Balinese spice mixtures known as Base Genep, which exhibit substantial variations in color and texture (Prasetia & Sunarya, 2024).

The successful classification of such complex local objects demonstrates the flexibility and robustness of deep learning architectures. Therefore, this approach is highly relevant and suitable for identifying cayenne pepper leaf diseases. Deep learning models are expected to effectively detect and distinguish disease features such as leaf spots, pigment changes caused by viral infections, and overlapping physical damage to leaf tissues.

The rapid advancement of convolutional neural networks (CNNs) has enabled computer vision systems to recognize highly complex visual patterns with remarkable accuracy. CNNs allow features to be extracted automatically from raw image pixels, making them particularly effective for monitoring plant phenotypes and classifying leaf health conditions. The application of deep learning-based computer vision techniques has demonstrated the capability to achieve very high accuracy in detecting and classifying various diseases and pests affecting chili leaves, potentially reducing the reliance on manual field diagnosis processes (Nurokhman et al., 2024).

Furthermore, modern computer vision systems are expected to perform optimally not only on controlled laboratory datasets but also under dynamic real-world environmental conditions. The robustness of computer vision methods in handling image variability in outdoor environments has been demonstrated in numerous field implementations. For example, hyperparameter optimization of neural network architectures such as MobileNet has been successfully applied to classify highly complex visual data captured by outdoor cameras in Bali (Suputra et al., 2025). This finding confirms that contemporary computer vision algorithms are highly adaptive and robust in processing local images characterized by varying lighting conditions and natural backgrounds, making them a suitable technological foundation for agricultural image analysis in open-field environments.

Deep learning is a subfield of machine learning that utilizes multilayer artificial neural networks to automatically learn data representations. Its fundamental concept is the hierarchical extraction of features from low-level to high-level representations. According to (LeCun et al., 2015; Somvanshi et al., 2026), the initial layers typically extract simple features such as edges and lines, while deeper layers integrate these features into more meaningful and complex visual representations. These deep learning architectures can enhance computational performance and accuracy by increasing the depth of the network, and their robustness extends to various fields, including the processing of complex remote-sensing images for broad environmental analysis (Dewi et al., 2025).

Convolutional neural networks (CNNs) are deep learning architectures specifically designed to process two-dimensional grid-structured data such as digital images. Unlike conventional neural networks that treat each pixel as an independent feature, CNNs exploit local connectivity and weight-sharing mechanisms. These characteristics enable CNNs to preserve spatial information and recognize local patterns hierarchically, ranging from low-level features such as edges and lines to high-level features such as object shapes and textures.

The fundamental principles of CNN architectures were first comprehensively demonstrated through the classic LeNet-5 model for document and character recognition (Bitar et al., 2023; LeCun et al., 1998; Nugraha et al., 2023). Based on this original concept, CNNs are constructed using several core components, including convolutional layers, pooling layers, activation functions, and fully connected layers. Convolutional layers apply filter matrices (kernels) across input images to generate feature maps representing specific visual characteristics. These feature maps are subsequently processed through nonlinear activation functions, enabling the network to learn complex patterns. Pooling layers then reduce the spatial dimensions of feature maps while preserving essential information. Finally, the extracted features can be passed through fully connected layers to perform classification into target categories.

CNNs have demonstrated exceptional capability in extracting subtle and complex spatial features, making them highly effective for recognizing objects with visually similar morphological characteristics. For instance, CNNs combined with the ResNet-50 architecture have been successfully employed to identify different species of little tuna based on comparative visual patterns in body structures (Pusparani et al., 2024). Furthermore, specialized CNN-based network architectures, such as the U-Net architecture, have also demonstrated high precision in segmenting intricate visual features and complex morphological shapes within clinical imaging datasets (I Made Angga Darma Putra et al., 2023). This success in recognizing intricate spatial patterns in biological objects confirms that layered CNN architectures are highly effective for handling variations in natural images. Therefore, CNN-based approaches are particularly relevant to this study for extracting and classifying disease-related features in cayenne pepper leaves, including mosaic virus-induced discoloration and necrotic textures associated with leaf spot symptoms.

You Only Look Once version 8 (YOLOv8) is a member of the YOLO architecture family developed by Ultralytics for image detection, segmentation, classification, and related computer vision tasks, with an emphasis on high inference speed and accuracy. A key distinction between YOLOv8 and earlier YOLO versions is its use of an anchor-free detection approach and a decoupled detection head. The anchor-free approach enables the network to predict object locations without relying on predefined anchor boxes, reducing the need for manually designed anchor configurations and improving computational efficiency (Terven et al., 2023). Meanwhile, the decoupled detection head separates the classification and bounding-box regression branches, allowing these tasks to be optimized more independently and contributing to improved training and detection performance (Terven et al., 2023).

The reliability of YOLO architectures in recognizing objects with complex and specific visual characteristics has been widely demonstrated in previous studies. For example, YOLOv5 was successfully applied to detect traditional Balinese shadow puppet (*Wayang Kulit*) characters in *Wayang Peteng* performances (Asmara et al., 2023; Solichin et al., 2025; Widaningrum et al., 2025). The study showed that YOLO architectures are highly robust in extracting detailed and varied shape features even under challenging environmental conditions with dynamic lighting. Such success in modeling culturally significant local objects provides evidence that YOLO-based algorithms possess strong generalization capabilities for handling visual complexity in real-world environments.

By incorporating architectural improvements, the YOLOv8 classification variant (YOLOv8-cls) offers efficient image classification with a relatively lightweight parameter structure. Its deep feature extraction capability makes it particularly suitable for classifying cayenne pepper leaf disease images collected directly from agricultural fields. The model is expected to effectively suppress natural background variability, such as soil and branches, while focusing on disease-related visual features with high accuracy and low computational overhead.

MobileNetV2 is a lightweight convolutional neural network (CNN) architecture specifically designed to address computational and memory limitations in mobile and embedded devices. The primary innovation of this architecture, as introduced in its original design (Sandler et al., 2018; Yu et al., 2026), lies in the use of depthwise separable convolutions combined with inverted residual blocks and linear bottlenecks.

Depthwise separable convolution efficiently decomposes standard convolution into two separate operations: depthwise convolution, which filters each channel independently, and pointwise (1×1) convolution, which recombines the channels. This mechanism drastically reduces the number of parameters and multiply–accumulate (MAC) operations without substantially sacrificing feature representation capacity.

Furthermore, inverted residual blocks enable the network to expand feature-map dimensions in intermediate layers for nonlinear transformations before compressing them through linear bottlenecks, thereby helping preserve useful information while reducing computational complexity. This lightweight yet powerful architecture makes MobileNetV2 highly competitive for agricultural image classification tasks. The effectiveness of the MobileNet family in detecting visual anomalies in leaves has been demonstrated through transfer learning studies for disease detection in *Eucalyptus pellita* leaves (Wita & Subekti, 2023). Recent comparative studies further reinforce this finding, demonstrating that MobileNetV2 can outperform heavier architectures such as VGG16 in computational efficiency, generalization, and classification accuracy for mangrove leaf diseases without overfitting (Yudhantara et al., 2026). The results showed that this lightweight architecture can identify plant pathological features with competitive accuracy while maintaining a significantly smaller model size than conventional large-scale CNN architectures.

The efficiency characteristics of MobileNetV2 make it highly suitable for cayenne pepper leaf disease classification in this study. Beyond evaluating classification accuracy, the model is also assessed for its potential deployment on low-specification smartphones commonly used by farmers, potentially eliminating the need for high-performance server infrastructure.

Numerous previous studies have explored the application of deep learning architectures for detecting plant diseases and pests. The use of YOLOv8 in agricultural applications has shown highly promising results. (Dewi et al., 2024) evaluated YOLOv8 for rice leaf disease classification and achieved an mAP50 score of 97.9%. The robustness of the architecture in recognizing small objects in agricultural environments was further supported by Manurung et al. (2024), who implemented YOLOv8 for potato beetle detection and obtained a precision score of 88.2% (Manurung et al., 2024). These studies confirm the strong capability of the YOLOv8 family in extracting complex visual features from plant imagery.

On the other hand, lightweight architectures have become a major focus in the development of intelligent mobile-based systems. Lu et al. (2023) evaluated MobileNetV2 for plant disease classification and demonstrated an accuracy of 99.53% while maintaining a very low number of model parameters (Lu et al., 2023). Similarly, Njoroge et al. (2025) highlighted the superiority of MobileNetV2 in terms of inference speed and memory efficiency when deployed directly on resource-constrained edge devices (Njoroge et al., 2025). The reliability of MobileNetV2 for agricultural applications was further reinforced by Tamayasa & Dewi (2026), who found that it effectively classified corn leaf diseases with an optimal testing accuracy of 96.50%, demonstrating greater effectiveness and stability than the EfficientNetB3 architecture (Tamayasa & Dewi, 2026).

Although these studies reported impressive results, a critical examination reveals several important limitations. First, most evaluations were conducted on different crop species (such as rice, potato, and corn) and frequently relied on secondary public datasets (e.g., PlantVillage) collected under controlled environmental conditions, which substantially reduces real-world generalizability. Second, few studies have employed stratified k-fold cross-validation; most relied on a single train-test split, which may introduce data-partitioning bias. Third, performance evaluations in prior work were predominantly limited to accuracy alone, without comprehensive multi-metric reporting (precision, recall, F1-score, MCC) or computational efficiency metrics such as model size and training time. Few studies have comprehensively compared the performance trade-offs between YOLOv8-cls, which emphasizes detailed feature extraction, and MobileNetV2, which prioritizes computational efficiency, particularly for cayenne pepper plants cultivated in open-field environments.

This study addresses these research gaps by: (1) using a primary field-collected dataset from Bali that captures natural agricultural variability in lighting, background, and disease severity; (2) employing stratified 5-fold cross-validation to ensure robust and unbiased evaluation; (3) reporting comprehensive multi-metric performance analysis, including MCC alongside standard classification metrics; and (4) directly comparing YOLOv8-cls and MobileNetV2 using hundreds of primary images collected from local agricultural fields. The evaluation specifically focuses on distinguishing four key conditions: healthy leaves, mosaic virus infection, leaf spot disease, and pest-induced damage. Consequently, the resulting models are expected to demonstrate higher ecological validity and practical applicability under real-world agricultural conditions.

This research is motivated by the need for accurate and efficient image-based classification methods for cayenne pepper leaf diseases to support digital agriculture practices. Although the YOLO architecture is primarily known as a real-time object detection framework, Ultralytics provides YOLOv8 classification models, denoted by the -cls suffix, specifically designed for image classification tasks. The classification variant uses a classification head rather than the detection head employed in object detection models, enabling it to function as an image classifier while retaining the YOLOv8 backbone's advantages in efficient feature extraction and relatively compact model size. This makes YOLOv8-cls directly comparable to MobileNetV2 for leaf disease classification. Accordingly, the study addresses several key research questions: (1) How well does the YOLOv8-cls model perform in classifying cayenne pepper leaf diseases? (2) How well does the MobileNetV2 model perform in the same task? (3) How do the two architectures compare based on evaluation metrics such as accuracy, precision, recall, F1-score, Matthews correlation coefficient (MCC), model size, and inference time?

In line with these research questions, the objective of this study is to evaluate the performance of YOLOv8-cls and MobileNetV2 in classifying cayenne pepper leaf diseases using digital imagery. Furthermore, the study seeks to comprehensively compare both models to identify their respective strengths and limitations, thereby determining the most effective and efficient model for supporting the implementation of digital agriculture technologies.

The contributions of this study include: (1) the development of a primary cayenne pepper leaf image dataset collected under real open-field conditions in Bali, enabling a more ecologically valid representation of disease variations than public benchmark datasets; (2) a rigorous comparative evaluation employing stratified 5-fold cross-validation with multi-metric assessment (accuracy, precision, recall, F1-score, and MCC), providing a more robust and less biased performance analysis than a single train-test split; (3) a systematic efficiency analysis encompassing model size and training time, in addition to classification accuracy; and (4) evidence-based recommendations regarding the most suitable deep learning architecture for deployment in mobile agriculture systems to support rapid, accurate, and practical plant disease detection in modern farming environments.

METHOD

This study employed a quantitative approach using an experimental research method. The experiment was conducted by training, testing, and objectively comparing two different deep learning architectures, namely YOLOv8-cls and MobileNetV2. Both models were trained using the same dataset under controlled hyperparameter settings. The objective was to empirically analyze the trade-off between classification accuracy and computational resource efficiency of each architecture in recognizing cayenne pepper leaf diseases.

Images were acquired under natural field conditions, including varying lighting environments, backgrounds, and disease severity levels. Data acquisition was conducted using a smartphone camera (12 MP resolution, f/1.8 aperture) at a shooting distance of approximately 15–30 cm from the leaf surface. Photographs were taken during morning and afternoon hours (07:00–09:00 and 15:00–17:00 WITA) to capture diverse natural lighting conditions. All images were taken outdoors in actual cayenne pepper cultivation areas without any artificial lighting or controlled backdrop, ensuring that the dataset authentically represents the visual complexity encountered during real field deployment.

Table 1
The Dataset

No	Leaf Condition Class	Visual Description	Number of Images
1	Healthy	Healthy leaves without spots or physical deformities	200
2	Mosaic Virus	Leaves exhibiting yellow mosaic patterns and curling symptoms	200
3	Leaf Spot	Leaves with circular necrotic lesions caused by fungal or bacterial infection	200
4	Chlorosis	Leaves showing pale yellow discoloration due to chlorophyll deficiency or disruption	200
		Total	800

To optimize feature extraction and accelerate model convergence during training, several image preprocessing stages were performed:

1. Cropping
   Irrelevant image regions were removed to reduce background noise, such as soil and agricultural mulch, allowing the model to focus on the primary leaf object.
2. Resizing
   Image dimensions were adjusted according to the input requirements of each architecture. Images were resized to 224 × 224 pixels for MobileNetV2 and 640 × 640 pixels for YOLOv8-cls.
3. RGB Normalization
   Pixel intensity values were transformed from the original range of 0–255 to a normalized range of 0–1 to stabilize gradient calculations during backpropagation.
4. Data Augmentation
   To reduce the risk of overfitting due to the limited dataset size, several augmentation techniques were applied to the training data, including random rotation, horizontal and vertical flipping, zoom scaling, and brightness adjustment. These transformations simulate variations commonly encountered in real agricultural environments.

Instead of employing a single static train-test split, this study adopted a Stratified 5-Fold Cross Validation strategy. This approach was selected to ensure a more objective model evaluation, minimize partitioning bias, and guarantee that each image would be used as both training and validation data during different iterations.

The complete dataset of 800 images was randomly divided into five equal folds, each containing 160 images while maintaining balanced class distributions. During each iteration, four folds (640 images or 80%) were used for training, while the remaining fold (160 images or 20%) served as validation data. This process was repeated five times, rotating the validation fold in each iteration. The data distribution for each fold is presented in Table 2.

Table 2
Data Distribution

Class	Training Data (4 fold /80%)	Validation Data (1 fold /20%)	Total Data
Healthy	160	40	200
Mosaic Virus	160	40	200
Leaf Spot	160	40	200
Chlorosis	160	40	200
Total	640	160	800

Both models were trained using a transfer learning approach by initializing the networks with pre-trained weights obtained from the ImageNet dataset. This strategy allows the models to leverage previously learned visual representations before being fine-tuned for cayenne pepper leaf disease classification.

To ensure objectivity and avoid biases associated with a fixed train-test split, model training was conducted using the 5-Fold Cross Validation scheme. The dataset was randomly divided into five equal folds. Training was performed in five separate iterations, where four folds were used as training data and one-fold was used for validation. The validation fold rotated across iterations until all folds had been used as validation data exactly once. Final model performance was calculated as the average performance across all five iterations.

The experiments were conducted under identical computational conditions using the following configuration parameters:

Table 3
Configuration Parameters

Training Parameter	Configuration Value
Validation Strategy	5-Fold Cross Validation
Epoch	100 per Fold (Total 500 Epoch per Arsitektur)
Batch Size	16 (MobileNetV2) / 32 (YOLOv8-cls)
Learning Rate	0.001
Optimizer	Adam (Adative Moment Estimation)
Weight Initialization	ImageNet

The final performance evaluation was conducted using validation data that had not been utilized during model training. Evaluation metrics were calculated quantitatively based on the confusion matrix, which consists of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values.

Accuracy

Accuracy measures the proportion of correctly classified instances, both positive and negative, relative to the total number of observations. This metric provides a general indication of how well the model classifies all cayenne pepper leaf conditions.

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

Precision

Precision measures the proportion of correctly predicted positive cases among all positive predictions generated by the model. In plant pathology applications, high precision indicates that when the model predicts a leaf to be infected with Mosaic Virus, the probability of the prediction being correct is very high, thereby minimizing false alarms.

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

Recall (Sensitivity)

Recall measures the proportion of actual positive cases that are correctly identified by the model. This metric is particularly important in disease detection because high recall indicates that the model successfully identifies most diseased leaves, minimizing the risk of infected leaves being misclassified as healthy.

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

F1-Score

The F1-Score is the harmonic mean of Precision and Recall. This metric is particularly useful when balancing the trade-off between false positives and false negatives, especially when slight class imbalances exist within the dataset.

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

RESULTS AND DISCUSSION

Research Results

Model Training Convergence Dynamics

To understand the learning behavior of the models throughout the 100 training epochs, the experiment conducted on Fold 3 was selected as a representative sample. The convergence process was analyzed by observing the trends of the Loss and Accuracy curves during both the training and validation phases.

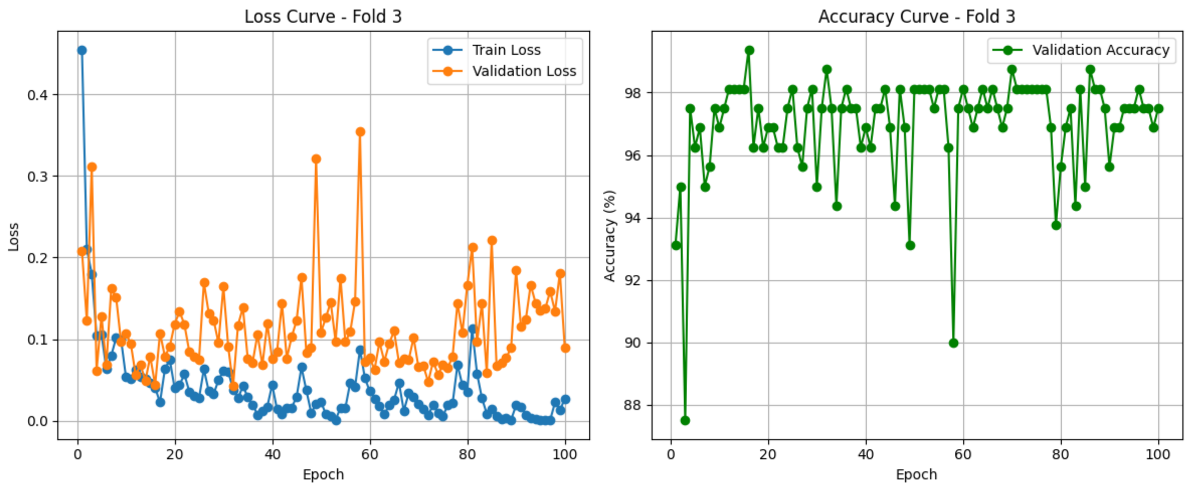

Figure 1. Training Loss and Accuracy of MobileNetV2 in Fold 3

Matthews Correlation Coefficient (MCC)

MCC provides a balanced measure of binary and multi-class classification quality, particularly robust when classes are imbalanced. It takes into account all four confusion matrix values (TP, TN, FP, FN) and returns a coefficient between −1 and +1, where +1 represents a perfect prediction, 0 an average random prediction, and −1 an inverse prediction. For multi-class problems, the multi-class MCC (MMCC) is computed recently validated for plant disease classification tasks (Obafemi-Ajayi et al., 2025; Hameed et al., 2026).

$$MCC = (TP \times TN - FP \times FN) / \sqrt{[(TP+FP)(TP+FN)(TN+FP)(TN+FN)]}$$

In the MobileNetV2 architecture (Figure 1 – Fold 3 Learning Curve), the Train Loss exhibited a highly stable downward trend approaching zero. However, the Validation Loss displayed noticeable fluctuations (spikes) at several epochs, particularly around epochs 50, 60, and 80. These fluctuations can be attributed to several possible factors: (1) mild overfitting tendencies in which the model memorizes training patterns but responds sensitively to new validation samples; (2) variation in the composition of each validation fold's batch, given the relatively small dataset size of 800 images; and (3) sensitivity of the depthwise separable convolution mechanism to aggressive augmentation transformations such as extreme rotation and brightness adjustment. Nevertheless, the Validation Accuracy remained consistently high, fluctuating between 94% and 99%, indicating that the model maintained strong generalization capabilities despite the observed variations in loss values.

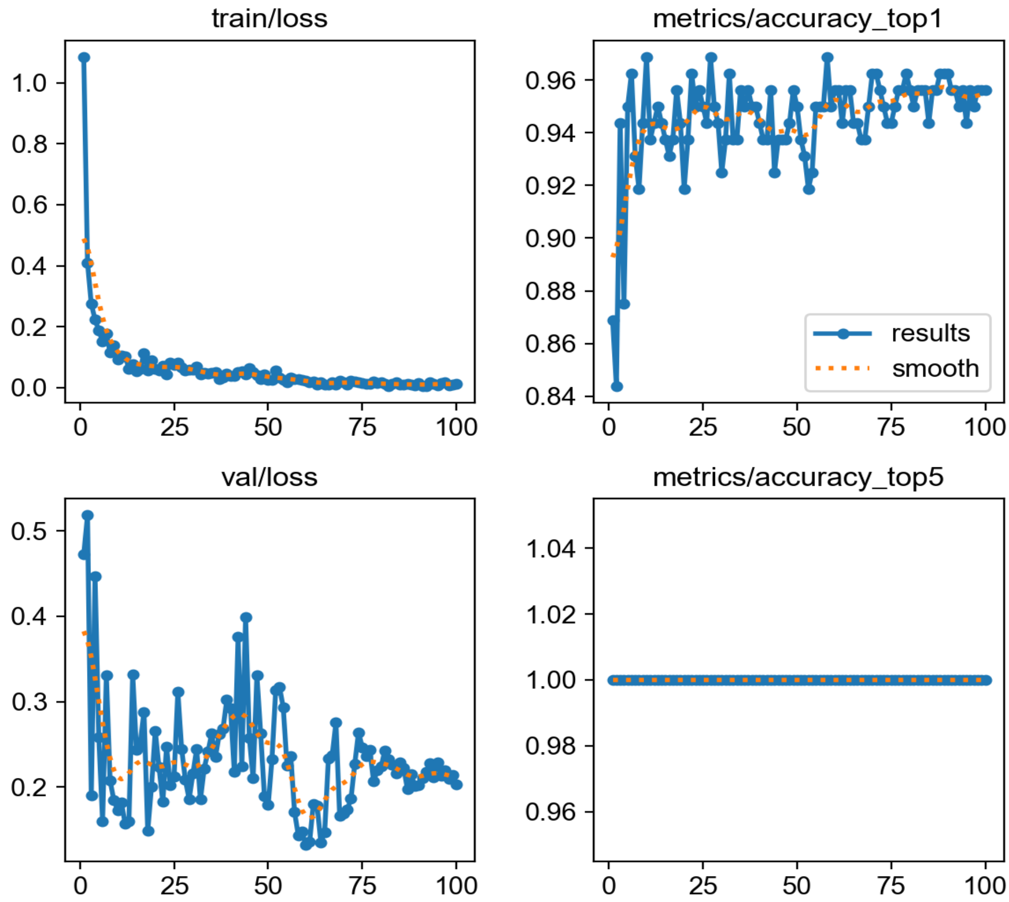

Figure 2. Training Loss and Accuracy of YOLOv8 in Fold 3

In contrast, the YOLOv8-cls architecture (Figure 2 – Results Graph) demonstrated a smooth exponential decline in train/loss during the first 25 epochs before reaching a stable state. The val/loss metric exhibited a reasonable oscillatory pattern and converged below 0.3 by the end of training. The accuracy_top1 metric consistently increased rapidly during the early training stages and stabilized at a highly optimal level (above 0.94), highlighting the strong adaptation capability of YOLOv8's built-in transfer learning features.

Classification Error Analysis (Confusion Matrix)

Classification performance was further examined using the Confusion Matrix from Fold 3 to identify weaknesses in distinguishing among disease classes, particularly in visually ambiguous cases.

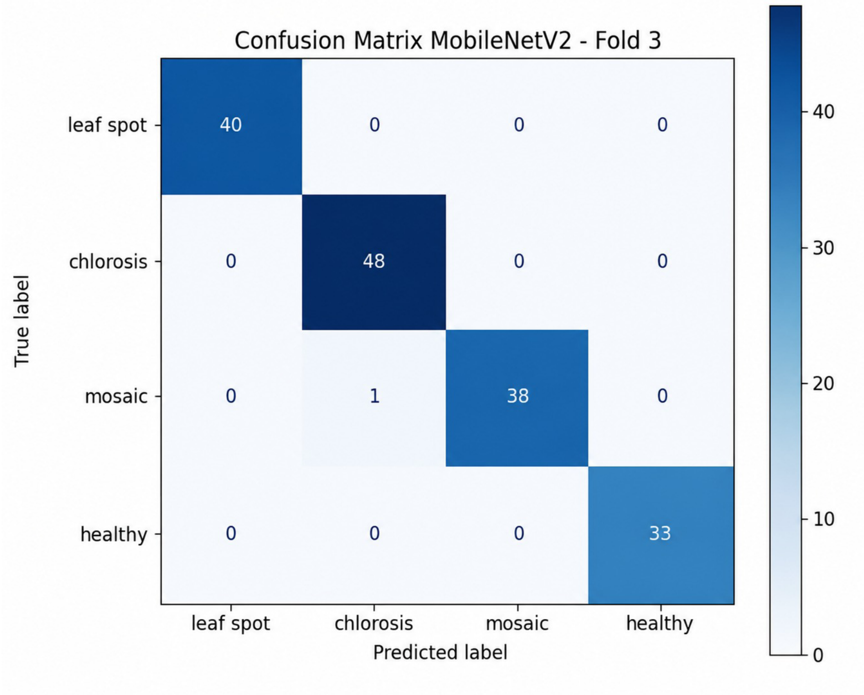

Figure 3. Confusion Matrix of MobileNetV2 in Fold 3

Based on Figure 3 (MobileNetV2 Confusion Matrix – Fold 3), the model demonstrated highly precise performance with minimal classification errors. Out of all test images in the fold, only one misclassification occurred, where a Mosaic Virus image was incorrectly classified as Chlorosis.

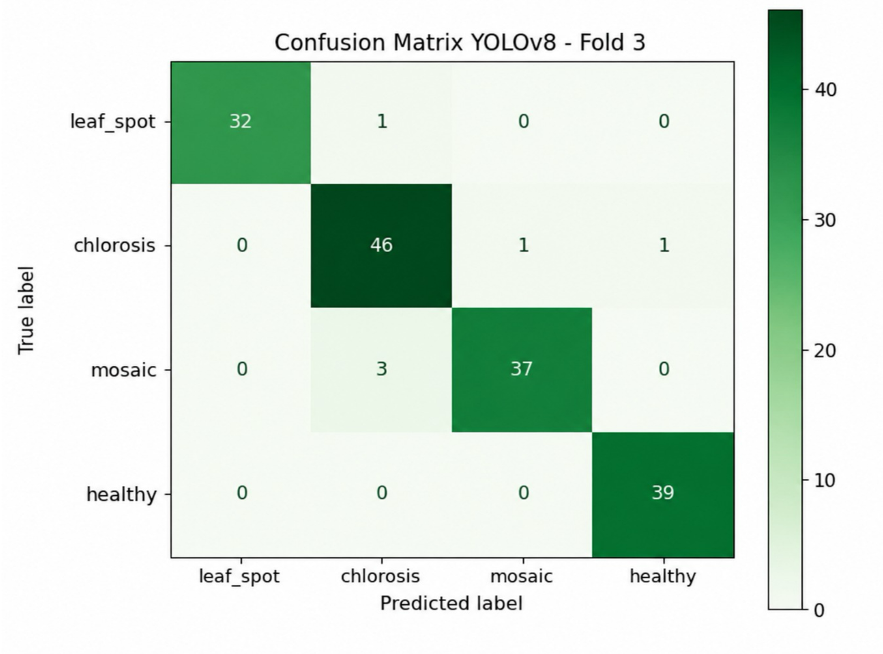

Figure 4. Confusion Matrix of YOLOv8 in Fold 3

Meanwhile, Figure 4 (YOLOv8 Confusion Matrix – Fold 3) revealed a slightly broader distribution of classification errors. YOLOv8 incorrectly classified three Mosaic Virus images as Chlorosis, one Leaf Spot image as Chlorosis, one Chlorosis image as Mosaic Virus, and one Chlorosis image as Healthy.

The analytical findings from both confusion matrices reveal a consistent pattern: classification errors most frequently occurred between the Mosaic Virus and Chlorosis classes. This observation is agronomically reasonable because both diseases manifest symptoms associated with chlorophyll degradation, resulting in yellowish discoloration or mosaic-like leaf patterns. The similarity of these macroscopic visual features presents a fine-grained classification challenge for feature extraction algorithms regardless of the underlying architecture. This inter-class confusion pattern aligns with findings reported in previous plant disease classification studies; for example, Nurokhman et al. (2024) observed analogous misclassification between symptomatically similar disease categories in chili leaf datasets, and Lu et al. (2023) similarly reported that chlorosis-adjacent disease classes posed the greatest classification challenge in MobileNetV2-based plant disease systems. The visual overlap between Mosaic Virus (characterized by irregular yellow-green mottling with vein banding) and Chlorosis (characterized by uniform interveinal yellowing) is a well-recognized agronomic challenge that necessitates either fine-grained feature augmentation or hyperspectral imaging for definitive differentiation.

Discussion

Comparative Analysis of Overall Performance and Efficiency

Based on the complete results obtained through 5-Fold Cross Validation, both architectures demonstrated excellent potential for real-world implementation. A detailed comparison of evaluation metrics and computational efficiency for each model is presented in Tables 4 and 5.

Table 4
MobileNetV2 Performance Results

Experiment	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Size (MB)	Time (min)
Fold 1	98.75%	98.88%	98.77%	98.81%	8.74	35.54
Fold 2	98.12%	98.22%	98.03%	98.12%	8.74	32.56
Fold 3	99.38%	99.49%	99.36%	99.42%	8.74	32.50
Fold 4	98.12%	98.10%	97.91%	97.99%	8.74	32.40
Fold 5	96.88%	97.10%	96.74%	96.87%	8.74	32.95
Avarage	98.25%	98.36%	98.16%	98.24%	8.74	33.19

Table 5
YOLOv8-cls Performance Results

Experiment	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Size (MB)	Time (min)
Fold 1	99.38%	99.43%	99.32%	99.37%	2.84	55.65
Fold 2	98.75%	98.81%	98.71%	98.75%	2.84	55.82
Fold 3	96.25%	96.72%	96.33%	96.49%	2.84	55.38
Fold 4	100.00%	100.00%	100.00%	100.00%	2.84	55.63
Fold 5	97.50%	97.78%	97.37%	97.45%	2.84	55.40
Avarage	98.38%	98.55%	98.35%	98.41%	2.84	55.58

Based on Tables 4 and 5, a clear trade-off exists among classification performance, computational efficiency, and storage requirements. YOLOv8-cls achieved a marginal advantage in the evaluation metrics, recording an average accuracy of 98.38%, compared with 98.25% for MobileNetV2. However, it should be noted that the absolute difference in accuracy of 0.13 percentage points is very small. Although this result consistently favors YOLOv8-cls across all five folds, no formal statistical significance test (e.g., a paired t-test or McNemar's test) was conducted; therefore, the practical significance of this marginal gap should be interpreted cautiously. The most significant advantage of YOLOv8-cls lies in its exceptionally compact model size of only 2.84 MB, making it highly suitable for deployment on hardware platforms with limited storage capacity.

On the other hand, MobileNetV2 demonstrated superior computational efficiency during training. The average time required to complete 100 epochs was only 33.19 minutes, substantially faster than YOLOv8-cls, which required 55.58 minutes on average, approximately 40% longer. This advantage makes MobileNetV2 a more time-efficient architecture and may also reduce energy consumption during training, particularly in scenarios where periodic retraining with newly collected datasets is required.

Overall, the findings suggest that YOLOv8-cls is preferable when maximizing classification performance and minimizing storage requirements are the primary objectives, whereas MobileNetV2 is a more practical choice when computational efficiency and rapid retraining capabilities are prioritized. These results are contextually aligned with, and extend beyond, previous comparative studies in the literature. Dewi et al. (2024) achieved an mAP50 of 97.9% with YOLOv8 for rice leaf disease, whereas this study demonstrates that YOLOv8-cls achieves an even higher classification accuracy of 98.38% specifically for cayenne pepper using field-collected data. Furthermore, Lu et al. (2023) reported that MobileNetV2 achieved 99.53% accuracy in crop disease classification using a controlled public dataset, while Tamayasa & Dewi (2026) reported an accuracy of 96.50% for corn leaf diseases. The 98.25% accuracy achieved in this study using a real-field dataset with greater visual variability suggests that MobileNetV2 maintains robust performance even under challenging open-field conditions. Regarding inference efficiency on mobile devices, although direct on-device benchmarking was not conducted in this study, the 2.84 MB model size of YOLOv8-cls is substantially smaller than MobileNetV2's 8.74 MB, suggesting a clear advantage for deployment in memory-constrained environments such as Android smartphones and Raspberry Pi systems. Future work should validate this advantage through empirical measurements of on-device inference time, including frames per second (FPS) and latency, to provide stronger evidence for deployment claims.

CONCLUSION

This study compared the performance of YOLOv8-classification (YOLOv8-cls) and MobileNetV2 for bird's eye chili leaf disease classification using a primary dataset collected under real-field conditions and evaluated through Stratified 5-Fold Cross-Validation. Both models demonstrated excellent classification performance, achieving average accuracies above 98%, confirming their effectiveness in identifying healthy leaves, Mosaic Virus, Leaf Spot, and Chlorosis. Among the two architectures, YOLOv8-cls achieved the highest overall performance, with an accuracy of 98.38%, precision of 98.55%, recall of 98.35%, and F1-score of 98.41%, while requiring only 2.84 MB of storage. These results indicate that YOLOv8-cls offers an optimal balance between predictive performance and model compactness, making it particularly suitable for deployment on mobile and edge devices. In contrast, MobileNetV2 required approximately 40% less training time, making it a practical alternative for applications requiring frequent model retraining under limited computational resources.

Although both models achieved high performance, confusion matrix analysis revealed that most classification errors occurred between the Mosaic Virus and Chlorosis classes because of their similar visual characteristics. This limitation suggests that fine-grained disease discrimination remains challenging under natural field conditions. Future research should expand the dataset by including additional disease categories and more diverse environmental conditions; investigate advanced approaches, such as attention-based models or hyperspectral imaging, to improve the discrimination of visually similar diseases; and evaluate model quantization and on-device inference performance to optimize real-world deployment. These improvements will strengthen the development of accurate, efficient, and accessible intelligent disease detection systems for digital agriculture.

ACKNOWLEDGEMENT

The authors would like to thank the cayenne pepper farmers in Bali who permitted access to their agricultural fields for primary data collection. The authors also express gratitude to the research institution and laboratory facilities that supported the computational experiments conducted in this study.

AUTHOR CONTRIBUTION STATEMENT

All authors contributed equally to this study. Author 1: conceptualization, methodology, data collection, and manuscript writing. Author 2: model training, experimental implementation, data analysis, and manuscript revision. Both authors have read and approved the final version of the manuscript.

REFERENCES

Asmara, I. G. N. B. P., Kesiman, M. W. A., & Indrawan, G. (2023). Balinese Shadow Puppet Characters Detection In The Wayang Peteng Performance Using The Yolov5 Algorithm. *Jurnal Nasional Pendidikan Teknik Informatika: Janapati*, *12*(3), 388–397.

Bitar, A., Rosales, R., & Paulitsch, M. (2023). Gradient-Based Feature-Attribution Explainability Methods For Spiking Neural Networks. *Frontiers In Neuroscience*, *17*. Https://Doi.Org/10.3389/Fnins.2023.1153999

Dewi, C., Bilaut, F. Y., Christanto, H. J., & Dai, G. (2024). *Deep Learning For The Classification Of Rice Leaf Diseases Using Yolov8*.

Hameed, S., Al-Shammari, E. T., & Ramzan, N. (2026). Evaluation of deep learning models for plant disease classification using MCC and balanced accuracy metrics. Results in Engineering, 26, 111212. https://doi.org/10.1016/j.rineng.2026.111212

I Made Angga Darma Putra, Maysanjaya, I. Md. D., & Kesiman, M. W. A. (2023). Pendekatan Berbasis U-Net Untuk Segmentasi Hard Exudate Dalam Citra Fundus Retina. *Insert : Information System And Emerging Technology Journal*, *4*(1), 26–36. Https://Doi.Org/10.23887/Insert.V4i1.59034

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, *521*(7553), 436–444.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied To Document Recognition. *Proceedings Of The Ieee*, *86*(11), 2278–2324.

Lu, J., Liu, X., Ma, X., Tong, J., & Peng, J. (2023). Improved Mobilenetv2 Crop Disease Identification Model For Intelligent Agriculture. *Peerj Computer Science*, *9*, E1595.

Manurung, D. G., Pinasthika, M. R., Vasya, M. A. O., Putri, R. A. D. S., Tampubolon, A. P., Prayata, R. F., Nisa, S. K., & Yudistira, N. (2024). Deteksi Dan Klasifikasi Hama Potato Beetle Pada Tanaman Kentang Menggunakan Yolov8. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(4), 723–734.

Natih, I. D. G. A. W., Kesiman, M. W. A., & Sunarya, I. M. G. (2026). Analisis Perbandingan Arsitektur Dan Optimizer Yolov11 Untuk Estimasi Buah Kelapa. *Riggs: Journal Of Artificial Intelligence And Digital Business*, *4*(4), 12–19.

Njoroge, T., Kibuku, R., & Mugoye, K. (2025). Comparative And Edge-Hybrid Modeling Of Efficientnetv2 And Mobilenetv2 For Multi-Classcrop Disease Classification With Statistical Validation. *Journal Of Edge Computing*, *4*(2), 234–262.

Nugraha, G. S., Wijaya, I. G. P. S., Bimantoro, F., Husodo, A. Y., & Hamami, F. (2023). Arabic Character Recognition Using CNN Lenet-5. *Joiv: International Journal On Informatics Visualization*, *7*(4), 2183–2188.

Nurokhman, A., Surorejo, S., Kurniawan, R. D., & Gunawan, G. (2024). Application Of Computer Vision Techniques To Detect Diseases And Pests Of Chili Plants. *Journal Of Intelligent Decision Support System (IDSS)*, *7*(1), 10–18.

Obafemi-Ajayi, T., Eichler, R., Walsh, B., Knisley, D., & Knisley, J. (2025). Multi-class plant disease classification using Matthews Correlation Coefficient as primary evaluation metric for imbalanced agricultural datasets. Engineering Applications of Artificial Intelligence, 139, 113347. https://doi.org/10.1016/j.engappai.2025.113347

Prasetia, I. P. W., & Sunarya, I. M. G. (2024). Image Classification Of Balinese Seasoning Base Genep Based On Deep Learning. *Jurnal Nasional Pendidikan Teknik Informatika: Janapati*, *13*(1), 79–90.

Pusparani, D. A., Kesiman, M. W. A., & Aryanto, K. Y. E. (2024). Identification Of Little Tuna Species Using Convolutional Neural Networks (Cnn) Method And Resnet-50 Architecture. *Indonesian Journal Of Artificial Intelligence And Data Mining*, *8*(1), 86–93.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted Residuals And Linear Bottlenecks. *Proceedings Of The Ieee Conference On Computer Vision And Pattern Recognition*, 4510–4520.

Saputra, I. P. G. D., Kesiman, M. W. A., & Sunarya, I. M. G. (2026). Deteksi Pemalsuan Qris Mpm Statis Menggunakan Yolo, Paddleocr Dan Metode Berbasis Aturan. *Bulletin Of Computer Science Research*, *6*(2), 763–775.

Solichin, A., Pebrianti, D., & Riyanto, S. (2025). Modified Mobilenet-V2 Convolution Neural Network (CNN) For Character Identification Of Surakarta Shadow Puppets. In *Ijacsa) International Journal Of Advanced Computer Science And Applications* (Vol. 16, Number 5).

Www.Ijacsa.Thesai.Org
Somvanshi, S., Islam, M. M., Chhetri, G., Chakraborty, R., Mimi, M. S., Shuvo, S. A., Islam, K. S., Javed, S., Rafat, S. A., Dutta, A., & Das, S. (2026). From Tiny Machine Learning To Tiny Deep Learning: A Survey. *Acm Computing Surveys*, *58*(7), 1–33. Https://Doi.Org/10.1145/3776588
Sumantara, I. G. L. T., Kesiman, M. W. A., & Sunarya, I. M. G. (2024). Comparative Analysis Of Cnn Methods For Periapical Radiograph Classification. *Jurnal Nasional Pendidikan Teknik Informatika: Janapati*, *13*(2), 204–214.
Suputra, I. P. A., Gunadi, I. G. A., & Sunarya, I. M. G. (2025). Hyperparameter Optimization With Mobilenet Architecture And Vgg Architecture For Urban Traffic Density Classification Using Bali Camera Image Data. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *9*(3), 1132–1145.
Tamayasa, K. A., & Dewi, L. J. E. (2026). Analisis Perbandingan Model Arsitektur Mobilenetv2 Dan Efficientnetb3 Dalam Klasifikasi Penyakit Daun Jagung. *Jurnal Informatika Dan Teknik Elektro Terapan*, *14*(1). Https://Doi.Org/10.23960/Jitet.V14i1.8624
Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A Comprehensive Review Of Yolo Architectures In Computer Vision: From Yolov1 To Yolov8 And Yolo-Nas. *Machine Learning And Knowledge Extraction*, *5*(4), 1680–1716.
Widaningrum, I., Astuti, I. P., Mustikasari, D., Nurfitri, K., Az-Zahra, R. R., Vidyastari, R. I., Selamat, A., & Ponorogo, U. M. (2025). The Lima Pandawa Shadow Puppet Characters Utilizing Principal Component Analysis (PCA) For Feature Extraction And K-Nearest Neighbor (Knn) For Classification) Recognition Of The Lima Pandawa Shadow Puppet Characters Utilizing Principal Component Analysis (PCA) For Feature Extraction And K-Nearest Neighbor (Knn) For Classification. In *Indonesian Journal Of Information Systems (IJIS)* (Vol. 8, Number 1).
Wita, D. S., & Subekti, A. (2023). Mobilenet-Based Transfer Learning For Detection Of Eucalyptus Pellita Diseases. *Jurnal Nasional Pendidikan Teknik Informatika: Janapati*, *12*(1), 1–7.
Yudhantara, I. K. P., Kertiasih, N. K., & Wahyu Wijaya, I. N. S. (2026). Perbandingan Model Klasifikasi Penyakit Daun Bakau Menggunakan Arsitektur Vgg16 Dan Mobilnetv2. *Jurnal Informatika Dan Teknik Elektro Terapan*, *14*(1). Https://Doi.Org/10.23960/Jitet.V14i1.8816
Yu, H., Hu, X., Guo, J., Bi, C., Xue, M., & Chen, H. (2026). An Enhanced Multi-Scale Lw-Mobilenetv2 For Traditional Chinese Medicine Decoction Pieces Recognition. *Biomedical Signal Processing And Control*, *113*, 109200. Https://Doi.Org/10.1016/J.Bspc.2025.109200