

# PERFORMANCE ANALYSIS OF CNN AND TRANSFER LEARNING ARCHITECTURES FOR BINARY CLASSIFICATION OF BANANA LEAF DISEASES

**Sajid A. Shaikh**

PG & Research Department of Electronics,  
DDSGP College, Chopda Dist. Jalgaon(MH)  
India 425107

**Kunal D. Gaikwad**

PG & Research Department of Electronics,  
DDSGP College, Chopda Dist. Jalgaon(MH)  
India 425107

Crossref DOI - <https://doi.org/10.63665/rh.v7i2.50>

## **Abstract :**

*Timely and accurate recognition of banana leaf diseases plays a vital role in improving agricultural yield and ensuring sustainable crop management. This study presents a comparative evaluation of a basic Convolutional Neural Network (CNN) and three transfer learning models VGG16, ResNet50, and MobileNetV2 for binary classification of healthy and diseased banana leaves. A publicly available multi-class banana leaf dataset was restructured into a binary dataset comprising 360 training images and 90 testing images to assess model performance under limited data conditions.*

*Each model was analyzed in two configurations: baseline (without optimization) and optimized (with data augmentation and fine-tuning). Model performance was evaluated using accuracy, precision, recall, F1-score, and learning curve analysis. Experimental results revealed that optimized models consistently outperformed their non-optimized counterparts. Among all evaluated architectures, the optimized VGG16 model achieved the highest performance, with a testing accuracy of 91.11%, precision of 98.15%, recall of 88.33%, and an F1-score of 92.98%. In contrast, deeper architectures such as ResNet50 exhibited reduced effectiveness due to the limited dataset size.*

*Overall, the findings suggest that moderately deep transfer learning models, particularly VGG16, are more suitable for small-scale agricultural image datasets. This study provides valuable insights for model selection in banana leaf disease detection and establishes a reliable baseline for future multi-class classification and custom dataset development.*

**Keywords :** Banana leaf disease detection, Convolutional Neural Networks, Transfer learning, Binary classification, VGG16, MobileNetV2, ResNet50, Precision agriculture

## **Introduction :**

Banana (*Musa spp.*) is among the most extensively grown fruit crops in tropical and subtropical regions, contributing significantly to global food security and agricultural economies [1,4,12,14]. However, banana production is highly vulnerable to various diseases,



including Black Sigatoka, Yellow Sigatoka, Panama disease, Moko disease, and Fusarium wilt, which lead to substantial yield losses and deterioration in fruit quality [4,5,10]. Therefore, timely and accurate disease identification is crucial to reducing economic losses and supporting sustainable farming practices [1,8].

Conventional approaches to banana disease diagnosis primarily depend on manual visual inspection by experienced agricultural professionals. Such methods are labor-intensive, subjective, and difficult to scale for large agricultural areas [9,10]. In recent years, advances in image processing and artificial intelligence—particularly deep learning-based Convolutional Neural Networks (CNNs)—have enabled automated and reliable plant disease detection [1–3]. These models are capable of learning complex visual patterns from leaf images and have demonstrated superior performance compared to traditional machine learning techniques [3,9].

Several studies have explored CNN-based, hybrid, transfer learning, segmentation-based, mobile-based, and UAV-assisted frameworks for banana disease detection and crop monitoring [4–8,11–14]. Transfer learning architectures such as VGG16, ResNet50, and MobileNetV2 have been widely adopted due to their ability to leverage pre-trained feature representations, especially when labeled agricultural datasets are limited [1,4,11,14]. Nevertheless, issues related to dataset size, class imbalance, computational complexity, and real-world generalization continue to pose significant research challenges [2,8,9].

Motivated by these challenges, this work examines the performance of deep learning models under limited data conditions by converting an existing publicly available multi-class banana leaf dataset into a binary classification problem consisting of healthy and diseased leaf categories. The dataset was divided into 360 training images and 90 testing images to ensure a consistent and fair evaluation framework. Four deep learning architectures—Basic CNN, VGG16, ResNet50, and MobileNetV2—were systematically analyzed using two training strategies: models trained without optimization and models trained with optimization techniques such as data augmentation and fine-tuning.

Model performance was evaluated using multiple criteria, including accuracy, precision, recall, F1-score, and learning behavior analysis through training and validation accuracy–loss curves, following established evaluation protocols reported in previous studies [8,11,13]. The experimental results indicate that optimization strategies substantially improve classification performance for moderately deep models. In particular, the optimized VGG16 model achieved the highest accuracy and F1-score, whereas deeper architectures like ResNet50 exhibited reduced performance due to overfitting when trained on limited data. These findings emphasize the importance of selecting suitable model complexity and optimization approaches for effective banana leaf disease detection in real-world, data-constrained agricultural applications.

## Literature Survey :

BananaSqueezeNet, a lightweight CNN architecture intended for the diagnosis of



three significant banana leaf diseases, was presented by Bhuiyan et al. [1]. With substantially fewer parameters, the model obtained high classification accuracy, making it appropriate for real-time and resource-constrained agricultural settings. The resilience of CNN models in different lighting circumstances was investigated by Chuquimarca et al. [2]. Their research showed that deep learning models perform consistently in real-world situations, which is essential for deployment at the field level. For the purpose of identifying banana breeds, Rahman et al. [3] compared deep learning and standard machine learning methods. By successfully learning texture, color, and shape information, customized CNN architectures greatly outperformed conventional classifiers. A CNN mobile application was created by Elinisa and Mduma [4] to detect banana illnesses like Fusarium Wilt and Black Sigatoka early on. In real-time diagnosis, the system attained over 90% confidence and offered practical mitigation suggestions.

A hybrid CNN–FSVM framework with 99% accuracy was proposed by Narayanan et al. [5]. The efficiency of hybrid models was demonstrated by the improvement in precision, recall, and F1-score that resulted from the combination of deep feature extraction with classical classifiers. Using Faster R-CNN with high-resolution RGB data, Neupane et al. [6] investigated UAV-based banana plant detection and counting. Plantation-scale monitoring was made possible by their multi-altitude data fusion method, which yielded detection accuracy of about 99%. A U-Net-based semantic segmentation model for early banana disease diagnosis was presented by Elinisa et al. [7]. Diseased areas on leaves and stalks may be precisely located thanks to the model's high Dice and IoU scores. Using supplemented datasets, Jiménez et al. [8] assessed EfficientNetB0, ResNet50, and VGG19 for the identification of banana leaf disease. The optimal balance between accuracy and processing efficiency was provided by EfficientNetB0.

Thiagarajan et al. [9] looked into the use of feature extraction techniques including SIFT, HOG, and LBP in conjunction with traditional machine learning techniques. Their study suggested ANN-based hybrid alternatives to CNNs and emphasized their shortcomings with regard to scale and rotation invariance. With an accuracy of 94.89%, Rohini and Raghavendra [10] suggested a CNN-based system for identifying banana leaf disease that incorporates image preprocessing techniques. In their comparison of the ResNet50 and VGG19 architectures, Syihad et al. [11] found that the former performed better, achieving 94% accuracy along with higher precision, recall, and F1-score. A review and hybrid framework combining CNNs with region-based segmentation and thresholding techniques was presented by Jadhaw and Bhandari [12], with a focus on models like ResNet-50, VGG-16, and AlexNet. A real-time comparison of InceptionV3, ResNet50V2, VGG16, and MobileNet was carried out by Rivankar et al. [13]. While MobileNet demonstrated superior real-time performance, ResNet50V2 attained 99.52% accuracy. Using deep CNNs with transfer learning, Selvaraj et al. [14] created an AI-powered banana disease and pest detection system. Models based on ResNet50 and Inception showed potential for mobile deployment and obtained accuracies above 90%. Even though deep learning and hybrid approaches have made great strides in the detection of banana disease, issues with model generalization on small and real-world datasets, computational efficiency for edge deployment, thorough



performance evaluation beyond accuracy, and robustness under various field conditions are still not adequately addressed.

Table.1: Comparative Analysis of Banana Leaf Disease Detection Techniques

Ref.	Author(s)	Method / Model Used	Application Focus	Key Findings / Performance
[1]	Elinisa & Mduma	Mobile-based CNN system	Real-time banana disease diagnosis	>90% confidence; realistic mobile implementation
[2]	Narayanan et al.	CNN + FSVM (Hybrid)	Disease classification	99% accuracy with improved precision, recall, and F1-score
[3]	Neupane et al.	Faster R-CNN (UAV-based)	Plant detection and counting	Using multi-altitude RGB pictures, detection accuracy is approximately 99%.
[4]	Rohini & Raghavendra	CNN with preprocessing	Banana leaf disease identification	Achieved 94.89% accuracy
[5]	Syihad et al.	ResNet50 vs VGG19	Leaf disease classification	ResNet50 superior with 94% accuracy
[6]	Rivankar et al.	InceptionV3, ResNet50V2, MobileNet	Real-time detection	ResNet50V2: 99.52% accuracy; MobileNet best speed
[7]	Selvaraj et al.	Deep CNN + Transfer Learning	Disease & pest detection	>90% accuracy; mobile-friendly models

## Materials and Methods :

### 1. Dataset Description :

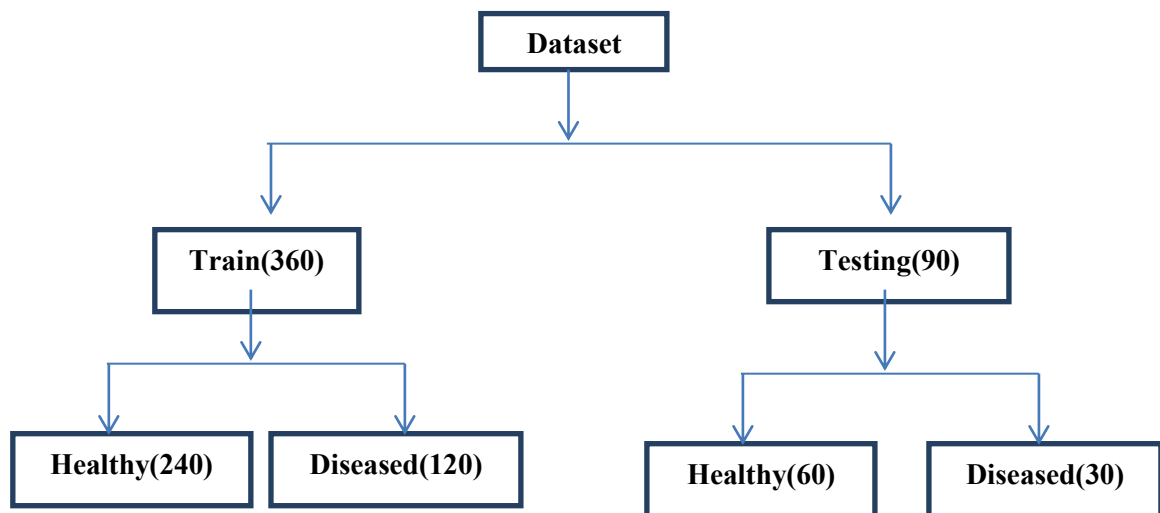
- Source of the publicly available banana leaf dataset:  
<https://data.mendeley.com/datasets/7vpdrbdkd4/2>
- Original multi-class nature of the dataset:
  - Version-2- Augmented Images of Banana leaves deficient in Nutrients
  - Version-2-RAW Images of Banana leaves deficient in Nutrients
    - Boron
    - Calcium
    - Healthy
    - Iron
    - Magnesium
    - Manganese
    - Potassium
    - Sulphur
    - Zinc
- Restructuring into binary classes:( Version-2-RAW)



**Healthy** : Pictures were chosen from the Healthy folder.

**Diseased** : Based on obvious symptom criteria, images were gathered from several folders that corresponded to different disease and nutritional deficit categories, such as manganese and potassium shortages.

- Dataset split:
  - 360 images for training:
  - 90 images for testing/validation



*Fig. 1: Dataset Folder tree diagram*

Representative samples from the dataset utilized in this investigation, divided into Category I: Healthy and Category II: Diseased banana leaf photos, are shown in Figure 2. The diseased class includes leaves with various pathological or nutrient-related disorders, such as discolouration, spots, and uneven surface textures, whereas the healthy class includes leaf samples with normal color, texture, and structural patterns without obvious symptoms. These sample photos support the problem's framing as a binary classification job by highlighting the visual distinctions between the two groups.

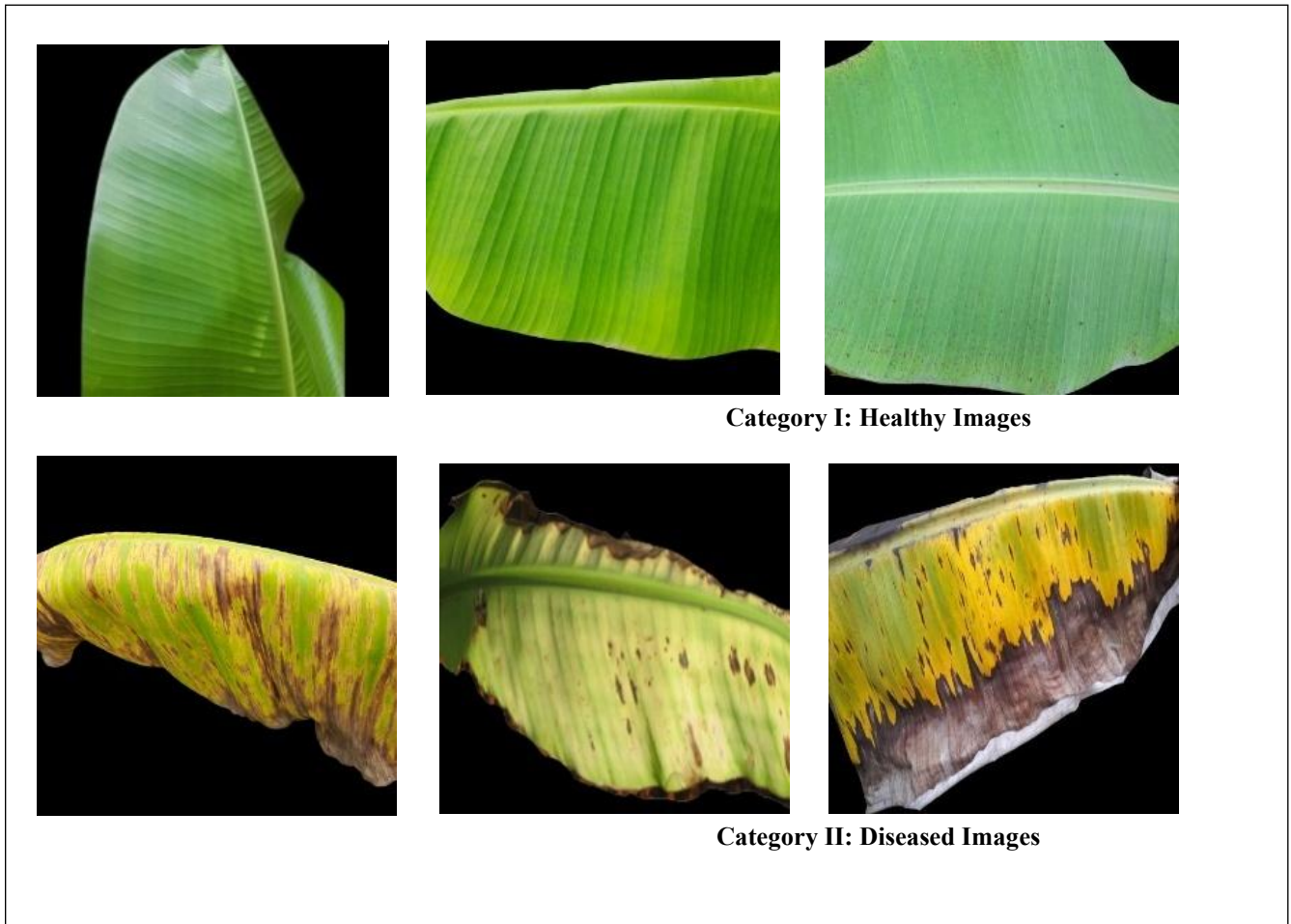


Fig.2: Dataset Samples

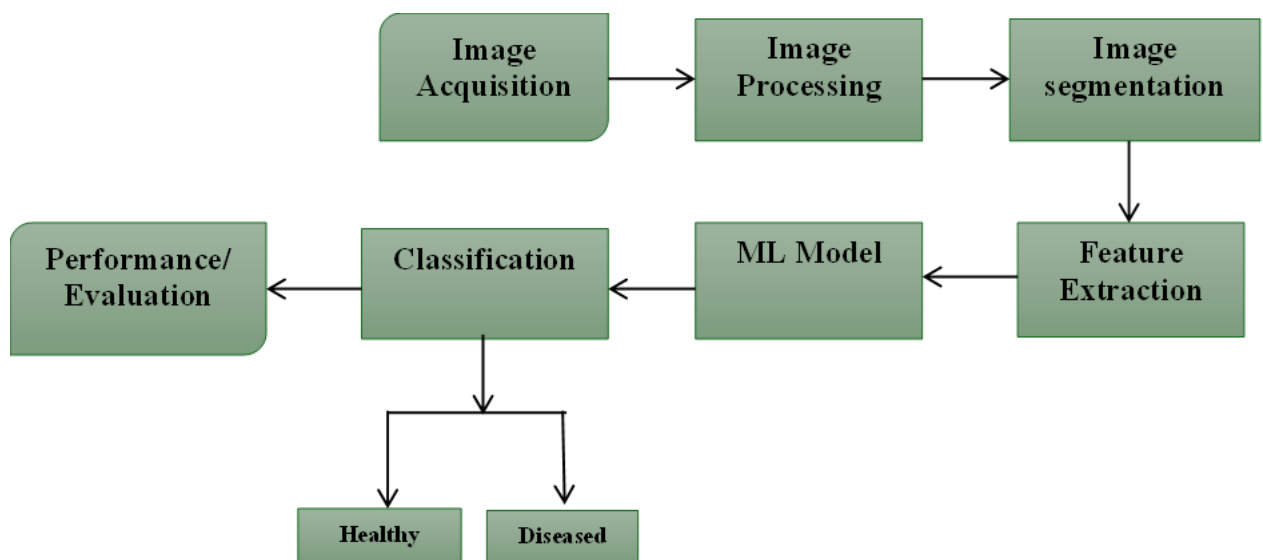


Fig. 3: Machine Learning Architecture for image Classification [22]

**Methodology :**

1. **Image Acquisition** : Devices like digital cameras, scanners, and cell phone cameras



are used to take pictures. With the use of these resources, a collection of leaf photos of both diseased and healthy plants is created.

2. **Image Pre-processing** : At this point, a number of methods are used to improve the overall quality of the photos by removing noise and undesired components.
3. **Image Segmentation** : This stage attempts to improve the image's interpretability by dividing significant areas, which facilitates and improves analysis. Since segmentation serves as the foundation for feature extraction, it is an essential stage in image processing. It is possible to use a variety of techniques, including the Random Forest algorithm, Otsu's thresholding, and thresholding techniques. The Random Forest algorithm, which offers better classification accuracy than many other learning techniques, is used in this work to segment and classify images.
4. **Feature Extraction** : Relevant color, shape, and texture features are extracted from the sick areas of plant leaves using methods like the Grey Level Co-occurrence Matrix (GLCM), computer vision techniques, and artificial intelligence approaches.
5. **Classification and Testing**: Using machine learning models like Support Vector Machines (SVM), Random Forests, K-Nearest Neighbors (KNN), and other popular classifiers, the retrieved features are examined and categorized. Either a reserved set of characteristics is used to test the model's performance, or new photos linked to the disease are applied and processed through all of the aforementioned steps to see if the classifier can correctly identify the disease.[22].

## 2. Data Preprocessing :

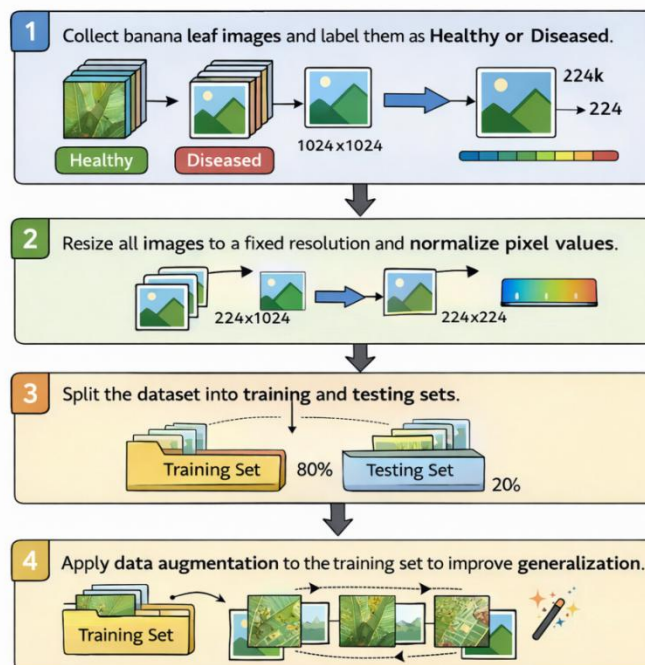


Fig. 4: Preprocessing flow chart

- 1) Image Resizing & Normalization : Standardizes leaf image size and pixel values to ensure stable and efficient model training.
- 2) Label Encoding (Binary): Assigns labels to classify plant leaves as healthy or

diseased.

3) Data Augmentation: Image augmentation is an integral component of the data preprocessing stage. The fig.4 shows Pre-processing techniques are applied to refine the dataset by eliminating irrelevant or redundant information, such as complex background elements, which may otherwise degrade the performance of convolutional neural network (CNN) models[23].

- **Rotation** : Helps the model recognize plant diseases at different leaf orientations.
- **Flipping** : Enables learning of disease patterns irrespective of left–right or up–down leaf positions.
- **Zooming** : Allows the model to focus on disease-affected regions at varying scales.

### 3. Model Architectures :

- **Basic CNN** : Overview of the layer structure: Figure 5 illustrates how a sequential model was used for implementation. The model consisted of four convolutional layers. A max-pooling layer came after each convolutional layer, and then a dropout layer with a rate of 0.2 came after that. While the first and second convolutional layers had 16 and 32 filters, respectively, the third and fourth convolutional layers each had 64 filters. Three-by-three convolutional filters and the ReLu activation function were used. After being flattened by a flattened layer, the results entered a dense layer with 512 neurons and the ReLu activation function. Four neurons with a softmax activation function and the number of classes made up the output dense layer. In order to normalize the photos to a range of zero to one, they were rescaled and reduced to  $512 \times 512$  pixels [4].

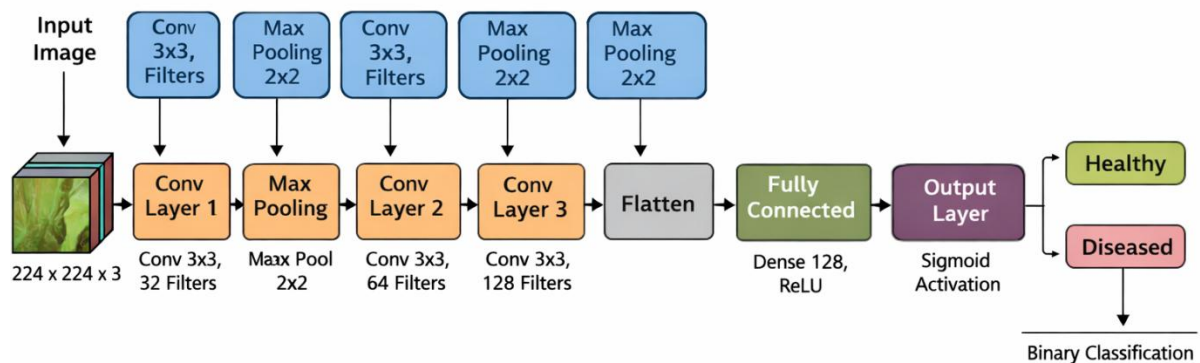


Fig.5: Architecture of CNN model

### • Transfer Learning Models :

- **ResNet50** : ResNet is a 50-layer convolutional neural network with the main advantage of using less memory than earlier versions. This is because a layer called GlobalAveragePooling was employed during the classification step rather than thick layers. This layer transforms the two-dimensional feature maps generated in the preceding feature extraction stage into a vector of n-classes, as illustrated in fig. 6, to calculate the probability of falling into each category[8].

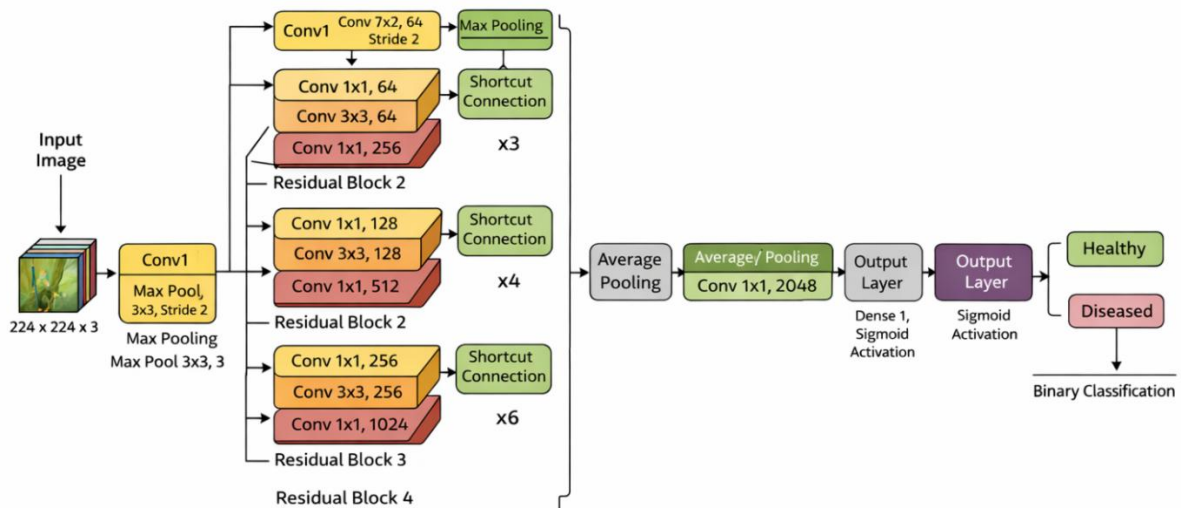


Fig.6: Architecture of ResNet 50

- VGG16** : A straightforward and well-organized deep learning model, VGG16 uses tiny  $3 \times 3$   $\times$   $33 \times 33$  convolution filters and has 16 layers. Its architecture is clearly arranged as three completely connected layers for classification come after stacked convolution and pooling layers (Fig. 7). Strong performance in picture classification tasks is facilitated by the model's consistent design, which also makes it simple to deploy and promotes efficient feature learning across layers. In contrast to more lightweight architectures, VGG16 necessitates a lot of processing and has many parameters [15].

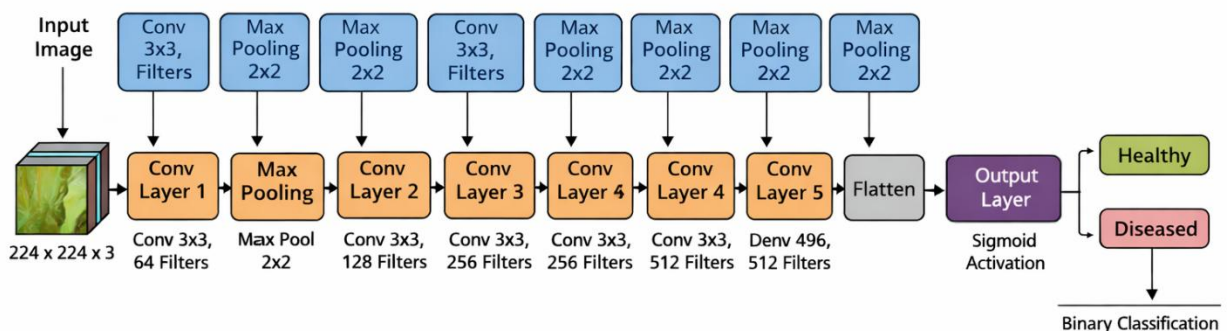


Fig.7: VGG16 Architecture

- MobileNetV2** : With an emphasis on lightweight structures, MobileNet was created especially for mobile and edge devices. It makes use of Depthwise separable convolution, which splits the convolution process into two phases: depth-wise convolution, which independently applies a filter to each input channel, and pointwise convolution, which aggregates the results. Because it reduces computation and model size, this factorization is ideal for embedded and mobile systems. Because it balances model size and accuracy, MobileNet's fig.8 is perfect for resource-constrained scenarios [15].

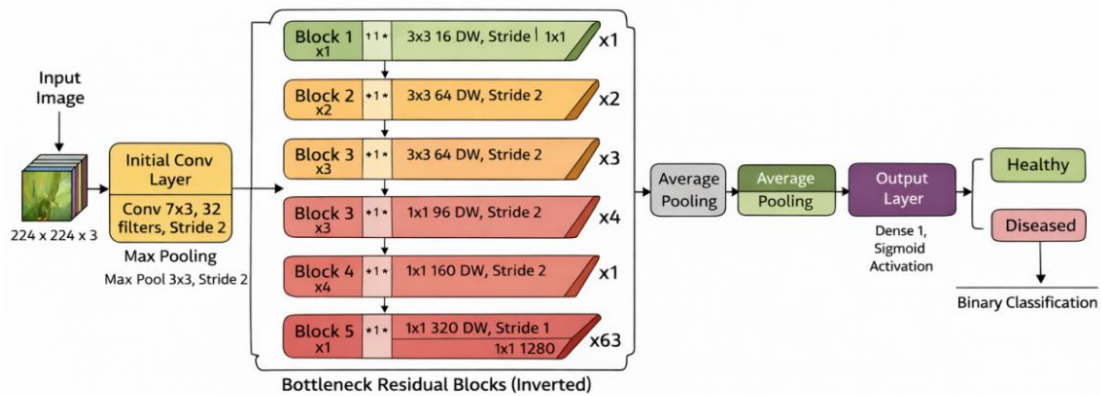


Fig.8: MobileNetV2 Architecture

- **Frozen layers** : The layers of a deep learning model that have already been trained and whose weights remain constant throughout training are known as frozen layers. These layers preserve learned low-level and mid-level characteristics from sizable benchmark datasets like ImageNet by acting as fixed feature extractors. Layer freezing lowers computing costs, avoids overfitting on small datasets, and speeds up training.
- **Fine-tuned layers (optimal setup)**: A pre-trained model's chosen layers have their weights unfrozen and retrained on a target dataset. When enough data and appropriate regularization are employed, this optimization enables the network to adjust high-level features to task-specific parameters, boosting classification performance and generalization [19,20].

#### 4. Experimental Setup :

Table.2 Common Parameter of models:

Parameter	Value
Batch Size	32
Epochs	10-20
Optimizer	Adam
Loss Function	Binary Cross-Entropy

Experimental Configurations: Two experimental configurations were taken into consideration in this study in order to assess how training tactics affected model performance.

#### Un-optimized Configuration :

In this configuration, models were trained with default hyperparameters without the use of fine-tuning or regularization methods. This setup provides a baseline for comprehending each architecture's innate capacity for learning with little assistance.

## Optimized Configuration :

On the other hand, to improve generalization and lessen overfitting, the optimized configuration includes training improvements such learning rate adjustment, regularization, and enhanced convergence techniques. It is commonly known that optimization methods greatly enhance deep learning performance, especially when dealing with small datasets [17].

## 5. Performance Evaluation Metrics :

To provide a thorough and trustworthy evaluation of classification efficacy, the performance of the suggested deep learning models was assessed using a variety of quantitative indicators. Metrics beyond overall accuracy were taken into consideration to capture class-wise behavior because the challenge requires the binary classification of healthy and unhealthy banana leaves.

**Accuracy :** Accuracy provides a broad overview of the model's effectiveness by measuring the proportion of correct predictions among all test samples. Nonetheless, in situations with imbalanced classes, relying solely on accuracy can be insufficient.

**Precision :** Precision measures the proportion of correctly identified diseased samples out of all samples the model predicts as diseased. High precision indicates a low rate of false positives, which is essential for avoiding false alarms in disease detection.

**Recall :** Recall, also known as sensitivity, measures the model's effectiveness in detecting true cases of disease. For monitoring crop health, achieving high recall is vital to minimize the risk of overlooking infected leaves.

**F1-score :** The F1-score, being the harmonic mean of precision and recall, delivers a balanced assessment when both false positives and false negatives are significant. This metric is especially valuable for judging the performance of binary classification models when data is scarce.

A confusion matrix offers detailed insight into how the model predicts each class by displaying true positives, true negatives, false positives, and false negatives.

Furthermore, learning dynamics, convergence behavior, and possible overfitting or underfitting during training were examined by analyzing training and validation accuracy and loss curves. [18,19].

## Experimental Results and Analysis :

### 1. Basic CNN Results :

Both baseline (unoptimized) and optimized setups were used to assess the Basic CNN. With a test accuracy of 81.11%, precision of 0.9057, recall of 0.8000, and F1-score of 0.8496 in the baseline configuration, the model demonstrated a respectable capacity for



classification but a restricted sensitivity to diseased samples.

The model's performance greatly improved once optimization strategies like data augmentation were used. In addition to improving precision (0.9310), recall (0.9000), and F1-score (0.9153), the modified CNN attained a higher accuracy of 88.89%. These enhancements show improved recall and accuracy balancing as well as improved generalization.

Overall, the findings demonstrate that optimization methods significantly increase a simple CNN architecture's efficacy for binary banana leaf disease classification on limited datasets.

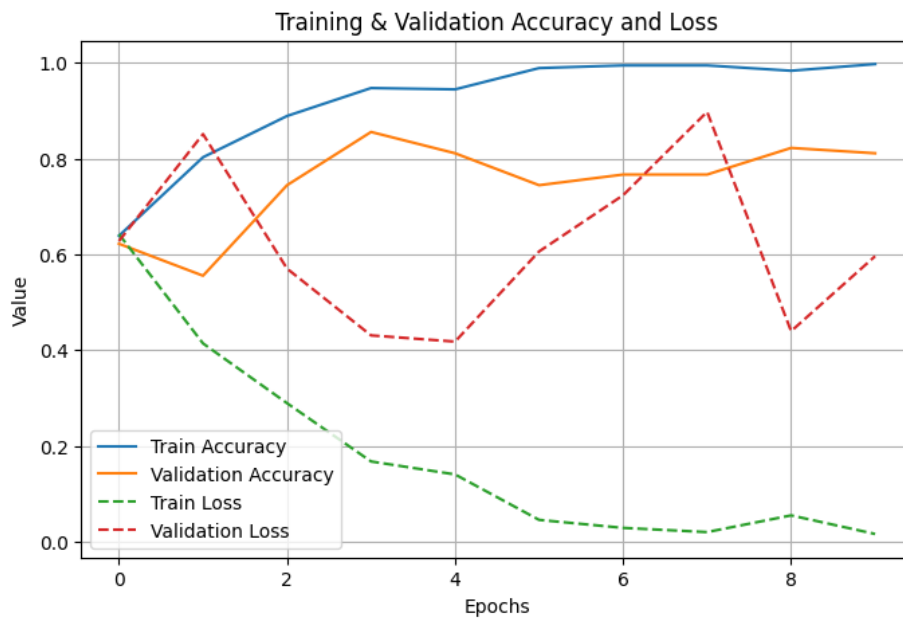


Figure 9: CNN model (Un-Optimized)

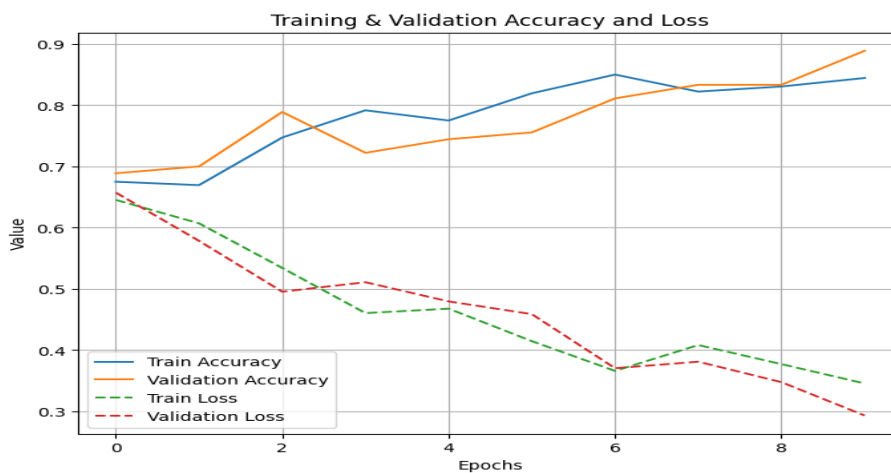


Fig. 10: CNN model Optimized



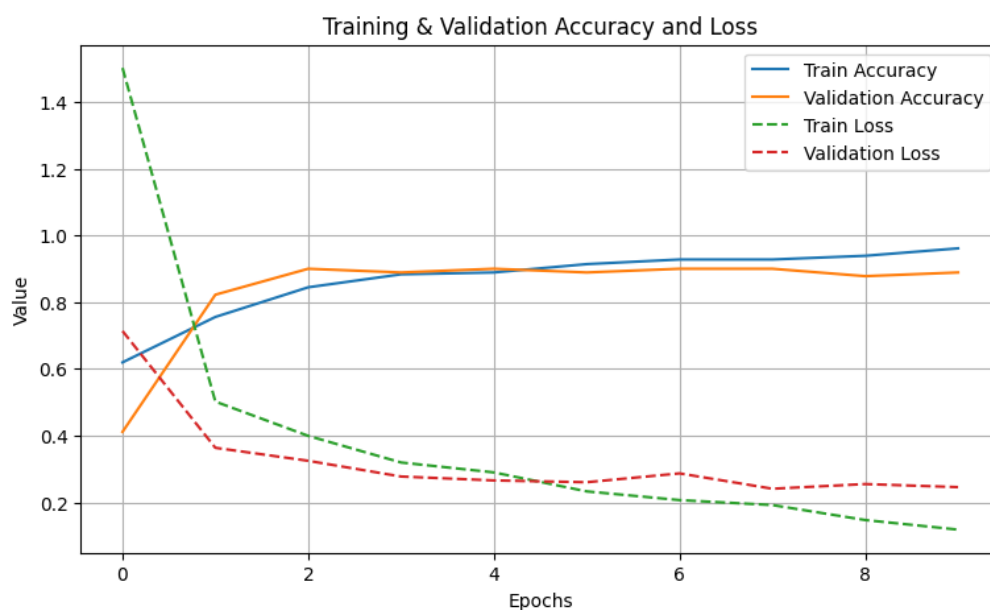
**Table 3: Performance Comparison of CNN Model**

Configuration	Test Accuracy (%)	Precision	Recall	F1-Score
CNN (Un-optimized)	81.11	0.9057	0.8000	0.8496
CNN (Optimized)	88.89	0.9310	0.9000	0.9153

**Observation :** Every statistic clearly improves with the optimized CNN. Generalization was greatly improved by data augmentation and better training techniques, demonstrating the significance of optimization for small datasets.

## 2. VGG16 Results :

With precision, recall, and F1-score values of 0.8886, 0.8889, and 0.8867, respectively, the VGG16 model attained a test accuracy of 88.89% in the un-optimized setup. These findings show that transfer learning is useful for classifying banana leaf diseases on sparse data, even in the absence of further improvement. The performance of VGG16 further improved after optimization approaches including data augmentation and selective fine-tuning were applied. The best test accuracy of 91.11%, precision of 0.9815, recall of 0.8833, and F1-score of 0.9298 were all attained by the optimized model. Improved class discrimination and general dependability are reflected in the significant rise in accuracy and F1-score. The optimized VGG16 architecture produced the best overall performance across the majority of assessment metrics out of all the models that were assessed. Its reasonable depth prevented overfitting on the limited dataset while enabling efficient feature learning. These findings demonstrate that, in situations with limited data, optimized VGG16 is the best model for binary banana leaf disease classification.



*Fig. 11: VGG16 Un-Optimized*

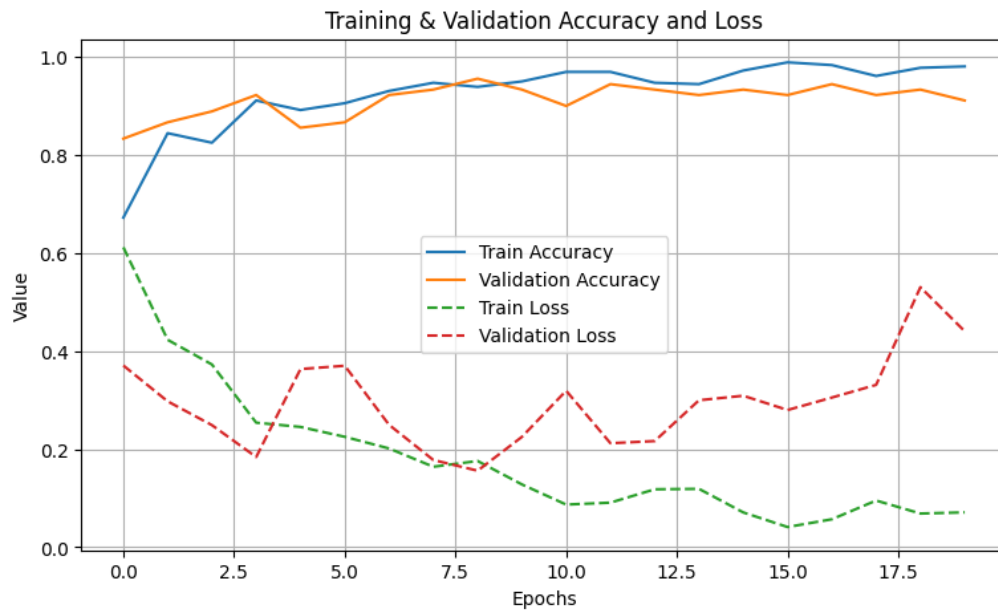


Fig.12: VGG16 Optimized

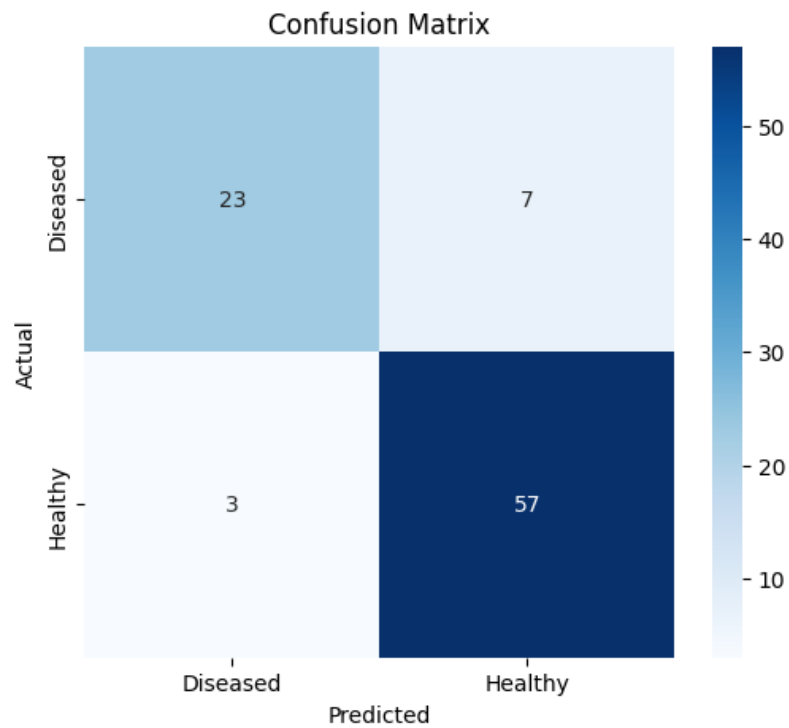


Fig 13: VGG16 Un-Optimized Matrix

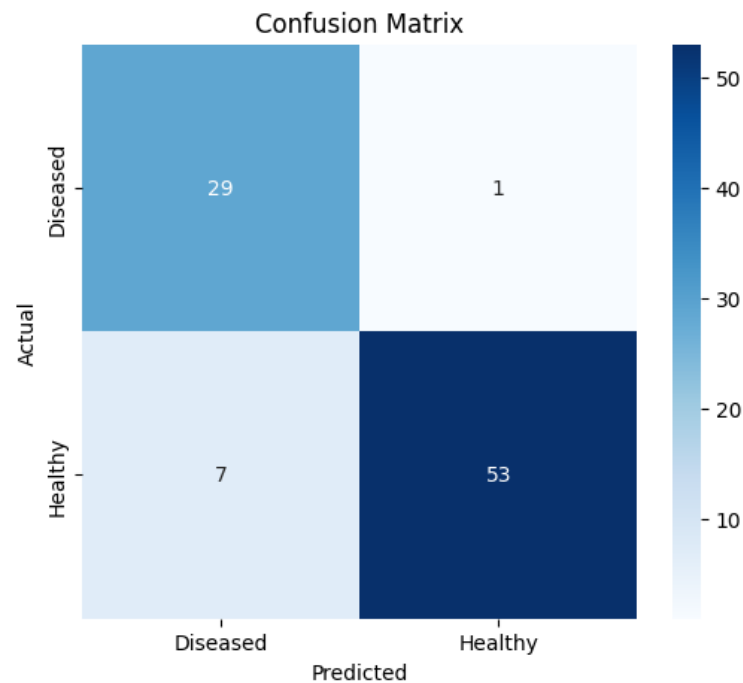


Fig.14: VGG16 Optimized Matrix

**Table 4: Performance Comparison of VGG16 Model**

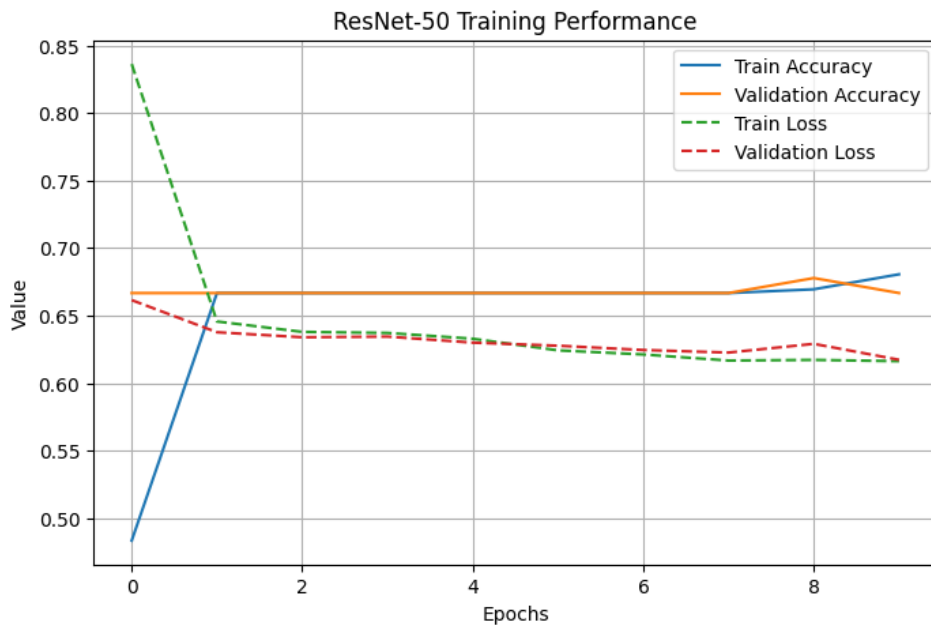
Configuration	Test Accuracy (%)	Precision	Recall	F1-Score
VGG16 (Un-optimized)	88.89	0.8886	0.8889	0.8867
VGG16 (Optimized)	91.11	0.9815	0.8833	0.9298

### 3. ResNet50 Results :

ResNet50 obtained an F1-score of 0.8000, precision of 0.6667, recall of 1.0000, and test accuracy of 66.67% in the non-optimized environment. The poor precision and accuracy reflect a significant proportion of false positives, reducing overall reliability, even though the recall value shows that the majority of diseased samples were accurately diagnosed.

Using optimization techniques did not result in any discernible increase in performance. The optimized ResNet50 model achieved the same results as the non-optimized version in terms of accuracy (66.67%), precision (0.6667), recall (1.0000), and F1-score (0.8000). This suggests that deep architectures, like ResNet50, are prone to overfitting when trained on little amounts of data and are extremely sensitive to dataset size.

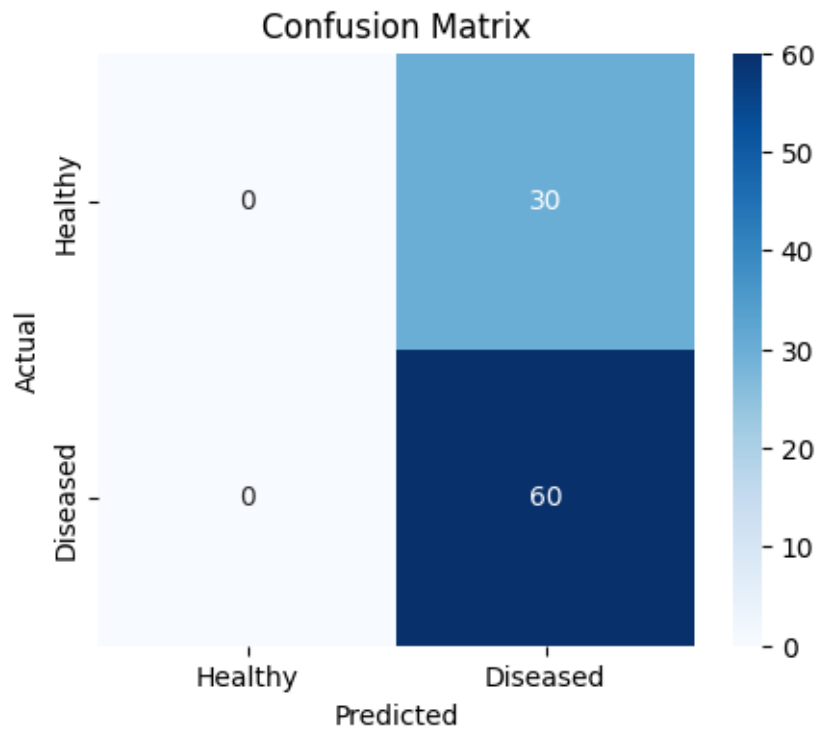
The findings show that small-scale binary banana leaf datasets are not a good fit for ResNet50 due to its high model complexity. The network's depth limits effective generalization under data-constrained settings, despite its excellent theoretical capabilities.



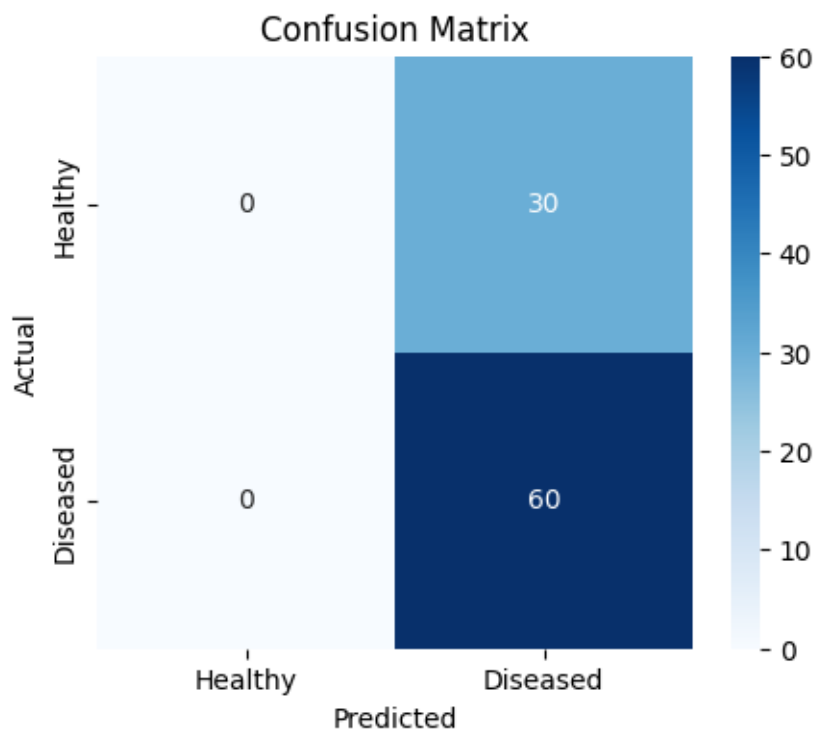
*Fig.15: ResNet50 Un-optimized*



*Fig.16: ResNet50 Optimized*



*Fig.17: ResNet50 Un-Optimized Matrix*



*Fig.18: ResNet50 Optimized Matrix*

**Table 5: Performance Comparison of ResNet50 Model**

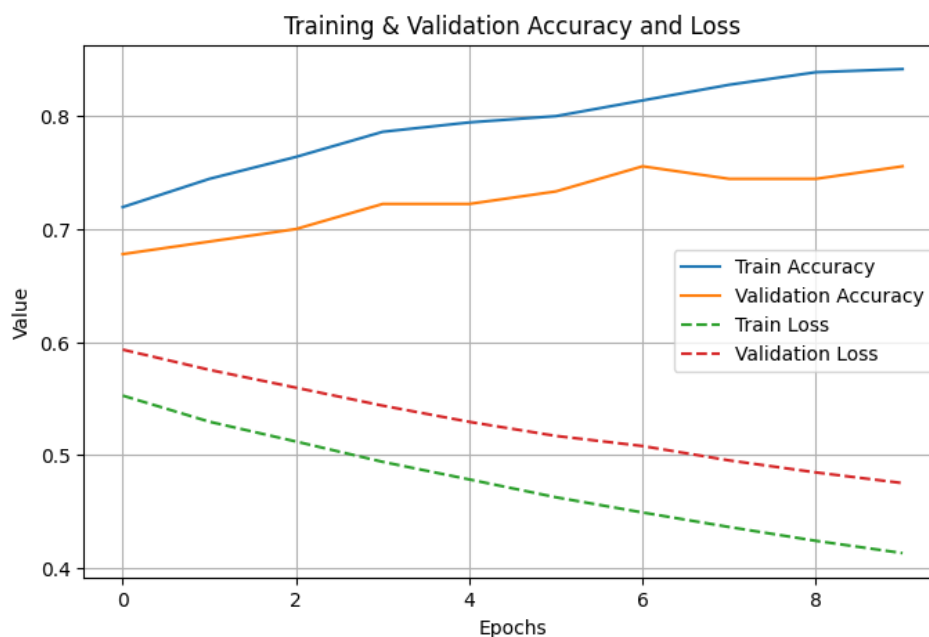
Configuration	Test Accuracy (%)	Precision	Recall	F1-Score
ResNet50 (Un-optimized)	66.67	0.6667	1.0000	0.8000
ResNet50 (Optimized)	66.67	0.6667	1.0000	0.8000

**Observation:** After optimization, ResNet50 exhibits no difference. Due to the short dataset size, the deep residual architecture had trouble generalizing and tended to overfit.

#### 4. MobileNetV2 Results :

MobileNetV2 obtained an F1-score of 0.8330, precision of 0.7639, recall of 0.9167, and test accuracy of 75.56% in the unoptimized setup. While the intermediate precision implies some misclassification of healthy samples, the high recall value shows significant sensitivity in identifying diseased leaves. Recall increased to 1.0000 after using enhanced procedures, however test accuracy marginally declined to 71.11%. Nevertheless, the precision decreased to 0.6977, leading to an F1-score of 0.8219. This result implies that false positives increased as the improved model got increasingly skewed toward predicting sick samples.

MobileNetV2 is appropriate for disease screening assignments where missing diseased samples is crucial due to its great memory and computational efficiency. However, when employing lightweight architectures on small datasets, careful tuning is necessary to balance accuracy and recall due to its decreased precision and susceptibility to optimization.



*Fig.19: MobileNet V2 Un-optimized*

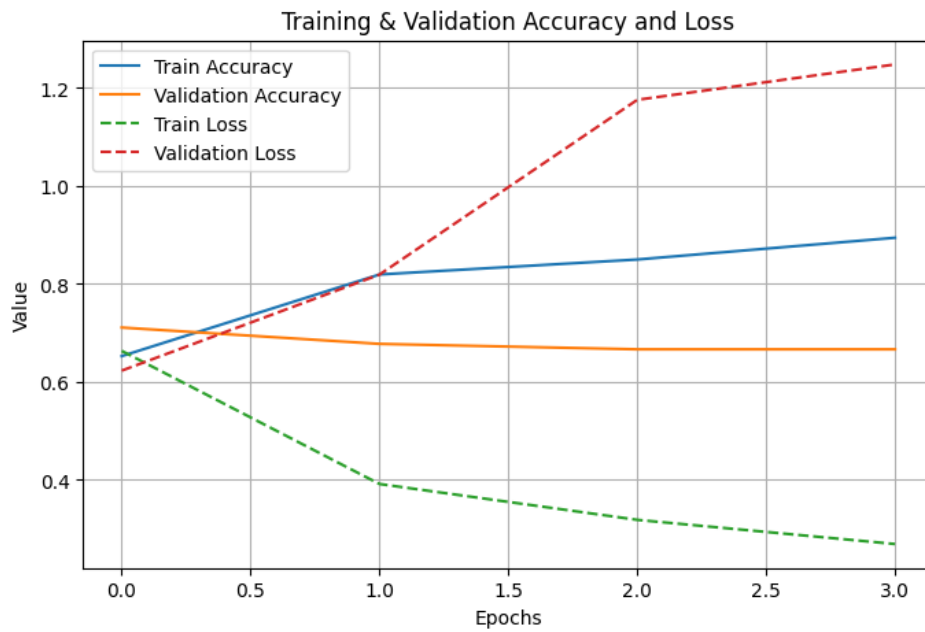


Fig.20: MobileNet V2 Optimized

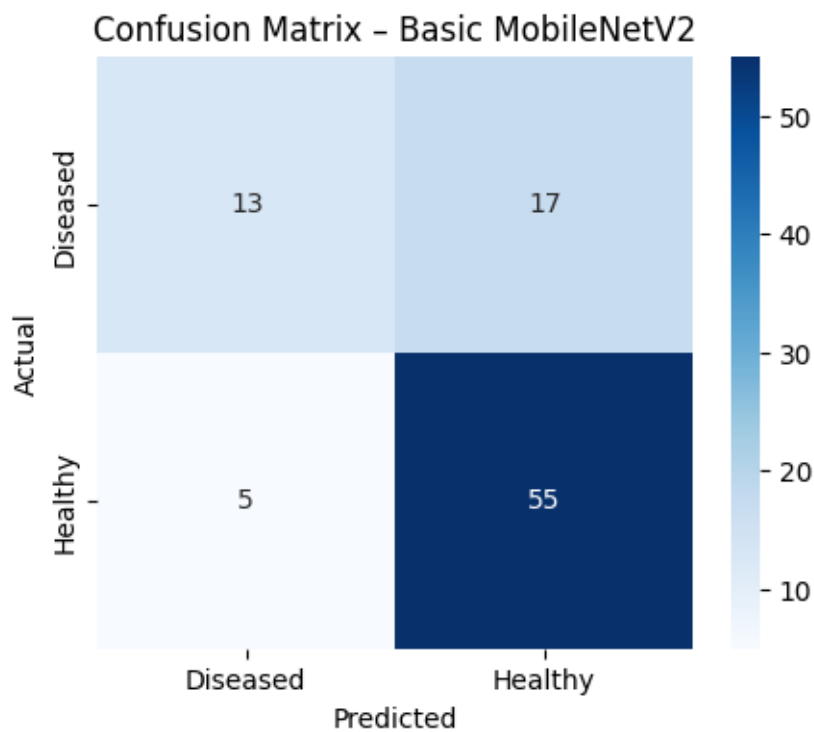


Fig.21: MobileNet V2 Un-Optimized Matrix

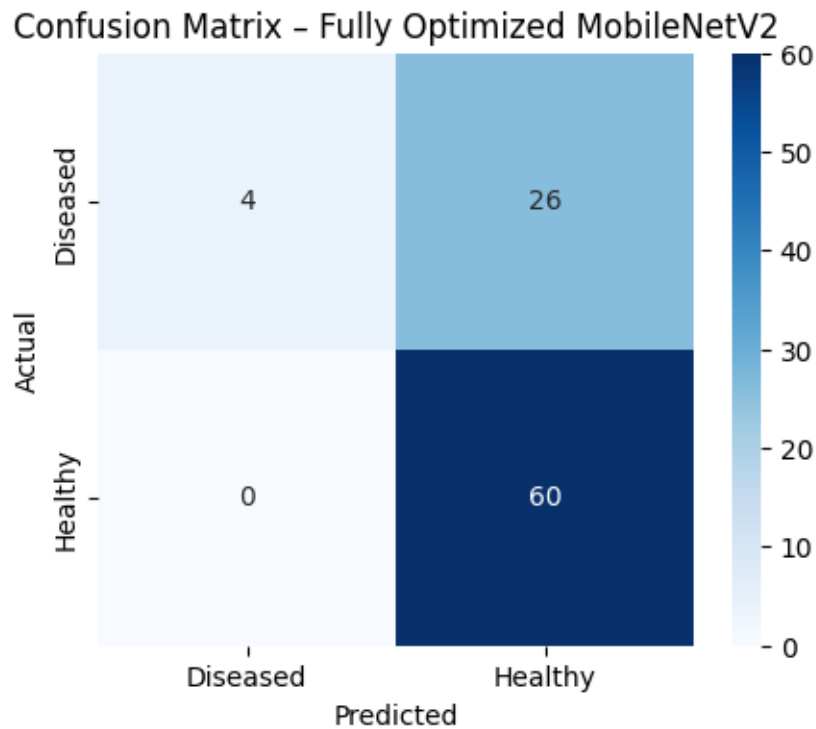


Fig.22: MobileNetV2 Optimized

**Table 6: Performance Comparison of MobileNetV2 Model**

Configuration	Test Accuracy (%)	Precision	Recall	F1-Score
MobileNetV2 (Un-optimized)	75.56	0.7639	0.9167	0.8330
MobileNetV2 (Optimized)	71.11	0.6977	1.0000	0.8219

**Observation :** MobileNetV2 showed a bias towards predicting sick samples, as evidenced by its high recall but low precision. Optimization revealed a precision–recall tradeoff by increasing recall while somewhat decreasing accuracy.

### 5. Comparative Performance Analysis :

The performance of all assessed deep learning models—including a simple CNN and three transfer learning architectures (VGG16, ResNet50, and MobileNetV2)—under both non-optimized and optimized setups is thoroughly compared in this section. To comprehend model behavior on a small-scale binary banana leaf disease dataset, the analysis focuses on classification accuracy, precision, recall, and F1-score.

#### 5.1 Tabular Comparison of All Models :

**Table 7 summarizes the quantitative performance of all models with and without optimization.**

**Table 7: Comparative Performance of Deep Learning Models**



Model	Optimization	Accuracy (%)	Precision	Recall	F1-Score
Basic CNN	Un-Optimized	81.11	0.9057	0.8000	0.8496
Basic CNN	Optimized	88.89	0.9310	0.9000	0.9153
VGG16	Un-Optimized	88.89	0.8886	0.8889	0.8867
VGG16	Optimized	91.11	0.9815	0.8833	0.9298
ResNet50	Un-Optimize	66.67	0.6667	1.0000	0.8000
ResNet50	Optimized	66.67	0.6667	1.0000	0.8000
MobileNetV2	Un-Optimize	75.56	0.7639	0.9167	0.8330
MobileNetV2	Optimized	71.11	0.6977	1.0000	0.8219

The results clearly indicate that optimization techniques significantly improve performance for most models, particularly the basic CNN and VGG16 architectures.

## 5.2. Effect of Optimization Techniques :

Optimization techniques such as data augmentation and selective fine-tuning had a notable impact on model performance, especially for architectures with moderate depth.

- Following optimization, the basic CNN's accuracy increased from 81.11% to 88.89% and its F1-score increased from 0.8496 to 0.9153, indicating that even basic architectures may perform competitively with proper training.
- All the models that were examined, VGG16 profited the most from optimization, attaining the best overall accuracy (91.11%) and F1-score (0.9298). Improved confidence in disease prediction is indicated by the notable increase in precision (from 0.8886 to 0.9815).
- In contrast, ResNet50 did not get better with optimization. The metrics and accuracy (66.67%) stay the same, suggesting that a too deep model relative to the size of the dataset leads to overfitting and poor generalization.
- Optimization for MobileNetV2 increased recall to 1.0000, but overall accuracy and precision suffered as a result. This suggests a tendency to predict diseased samples, which might not be ideal for making balanced agricultural decisions.

These findings highlight that optimization is effective only when model complexity aligns with dataset size.

## 5.3 Impact of Model Depth on Small Datasets :

Model depth plays a critical role in determining classification performance under limited data conditions.

- Moderately deep architectures, such as VGG16, performed better since they could extract useful features without using too many parameters.
- Even though deeper architectures, like ResNet50, had flawless recall, they failed to generalize well and showed overfitting. Instead of extensive feature learning, this behavior suggests memory.



- Although lightweight models, like MobileNetV2, offered excellent recall and computational efficiency, they were less dependable for accurate illness discriminating because to their decreased precision and erratic accuracy trends..
- In limited datasets, the simple CNN outperformed transfer learning models when optimized, highlighting the significance of training approach over architectural complexity.

These findings are further supported by the training–validation accuracy and loss curves, which showed that optimized CNN and VGG16 models had less overfitting and smoother convergence than deeper architectures.

### **Discussion :**

The experimental results offer important insights into how various deep learning architectures behave when classifying binary banana leaf diseases with sparse data. The findings demonstrate how model complexity, optimization techniques, and dataset properties affect classification performance.

### **1. Superior Performance of VGG16 :**

When optimization approaches were used, VGG16 consistently obtained the highest accuracy and F1-score among all examined models. Its balanced architectural depth, which permits efficient feature extraction without undue parameter complexity, is responsible for this. In contrast to deeper networks, VGG16 is capable of good generalization on limited datasets, particularly when selective fine-tuning and data augmentation are used. VGG16 is the most dependable model for the dataset since it was able to learn discriminative characteristics between healthy and diseased leaves, as seen by its high precision and recall scores.

### **2. Limitations of Deeper Architectures (ResNet50) :**

Despite being a strong deep architecture, ResNet50's efficacy in this investigation was just moderate. The model performed the same in both optimized and non-optimized settings, demonstrating that optimization was ineffective. This trend implies that in order to fully utilize their representational potential, deeper networks need larger and more varied information. ResNet50 frequently overfits or converges to less-than-ideal solutions when trained on tiny datasets, which results in decreased accuracy and erratic precision–recall behavior.

### **3. Suitability of Binary Classification for Agricultural Applications :**

The choice to reorganize the dataset from multi-class to binary classification (healthy vs. diseased) is in line with practical agricultural requirements. Early disease detection is frequently more important in real-world farming situations than determining the precise type of sickness. Binary classification is appropriate for small farms and early-stage disease monitoring systems because it streamlines model learning, lowers data requirements, and



facilitates quicker decision-making.

#### **4. Precision–Recall Tradeoffs and Their Implications :**

Important precision–recall tradeoffs between various models are shown by the study. Strong sensitivity to sick samples was demonstrated by models like MobileNetV2, which achieved very high recall at the expense of decreased precision. A larger false-positive rate is implied by this trade-off, which may be acceptable in early disease screening but undesirable for automated treatment decisions. VGG16, on the other hand, achieved greater overall reliability by maintaining a better balance between precision and recall.

#### **5. Practical Insights for Small-Scale Datasets :**

The findings highlight the importance of considering dataset size and application needs when choosing a model. When training data is scarce, moderately deep models using optimization techniques outperform highly deep structures. Although lightweight models are efficient, they must be carefully adjusted to prevent biased predictions. Overall, the results indicate that steady and accurate performance on small-scale agricultural picture datasets requires a combination of optimization strategies and a suitable model complexity.

#### **Conclusion :**

Using a generated dataset, this study compared several transfer learning architectures and a basic Convolutional Neural Network (CNN) for the binary classification of banana leaf illnesses. The experimental findings show that both architectural design and the use of optimization techniques have a significant impact on model performance, especially when training data is scarce. With the highest accuracy and F1-score, the optimized VGG16 architecture proved to be the best-performing model out of all those assessed. It was highly suited for small-scale agricultural picture datasets because of its balanced depth, which allowed for efficient feature extraction without overfitting.

Out of all the models evaluated, the improved VGG16 architecture demonstrated the highest accuracy and F1-score, making it the most effective model. Because of its balanced depth, which enabled effective feature extraction without overfitting, it was ideal for small-scale agricultural image datasets. The results unequivocally demonstrate how crucial optimization strategies like data augmentation and selective fine-tuning are for enhancing model generalization and classification accuracy. Optimization is crucial when working with limited datasets, as demonstrated by the fact that optimized models consistently outperformed their baseline counterparts across all evaluation metrics.

One important addition of this work is the rearrangement of an existing multi-class banana leaf dataset into a binary classification framework. This change simplifies the learning process and aligns with real-world agricultural requirements, where early disease detection is often more crucial than identifying specific disease types. The proposed methodology provides a practical and effective foundation for future research that includes multi-class classification, the development of custom datasets, and real-time agricultural



application.

### **Future Scope :**

The suggested framework can be expanded to multi-class illness classification, allowing the diagnosis of particular banana diseases and nutritional deficits, even if the current study concentrates on binary classification of banana leaf health. This addition would enable more focused agricultural operations and enhance the system's diagnostic capabilities.

Future research will also include gathering a unique real-world dataset that was recorded in the field with natural lighting, background, leaf orientation, and disease severity variations. In real-world deployment circumstances, such data will improve generalization and strengthen the model's resilience.

By merging predictions from several optimal models, cross-validation techniques and ensemble learning methodologies can be investigated to further improve model dependability. Especially when working with sparse or unbalanced datasets, these methods should lessen bias, increase stability, and improve overall classification performance.

Future studies will also look into how to put models on mobile and edge devices, including smartphones and embedded systems, in a lightweight and effective manner, allowing for on-site illness detection without the need for cloud infrastructure. Small-scale farmers in areas with limited resources will particularly benefit from this.

Lastly, real-time field testing will be used to evaluate the constructed models, evaluating their performance in actual agricultural settings. In order to facilitate the actual implementation of deep learning-based disease detection systems in precision agriculture, this stage is crucial for assessing system usability, response speed, and prediction reliability.

### **References :**

- Bhuiyan, M. A. B., Abdullah, H. M., Arman, S. E., Saminur Rahman, S., & Al Mahmud, K. (2023). *BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases*. *Smart Agricultural Technology*, 4, 100214.
- Chuquimarca, L. E., Vintimilla, B. X., & Velastin, S. A. (2025). Assessing deep learning model robustness for banana ripeness classification under varying illumination conditions. *Smart Agricultural Technology*, 101333.
- Rahman, M. A., Raihan, M. A., Uddin, M. S., Hasan, M., & Habib, M. T. (2025). A Comprehensive Analysis of Classic Machine Learning and Deep Learning Modeling for Breed Recognition of Bananas. *IEEE Access*, 13, 194562-194579.
- Elinisa, C. A., & Mduma, N. (2024). Mobile-based convolutional neural network model for the early identification of banana diseases. *Smart Agricultural Technology*, 7, 100423.
- Narayanan, K. L., Krishnan, R. S., Robinson, Y. H., Julie, E. G., Vimal, S., Saravanan, V., & Kaliappan, M. (2022). Banana plant disease classification using



- hybrid convolutional neural network. *Computational Intelligence and Neuroscience*, 2022(1), 9153699.
- Neupane, B., Horanont, T., & Hung, N. D. (2019). Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PloS one*, 14(10), e0223906.
  - Elinisa, C. A., Wa Maina, C., Vodacek, A., & Mduma, N. (2025). Image Segmentation Deep Learning Model for Early Detection of Banana Diseases. *Applied Artificial Intelligence*, 39(1), 2440837.
  - Jiménez, N., Orellana, S., Mazon-Olivo, B., Rivas-Asanza, W., & Ramírez-Morales, I. (2025). Detection of leaf diseases in banana crops using deep learning techniques. *AI*, 6(3), 61.
  - Thiagarajan, J. D., Kulkarni, S. V., Jadhav, S. A., Waghe, A. A., Raja, S. P., Rajagopal, S., ... & Subramaniam, S. (2024). Analysis of banana plant health using machine learning techniques. *Scientific Reports*, 14(1), 15041.
  - Rohini, R. and Raghavendra, S., “CNN-based banana leaf disease identification using image preprocessing techniques,” *International Journal of Engineering Research and Technology*, vol. 10, no. 6, pp. 94–101, 2021.
  - Syihad, A. M., Zaki, W. M. D. W., and Muhamad, N., “Performance comparison of deep CNN architectures for banana leaf disease classification,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 345–352, 2020.
  - Jadhav, A. and Bhandari, S., “A review on banana leaf disease detection using deep learning techniques,” *Procedia Computer Science*, vol. 171, pp. 937–945, 2020.
  - Rivankar, A. A., Jadhav, P., and Kulkarni, A., “Real-time banana disease detection using deep convolutional neural networks,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 3125–3136, 2022.
  - Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., and Blomme, G., “AI-powered banana disease and pest detection using deep learning,” *Plant Methods*, vol. 15, no. 92, pp. 1–12, 2019.
  - Rivankar, R. R., Pai, S. N., Rhisheekesan, A., Prakash, L., VG, S., & Joseph, A. P. (2025). Real-Time Disease Identification of Banana Plants Using Neural Network Models. *IAENG International Journal of Computer Science*, 52(8).
  - Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
  - Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), 1-48.
  - Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4), 427-437.
  - Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and electronics in agriculture*, 145, 311-318.
  - Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 161, 272-279.
  - Kamilaris, A., & Prenafeta-Boldu, F. X. (2018). Deep learning in agri-culture: a



- survey. computers and electronics in agriculture 147: 70–90.
- S.A Shaikh and K.D Gaikwad,(2025). Plant Disease Detection using machine learning: Challenges, Innovation and future direction Machine Learning Approach 2348-7143
  - Liu, M. (2024). Analysis of image data augmentation in smart agriculture. Highlights Sci., Eng. Technol., 85, 750- 756.
  - Antwi, K., Bennin, K. E., Asiedu, D. K. P., & Tekinerdogan, B. (2024). On the application of image augmentation for plant disease detection: A systematic literature review. Smart Agricultural Technology, 9, 100590.
  - Khan, R. U., Khan, K., Albattah, W., & Qamar, A. M. (2021). Image-based detection of plant diseases: from classical machine learning to deep learning journey. Wireless Communications and Mobile Computing, 2021, 1-13.
  - Ahmad, A., Saraswat, D., & El Gamal, A. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. Smart Agricultural Technology, 3, 100083.
  - Lanjewar, M. G., & Panchbhai, K. G. (2023). Convolutional neural network based tea leaf disease prediction system on smart phone using paas cloud. Neural Computing and Applications, 35(3), 2755-2771.
  - Ahmed, I., & Yadav, P. K. (2023). A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases. Sustainable Operations and Computers, 4, 96-104.
  - Gosai, D., Kaka, B., Garg, D., Patel, R., & Ganatra, A. (2022, January). Plant disease detection and classification using machine learning algorithm. In 2022 International Conference for Advancement in Technology (ICONAT) (pp. 1-6). IEEE.
  - Amara, J., Bouaziz, B., & Algergawy, A. (2017). A deep learning-based approach for banana leaf diseases classification. In Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband (pp. 79-88). Gesellschaft für Informatik eV.

