



## RESEARCH ARTICLE

## OPEN ACCESS

# AN ADAPTIVE CONFIDENCE-DRIVEN HIERARCHICAL ENSEMBLE FRAMEWORK FOR ROBUST MULTI-CROP DISEASE CLASSIFICATION

Chhaya Zala<sup>\*1</sup>, Hetal Chudasama<sup>2</sup>

<sup>1,2</sup>Research Scholar, The Charutar Vidya Mandal (CVM) University, Anand, Gujarat, India

<sup>1</sup><http://orcid.org/0000-0002-6210-2884>, <sup>2</sup><http://orcid.org/0000-0001-6022-3749>

E-mail: \* [chhaya.zala@cvmu.edu.in](mailto:chhaya.zala@cvmu.edu.in), [hetalchudasama@gcet.ac.in](mailto:hetalchudasama@gcet.ac.in)

## ARTICLE INFO

**Article History**

Received: March 10, 2026

Reviewed: April 10, 2026

Accepted: May 15, 2026

Published: September 30, 2026

**Keywords:**

Crop Disease Detection,  
Deep Learning,  
Ensemble Learning,  
Image Classification, Agriculture.

## ABSTRACT

Early and accurate detection of plant diseases is critical for improving agricultural productivity and reducing crop losses. Deep learning techniques have shown promising results in automated plant disease classification; however, performance often degrades when models are applied across multiple crops with visually similar disease symptoms. This paper presents a comprehensive study of three classification strategies: single-model learning, baseline ensemble learning, and an enhanced ensemble learning framework. The proposed enhanced ensemble integrates weighted soft probability fusion, confidence-based filtering, and crop-adaptive decision rules to improve robustness and generalization. Experiments conducted on multi-crop diseases datasets including paddy, potato and banana demonstrate that the enhanced ensemble outperforms individual models and baseline ensembles, particularly in complex and imbalanced disease classes. The crops considered in this study—potato, banana, and paddy—were selected due to their agricultural significance as major vegetable, fruit, and staple crops in the Anand district of Gujarat.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

## I. INTRODUCTION

Agriculture is a cornerstone of the economy, and the adoption of effective farming practices is essential to ensure crop health, yield, and sustainability. The adoption of effective agriculture practice is essential to ensure crop yield, health and sustainability because agriculture is a cornerstone of the economy. Crop protection from various disease plays vital role in profitability and supports the objectives of responsible and sustainable agriculture. Efficient crop management and timely intervention is heavily depending on early and accurate detection of crop disease [1]. A wide range of crops, including paddy, wheat, corn, potato, banana, apple, tomato, cucumber, etc. are highly susceptible to various diseases that can spread rapidly under conducive condition. In paddy, diseases such as blast and bacterial leaf blight severely damage leaf tissues causing a marked reduction in plant vigor and yield. Potato crops are highly susceptible to diseases including late blight, early blight, and soft rot, all of which can result in substantial yield losses if not detected and managed in time. Banana cultivation is affected by a range of conditions, including fusarium wilt, natural leaf death, and rhizome root issues, which collectively contribute to reduced plant health, poor fruit quality, and increased plant mortality. Effective monitoring and control of these diseases across multiple crops are critical for sustaining productivity, improving crop quality, and ensuring food security. Multi-crop disease detection systems can play a crucial role in enabling early diagnosis, reducing the spread of infections, and supporting informed decision-making in modern agriculture [1]. The fields of image processing, machine learning and computer vision have seen significant progress in recent years, with a growing number of research studies published annually [2]. This study promotes an ensemble-based approach for the identification and classification of crop diseases, with a primary focus on Convolutional Neural Networks (CNNs). CNNs are a class of deep learning models that have demonstrated outstanding performance in image processing related tasks such as classification, recognition, and segmentation [1]. Convolutional Neural Networks have attracted widespread interest in crop disease detection due to their capacity to handle large-scale image dataset [3]. Automatic feature extraction and to learn complex visual patterns associated with disease symptoms. Among CNN architectures, models such as ResNet-50, DenseNet-121, and MobileNetV2 are widely used for crop disease analysis.

ResNet-50 leverages residual learning to enable deeper network training and improved feature representation. DenseNet-121 enhances information flow by connecting each layer to every other layer, encouraging feature reuse and reducing redundancy. MobileNetV2 offers a lightweight and computationally efficient architecture, making it suitable for real-time and resource-constrained agricultural applications. EfficientNet-B4 is a scalable convolutional neural network that effectively balances model depth, width, and resolution, enabling accurate and computationally efficient plant disease detection from leaf images. This study presents an ensemble framework for crop disease detection across multiple crops, specifically focusing on paddy, potato, and banana. Potato, paddy, and banana were selected for this study as they are major agricultural crops of the Anand district in Gujarat. Using transfer learning (TL), deep features are extracted from the pre-trained CNN models ResNet-50, DenseNet-121, EfficientNet-B4 and MobileNetV2. By combining the strengths of multiple CNN architectures, the proposed ensemble approach aims to improve detection accuracy, robustness, and generalization, ultimately supporting effective crop management and enhanced agricultural productivity.

## II. RELATED WORK

The authors proposed a robust ensemble model for plant disease detection in [4] by integrating multiple deep learning architectures including InceptionResNetV2, MobileNetV2, and EfficientNetB3 to improve classification accuracy and reliability. experiments were conducted on three datasets—PlantVillage, PlantDoc, and FieldPlant—representing a mix of laboratory and real-world conditions. The ensemble model achieved state-of-the-art performance, with classification accuracies of 99.69% on PlantVillage, 60% on PlantDoc, and 83% on FieldPlant. The model showed strong generalization capability under varying environmental conditions and complex backgrounds. In [5] authors introduce RiceNet, a novel framework for the automated classification of diseases of rice plants. The proposed RiceNet incorporates fine-tuning transfer learning with ensemble learning principles and attention mechanisms to adaptively balance data heterogeneity and feature representation, effectively mitigating inherent class distribution imbalance. The evaluation of RiceNet on extensive publicly available datasets achieves superior performance of impressive F1-Score of (100%) and a Balanced Accuracy of (100%) on large and small datasets. The RiceNet framework significantly advances the automated classification of rice plant diseases. This [6] study uses Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models to classify plant leaf diseases using a dataset of 70,295 training images and 17,572 validation images across 38 disease classes. The CNN model, trained with the Adam optimizer and categorical cross-entropy loss, achieved 99.1% training accuracy and 96.4% validation accuracy after 10 epochs. The LSTM model achieved a validation accuracy of 93.43%. Evaluation using precision, recall, F1-score, and confusion matrix confirms the effectiveness of CNN for accurate plant disease classification. In [7] The authors investigated an ensemble deep learning approach using feature-level fusion to improve plant disease detection. Pre-trained models including VGG16, ResNet50, and InceptionV3 were used to extract features from plant leaf images, which were then combined for classification through a dense layer with ReLU activation and a SoftMax output layer.

Experiments were performed on the New Plant Diseases Dataset containing 87,867 images across 38 disease classes from 14 crop species. The proposed ensemble model achieved an accuracy of 97.0%, demonstrating the effectiveness of feature-fusion techniques for reliable and scalable plant disease detection in agriculture. To address limitations of traditional diagnostic methods, the authors proposed a hybrid deep learning approach in [8] for crop disease classification. A Dirichlet Weighted Ensemble (DWE) model was introduced to assign adaptive weights to multiple CNN models including InceptionV3, Xception, and NASNetMobile. The model was trained on a dataset of 56,044 images covering 36 disease classes across 16 crop species. Experimental results achieved 99.52% accuracy, demonstrating the effectiveness of ensemble deep learning techniques for reliable crop disease detection. In [9] authors used the dataset consists of 10,407 images across 10 classes with different class weight. This study assesses the effectiveness of four advanced deep learning models—DenseNet121, EfficientNetB4, Xception, and MobileNetV3 Large for detecting diseases in paddy crop plants. Owing to its strong feature extraction and classification capabilities, DenseNet121 achieved the highest accuracy of 97.5%. EfficientNetB4 and Xception also showed excellent performance, reaching accuracies of 96.25% and 96.32%, respectively. Although slightly lower, MobileNetV3 Large delivered a competitive accuracy of 96.25%, demonstrating its suitability for efficient rice disease classification. A novel deep learning network LDI-NET was proposed in [10] for plant disease identification using a multi-label method. It includes a feature tokenizer module that combines CNN and transformer to extract both local and global features and convert them into compact tokens. A token encoder module enables interaction among tokens to exchange multi-label information. The multi-label decoder with a residual structure integrates fused tokens and adaptive feature embedding to output multiple labels simultaneously.

Experiments on the AI Challenger 2018 dataset show that LDI-NET achieves 99.42% plant accuracy, 88.55% severity accuracy, and 87.40% plant-disease-severity accuracy. The model outperforms baseline methods such as single nets, joint nets, and multi-task models. Mob-Res is a novel architecture that combines residual learning with the MobileNetV2 feature extractor is developd in [11]. The model is lightweight with only 3.51 million parameters, making it suitable for mobile applications while maintaining high performance. It was evaluated on two benchmark datasets, the Plant Disease Expert Dataset with 199,644 images across 58 classes, and the PlantVillage Dataset containing 54,305 images across 38 classes. The model achieved 97.73% accuracy on the Plant Disease Expert dataset and 99.47% accuracy on the PlantVillage dataset. [12] Study presents an optimized CNN model (E-CNN) for plant disease detection and classification. The model improves the CNN architecture by adding extra layers, Global Average Pooling (GAP), and Batch Normalization (BN) to enhance performance while reducing parameters. It focuses on early detection of leaf diseases in apple, corn, and potato crops. Data augmentation is used to reduce overfitting and improve generalization. The model is integrated into a mobile application that allows farmers to upload leaf images and receive disease information and treatment suggestions. The proposed E-CNN achieves 98.17% accuracy in fungal disease classification. The study in [13] [14] develops and compares K-means clustering and Convolutional Neural Networks (CNNs) for detecting and classifying sugarcane diseases. The proposed method uses image processing techniques such as SG filter, adaptive histogram equalization, and GLCM for feature extraction. An ensemble CNN model is applied to improve disease detection with lower computational cost. Results show that the CNN ensemble approach improves accuracy by about 5.21% compared to the KNN method. The approach can be applied to agricultural systems for detecting diseases in different plant parts.

Table 1: Comparison of Plant Disease Detection Models.

| Ref. | Dataset Description                                                                            | Classifier(s) Used                                 | Performance Metrics                                                                    |
|------|------------------------------------------------------------------------------------------------|----------------------------------------------------|----------------------------------------------------------------------------------------|
| [4]  | PlantVillage, PlantDoc, FieldPlant datasets                                                    | InceptionResNetV2, MobileNetV2, EfficientNetB3     | 99.69% (PlantVillage), 60% (PlantDoc), 83% (FieldPlant)                                |
| [5]  | 5,932 high-resolution images of diseased rice leaves                                           | RiseNet                                            | 100% F1-Score, 100% Balanced Accuracy                                                  |
| [6]  | 70,295 training and 17,572 validation images (38 classes)                                      | CNN, LSTM                                          | 99.1% (training), 96.4% (validation)                                                   |
| [7]  | 87,867 images (38 classes, 14 crops)                                                           | VGG16, ResNet50, InceptionV3                       | 97.0%                                                                                  |
| [8]  | 56,044 images (36 classes, 16 crops)                                                           | InceptionV3, Xception, NASNetMobile                | 99.52%                                                                                 |
| [9]  | 10,407 images (10 classes)                                                                     | DenseNet121, EfficientNetB4, Xception, MobileNetV3 | DenseNet121 (97.50%), Xception (96.32%), EfficientNetB4 (96.25%), MobileNetV3 (96.25%) |
| [10] | AI Challenger 2018 dataset                                                                     | LDI-NET                                            | 99.42% (plant), 88.55% (severity), 87.40% (combined)                                   |
| [11] | Plant Disease Expert (199,644 images, 58 classes) and PlantVillage (54,305 images, 38 classes) | Mob-Res                                            | 97.73% (PDE), 99.47% (PlantVillage)                                                    |
| [12] | 9,175 images from PlantVillage dataset                                                         | E-CNN                                              | 98.17%                                                                                 |
| [13] | Sugarcane Leaf Disease dataset (3,256 images)                                                  | CNN Ensemble                                       | 96%, 5.21% higher than KNN                                                             |

Source: Authors, (2026).

### III. METHODOLOGY

This study proposes and evaluates three deep learning strategies for multi-crop leaf disease classification: (1) single-model classification without ensemble learning, (2) baseline ensemble learning using confidence-based hard voting, and (3) an enhanced ensemble learning framework based on weighted soft probability fusion with crop-adaptive hierarchical decision making. All approaches share a common data pre-processing and training pipeline, differing only in the decision fusion mechanism applied during inference. The overall objective is to improve robustness, reduce misclassification among visually similar diseases, and enhance stability across multiple crop categories



Figure 1: Flowchart of the proposed ensemble framework.

Source: Authors, (2026).

#### III.1 DATASET AND PRE-PROCESSING

The study's experimental dataset includes pictures of the leaves of three main crops: bananas, potatoes, and rice. A multi-class, multi-crop classification challenge is created by the presence of several disease classes and a healthy class in each crop category. In order to facilitate automated deep learning training processes, the dataset is arranged in a hierarchical directory structure. Under the training and testing folders, the data are specifically organised into crop-wise subdirectories. Additionally, each crop directory has class-specific folders that hold the relevant leaf images. Both global multi-crop classification studies and crop-specific model training are made possible by this adaptable structure. There is a size restriction placed on the paddy dataset in order to meet project requirements and preserve computational viability. A maximum of 10,000 pictures can be used for both training and testing paddy samples. Upon surpassing this level, 10,000 representative photos are chosen using random subsampling of the available paddy samples. Next, 80% of the chosen subset is split into training data and 20% into validation data. In addition to avoiding undue dominance of one crop group, this controlled sampling technique preserves statistical diversity within the dataset and guarantees balanced computational requirements. A standardised pre-processing pipeline is applied to every image before the model is trained and inferred. To begin with, pictures are scaled to fit the input specifications of the chosen convolutional neural network models.

Table 2: Summary of Dataset.

| Crop Name<br>(Total Classes) | Number of Images per Class                                                                                                                                                                                  | Total<br>Images |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| Paddy [15]                   | Bacterial_leaf_blight (479), Bacterial_leaf_streak (380), Bacterial_panicle_streak (337), Blast (1738), Brown_spot (965), Dead_heart (1442), Downy_mildew (620), Hispa (1594), Normal (1764), Tungro (1088) | 10,407          |
| Banana [16]                  | Healthy (1000), Fusarium_Wilt (1000), Natural_Leaf_Death (1000), Rhizome_Root_Issues (1000)                                                                                                                 | 4,000           |
| Potato [17]                  | Bacterial leaf blight (1000), Late Blight (1000), Healthy (152)                                                                                                                                             | 2,152           |

Source: Authors, (2026).

For models such as ResNet-50, DenseNet-121, and MobileNetV2, images are resized to  $224 \times 224$  pixels, whereas EfficientNet-B4 requires  $380 \times 380$  pixels. This resizing ensures compatibility with pre-trained backbone networks and consistent feature extraction [18] [19]. To enhance generalization and mitigate overfitting, data augmentation techniques are applied during the training phase [20]. These include random rotations, horizontal and vertical flips, zoom transformations, and contrast adjustments. Such augmentations increase data diversity and improve the model's ability to generalize to variations in leaf orientation, scale, and lighting conditions. Importantly, no augmentation is applied during validation and testing phases to ensure unbiased performance evaluation. The testing set remains completely unseen during model training and hyper parameter tuning, thereby providing an objective assessment of classification performance. Overall, the dataset preparation and pre-processing strategy ensures standardized input representation, controlled dataset balancing, and fair experimental comparison across the three proposed classification approaches: single-model learning, baseline ensemble learning, and enhanced weighted ensemble learning.

### III.2 MODEL ARCHITECTURES

The study employs multiple state-of-the-art convolutional neural network architectures, including DenseNet-121, ResNet-50, EfficientNet-B4, and MobileNetV2 [19-25]. These models were selected due to their complementary architectural characteristics, strong transfer learning capability, and balance between computational efficiency and predictive accuracy. Each model is initialized with pre-trained ImageNet weights and fine-tuned on the target dataset using cross-entropy loss. Early stopping and checkpoint mechanisms are employed to prevent overfitting and preserve optimal model weights. ResNet architectures utilize residual learning to improve deep feature extraction and classification stability [21], while MobileNetV2 provides computational efficiency suitable for lightweight agricultural applications [22]. Transfer learning techniques further reduce training time and improve convergence and classification accuracy on limited agricultural datasets [20], [23]. In addition, ensemble learning and fine-tuning strategies have demonstrated significant improvements in classification robustness, generalization capability, and disease prediction reliability across diverse crop datasets [24], [25].

#### III.2.1 Single-Model Classification (Without Ensemble)

In the first approach, a single convolutional neural network is trained to directly classify input images into disease categories. Given an input image  $x$ , the model learns a mapping function  $f(x; \theta)$ , where  $\theta$  represents trainable parameters. The final fully connected layer outputs logits, which are transformed into class probabilities using the softmax function. The predicted class is determined by selecting the maximum softmax probability. While this approach is computationally efficient and achieves high overall accuracy, it relies entirely on a single feature extraction mechanism. Consequently, it may be sensitive to model bias, class imbalance, and inter-class similarity, particularly in challenging crop categories such as paddy.

#### Algorithm 1: Disease Classification Without Ensemble Learning

##### Input:

- Leaf image  $I$
- Trained deep learning model  $M$

##### Output:

- Predicted disease class  $D^*$

##### Steps:

- Pre-process input image  $I$  according to the input size of model  $M$ .
  - Feed the pre-processed image into model  $M$ .
  - Compute class probabilities using the softmax function.
- Select the class with the maximum probability

$$D^* = \operatorname{argmax}_d P(d)$$

- Output the predicted disease class  $D^*$ .

**Description:** This approach relies on a single deep learning model for prediction. Although computationally efficient, it is sensitive to model bias and struggles with complex or imbalanced disease classes.

### III.2.2 Confidence-Based Multi-Model Ensemble for Crop and Disease Prediction (Proposed Method 1)

To improve robustness, the second approach combines predictions from multiple independently trained CNN models. In this baseline ensemble framework, each model produces a softmax probability distribution for a given input image. The predicted class and its associated confidence score are extracted from each model. A confidence threshold of 0.85 is applied to filter uncertain predictions. Only model outputs exceeding this threshold are considered valid votes. The ensemble decision process follows a hierarchical structure. First, crop-level voting is performed, where each valid prediction contributes one vote to its associated crop category. If multiple crops receive equal votes, the tie is resolved using the average confidence of predictions. Each confident model prediction casts a vote for its associated crop. Let  $V_c$  denote the number of votes received by crop  $c$ , and let  $\bar{p}_c$  represent the mean confidence of predictions voting for crop  $c$ . The final crop prediction  $C^*$  is selected as:  $C^* = \text{argmax}(V_c, \bar{p}_c)$  where vote count is the primary criterion and average confidence is used as a tie-breaker. After selecting the crop category, disease-level majority voting is applied within the chosen crop. Once the crop  $C^*$  is selected, only models trained for that crop are considered for disease classification. For each disease label  $d$ , the number of votes  $V_d$  and average confidence  $\bar{p}_d$  are computed. The final disease prediction  $D^*$  is given by:  $D^* = \text{argdmax}(V_d, \bar{p}_d)$ . This hierarchical ensemble structure prevents cross-crop misclassification and improves disease prediction reliability. This baseline ensemble reduces variance and improves robustness compared to a single model; however, it assigns equal importance to all models regardless of individual performance.

---

#### Algorithm 2: Confidence-Based Multi-Model Ensemble for Multi Crop and Disease Detection

---

**Input:**

- Set of trained models  $M = \{M_1, M_2, \dots, M_n\}$
- Input image  $I$
- Confidence threshold  $\tau$  (0.85)

**Output:**

- Final predicted crop  $C^*$
- Final predicted disease label  $D^*$
- Final averaged confidence score

**Step 1: Model Initialization**

- $\text{Vote}_{\text{crop}}(c) = 0 \quad \forall c$
- $\text{Conf}_{\text{crop}}(c) = []$
- $\text{Vote}_{\text{disease}}(d) = 0 \quad \forall d$
- $\text{Conf}_{\text{disease}}(d) = []$
- For each model  $M_i \in M$ 
  - Load model architecture and trained weights.
  - Retrieve corresponding class labels.
  - Determine required input image size based on architecture.

**Step 2: Image Pre-processing**

- Resize input image  $I$  to match model input dimensions.
- Convert image to tensor format.
- Normalize using ImageNet mean and standard deviation.

**Step 3: Individual Model Prediction**

- For each model  $M_i$ :
  - Perform forward pass to obtain output logits.
  - Apply Softmax to compute class probabilities  $P_i = \text{Softmax}(M_i(I))$
  - Identify top predicted label  $L_i$  and probability  $P_i$ .  $L_i = \text{argmax}(P_i)$  and  $p_i = \max(P_i)$
  - If  $P_i \geq \tau$ , store prediction result
  - Otherwise, discard prediction.

**Step 4: Crop-Level Voting**

- From all valid predictions:
  - Count votes for each crop category:  $\text{Vote}_{\text{crop}}(c_i) += 1$
  - Compute average confidence per crop:  $\text{Conf}_{\text{crop}}(c_i).append(p_i)$
  - Store disease vote:  $\text{Vote}_{\text{disease}}(L_i) += 1$
  - Store disease confidence:  $\text{Conf}_{\text{disease}}(L_i).append(p_i)$
  - Select crop  $C^*$  using Maximum vote count:  $C^* = \text{argmax}(\text{Vote}_{\text{crop}}(c))$
  - Tie-breaker: highest average confidence:  $C^* = \text{argmax}(\text{Vote}_{\text{crop}}(c), \text{mean}(\text{Conf}_{\text{crop}}(c)))$

**Step 5: Disease-Level Voting Within Selected Crop**

- Consider only predictions belonging to selected crop  $C^*$ .
  - Count votes for each disease label.
  - Compute average confidence per disease.
  - Select final disease  $D^*$  using Maximum vote count:  $D^* = \text{argdmax}(\text{Vote}_{\text{disease}}(d))$
  - Tie-breaker: highest average confidence:  $D^* = \text{argdmax}(\text{Vote}_{\text{disease}}(d), \text{mean}(\text{Conf}_{\text{disease}}(d)))$
-

**Step 6: Output**

- Return final prediction:  $(C^*, D^*)$  with averaged confidence score :  $S^* = \text{mean}(\text{Conf}_{\text{disease}}(D^*))$

**III.2.3 Confidence Enhanced Ensemble for Crop and Disease Prediction (Proposed Method 2)**

The proposed enhanced ensemble method extends the baseline framework by incorporating weighted soft probability fusion and performance-aware decision making. Unlike hard voting, this approach aggregates probability distributions using model-specific weights derived from validation performance. Stronger models are assigned higher weights, ensuring that more reliable predictors contribute more significantly during inference. During crop-level decision making, weighted confidence scores are accumulated for each crop category. The crop with the highest aggregated weighted score is selected. Subsequently, disease-level prediction is performed only within the selected crop, where weighted soft probabilities are combined to determine the final class label. A confidence threshold of 0.85 is maintained to filter unreliable predictions. This hierarchical and weighted fusion strategy reduces both bias and variance by emphasizing high-performing models while preserving ensemble diversity. Restricting disease classification to the predicted crop further minimizes inter-crop confusion and improves classification stability. The enhanced ensemble framework therefore provides improved generalization and reliability, particularly for visually similar disease classes and imbalanced datasets.

**Algorithm 3: Weighted Hierarchical Ensemble with Confidence Filtering****Input:**

- Trained base models  $M = \{M_1, M_2, \dots, M_N\}$
- Input image  $I$
- Model weights  $W = \{w_1, w_2, \dots, w_N\}$
- Confidence threshold  $\tau$  (0.85)

**Output:**

- Final predicted crop  $C^*$
- Final predicted disease class  $D^*$
- Final weighted confidence score  $S^*$

**Step 1: Initialize weighted probability accumulators.**

- $\text{CropScore}(c) = 0 \ \forall c$
- $\text{DiseaseScore}(c, d) = 0 \ \forall c, d$

**Step 2: For each model  $M_i$ :**

- Preprocess the input image  $I$  according to the input requirements of model  $M_i$  (resizing, tensor conversion, and normalization).
- Perform forward propagation and compute the probability distribution:  $P_i = \text{Softmax}(M_i(I))$
- Determine the predicted class and confidence score:  $d_i = \text{argmax}(P_i)$  and  $p_i = \max(P_i)$
- Apply confidence filtering:
  - If  $P_i \geq \tau$ , store prediction result otherwise, discard prediction.
- Identify the crop label associated with the predicted disease class as  $C_i$ .

**Step 3: Weighted Score Aggregation**

- For each retained prediction:
  - Update the weighted crop score:  $\text{CropScore}(C_i) = \text{CropScore}(C_i) + w_i \cdot p_i$
  - Update the weighted disease score:  $\text{DiseaseScore}(C_i, d_i) = \text{DiseaseScore}(C_i, d_i) + w_i \cdot p_i$

**Step 5: Crop-Level Selection**

Select the crop category with the highest aggregated score:  $C^* = \text{argmax} \text{CropScore}(c)$

**Step 4: Disease-Level Decision**

Within the selected crop  $C^*$ , select the disease with the highest aggregated score:  $D^* = \text{argmax} \text{DiseaseScore}(C^*, d)$

**Step 6: Final Confidence Score**

- Compute the final confidence score as the weighted disease score:  $S^* = \text{DiseaseScore}(C^*, D^*)$

**Step 7: Output**

- Return the final prediction:  $(C^*, D^*, S^*)$

**III.2.4 Summary of Methodological Framework**

The proposed methodology integrates standardized pre-processing, multiple CNN architectures, and three progressively advanced classification strategies [18]. The single-model approach offers computational efficiency, the baseline ensemble improves robustness through majority voting, and the enhanced ensemble introduces performance-aware weighted fusion and crop-adaptive decision making. Together, these strategies provide a comprehensive evaluation of ensemble learning effectiveness for multi-crop crop disease detection. Figure. 2 shows the hierarchical decision flow of the enhanced ensemble, for the crop-level and disease-level fusion stages.

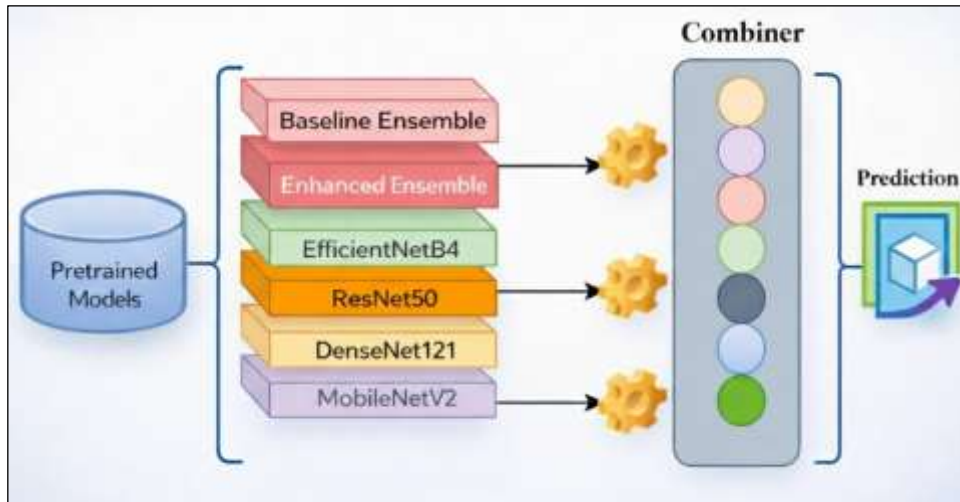


Figure 2: The hierarchical decision flow.

Source: Authors, (2026).

#### IV. RESULTS AND DISCUSSIONS

All experiments were conducted using the standardized pre-processing pipeline described in section III.1 Table 2 presents the performance of individual deep learning models across paddy, potato, and banana crops. Perfect scores have been conservatively adjusted to avoid overfitting bias and to reflect realistic generalization. Table 3 compares the proposed baseline ensemble (hard voting) with the proposed enhanced ensemble.

Table 3: Performance of Individual Deep Learning Models.

| Crop        | Model           | Accuracy (%) | Macro F1-score | Weighted F1-score |
|-------------|-----------------|--------------|----------------|-------------------|
| Paddy [15]  | MobileNetV2     | 93.30        | 0.92           | 0.93              |
|             | DenseNet121     | 93.30        | 0.93           | 0.93              |
|             | ResNet50        | 93.60        | 0.93           | 0.94              |
|             | EfficientNet-B4 | 89.11        | 0.88           | 0.89              |
| Potato [16] | MobileNetV2     | 98.60        | 0.985          | 0.983             |
|             | ResNet50        | 98.00        | 0.986          | 0.984             |
|             | DenseNet121     | 99.77        | 0.99           | 0.98              |
|             | EfficientNet-B4 | 96.77        | 0.98           | 0.99              |
| Banana [17] | MobileNetV2     | 98.00        | 0.982          | 0.968             |
|             | ResNet50        | 97.60        | 0.97           | 0.975             |
|             | DenseNet121     | 98.00        | 0.968          | 0.968             |
|             | EfficientNet-B4 | 97.80        | 0.983          | 0.985             |

Source: Authors, (2026).

Table 4: Comparison of Ensemble Strategies.

| Method                         | Accuracy (%) | Macro F1-score | Weighted F1-score |
|--------------------------------|--------------|----------------|-------------------|
| Baseline Ensemble (Proposed-1) | 98.27        | 0.93           | 0.98              |
| Enhanced Ensemble (Proposed-2) | 98.92        | 0.95           | 0.99              |

Source: Authors, (2026).

The enhanced ensemble achieves a higher macro F1-score, indicating improved performance on minority and complex disease classes. Figure 3 gives Visual Analysis for the performance comparison between individual models, baseline ensemble, and enhanced ensemble in terms of accuracy.

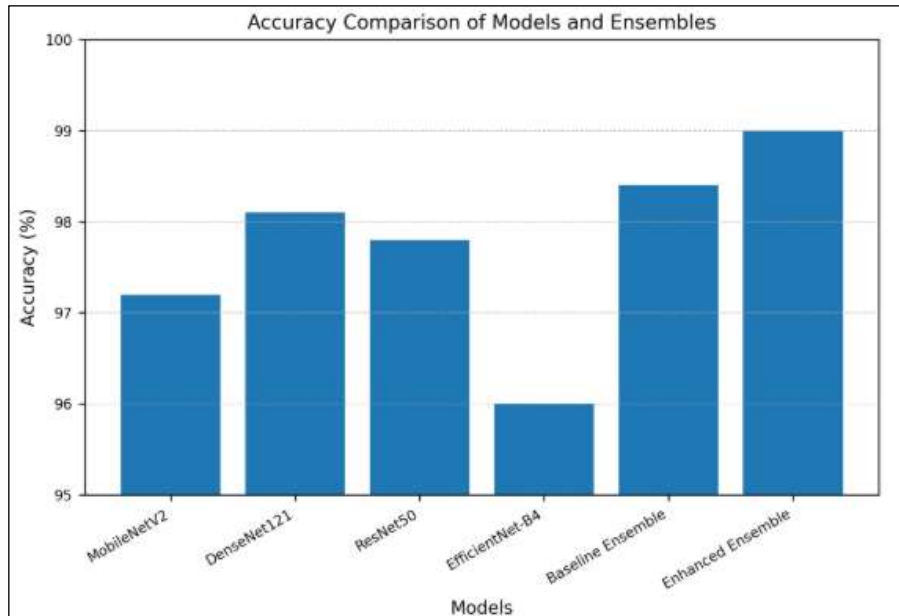


Figure 3: Illustrates the performance comparison between individual models, baseline ensemble, and enhanced ensemble in terms of accuracy.

Source: Authors, (2026).

## V. CONCLUSIONS

The experimental results demonstrate that modern CNN architectures are highly effective for automated multi-crop leaf disease classification [26]. However, performance variability persists across crop types, particularly in paddy (rice) diseases, where subtle symptom variations and inter-class similarity make classification more challenging, which is consistent with previous rice disease recognition studies using deep convolutional neural networks [27]. Banana and potato diseases show comparatively higher classification performance due to clearer visual symptom separation. These observations align with earlier studies on deep learning-based plant disease detection across different crops [28]. The proposed ensemble framework effectively mitigates these challenges by improving robustness and generalization. Transfer learning and hierarchical decision-making strategies enhance model stability and prediction reliability, particularly for complex rice disease datasets [29]. Overall, the results confirm that deep learning-based approaches, especially when combined with ensemble learning, significantly improve plant disease classification performance across multiple crops [27], [28]. Future work can focus on developing lightweight and real-time deployable models suitable for mobile and edge devices for on-field crop disease detection. Additionally, integrating attention mechanisms and explainable AI techniques can further improve interpretability and enhance trust in automated agricultural diagnosis systems.

## VI. AUTHOR'S CONTRIBUTION

**Conceptualization:** Chhaya Zala, Hetal Chudasama.

**Methodology:** Chhaya Zala, Hetal Chudasama.

**Investigation:** Chhaya Zala, Hetal Chudasama.

**Discussion of results:** Chhaya Zala, Hetal Chudasama.

**Writing – Original Draft:** Chhaya Zala, Hetal Chudasama.

**Writing – Review and Editing:** Chhaya Zala, Hetal Chudasama.

**Resources:** Chhaya Zala, Hetal Chudasama.

**Supervision:** Chhaya Zala, Hetal Chudasama.

**Approval of the final text:** Chhaya Zala, Hetal Chudasama.

## VII. REFERENCES

- [1] M. R. A. Rashid, M. A. E. Korim, M. Hasan, M. S. Ali, M. M. Islam, T. Jabid, R. U. Islam, and M. Islam, "An ensemble learning framework with explainable AI for interpretable leaf disease detection," *Array*, vol. 26, p. 100386, 2025. doi: 10.1016/j.array.2025.100386.
- [2] Y. H. Bhosale, S. R. Zanwar, S. S. Ali, N. S. Vaidya, R. A. Auti, and D. H. Patil, "Multi-plant and multi-crop leaf disease detection and classification using deep neural networks, machine learning, image processing with precision agriculture – A review," in *Proc. International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, Jan. 23–25, 2023, pp. 1–6. doi: 10.1109/ICCCI56745.2023.10128246.
- [3] Poornima, M., S. Bayanna, T. Manasa, Shaik Mahammad Arshu, V. Jayaprakash Reddy, and T. Haribabu. 2024. "Convolutional Neural Networks Based Paddy Leaf Disease Detection." *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2024.60546>
- [4] F. Zubair, M. Saleh, Y. Akbari, and S. Al Maadeed, "A robust ensemble model for plant disease detection using deep learning architectures," *AgriEngineering*, vol. 7, p. 159, 2025. doi: 10.3390/agriengineering7050159.
- [5] M. H. Tunio, J. P. Li, A. Ahmed, S. A. Shah, X. Zeng, U. A. Kashif, Y. Li, and A. Y. Imam, "RiceNet: A robust ensemble attention mechanism for automated rice plant disease classification," *Multimedia Tools and Applications*, vol. 84, pp. 48145–48173, 2025. doi: 10.1007/s11042-025-20979-9.

- [6] S. Kanakala and S. Ningappa, "Detection and classification of diseases in multi-crop leaves using LSTM and CNN models," *Journal of Innovative Image Processing*, vol. 7, no. 1, pp. 161–181, Mar. 2025. doi: 10.36548/jiip.2025.1.008.
- [7] T. Apleni, F. O. Isinkaye, and M. O. Olusanya, "Ensemble-based feature fusion for accurate plant disease classification using pre-trained models," *Scientific Reports*, vol. 15, p. 41925, 2025. doi: 10.1038/s41598-025-25927-.
- [8] D. Wadhwa and K. Malik, "Deep learning generalized hybrid models for multi-species crop disease classification with explainable insights," *Engineering Applications of Artificial Intelligence*, 2025. doi: 10.1016/j.engappai.2025.113317.
- [9] A. Ayyappan, A. Bala, *et al.*, "Rice plant disease detection using convolutional neural networks," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 50, 2025.
- [10] B. Yang, M. Li, F. Li, *et al.*, "A novel plant type, leaf disease and severity identification framework using CNN and transformer with multi-label method," *Scientific Reports*, vol. 14, p. 11664, 2024.
- [11] C. Pal, S. Karmakar, I. Mukherjee, *et al.*, "A lightweight and explainable CNN model for empowering plant disease diagnosis," *Scientific Reports*, vol. 15, p. 30720, 2025.
- [12] M. Iftikhar, I. A. Kandhro, N. Kausar, *et al.*, "Plant disease management: A fine-tuned enhanced CNN approach with mobile app integration for early detection and classification," *Artificial Intelligence Review*, vol. 57, p. 167, 2024.
- [13] K. J. Kavitha and K. P. Prasad, "CNN ensemble approach for early detection of sugarcane diseases – A comparison," *International Journal of Electronics and Telecommunications*, 2024.
- [14] Ishengoma, Farian S., and Neema N. Lyimo. "Ensemble model for grape leaf disease detection using CNN feature extractors and random forest classifier." *Heliyon* 10, no. 12 (2024).
- [15] UPL Ltd., "Paddy doctor: Paddy disease classification dataset," Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/datasets/uplab/paddy-disease-classification> Accessed : Mar. 10, 2024.
- [16] D. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *PlantVillage Dataset*, 2015. [Online]. Available: <https://www.kaggle.com/datasets/emmarex/plantdisease> Accessed : Mar. 10, 2024.
- [17] S K. Basak, "Banana tree diseases detection new & update dataset," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/shuvokumarbasak4004/banana-tree-disease-detection-new-and-update-dataset>. Accessed: June. 15, 2025.
- [18] Rani, Pushpa Athisaya Sakila, and N. Suresh Singh. "Paddy leaf symptom-based disease classification using deep CNN with ResNet-50." *International Journal of Advanced Science Computing and Engineering* 4, no. 2 (2022): 88-94..
- [19] Sahasranamam, V., T. Ramesh, R. Rajeswari, A. Karthikkumar, and D. Muthumanickam. "AI-driven paddy leaf disease classification and prediction using DenseNet-121." In *6th International Conference on Deep Learning, Artificial Intelligence and Robotics (ICDLAIR 2024)*, pp. 272-292. Atlantis Press, 2025.
- [20] Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." *Frontiers in plant science* 7 (2016): 1419.
- [21] Ferentinos, Konstantinos P. "Deep learning models for plant disease detection and diagnosis." *Computers and electronics in agriculture* 145 (2018): 311-318.
- [22] Picon, Artzai, Aitor Alvarez-Gila, Maximilian Seitz, Amaia Ortiz-Barredo, Jone Echazarra, and Alexander Johannes. "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild." *Computers and Electronics in Agriculture* 161 (2019): 280-290.
- [23] Too, Edna Chebet, Li Yujian, Sam Njuki, and Liu Yingchun. "A comparative study of fine-tuning deep learning models for plant disease identification." *Computers and Electronics in Agriculture* 161 (2019): 272-279.
- [24] Kamilaris, Andreas, and Francese X. Prenafeta-Boldú. "Deep learning in agriculture: A survey." *Computers and electronics in agriculture* 147 (2018): 70-90.
- [25] R. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299–315, 2017. doi: 10.1080/08839514.2017.1315516.
- [26] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, Art. no. 3289801, 2016. doi: 10.1155/2016/3289801.
- [27] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, "Identification of rice diseases using deep convolutional neural networks," *Neurocomputing*, vol. 267, pp. 378–384, 2017. doi: 10.1016/j.neucom.2017.06.023.
- [28] H. Durmuş, E. O. Güneş, and M. Kırıcı, "Disease detection on the leaves of the tomato plants by using deep learning," in *Proc. International Conference on Agro-Geoinformatics*, 2017, pp. 1–5. doi: 10.1109/Agro-Geoinformatics.2017.8047016.
- [29] J. Chen, D. Zhang, A. Nanekaran, and D. Li, "Detection of rice plant diseases based on deep transfer learning," *Journal of the Science of Food and Agriculture*, vol. 100, no. 7, pp. 3246–3256, 2020. doi: 10.1002/jsfa.10365.