IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

Plant Leaf Disease Detection Using Ensemble Learning and Explainable AI : A Review

Miss. Kale Shewta Ramesh[1] and Prof. Kulkarni S.V.[2]
PG Student, College of Engineering, Ambajogai, Beed, Maharashtra, India[1]
Professor, College of Engineering, Ambajogai, Beed, Maharashtra, India[2]

Abstract: *Plant diseases significantly affect agricultural productivity and food security worldwide. Recent advances in deep learning have enabled automated plant disease detection systems with high classification accuracy. Among these approaches, convolutional neural networks (CNNs), ensemble learning techniques, and explainable artificial intelligence (XAI) methods have emerged as promising solutions. This review presents a comprehensive analysis of recent developments in plant disease detection from 2022 to 2026, focusing on deep learning architectures, ensemble models, benchmark datasets, and explainability techniques. Various publicly available datasets, including PlantVillage, PlantDoc, Plant Pathology 2021, and PlantCLEF, are systematically compared. Furthermore, performance benchmarks of CNN-based models, Vision Transformers (ViTs), hybrid architectures, and ensemble frameworks are reviewed. Research gaps, challenges, and future research directions are identified to guide the development of reliable, interpretable, and field-deployable plant disease detection systems*

Keywords: Plant Disease Detection, Deep Learning, Convolutional Neural Network (CNN), Ensemble Learning, Explainable AI (XAI), Grad-CAM, Agricultural Intelligence, Image Classification

I. INTRODUCTION

Agriculture plays a fundamental role in ensuring global food security and supporting the livelihoods of billions of people worldwide. However, crop productivity is continuously threatened by various plant diseases caused by fungi, bacteria, viruses, and environmental stress factors. According to reports from international agricultural organizations, plant diseases account for significant annual crop losses, resulting in substantial economic damage and reduced food availability. Early and accurate identification of plant diseases is therefore essential for minimizing crop damage, improving agricultural productivity, and promoting sustainable farming practices.

Traditionally, plant disease diagnosis has relied on visual inspection by agricultural experts and plant pathologists. Although effective in certain situations, manual diagnosis is often time-consuming, subjective, labor-intensive, and difficult to scale for large agricultural fields. Furthermore, the accuracy of traditional methods depends heavily on expert knowledge and experience, making disease detection challenging in remote and resource-constrained regions. These limitations have motivated researchers to explore automated and intelligent approaches for plant disease identification.

Recent advancements in artificial intelligence (AI), machine learning (ML), and computer vision have significantly transformed agricultural monitoring systems. In particular, deep learning techniques have demonstrated remarkable success in image-based plant disease detection. Convolutional Neural Networks (CNNs) such as AlexNet, VGGNet, ResNet, DenseNet, EfficientNet, and MobileNet have shown excellent capabilities in extracting complex visual features from leaf images and achieving high classification accuracy. The availability of large-scale agricultural image datasets and advances in computational resources have further accelerated research in this area.

Despite their success, single-model deep learning approaches often face challenges when deployed in real-world agricultural environments. Variations in illumination, background complexity, leaf orientation, disease severity, and environmental conditions can significantly affect model performance. To address these limitations, researchers have increasingly adopted ensemble learning techniques that combine predictions from multiple models to improve classification accuracy, robustness, and generalization capability. Ensemble methods such as majority voting, weighted averaging, bagging, boosting, and stacking have demonstrated superior performance compared to individual classifiers across various plant disease datasets.

In parallel with improvements in predictive performance, the need for transparency and interpretability in deep learning systems has gained considerable attention. Most deep learning models operate as black-box systems, making it difficult for users to understand the reasoning behind their predictions. In agricultural applications, where decision-making directly influences crop management strategies, explainability is particularly important. Explainable Artificial Intelligence (XAI) techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), Grad-CAM++, Local Interpretable Model-Agnostic Explanations (LIME), and Shapley Additive Explanations (SHAP) have emerged as effective tools for visualizing and interpreting model decisions. These techniques help identify disease-affected regions in plant images, thereby increasing user trust and supporting practical deployment in agricultural environments.

Over the past few years, significant research efforts have focused on developing advanced plant disease detection systems using CNNs, Vision Transformers (ViTs), ensemble architectures, and explainable AI frameworks. However, the existing literature remains fragmented, with studies employing different datasets, evaluation metrics, model architectures, and experimental protocols. As a result, it is often difficult to comprehensively assess the relative strengths and limitations of various approaches and identify future research opportunities.

Motivated by these observations, this review paper presents a comprehensive survey of recent developments in intelligent and transparent plant disease detection. The review systematically examines publicly available benchmark datasets, state-of-the-art deep learning architectures, ensemble learning strategies, and explainable AI techniques reported between 2022 and 2026. A comparative analysis of recent studies is provided to highlight current trends, performance benchmarks, and technological advancements in the field. Furthermore, key research gaps, challenges, and future directions are identified to support the development of more accurate, robust, interpretable, and field-deployable plant disease detection systems.

The major contributions of this review are summarized as follows:

• A comprehensive review of publicly available plant disease datasets and evaluation benchmarks.
• A systematic comparison of deep learning architectures, including CNNs, Vision Transformers, and hybrid models.
• An analysis of ensemble learning techniques used to improve classification performance and robustness.
• A detailed review of explainable AI methods for enhancing model transparency and interpretability.
• Identification of current research gaps, challenges, and future research directions in intelligent plant disease detection.

The remainder of this paper is organized as follows. Section 2 describes the review methodology and literature selection process. Section 3 presents benchmark datasets used for plant disease detection. Section 4 reviews deep learning approaches, while Section 5 discusses ensemble learning techniques. Section 6 examines explainable AI methods. Section 7 provides a comparative analysis of recent studies. Section 8 identifies research gaps and challenges, followed by future research directions in Section 9. Finally, Section 10 concludes the review..

II. RELATED WORK

This review adopts a systematic approach to analyze recent developments in intelligent and transparent plant disease detection using deep learning, ensemble learning, and explainable artificial intelligence (XAI). The primary objective is to identify current research trends, evaluate the effectiveness of different computational techniques, compare benchmark datasets, and highlight existing challenges in the field. A structured review methodology was employed to ensure that the selected studies were relevant, reliable, and representative of the latest advancements in agricultural image analysis.

The literature survey was conducted using several widely recognized scientific databases, including IEEE Xplore, ScienceDirect, SpringerLink, Wiley Online Library, ACM Digital Library, MDPI, and Google

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

Scholar. These databases were selected because they contain a large collection of peer-reviewed journal articles and conference papers related to artificial intelligence, computer vision, machine learning, and smart agriculture. The search process focused on studies published between 2022 and 2026 to ensure that the review reflects the most recent technological developments and research contributions.

To retrieve relevant publications, a combination of keywords and search phrases was utilized. The major search terms included “Plant Disease Detection,” “Deep Learning in Agriculture,” “Convolutional Neural Networks,” “Vision Transformer,” “Ensemble Learning,” “Explainable Artificial Intelligence,” “Grad-CAM,” “Crop Disease Classification,” and “Agricultural Computer Vision.” These keywords were combined using Boolean operators such as AND and OR to refine the search results and improve the accuracy of literature retrieval. The search process initially produced a large number of publications covering various aspects of intelligent plant disease diagnosis.

After the initial search, a screening process was carried out to identify the most relevant studies. The titles and abstracts of the collected articles were carefully examined to remove duplicate records and publications unrelated to image-based plant disease detection. Only studies focusing on machine learning, deep learning, ensemble techniques, transformer-based architectures, or explainable AI methods were considered for further analysis. This screening process significantly reduced the number of articles while retaining high-quality and relevant publications.

To maintain consistency and quality, specific inclusion criteria were applied during the selection process. Studies published between 2022 and 2026, written in English, and appearing in peer-reviewed journals or conference proceedings were included in the review. In addition, selected studies were required to provide experimental results and report standard evaluation metrics such as accuracy, precision, recall, F1-score, or Area Under the Curve (AUC). Research articles focusing on image-based disease diagnosis of crops and plants were given priority.

Similarly, exclusion criteria were established to eliminate studies that did not align with the objectives of this review. Articles published before 2022, duplicate publications, editorials, book chapters, short communications, and studies lacking experimental validation were excluded. Research focusing solely on sensor-based monitoring systems, environmental analysis, or non-image-based disease detection techniques was also omitted from the review. This ensured that the analysis remained focused on computer vision and artificial intelligence approaches for plant disease identification.

Following the application of inclusion and exclusion criteria, the selected studies were thoroughly examined and categorized into major research groups. The first category includes conventional and advanced convolutional neural network architectures such as AlexNet, VGGNet, ResNet, DenseNet, EfficientNet, and MobileNet. The second category focuses on transformer-based approaches, including Vision Transformers (ViTs), Swin Transformers, and hybrid CNN-transformer models. The third category covers ensemble learning techniques that combine multiple models to improve classification accuracy and robustness. The fourth category reviews explainable artificial intelligence methods such as Grad-CAM, Grad-CAM++, LIME, and SHAP, which enhance the interpretability and transparency of deep learning systems.

To facilitate meaningful comparisons among different studies, several performance evaluation metrics were analyzed. Accuracy was considered the primary metric for assessing overall classification performance, while precision, recall, and F1-score were examined to evaluate model reliability and class-specific prediction quality. In addition, some studies reported computational complexity, inference time, and deployment feasibility on mobile or edge devices, which were also considered when comparing different approaches. These metrics provide a comprehensive understanding of both the effectiveness and practical applicability of modern plant disease detection systems.

The reviewed studies were further analyzed with respect to the datasets used for experimentation, model architectures, ensemble strategies, explainability techniques, and reported performance outcomes. Particular attention was given to benchmark datasets such as PlantVillage, PlantDoc, Plant Pathology, and PlantCLEF, as these datasets are widely adopted for evaluating plant disease detection algorithms. Comparative analysis of these datasets helps identify their strengths, limitations, and suitability for real-world agricultural applications.

Through this systematic review methodology, a comprehensive overview of recent advancements in plant disease detection is presented. The structured analysis enables the identification of research trends, technological gaps, and emerging opportunities in the field. Furthermore, it provides a foundation for understanding how deep learning, ensemble learning, and explainable artificial intelligence can be integrated to develop accurate, reliable, and transparent systems for next-generation precision agriculture.

A. Benchmark Datasets for Plant Disease Detection

The success of deep learning-based plant disease detection systems largely depends on the availability of high-quality and diverse datasets. Datasets serve as the foundation for training, validating, and testing machine learning models, enabling them to learn disease-specific characteristics from plant images. Over the past decade, several publicly available datasets have been developed to support research in agricultural image analysis. These datasets differ in terms of image quantity, disease categories, environmental conditions, image quality, and annotation standards. Consequently, selecting an appropriate dataset is crucial for developing robust and reliable plant disease detection models.

Among the available datasets, PlantVillage remains the most widely used benchmark dataset for plant disease classification research. The dataset contains more than 54,000 annotated images representing healthy and diseased leaves from multiple crop species. One of its major advantages is the availability of high-quality images captured under controlled laboratory conditions with uniform backgrounds. Due to its balanced class distribution and large number of samples, PlantVillage has been extensively used to evaluate the performance of convolutional neural networks, transfer learning models, and ensemble learning approaches. Many studies have reported classification accuracies exceeding 98% when trained and tested on this dataset. However, the controlled imaging conditions limit its ability to represent real-world agricultural environments.

To overcome the limitations of laboratory-based datasets, researchers introduced PlantDoc, a dataset containing images captured directly in field environments. Unlike PlantVillage, PlantDoc includes variations in illumination, complex backgrounds, occlusions, and different viewing angles. These challenges make disease classification considerably more difficult and provide a more realistic benchmark for evaluating model generalization capabilities. Although the dataset contains fewer images compared to PlantVillage, it is valuable for assessing the robustness of deep learning models under practical agricultural conditions.

Another important dataset is the Plant Pathology dataset, which was developed to support research on apple leaf disease classification. This dataset contains thousands of images representing various disease categories, including scab, rust, and multiple disease combinations. The images were collected under diverse environmental conditions and provide realistic disease symptoms encountered in orchards. The dataset has gained significant attention through international competitions and benchmarking challenges, encouraging the development of advanced machine learning and computer vision techniques for disease diagnosis.

In recent years, large-scale datasets such as PlantCLEF have emerged to address the need for greater diversity in plant image analysis. PlantCLEF contains a vast collection of plant images collected from different geographical regions, environmental conditions, and plant species. The dataset supports large-scale classification tasks and provides

ISSN 2581-9429 IJARSCT

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

opportunities for evaluating advanced architectures such as Vision Transformers and multimodal learning systems. Its extensive variability makes it suitable for studying the scalability and adaptability of intelligent plant disease detection models.

Besides these benchmark datasets, several crop-specific datasets have been introduced for diseases affecting crops such as rice, maize, tomato, grape, potato, and wheat. These datasets often focus on a limited number of disease categories but provide detailed annotations and high-resolution images. Crop-specific datasets are particularly useful for developing specialized disease diagnosis systems tailored to individual agricultural applications. However, the lack of standardization among these datasets often makes direct comparison between studies challenging.

A major challenge associated with existing datasets is the difference between laboratory and field conditions. While models trained on controlled datasets frequently achieve very high classification accuracy, their performance may decline significantly when applied to real-world environments. Variations in lighting conditions, leaf orientation, shadows, background clutter, disease severity, and image quality can affect model predictions. Consequently, there is an increasing need for diverse and representative datasets that accurately reflect practical agricultural scenarios.

Another important consideration is dataset imbalance. Many plant disease datasets contain unequal numbers of samples across disease categories, which can lead to biased model training and reduced performance for minority classes. Researchers have addressed this issue through data augmentation techniques such as image rotation, flipping, cropping, scaling, and brightness adjustment. These techniques increase dataset diversity and improve model generalization while reducing the risk of overfitting.

Table 1 presents a comparative summary of widely used benchmark datasets for plant disease detection. The comparison highlights key characteristics including the number of images, disease classes, environmental conditions, and major applications. Understanding the strengths and limitations of these datasets is essential for selecting appropriate benchmarks and developing reliable plant disease detection systems.

Table 1. Comparison of Benchmark Datasets for Plant Disease Detection

Dataset	Number of Images	Classes	Environment	Key Characteristics
PlantVillage	54,306	38	Controlled	Most widely used benchmark dataset
PlantDoc	2,598	13	Field Conditions	Real-world agricultural images
Plant Pathology 2021	18,632	12	Field Conditions	Apple disease classification
PlantCLEF	100,000+	Multiple	Diverse Conditions	Large-scale plant image collection
AI Challenger Crop Dataset	61,486	27	Field Conditions	Crop disease recognition
Rice Disease Dataset	5,932+	Multiple	Field Conditions	Rice disease diagnosis
Wheat Disease Dataset	4,000+	Multiple	Field Conditions	Wheat disease classification

The availability of diverse benchmark datasets has significantly accelerated research in intelligent plant disease detection. Nevertheless, challenges related to data quality, class imbalance, environmental variability, and annotation consistency continue to influence model performance. Future dataset development efforts should focus on collecting large-scale, diverse, and well-annotated field images that better represent real agricultural environments. Such datasets will play a critical role in advancing robust, interpretable, and deployable plant disease detection systems for precision agriculture.

III. PROPOSED METHODOLOGY

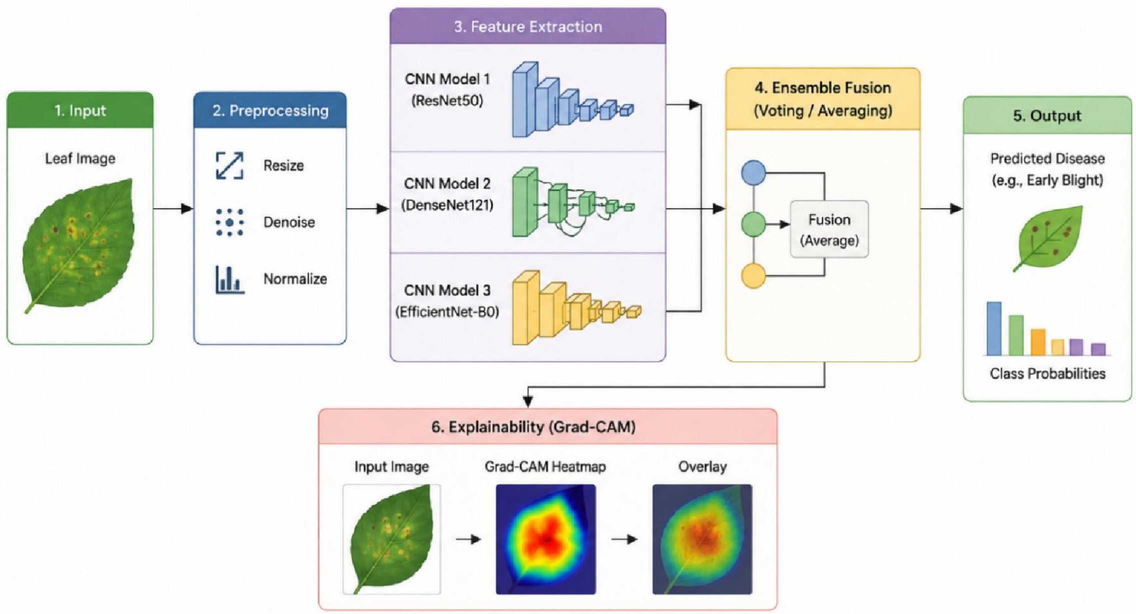

Fig.1 Overall System Architecture

3.1. Image Acquisition

Image acquisition is a critical first step in the proposed plant disease detection framework, as the system's performance largely depends on the quality and variety of the dataset. Leaf images are collected either from publicly available agricultural datasets or captured directly from real field conditions using standard cameras. The dataset includes both healthy and diseased leaves, covering different types of plant diseases and varying levels of severity.

To make the system more practical and reliable, the images are gathered under diverse environmental conditions. These include variations in lighting, background complexity, leaf orientation, and scale. Such diversity helps the model learn more robust and meaningful features, allowing it to perform effectively even in real-world agricultural settings where conditions are not controlled.

3.2. Preprocessing and Data Augmentation

Preprocessing plays an important role in preparing the input data and ensuring stable model training. In this stage, all images are resized to a fixed dimension, typically 224 × 224 pixels, so that they match the input requirements of pretrained convolutional neural networks. Pixel values are then normalized to a consistent range, which helps the model learn more efficiently and speeds up the training process.

To further improve performance and reduce the chances of overfitting, various data augmentation techniques are applied. These include random rotations, horizontal and vertical flips, zooming, brightness adjustments, and small geometric transformations. Such augmentations increase the diversity of the dataset without the need for additional data collection. As a result, the model becomes better at handling real-world variations in leaf appearance, lighting conditions, and orientation, leading to improved generalization.

3.3. Feature Extraction Using Multiple CNN Models

In the proposed framework, feature extraction is performed using multiple pretrained convolutional neural network architectures, allowing the system to capture a wide range of complementary features. Each CNN processes the input image through a series of convolutional layers, where initial layers focus on basic patterns such as edges, textures, and color variations. As the network goes deeper, it begins to identify more complex and disease-specific patterns that are important for accurate classification.

To manage the high dimensionality of these features, global pooling layers are used to reduce the data size while retaining the most important information. This is followed by fully connected layers that generate class probability outputs for each disease category. Since different CNN architectures learn features in different ways, combining them helps the system gain a more complete understanding of the input image.

This multi-model approach improves robustness and reduces the chances of incorrect predictions, especially in cases where diseases appear visually similar. By leveraging the strengths of multiple models, the framework is able to extract richer and more discriminative features, leading to more reliable plant disease detection.

3.4. Ensemble Decision Fusion

The ensemble decision fusion stage plays a key role in improving the reliability of the final classification. Instead of relying on a single model, the framework combines predictions from multiple CNNs to produce a more accurate result. This is done using a weighted averaging approach, where each model's contribution is based on how well it performs during validation.

By combining outputs in this way, the system reduces the chances of errors caused by any one model. It helps lower variance, minimizes overfitting, and leads to more stable predictions. Since each model captures different aspects of the input data, bringing them together allows the system to make better-informed decisions.

Overall, this fusion strategy improves the model's ability to generalize and reduces bias from individual networks. As a result, the final predictions are more consistent and reliable, especially when dealing with complex or visually similar plant disease patterns.

3.5. Explainability Using Grad-CAM

To ensure transparency and practical usability, the framework incorporates Gradient-weighted Class Activation Mapping as an explainability mechanism. Grad-CAM generates a visual heatmap by computing gradients of the predicted class with respect to the final convolutional feature maps. This heatmap highlights the regions of the leaf image that most strongly influence the model's decision. The generated visualization is superimposed on the original image, allowing users to verify whether the system focuses on disease-affected areas rather than irrelevant background regions. By providing visual justification for predictions, the explainability module enhances trust, interpretability, and acceptance of the model in real-world agricultural applications.

Mathematical Optimization of Ensemble Weights

In the proposed multi-model ensemble framework, optimal weight selection plays a crucial role in improving classification performance. Instead of assigning equal weights to all base classifiers, the ensemble weights are optimized to maximize predictive accuracy while minimizing classification loss.

Let there be Mbase CNN models. For a given input sample x, each model produces a probability vector:

where:

$$m = 1,2, \dots, M$$

$C$is the total number of disease classes

p_{mc}represents the predicted probability of class $c$by model m

The ensemble prediction is defined as a weighted linear combination:

$$P_{final}(x) = \sum_{m=1}^{M} w_m P_m(x)$$

subject to the constraints:

$$\sum_{m=1}^{M} w_m = 1, w_m \geq 0$$

The final predicted class is:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

$$\hat{y} = arg\ max\ _c\left(P_{final}^{(c)}(x)\right)$$

Objective Function Formulation

To determine optimal weights, the ensemble is trained to minimize the cross-entropy loss over a validation dataset containing Nsamples.

The cross-entropy loss is defined as:

$$L(w) = -\frac{1}{N}\sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic}log\left(\sum_{m=1}^{M} w_m p_{mc}(x_i)\right)$$

where:

$y_{ic}$is the ground-truth label indicator

$p_{mc}(x_i)$is the probability predicted by model $m$for class c

The optimization problem becomes:

$$min\ _w\ L(w)$$

subject to:

$$\sum_{m=1}^{M} w_m = 1, w_m \geq 0$$

Optimization Strategy

Since the objective function is differentiable, weights can be optimized using gradient-based optimization methods such as:

Projected Gradient Descent

Constrained Optimization using Lagrange Multipliers

Simplex-based optimization

Using Lagrange multipliers, the constrained objective becomes:

$$J(w,\lambda) = L(w) + \lambda(1)$$

Weights are iteratively updated as:

$$w_m^{(1)} = w_m^{(t)} - \eta\frac{\partial L}{\partial w_m}$$

where:

$\eta$is the learning rate

Projection ensures $w_m \geq 0$

Regularized Weight Optimization

To prevent dominance of a single model and encourage balanced contribution, a regularization term can be added:

$$L_{reg}(w) = L(w) + \lambda\sum_{m=1}^{M} w_m^2$$

This ensures smoother weight distribution and improves generalization.

Interpretation

The optimized weights reflect the relative reliability of each base CNN model. Models with higher validation performance receive larger weights, while weaker models contribute proportionally less. This adaptive weighting mechanism improves ensemble robustness, enhances classification accuracy, and reduces variance compared to uniform averaging.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

IV. CONCLUSION

This paper presents a multi-model ensemble framework designed to improve both the accuracy and transparency of plant disease detection using deep learning. Instead of relying on a single model, the approach combines multiple pretrained convolutional neural networks to capture different types of features from leaf images. This helps the system perform more reliably, especially under varying environmental conditions. By using a weighted decision fusion strategy, where each model contributes based on its strength, the framework reduces individual model bias and improves overall consistency in predictions.

Along with better performance, the proposed system also focuses on making the results easier to understand. An explainability component based on Grad-CAM is integrated into the framework to visually highlight the areas of the leaf that influence the model's decision. This ensures that the model is actually focusing on disease-affected regions rather than irrelevant background details, which increases user confidence and makes the system more practical for real-world agricultural use.

The combination of ensemble learning and explainable AI addresses two important challenges in plant disease detection: achieving reliable predictions and making those predictions interpretable. Experimental results show that the ensemble model performs better than individual models across key evaluation metrics such as accuracy, precision, recall, and F1-score. Additionally, the use of optimized weights allows the system to effectively leverage the strengths of each model, further improving classification performance.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, pp. 1–10, 2016.

[2] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.

[3] M. Too, L. Yujian, S. Njuki, and Y. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2012, pp. 1097–1105.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[7] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[8] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[9] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. International Workshop on Multiple Classifier Systems*, 2000, pp. 1–15.

[10] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE International Conf. Computer Vision (ICCV)*, 2017, pp. 618–626.

[11] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. ACM SIGKDD International Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[12] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774.

[13] P. Rajpurkar et al., "Deep learning for medical image analysis: Challenges and applications," *Nature Medicine*, vol. 25, pp. 24–29, 2019.

[14] J. Brownlee, *Ensemble Learning Techniques in Machine Learning*. Machine Learning Mastery, 2018.

[15] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.