International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 5, May 2025)

# Review of Machine Learning Approaches for Plant Leaf Disease Detection

[1]Jyoti Kumar, [2]Shivraj Singh

[1]*M.Tech Scholar, Department of Electronics and Communication, Sagar Institute of Research & Technology, Bhopal, India*
[2]*Assistant Professor, Department of Electronics and Communication, Sagar Institute of Research & Technology, Bhopal, India*

***Abstract*— The early and accurate detection of plant leaf diseases plays a pivotal role in ensuring global agricultural productivity and food security. Traditional methods of disease identification are labor-intensive, time-consuming, and highly dependent on expert knowledge. Recent advancements in machine learning (ML) have introduced automated, efficient, and scalable solutions for identifying plant diseases through image-based analysis of leaves. This review systematically explores various machine learning techniques—including Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), and deep learning methods such as Convolutional Neural Networks (CNNs)—used for plant leaf disease detection. The study examines each method's performance, datasets, pre-processing techniques, and challenges, offering a comparative analysis that highlights current trends, limitations, and future research directions.**

***Keywords*— ML, Alzheimer's Disease, Supervised, Prediction.**

## I. INTRODUCTION

Agriculture remains a cornerstone of the global economy, providing sustenance, raw materials, and employment to billions worldwide. However, the sector continually faces threats from various sources, among which plant diseases are particularly detrimental [1]. These diseases often begin with subtle changes in the appearance of leaves, which, if undetected or misdiagnosed, can lead to substantial crop losses, economic setbacks, and threats to food security. Traditionally, plant disease diagnosis has relied on manual inspection by experts, a method fraught with challenges such as subjectivity, high labor costs, and inconsistent accuracy, particularly in large-scale or resource-constrained farming environments [2].

In recent years, the integration of artificial intelligence, particularly machine learning, into the agricultural domain has revolutionized disease detection processes. Machine learning offers a promising alternative to traditional approaches by enabling the automatic analysis and classification of plant leaf images to detect specific diseases [3]. These methods learn from large datasets, identifying patterns and features that may be imperceptible to the human eye. Among the most widely used ML techniques are Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbors (KNN), and Decision Trees (DT), each bringing its strengths in classification and prediction tasks. More recently, deep learning models such as Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in handling high-dimensional image data, further pushing the boundaries of precision and scalability in plant disease detection [4] [5].

This review aims to provide a comprehensive overview of various machine learning approaches employed in plant leaf disease detection. It highlights the critical steps involved in building ML-based systems, including image acquisition, pre-processing, feature extraction, model training, and evaluation. The review also delves into the different datasets used in training and validating these models, analyzing their impact on model performance [6]. Furthermore, it discusses the limitations and challenges faced in real-world deployment—such as environmental variability, image noise, and dataset imbalance—while suggesting potential solutions and future research directions. By consolidating existing knowledge, this review seeks to guide researchers, agronomists, and technologists in designing robust, accurate, and practical disease detection systems that can significantly enhance agricultural resilience and productivity [7][8].

## II. LITERATURE SURVEY

A. Oad et al. (2024) introduces an AI (Artificial Intelligence) model that detects and explains plant diseases through image analysis. The proposed system, distinct from existing detectors, identifies numerous diseases in vegetables and fruits by employing our proposed ensemble learning classifier involving four deep learning models: VGG16, VGG19, ResNet101 V2, and Inception V3, achieving an accuracy exceeding 90%. The reason for using ensemble learning is to obtain accurate predictions. Furthermore, the system sets itself apart by providing explanations for predictions using LIME (Local Interpretable Model-Agnostic Explanations), applied to interpret the predictions of deep learning models [1].

PlantDoc is an online tool that is powered by artificial intelligence (AI) that use convolutional neural networks (CNNs) with the goal of identifying illnesses that affect plants. The invention was made by Khan and Srivastava in the year 2023. Image analysis of plant leaves is performed by the system in order to identify potential issues and provide suggestions for appropriate solutions. Rapid action is made possible as a result, which contributes to the prevention of the spread of illnesses [2].

In addition to this, Nagaraju et al. (2022) proposed a CNN-based model that was enhanced by the use of image augmentation techniques in order to improve the detection of leaf diseases. The study offered two methods, namely the Image Preprocessing and Transformation Algorithm (IPTA) and the Image Masking and REC-based Hybrid Segmentation Algorithm (IMHSA), with the intention of tackling the problems of overfitting and classification. Both of these algorithms were developed in order to solve the respective concerns. As a direct result of this, the performance of the model was successfully improved [3].

In order to accomplish the task of feature computation, Ahmed et al. (2021) presented a method that made use of the Gray-Level Co-occurrence Matrix (GLCM). This method combines the color and texture information that was obtained from photographs of leaves. A one-versus-one support vector machine (SVM) classifier that achieved 98.79% accuracy as determined by 10-fold cross-validation demonstrated the efficacy of feature-based approaches in the detection of plant diseases. This was shown by the fact that the classifier was found to be effective [4].

The Plant Pathology 2020 dataset was given by Thapa and colleagues in their paper from the year 2020. The dataset consisted of 3,651 high-quality photographs of apple leaves that had been affected by a variety of diseases. Using this dataset, a CNN model was able to achieve an accuracy of 97% in identifying illnesses. As a result, this dataset is an extremely significant resource for the creation and testing of algorithms that detect plant diseases [5].

In order to see the identification of plant illnesses, Shruthi et al. (2019) conducted a comparative analysis of machine learning techniques. These approaches included support vector machines (SVM), random forests (RF), and stochastic gradient descent (SGD). The end goal of this study was to observe the identification of plant diseases. As a result of the results of the study, support vector machines (SVM) performed much better than other classifiers. This highlights the relevance of selecting algorithms that are appropriate for certain applications [6].

Singh and Kaur (2018) created a method that combined region-based segmentation with K-Nearest Neighbor (KNN) classification in order to detect plant illnesses. This method was designed for the goal of identifying plant diseases. It was possible to effectively locate ill patches in leaf photos as a consequence of the method, which brought to light the prospect of combining segmentation techniques with machine learning classifiers [7].

In the year 2017, Mohanty and his colleagues conducted research on the use of deep learning, and more specifically convolutional neural networks (CNNs), for the purpose of diagnosing plant illnesses upon the basis of photographs. By reaching good accuracy across 14 crop species and 26 diseases, the study established the effectiveness of deep learning models in the management of varied plant disease datasets. This was accomplished by demonstrating the longevity of these models [8].

Through the use of edge detection techniques, Revathi and Hemalatha (2016) were able to successfully classify the illnesses that cause cotton leaf spots. The relevance of conventional image analysis methods in the area of plant pathology was highlighted by the fact that the study was able to achieve a highly accurate categorization of diseases via the use of image processing techniques [9].

In 2015, Pujari and his colleagues proposed a method for classifying plant illnesses that took into consideration both the color and the textural aspects of the disease. Through the use

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 5, May 2025)

of support vector machine (SVM) classifiers, the method demonstrated a high degree of accuracy in the diagnosis of diseases. The importance of feature selection in machine learning models for plant pathology is brought into focus by this [10].

## III. CHALLENGES

Challenges in Machine Learning Approaches for Plant Leaf Disease Detection-

1. **Limited and Imbalanced Datasets** Many ML models require large and diverse datasets to generalize effectively. However, publicly available plant disease datasets are often limited in size, biased towards common diseases, or imbalanced in class distribution (more healthy leaves than diseased ones). This can lead to overfitting and poor model generalization to real-world scenarios.
2. **Variability in Leaf Appearance** Leaves of the same species can appear drastically different due to age, lighting, orientation, seasonal changes, or environmental conditions. Disease symptoms can also manifest differently on different parts of the leaf or in different stages. These variations increase intra-class variability, making disease classification harder for ML algorithms.
3. **Complex Backgrounds and Noise** In practical agricultural settings, leaf images are often captured in uncontrolled environments with cluttered backgrounds (soil, other plants, tools, shadows, etc.). This background noise can confuse the model, especially if segmentation is not used to isolate the leaf region.
4. **Similarity Between Diseases** Some plant diseases exhibit visually similar symptoms (e.g., spots, discoloration, or mold growth), making it difficult even for expert human observers to distinguish them accurately. ML models may misclassify such diseases unless fine-grained features are extracted.
5. **Lack of Field-Level Validation** Most ML models are trained and tested in lab settings using curated datasets. When deployed in the field, these models often experience significant accuracy drops due to varying lighting, camera quality, or plant variety. Lack of robustness in real-world conditions remains a major hurdle.
6. **Need for High-Quality Image Data** Many advanced models, especially deep learning approaches, rely heavily on high-resolution and well-annotated image data. Low-quality images due to poor camera sensors or inadequate lighting can degrade the model's performance significantly.
7. **High Computational Requirements** Deep learning models such as CNNs require high computational power and GPU resources for training and inference. This poses a challenge for deploying such models on mobile devices or low-cost edge devices used by farmers in remote areas.
8. **Annotation and Labeling Challenges** Accurate labeling of plant diseases requires expert knowledge from agronomists or plant pathologists. Manual annotation is time-consuming and prone to human error, which can propagate errors into model predictions and evaluations.

## IV. PROPOSED STRATEGY

To address the prevailing limitations in current machine learning-based plant leaf disease detection systems, this study proposes a comprehensive and scalable strategy that integrates advanced data processing, robust model selection, and practical deployment mechanisms. The approach focuses on enhancing detection accuracy, model generalization, and field-level applicability by combining hybrid techniques with domain knowledge.

### 1. Data Acquisition and Preprocessing

The first step involves curating a large-scale, diverse dataset that includes images of healthy and diseased leaves from multiple crops, under various lighting, angle, and background conditions. To tackle noise and inconsistency in raw data, preprocessing techniques such as Gaussian filtering, histogram equalization, and leaf segmentation using color thresholding or semantic segmentation (e.g., U-Net) will be applied. Data augmentation methods like rotation, flipping, scaling, and brightness adjustment will further enhance the robustness of the training set.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 5, May 2025)

### 2. Feature Extraction and Hybrid Modeling

Instead of relying solely on raw image pixels, the proposed strategy emphasizes hybrid feature extraction. Handcrafted features (e.g., color histograms, texture descriptors like GLCM, and shape features) will be combined with deep features from pre-trained CNN models like ResNet50 or EfficientNet. This hybrid approach leverages both domain-specific features and deep learned representations, thus improving classification performance on complex disease patterns.

### 3. Model Selection and Ensemble Learning

For classification, the proposed strategy employs an ensemble of machine learning and deep learning models. Techniques such as Random Forest, SVM, and XGBoost will be combined with deep CNN architectures using voting or stacking methods. This ensemble framework aims to reduce individual model bias and variance, improving overall prediction accuracy. Attention mechanisms and transfer learning will be incorporated to further enhance deep learning models on limited annotated data.

### 4. Class Imbalance Handling

To address dataset imbalance, synthetic data generation techniques like SMOTE (Synthetic Minority Oversampling Technique) or GANs (Generative Adversarial Networks) will be used to generate new samples for under-represented disease classes. Additionally, cost-sensitive learning will be implemented to penalize the misclassification of rare disease categories more heavily during model training.

### 5. Model Generalization and Cross-Crop Transferability

To enhance the generalization ability of the model across different plant species and regions, domain adaptation techniques and cross-crop training will be applied. A portion of the dataset from various crops will be used to fine-tune the base models, enabling transfer learning and multi-crop disease detection capabilities.

### 6. User-Friendly Interface and Decision Support System

A farmer-friendly mobile application interface will be developed, incorporating real-time image capture, disease classification, and treatment recommendation based on agronomic guidelines. The app will also include a feedback system to allow users to correct model errors, which will help retrain and improve the model over time.

## V. CONCLUSION

The review of machine learning approaches for plant leaf disease detection highlights significant advancements in leveraging image processing, deep learning, and hybrid modeling techniques to automate and enhance disease diagnosis in agriculture. Despite promising results, current systems face challenges related to data quality, generalization, and real-world deployment. The proposed strategy addresses these issues through a comprehensive framework that integrates advanced preprocessing, ensemble learning, mobile deployment, and continuous learning, aiming to create a robust, scalable, and farmer-friendly solution. Such intelligent systems have the potential to revolutionize precision agriculture, reduce crop losses, and support sustainable farming practices.

## REFERENCES

1. A. Oad et al., "Plant Leaf Disease Detection Using Ensemble Learning and Explainable AI," in IEEE Access, vol. 12, pp. 156038-156049, 2024, doi: 10.1109/ACCESS.2024.3484574.
2. A. Khan and A. Srivastava, "PlantDoc-Plant Disease Detection using AI," Journal of Informatics Electrical and Electronics Engineering, vol. 4, no. 1, pp. 1–10, 2023.
3. M. Nagaraju et al., "Convolution network model based leaf disease detection using augmentation techniques," Expert Systems, vol. 39, no. 4, 2022.
4. N. Ahmed, H. M. S. Asif, and G. Saleem, "Leaf Image-based Plant Disease Identification using Color and Texture Features," arXiv preprint arXiv:2102.04515, 2021.
5. R. Thapa et al., "The Plant Pathology 2020 challenge dataset to classify foliar disease of apples," arXiv preprint arXiv:2004.11958, 2020.
6. U. Shruthi, V. Nagaveni, and B. K. Raghavendra, "A review on machine learning classification techniques for plant disease detection," in 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 281–284.
7. J. Singh and H. Kaur, "Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier," in International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB), 2018, pp. 1667–1675.

8. S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.

9. P. Revathi and M. Hemalatha, "Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques," in IEEE International Conference on Emerging Trends in Science, Engineering and Technology, 2012, pp. 169–173.

10. J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, "Image processing-based detection of fungal diseases in plants," Procedia Computer Science, vol. 46, pp. 1802–1808, 2015.