A Comprehensive Review of Machine Learning and Deep Learning Techniques for Crop Disease Detection

Aman Goyal
B.Tech Student, Department of CSE, Global Institute of Technology, Jaipur
23egjcs017@gitjaipur.com

Dr. Manju Mathur
Associate Professor, Department of CSE, Global Institute of Technology, Jaipur
manju.mathur@gitjaipur.com

Amit Bohra
Assistant Professor, Department of CSE, Global Institute of Technology, Jaipur
amit.bohra@gitjaipur.com

ABSTRACT: Crop diseases pose a serious threat to global agricultural productivity, resulting in significant yield losses and food security concerns. Conventional disease identification methods are often time-consuming, labor-intensive, and highly dependent on expert knowledge, making them unsuitable for large-scale monitoring. Recent advances in artificial intelligence have enabled machine learning (ML) and deep learning (DL) techniques to provide automated, accurate, and efficient solutions for crop disease detection using plant leaf images and sensor-based data.

This paper presents a comprehensive review of ML-based approaches for crop disease detection, covering the standard processing pipeline, including image acquisition, preprocessing, segmentation, feature extraction, and classification. Popular benchmark datasets such as PlantVillage, along with emerging real-field datasets, are critically examined. The performance of traditional ML classifiers, including Support Vector Machines (SVM) and Random Forests, is analyzed alongside advanced deep learning architectures such as Convolutional Neural Networks (CNNs), ResNet, and MobileNet. Evaluation metrics such as accuracy, F1-score, Intersection over Union (IoU), and confusion matrices are discussed to assess model effectiveness and robustness. The review reveals that although deep learning models achieve high accuracy on benchmark datasets, their real-world deployment is constrained by environmental variability, limited labeled data, and generalization issues. The study highlights future research directions toward developing scalable, interpretable, and field-deployable solutions for precision agriculture and sustainable farming.

KEYWORDS: Crop Disease Detection, Machine Learning, Deep Learning, Precision Agriculture, Convolutional Neural Networks (CNN), Image Processing.

1. INTRODUCTION

Agriculture remains the backbone of the global economy, employing a significant portion of the world's population and contributing heavily to food production and national GDPs, particularly in developing countries [1]. Despite its importance, agricultural productivity is severely affected by crop diseases, which account for an estimated 20–40% loss in yield every year. These losses not only impact farmers economically but also threaten global food security. Traditional approaches to plant disease diagnosis rely on expert knowledge and laboratory-based analysis. However, such methods are often slow, expensive, and inconsistent due to human subjectivity [2]-[4].

To address these challenges, machine learning has emerged as a transformative solution providing automated, accurate, and early detection of crop diseases. By analyzing visual symptoms through captured images or environmental data, machine learning models can classify diseases rapidly and with high precision [5], [6]. The availability of advanced imaging devices, including smartphones, IoT sensors, and drones, has further accelerated the adoption of intelligent agricultural monitoring systems. As a result, ML-based disease detection plays a crucial role in advancing precision agriculture and supporting farmers with timely decisions, reduced chemical usage and improved crop health [7]-[10].

Table 1: Datasets Used in Crop Disease Detection

Dataset Name	Description	Total Classes	Source
PlantVillage	Leaf images of 38 plant-disease categories	>54,000	Open dataset
PlantDoc	Real-field images	27	Kaggle
AI Challenger	Crop disease + pests	10,000+	Competition dataset
CropDoc	Rice crop field images	5000+	University-developed datasets

Table 2: Machine Learning Techniques Used in Crop Disease Detection

Technique	Pros	Cons
SVM	High performance on small datasets	Requires feature engineering
k-NN	Simple and interpretable	Low speed for large datasets
Random Forest	Handles non-linear data	Risk of overfitting
ANN	Learns complex features	Limited with small datasets

Table 3: Deep Learning Techniques Used in Crop Disease Detection

Model	Key Features	Performance
AlexNet	First widely used CNN for image detection	Good baseline
VGG16/VGG19	Deeper architecture, better feature extraction	Slow inference
GoogleNet/Inception	Efficient, handles multi-scale features	High accuracy
ResNet	Skip connections to reduce vanishing gradient	State-of-the-art detection

MobileNet	Designed for mobile deployment	Low computational cost

2. MACHINE LEARNING WORKFLOW FOR CROP DISEASE DETECTION

A typical machine learning system for crop disease detection follows a structured workflow beginning with image acquisition. Images of plants are captured in real-world farm conditions using mobile devices, cameras mounted on unmanned aerial vehicles, or collected from laboratory-based datasets such as PlantVillage. These images often vary in quality due to factors such as lighting conditions, background clutter, and leaf orientation, requiring robust processing methods.

The next step is preprocessing, where the acquired images undergo enhancement to improve clarity and usability. This involves techniques such as noise removal, resizing, contrast adjustment, and background normalization. By reducing unwanted variations in the images, preprocessing ensures reliable feature extraction.

Segmentation is then applied to isolate the infected portion of the leaf from the healthy region. Traditional segmentation approaches, including thresholding, clustering, and contour-based methods, are frequently used. Recently, deep learning-based segmentation architectures, such as U-Net and Mask R-CNN, have demonstrated strong performance in accurately separating disease-affected areas even under complex visual conditions.

Following segmentation, the processed images undergo feature extraction. In this stage, visual attributes such as color, texture, and shape are translated into measurable patterns. Classical feature extraction techniques like Gray Level Co-occurrence Matrix and Local Binary Patterns help in describing the texture and lesion characteristics, while deep neural networks automatically learn rich and discriminative features directly from image data.

Finally, classification is performed to identify the disease category. Traditional machine learning classifiers, including Support Vector Machines, Random Forests, and K-Nearest Neighbor, have been widely used for this purpose. However, deep learning has become the dominant methodology due to the ability of convolutional neural networks to model complex patterns and achieve higher accuracy. Transfer learning techniques, which reuse pre-trained models, further enhance performance, especially when training data is limited.

3. PERFORMANCE EVALUATION METRICS

Evaluating model performance is a crucial step in assessing the effectiveness of machine learning techniques for crop disease detection. Several statistical measures are used to analyze the accuracy, reliability, and robustness of the system.

- **Accuracy:** Accuracy is the most commonly used metric, representing the ratio of correctly predicted samples to the total number of samples. It indicates the model's overall correctness in classification. However, accuracy may be misleading in imbalanced datasets where healthy leaves outnumber diseased ones.

- **Precision:** Precision measures how many of the samples predicted as diseased are truly diseased. It focuses on reducing false positives. High precision ensures the system does not mistakenly classify healthy crops as infected, which avoids unnecessary treatment costs.
- **Recall:** Recall, also known as sensitivity or true positive rate, evaluates the model's ability to correctly identify actual disease cases. It emphasizes minimizing false negatives. High recall is vital in agriculture to prevent missed detections that could lead to rapid disease spread.
- **F1-Score:** F1-Score provides a balanced performance metric by combining precision and recall into a single harmonic mean. It is especially useful when dealing with uneven class distribution and offers better insight than accuracy alone.
- **Confusion Matrix:** A confusion matrix displays prediction results in a tabular format, showing correct and incorrect classifications for each disease category. It helps researchers identify which diseases are more prone to misclassification and guides improvements in the model.
- **Intersection over Union (IoU):** IoU is used particularly for segmentation tasks, where the objective is to locate the diseased part of the leaf. It measures the overlap between the predicted infected region and the actual ground-truth region. Higher IoU values indicate more accurate spatial detection.

Although many models exhibit high performance using these metrics on laboratory-controlled datasets, the same accuracy may not be achieved under real agricultural conditions. Variations in lighting, background noise, leaf orientation, and image quality often reduce model effectiveness. Hence, field testing and dataset diversity are essential for reliable deployment of crop disease detection systems.

5. CONCLUSION

Crop disease detection using machine learning represents a significant step toward achieving sustainable and intelligent agriculture. The transition from manual diagnosis to image-based automated detection has demonstrated remarkable improvements in speed, accuracy, and accessibility for farmers. Traditional ML models, when combined with handcrafted feature extraction, provide reliable outcomes for controlled scenarios, while deep learning models particularly convolutional neural networks have shown superior performance by directly learning distinguishing features from raw images.

However, despite advancements, challenges remain in applying these models to real farming conditions. Factors such as inconsistent lighting, background complexity, pest presence, and limited availability of annotated real-field datasets impact model generalization and scalability. Additionally, high computational demands restrict deployment on low-power agricultural devices.

To overcome these limitations, future research should focus on developing lightweight architectures, incorporating multimodal data (e.g., IoT sensor inputs), enhancing dataset diversity, and strengthening model interpretability. Integrating ML tools with precision agriculture platforms and mobile applications can support early intervention, reduce pesticide usage, and promote healthier crop production.

Thus, machine learning-driven crop disease detection holds transformative potential for global food security, and continued innovation will enable widespread adoption in practical agricultural environments.

REFERENCES

[1] P. Jha, D. Dembla, and W. Dubey, "Comparative analysis of crop diseases detection using machine learning algorithm," in Proceedings of the 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 569–574, 2023.

[2] RAKSHA and M. Mathur, "Analysis of various plant disease detection techniques using KNN classifier," International Journal of Computer Science and Mobile Computing, vol. 8, no. 7, pp. 65–70, 2019.

[3] P. Jha, D. Dembla, and W. Dubey, "Deep learning models for enhancing potato leaf disease prediction: Implementation of transfer learning based stacking ensemble model," Multimedia Tools and Applications, vol. 83, pp. 37839–37858, 2024.

[4] P. Jha, D. Dembla, and W. Dubey, "Crop disease detection and classification using deep learning-based classifier algorithm," in Emerging Trends in Expert Applications and Security (ICETEAS 2023), Lecture Notes in Networks and Systems, vol. 682, pp. 1–12 (online pages may vary), 2023.

[5] P. Jha, D. Dembla, and W. Dubey, "Implementation of machine learning classification algorithm based on ensemble learning for detection of vegetable crops disease," International Journal of Advanced Computer Science and Applications, vol. 15, no. 1, pp. 1–9, 2024.

[6] P. Jha, D. Dembla, and W. Dubey, "Implementation of transfer learning based ensemble model using image processing for detection of potato and bell pepper leaf diseases," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, pp. 69–80, 2024.

[7] G. Sharma, N. Hemrajani, S. Sharma, A. Upadhyay, Y. Bhardwaj, and A. Kumar, "Data management framework for IoT edge-cloud architecture for resource-constrained IoT application," Journal of Discrete Mathematical Sciences and Cryptography, vol. 25, no. 4, pp. 1093–1103, 2022.

[8] S. P. Chaturvedi, A. Yadav, A. Kumar, and R. Mukherjee, "Unlocking IoT security: Enabling the future with lightweight cryptographic ciphers," in Intelligent Computing Techniques for Smart Energy Systems (ICTSES 2023), Lecture Notes in Electrical Engineering, vol. 1277, pp. 189–199, 2025.

[9] A. Gautam, R. Ajmera, D. K. Dharamdasani, S. Srivastava, and A. Johari, "Improving climate change predictions using time series analysis and deep learning," Global and Stochastic Analysis, vol. 12, no. 4, Jul. 2025.

[10] 5. H. Sharma and R. Ajmera, "Comprehensive review and analysis of elderly fall detection system using machine learning," Tuijin Jishu/Journal of Propulsion Technology, vol. 44, no. 5, 2023.

[11] M. Kumar, R. Ajmera, and D. K., “Statistical analysis and accuracy assessment of improved machine learning based opinion mining framework,” Advances in Nonlinear Variational Inequalities, vol. 27, no. 1, 2024.

[12] M. Mathur and R. Jain, “A comparative analysis of deep learning algorithms for fruit disease classification,” Journal of Electrical Systems, vol. 20, no. 7 (Special Issue), ISSN: 1112-5209, 2024.

[13] M. Mathur and R. Jain, “Detection of fruit diseases with hybrid DWT–GLCM approach,” European Chemical Bulletin, vol. 12, Special Issue 7, pp. 613–624, ISSN: 2063-5346, 2023.

[14] M. Mathur and R. Jain, “Fruit detection using machine learning: A review,” Journal of Harbin Institute of Technology, vol. 54, no. 9, pp. 24–31, 2022.

[15] M. Mathur, “Analysis of various plant disease detection techniques using KNN classifier,” International Journal of Computer Science and Mobile Computing, vol. 8, no. 7, pp. 65–70, 2019.