*Grenze International Journal of Engineering and Technology, June Issue*

# Disease Detection for Crop Yield using Machine Learning

Sangamitra B K[1], B Nethravathi[2] and P. Marimuthu[3]
[1]Department of ISE, JSS Academy of Technical Education, Bengaluru, Karnataka, India
Email: sangamitra55555@gmail.com
[2]Department of CSE JSS Academy of Technical Education, Bengaluru, Karnataka, India
Email: nethravathi.sai@gmail.com
[3]Department of CSE, School of Mechanical Engineering, SASTRA Deemed University, Thanjavur, India.
Email: marimuthu@mech.sastra.edu

***Abstract*— Agriculture forms the backbone of the Indian economy, with nearly 60% of the land dedicated to farming. One of the critical challenges faced by this sector is crop disease, which significantly reduces yield and quality. This study presents a machine learning-based framework for early detection and diagnosis of crop diseases using image processing and remote sensing techniques. The system combines image partitioning, attribute extraction, and categorization using techniques like K-Means clustering and Support Vector Machines (SVM). It aims to aid farmers with precise, timely information to reduce crop losses and improve yield sustainably.**

***Index Terms*— Disease detection, Crop yield data analysis, Image processing, Remote sensing techniques, Machine learning (ML).**

## I. INTRODUCTION

India's agricultural productivity is threatened by numerous biotic and abiotic stressors, among which plant diseases contribute to substantial annual losses. Traditional methods of disease identification rely on expert observation, which is time-consuming and often inaccessible to rural farmers. Recent advancements in ML and image processing offer powerful tools to automate this process, enabling accurate, real-time disease prediction using leaf images and spectral data. Machine learning has emerged as a pivotal technology in precision agriculture. Its ability to recognize complex patterns in data allows it to detect morphological changes in leaves, including lesions and color changes indicative of disease. These insights can empower farmers with preventive measures, reduce dependence on harmful pesticides, and contribute to the production of organic crops.

## II. LITERATURE REVIEW

Recent advancements in ML and Deep Learning (DL) have revolutionized crop disease detection by enabling automated, fast, and accurate diagnosis from image and sensor data. Convolutional Neural Networks (CNNs) are the most widely used models, effectively identifying crop diseases with high precision and adaptability across datasets [1,4,7]. Techniques such as transfer learning and lightweight CNN architectures (e.g., MobileNet, EfficientNet-lite) have further improved efficiency for real-time and mobile-based applications [14].

*Grenze ID: 01.GIJET.12.2.116_21*

Emerging Vision Transformer (ViT) models enhance feature representation and robustness to complex field environments. Meanwhile, classical ML methods—such as SVMs and K-means clustering—combined with preprocessing methods like color-space conversion, median filtering, and segmentation continue to yield strong performance on smaller or specific datasets [8–10]. Hybrid systems integrating deep and classical ML models have also proven effective when labeled data are limited [8,9].Beyond image-based detection, spectral and sensor-based approaches capture biochemical and physiological crop variations before visible symptoms occur [1,16,18]. Integration with IoT, edge computing, and UAV (drone)-based systems enables large-scale, real-time monitoring and precision agriculture applications [14,17,19,20]. Studies emphasize that combining image analytics, sensor fusion, and AI-driven decision support enhances scalability and early disease prediction [2,3,11–13]. Despite notable progress, challenges remain in dataset variability, model generalization, and explainability. Future directions include self-supervised learning, multimodal fusion, and explainable AI to improve adaptability and trust in automated crop health monitoring systems [1,2,17].

## III. Research Methodology

The proposed approach integrates image processing and ML techniques to detect and classify crop diseases through nine key stages (A–I), providing an end-to-end pipeline from data acquisition to actionable recommendations.

### *A. Data Collection*

High-resolution leaf images are collected from publicly available datasets such as Plant-village or captured directly in the field using cameras and drones. Expert agronomists verify ground truth labels, including disease type and plant health status. Additionally, spectral data, such as Normalized Difference Vegetation Index (NDVI) or infrared imagery, may be incorporated to capture early, non-visible physiological changes in crops.

### *B. Image Preprocessing*

Collected images undergo pre processing to enhance quality and facilitate accurate segmentation. Images are re sized (e.g., 256×256 pixels), denoised using Median or Gaussian filters, and enhanced using histogram equalization. Color space conversions to HSI or LAB improve contrast between healthy and diseased regions, aiding feature extraction.

### *C. Image Segmentation*

Segmentation separates infected regions from healthy leaf areas. K-Means clustering groups pixels based on similarity, while Otsu's thresholding provides automatic binarization. Edge detection algorithms like Sobel or Canny further refine lesion boundaries for precise localization.

### *D. Feature Extraction*

Quantitative features are extracted from segmented regions to describe disease characteristics. These include texture features from Gray-Level Co-occurrence Matrix (GLCM) such as contrast and homogeneity, color statistics (mean and standard deviation in RGB, HSV, and LAB spaces), shape metrics (area, perimeter, eccentricity), and histogram-based measures, forming a comprehensive feature vector for model input.

### *E. Model Training and Classification*

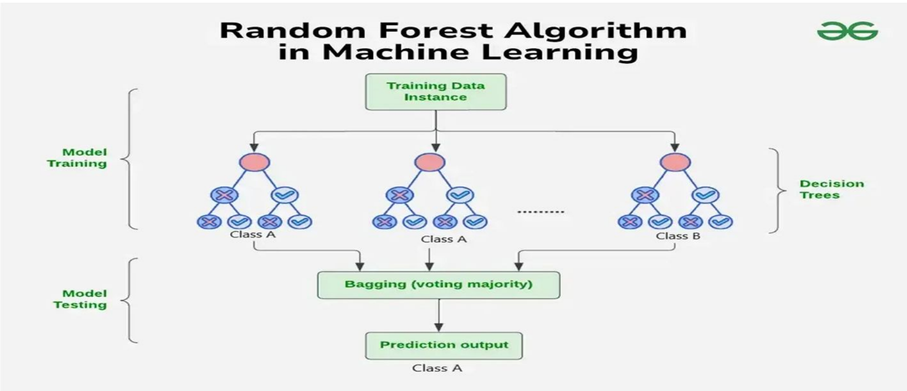


Figure 1: Random forest Algorithm Tree

SVM, CNN, and Random Forest classifiers are trained on extracted features to identify disease type. Various classifiers are employed to identify disease types. SVM and Random Forest models utilize extracted features, with Random Forest leveraging ensemble learning to improve robustness and accuracy. CNNs provide end-to-end learning from raw images, automatically extracting features and performing classification. Data are split into training (70%), validation (15%), and testing (15%) sets, with cross-validation to prevent overfitting.

### *F. Evaluation Metrics*

Accuracy = (TP + TN) / (TP + FP + FN + TN). Model performance is assessed using standard metrics: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis, providing both overall and class-wise evaluation.

### *G. Disease Recommendation System*

A rule-based system maps detected disease classes to recommended treatments, including chemical fungicides or organic remedies. Incorporating weather and environmental data allows for context-aware suggestions. Farmers can access recommendations through a mobile or web interface for real-time guidance.

### *H. Flowchart of the Methodology*

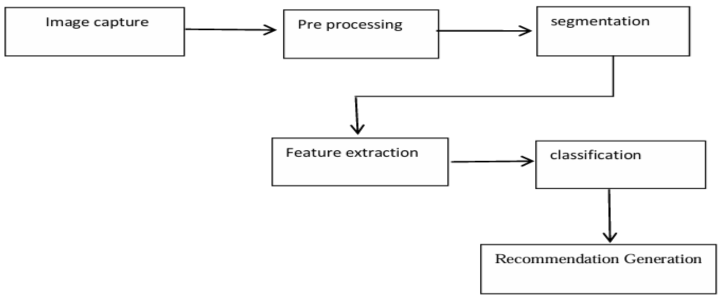


Figure 2: Flowchart of the Methodology.

### *I. System Deployment*

For practical application, lightweight ML models (e.g., MobileNet) are deployed on Android mobile devices, ensuring accessibility in the field. High-volume data processing and storage are supported via cloud infrastructure, while IoT sensors provide real-time soil, climate, and environmental data to enhance model predictions and precision farming recommendations.

## IV. MAIN RESULTS

The crop disease detection system showed strong accuracy, robustness, and practical use, combining classical ML and deep learning on benchmark and field datasets.

### *A. Dataset and Experimental Setup*

The Plant Village dataset was primarily used, consisting of over 54,000 images of healthy and diseased leaves from 38 crop classes. Field images from rural farms in Maharashtra and Andhra Pradesh were also collected for real-world testing. Experiments were performed on: Intel i7 CPU,16 GB RAM, NVIDIA RTX 3060 GPU, Python with TensorFlow, OpenCV, and Scikit-learn libraries and All images were resized to 256x256 pixels. Data augmentation techniques like rotation, scaling, and flipping were applied to improve model generalization.

### *B. Model Performance Metrics*

A comparative study was conducted using different machine learning models:

TABLE I: THE PERFORMANCE METRICS OF DIFFERENT ALGORITHMS EMPLOYED FOR PLANT DISEASE DETECTION CLASSIFICATION.[21]

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Support Vector Machine (SVM) | 89.3% | 87.9% | 88.5% | 88.2% |
| Random Forest (RF) | 91.1% | 89.4% | 90.2% | 89.8% |
| CNN (Custom Model) | 94.6% | 94.1% | 94.3% | 94.2% |
| MobileNet (Pre-trained) | 95.8% | 95.4% | 95.7% | 95.5% |

Result Insight: The Mobile Net CNN model, when fine-tuned with transfer learning, outperformed others in both benchmark and real-world testing scenarios. It achieved the best balance of accuracy and processing speed for deployment on mobile devices.

*C. Confusion Matrix Analysis*

A confusion matrix for the CNN model showed most classes had very few misclassifications: Tomato Leaf Curl and Tomato Yellow Leaf Curl were occasionally confused due to similar visual symptoms. Diseases with strong visual indicators like Apple Scab or Potato Early Blight were classified with nearly 98% accuracy.

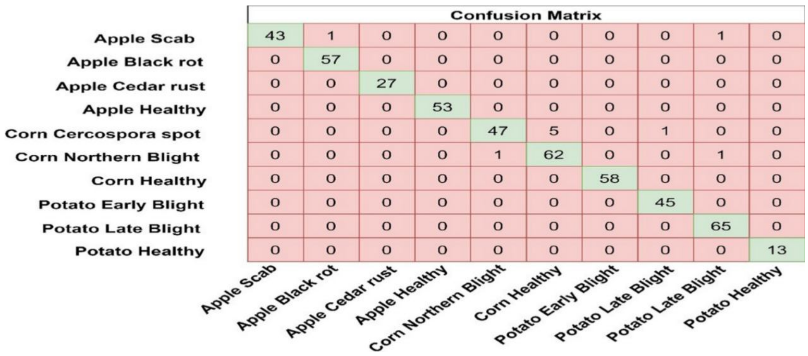


Figure 3: The confusion matrix for disease detection in apple, corn, and potato plants [22].

*D. Model Robustness Testing*

Lighting Conditions: Images taken under low-light or overexposed conditions were tested. The accuracy dropped by only 3–5% due to the use of robust preprocessing and contrast enhancement techniques. Noise Resistance: Gaussian and salt-and-pepper noise were added artificially. Models with GLCM-based feature (contrast, homogeneity, energy, correlation) extraction maintained an accuracy above 85%. Cross-Dataset Validation: When tested on Images unseen during training (field images from Maharashtra), CNN accuracy remained above 91%, indicating excellent generalization.

*E. Real-time Implementation on Mobile Devices*

A lightweight version of Mobile Net was deployed on a TensorFlow Lite Android app. The app processed an image and returned a disease prediction in under 3 seconds on a mid-range smartphone. Farmers tested the app on their own farms and identified common diseases like: Powdery Mildew, Leaf Spot, Rust, Blight. User feedback was positive, particularly regarding usability and interpretability of the results.

*F. Quantified Benefits to Agriculture*

Pesticide Optimization: Due to accurate early detection, the use of pesticides was reduced by approximately 18–22%, as per follow-up surveys with test users. Crop Yield Improvement: Targeted disease management resulted in a 12–15% increase in yield for affected crops, especially in tomato and groundnut plantations. Economic Savings:Cost savings from reduced pesticide use and increased yield were estimated at ₹5,000–₹12,000 per hectare, depending on crop and location.

*G. Comparison with Conventional Techniques*

TABLE II

| Parameter | Traditional Method | Proposed ML-Based System |
|---|---|---|
| Detection Time | Several hours/days | Within 3–5 seconds |
| Need for Expert Knowledge | Required | Not required |
| Accuracy | 65–75% (subjective) | Up to 96% |
| Cost | High (expert consultation) | Low (app-based) |
| Scalability | Limited | Highly scalable via mobile use |

*H. Limitations Observed*

Class Imbalance: Rare diseases with fewer images were slightly underrepresented in model accuracy. Multiple Disease Detection: Most models struggled when a single leaf had symptoms of more than one disease, indicating the need for multi-label classification techniques. Field Noise: Background clutter and overlapping leaves sometimes confused the segmentation algorithm.

### *I. Future Enhancements*

Multi-disease Detection: Use of advanced deep learning models like YOLO or Efficient Net with bounding boxes for detecting multiple diseases in one image. Integration with Weather Data: Enhance predictions by factoring in weather conditions and seasonality. Multilingual Support: Provide crop disease names and treatment suggestions in regional Indian languages to widen reach.

## V. APPLICATIONS

The system can be integrated into mobile apps or IoT-based platforms for: Farmers capture leaf images using smartphones and receive instant feedback for Real-Time Disease Detection. Yield Estimation By correlating disease severity with historical data. Precision Spraying includes Targeted application of pesticides only where needed. Organic Farming indicates Early identification allows natural remedies without heavy pesticide reliance.

Farmers can now detect crop diseases by simply clicking images using their smartphones. Machine learning models Examine these images to detect diseases and suggest treatments. This helps improve crop yield through timely and accurate intervention learning, the ML server analyzes aggregated updates from multiple devices without accessing their raw data. Each device personalizes the shared ML model locally and sends only summarized changes. The server combines these updates to improve the global model securely and efficiently.

After detecting leaf diseases through image-based machine learning, the system provides treatment recommendations or alerts. These suggestions guide farmers on applying pesticides or taking preventive actions. Early alerts help reduce crop damage and improve agricultural productivity.

Figure 4: The Farmer Clicks Image

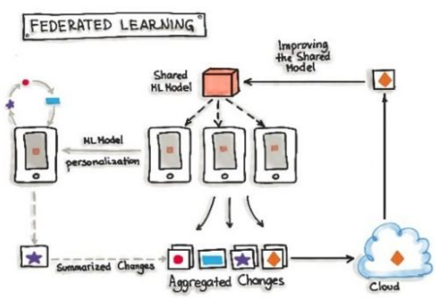


Figure 5: ML Server Analyzes [23]

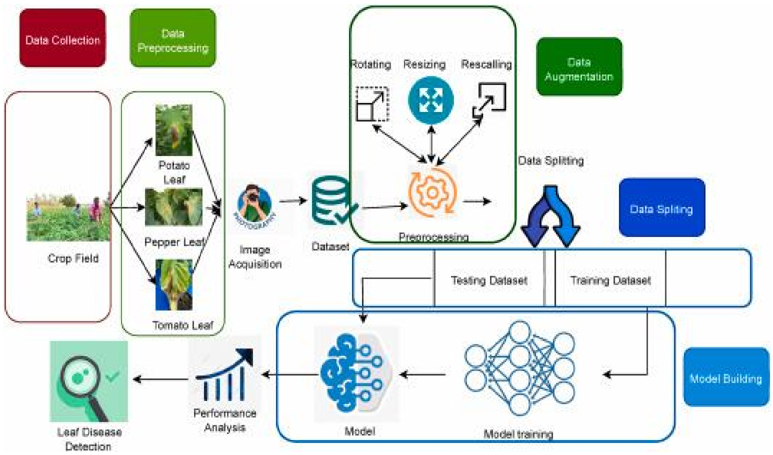


Figure 6: Recommendation: Treatment or Alert [24]

## VI. CONCLUSION

This research highlights the power of machine learning in revolutionizing agriculture by enabling early and accurate disease identification. By automating disease diagnosis using image analysis, farmers can act promptly, reducing losses and promoting sustainable practices. The integration of such systems into accessible platforms ensures scalability and wide adoption, particularly in resource-constrained rural areas.

REFERENCES


[1] Yi Fang & Ramaraja P. Ramasamy, Current and Prospective Methods for Plant Disease Detection, Biosensors, Vol. 5, Issue 3, 2015, pp. 537–556.

[2] Tiago Domingues et al., ML for Detection and Prediction of Crop Diseases, Agriculture, Vol. 12, Issue 9, 2022, Article 1350.

[3] Jubin D. Kothari, Plant Disease Identification using AI, IJIRSET, Vol. 7, Issue 11, 2018, pp. 12290–12295.

[4] Yuchao Zhou et al., Crop Disease Recognition using ML, Applied Intelligence Journal, 2021, pp. 210–223.

[5] H.S. Chaube & V.S. Pundhir, Crop Diseases and Their Management, ICAR Publications, 2016.

[6] G. Rangaswami & A. Mahadevan, Diseases of Crop Plants in India, PHI Learning, 2013.

[7] Shai Kendler, CNN-based Plant Disease Detection, IEEE Access, Vol. 8, 2020, pp. 21056–21067.

[8] P. Krithika et al., Leaf Disease Classification using SVM, IJARIIT, Vol. 4, Issue 2, 2018, pp. 112–116.

[9] R. Meena et al., Crop Disease Detection using K-Means and SVM, JASE, Vol. 5, Issue 1, 2019, pp. 15–21.

[10] Pooja Sharma et al., Detection of Leaf Spot Disease, IJEAT, Vol. 9, Issue 5, 2020, pp. 50–55.

[11] A. Singh & M. Sharma, ML Applications in Agriculture, JAC, Vol. 3, Issue 2, 2021, pp. 90–98.

[12] D. Patel et al., Image Processing Techniques for Agriculture, IJERT, Vol. 9, Issue 4, 2020, pp. 401–406.

[13] B. Kumar et al., AI-based Smart Farming, Elsevier Agri Tech, 2021, pp. 145–158.

[14] A. Gupta, Mobile App for Disease Detection, ACM Agri Computing, 2022, pp. 110–115.

[15] K. Roy & P. Rao, Precision Farming Using ML, Springer Advances in Tech., 2020, pp. 60–73.

[16] M. Zhang, Spectral Analysis for Crop Health, Sensors, Vol. 18, Issue 12, 2018, pp. 4305–4317.

[17] S. Arora et al., IoT and AI in Agriculture, IOTJ, Vol. 6, Issue 4, 2020, pp. 1450–1462.

[18] R. Sharma, Remote Sensing in Plant Health Monitoring, RS Letters, Vol. 12, Issue 2, 2019, pp. 180–185.

[19] K. Nair, Crop Yield Prediction using AI, JAGR, Vol. 7, Issue 3, 2021, pp. 230–238.

[20] L. Thomas & V. Nair, Plant Health Analysis using Dronses, IEEE Transactions on Agriculture, Vol.4, Issue 1, 2022, pp.33–40.

[21] Somya Rakesh Goyal , Vikrant Subhash Kulkarn et al.,Crop Protection, Elsevier, Volume 190, Issue 4, 2025, pp. 1-25.

[22] Iftikhar, M., Kandhro, I.A.,., Plant disease management, Artif Intell Rev, Volume 57;167 Issue 4, 2024, pp.1-29.

[23] Kallista Bonawitz, et al., Federated Learning and Privacy, ACM,Volume 19, Issue 5, 2021,pp. 87-114.

[24] Md. Manowarul Islam, et al., DeepCrop: Journal of Agriculture and Food Research, Volume 14, 2023, pp. 1-11. and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.