A Review on Machine and Deep Learning Approaches for Crop Disease Detection

Sushma B, Syed Rabbith

[1]*Assistant Professor, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru.*
[2]*U.G. Student, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru*

---***---

Abstract - Crop diseases significantly affect agricultural productivity and farmer livelihoods. Reducing crop losses and advancing sustainable farming require early detection and response. Recent advancements in Machine Learning (ML) and Deep Learning (DL) have shown promising results in crop disease detection using image analysis. This paper presents a detailed review of ML-based techniques for detecting crop diseases, with a focus on Convolutional Neural Networks (CNNs), transfer learning models, Realtime deployment through web applications like Streamlet, and the shortcomings of the methods used thus far. We also discuss datasets used in training models, performance metrics, and future trends such as Edge AI, IoT integration, and multilingual support for inclusive smart agriculture.

***Key Words*:** Machine Learning, Crop Disease Detection, Deep Learning, CNN, Streamlit, Smart Agriculture, Edge AI

1.INTRODUCTION

The agricultural sector faces a multitude of challenges ranging from climate change and pest invasions to crop diseases that result in reduced yield and food insecurity. Traditional methods of identifying illnesses are manual and often prone to errors. As agriculture embraces the digital revolution, machine learning and image processing offer significant advantages in automating disease detection. These technologies enable accurate and fast diagnosis through image-based classification models, empowering farmers with data-driven decisions.

II. Role of Machine Learning in Crop Health Monitoring

Machine Learning provides automated tools that learn patterns from datasets and predict outcomes with minimal human intervention. ML is utilized in agriculture for:

Crop disease detection

• Yield prediction

• Soil nutrient estimation

• Weed and pest recognition

Deep Learning, a subfield of ML, particularly CNNs, has outperformed traditional ML models in tasks involving images, such as identifying leaf patterns, texture, and color anomalies caused by diseases.

III. Datasets for Crop Disease Detection

ML models require large, labeled datasets to learn disease characteristics. Popular datasets include:

A. PlantVillage Dataset

• • More than 50,000 photos with annotations showing both healthy and sick leaves from 38 different crop classes.

The majority of agricultural disease detection research uses this dataset as a baseline.

B. Kaggle Crop Disease Datasets

• Provides disease images for regionspecific crops.

• Useful for real-world diversity and class imbalance testing.

C. Custom Datasets

• Sourced from agricultural universities and farmer networks.

• Include real-life noise such as varying lighting, background clutter, and partial leaves.

IV. Machine Learning Techniques for Disease Classification

A. Convolutional Neural Networks (CNNs)

CNNs are widely used because of their ability to extract spatial features from images. Common CNN models consist of:

• Convolutional layers

• Pooling layers

• Fully connected layers

• Softmax classifier

CNNs achieve high accuracy on image classification tasks and can generalize well with data augmentation.

B. Transfer Learning

Pre-trained models like:

• ResNet50

• InceptionV3 • MobileNet are fine-tuned on crop datasets for faster convergence and better performance on limited data.

C. Traditional ML Models

• Support Vector Machine (SVM)

• Random Forest (RF)

• K-Nearest Neighbor (KNN)

These are used with extracted color/shape features and are lighter than CNNs, making them suitable for mobile deployment.

V. Model Evaluation Metrics

• ML model performance is evaluated using:

• Accuracy = (TP + TN) / (TP + TN + FP + FN)

• Precision = TP / (TP + FP)

• Recall = TP / (TP + FN)

• F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

Confusion Matrix: A graphic depiction of the classification outcomes

Most CNN-based systems report over 90% accuracy on clean datasets like PlantVillage, but this drops with realworld data.

VI. Real-Time Deployment Tools

A. Streamlit

A lightweight Python framework called Streamlit enables quick ML app prototyping.

Its benefits include:

• Easy integration with CNN models

• File upload interface for leaf images

• Real-time prediction with confidence scores

• Multiplatform compatibility (desktop, mobile)

B. Backend and Database

• Firebase, PostgreSQL, or MongoDB used for storing user data and predictions.

• Flask/TensorFlow Serving handles backend inference.

VII. Challenges and Limitations

A. Image Quality Dependency

Poor lighting, occlusion, and overlapping leaves reduce prediction accuracy.

B. Class Imbalance and Data Scarcity

Some diseases have very few samples, which biases the model.

C. Overfitting

Models could memorize training data instead of generalizing if augmentation and regularization are not done correctly.

D. Internet Dependency

Most systems need cloud inference, which limits use in remote areas.

E. Visual Similarity Between Diseases

Different diseases may exhibit similar symptoms on leaves.

VIII. Future Directions

Edge AI and Offline Capability

Using lightweight models like MobileNet or TinyML can enable offline prediction on smartphones or Raspberry Pi.

A. IoT Integration

IoT sensors can improve disease forecasting by gathering data in real time, such as soil moisture, temperature, and humidity. .

B. Time-Series Analysis

Instead of static images, future systems will monitor disease progression over time.

C. Multilingual and Voice-Based Systems

To improve accessibility for farmers in different regions.

D. Ensemble and Hybrid Models

Combining CNNs with traditional classifiers to increase robustness.

IX. Conclusion

In particular, CNN-based models based on machine learning have greatly improved crop disease diagnosis. Even if there are still issues with practical implementations, deep learning combined with cloud and edge technologies will be the key to precision agriculture's future. Global food security may benefit greatly from ML-powered crop health systems as datasets and models become more readily available.

X. References

[1] Mohanty, S.P., Hughes, D.P., & Salathé, M. “Using Deep Learning for ImageBased Plant Disease Detection,” Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[2] Kamilaris, A., & Prenafeta-Boldú, F.X. “Deep Learning in Agriculture: A Survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70-90, 2018.

[3] Brahimi, M. et al., “Deep Learning for Tomato Diseases: Classification and Symptoms Visualization,” Applied Artificial Intelligence, vol. 31, no. 4, pp. 299-315, 2017.

[4] Sladojevic, S. et al., “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” Comp. Intelligence and Neuroscience, vol. 2016.

[5] Ferentinos, K.P., “Deep learning models for plant disease detection and diagnosis,” Comp. and Elec. in Agriculture, vol. 145, pp. 311-318, 2018.

[6] Streamlit Documentation, https://docs.streamlit.io/