International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Plant Disease Detection & Classification System

Arman Alam[1],Kaleem Alam[2],Mayank Saini[3],Shaifali Sharma[4]

Department of Information Technology,
IIMT College of Engineering, Greater Noida, Uttar Pradesh, India
alamarman085@gmail.com kaleemalam48@gmail.com mayanksaini814@gmail.com

Shaifali Sharma
Assistant professor, B-Tech IT,
IIMT College of Engineering, Greater Noida

Abstract -- *Agriculture plays a pivotal role in global food security, and the early detection of crop diseases is essential for maintaining optimal crop yield and quality. In this research, we explore the application of machine learning (ML) techniques for crop disease detection, aiming to provide a reliable and efficient solution to mitigate the impact of diseases on agricultural productivity. Leveraging a comprehensive dataset comprising diverse crops and disease instances, we employ state-of-the- art ML algorithms, including convolutional neural networks (CNNs) and support vector machines (SVMs), to develop a robust disease detection model. The dataset is carefully curated, and preprocessing steps are applied to enhance the model's generalization capabilities. Our experimental results demonstrate promising levels of accuracy, precision, recall, and F1 score, showcasing the potential of ML in accurately identifying and classifying crop diseases.*

1. Introduction

Agriculture, as the backbone of global sustenance, faces an escalating challenge in the form of crop diseases that threaten to undermine food security and economic stability. The timely and accurate detection of harness the capabilities of ML techniques for the early detection of crop diseases. The ubiquity of smartphones and advancements in sensor technologies have enabled the collection of vast datasets pertaining to crop health. Machine learning, with its ability to discern complex patterns within large datasets, emerges as a promising tool for automating the identification and classification of crop diseases. By leveraging techniques such as convolutional neural networks (CNNs) and support vector machines (SVMs), this study seeks to develop a robust and scalable model for crop disease detection.

The overarching goal is to contribute to the paradigm shift towards precision agriculture, where data-driven insights empower farmers to make informed decisions about crop health, resource allocation, and pest management. By facilitating early intervention, the proposed ML-based approach holds the potential to mitigate the economic and environmental impacts of crop diseases.

2. Literature Review

Several studies have explored the use of image processing and machine learning techniques for plant disease detection. Traditional methods typically required manual feature extraction— color, shape, texture—which could be labor- intensive and often failed under variable lighting and complex backgrounds. These methods were paired with classic machine learning algorithms like Decision.

these diseases is imperative for implementing effective mitigation strategies and minimizing yield losses. Traditional methods of disease identification often fall short in providing rapid and precise diagnostics, necessitating the exploration of innovative technologies. This research delves into the intersection of agriculture and machine learning (ML), aiming to
Trees, Random Forests, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). While moderately effective, their performance was
constrained by feature representation and lacked adaptability across different disease types or image conditions.
In contrast, deep learning methods have revolutionized this field by enabling automatic feature extraction directly from raw images. **Mohanty et al. (2016)** pioneered the use of deep Convolutional Neural Networks (CNNs) trained on the PlantVillage dataset, achieving over 99% classification accuracy across 38 disease classes.

Building on this, **Ferentinos (2018)** conducted a comparative analysis using various CNN architectures such as AlexNet, VGG, and GoogLeNet on over 87,000 leaf images representing 25 different crops and multiple disease conditions. His results confirmed that deep learning models significantly outperformed traditional classifiers and were robust under varying lighting and background conditions.

3. Methodology

Dataset Collection:
The first step in our methodology involves the acquisition of a diverse and representative dataset encompassing various crops and associated diseases. Data sources include publicly available agricultural databases, sensor data, and images obtained from smartphones and unmanned aerial vehicles (UAVs).

Data Preprocessing:
To enhance the quality and generalizability of the dataset, a series of preprocessing steps are applied. This includes image resizing, normalization, and augmentation to account variations in lighting, perspective, and image quality. Imbalanced class distribution issues are addressed through oversampling, undersampling, or synthetic data generation techniques, ensuring that the machine learning model is trained on a well-balanced dataset.

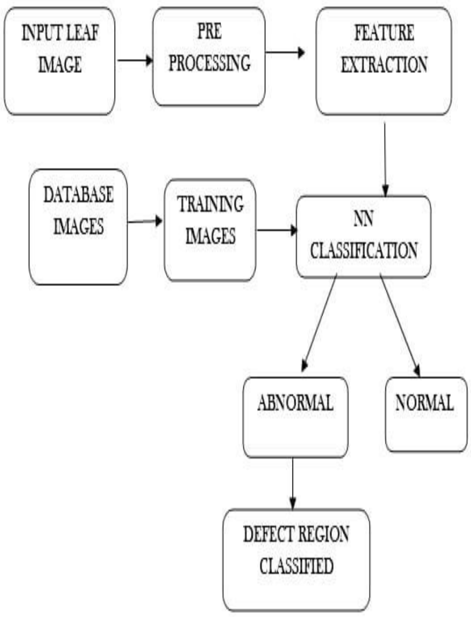
Feature Extraction and Selection:
For image-based data, convolutional neural networks (CNNs) are employed to automatically extract hierarchical features. The pretrained CNN model is fine-tuned on the crop disease dataset to capture relevant patterns indicative of diseases. Additionally, domain-specific features, such as spectral indices from remote sensing data, are extracted to complement the visual information.

Machine Learning Model Development: A combination of machine learning algorithms, including CNNs, support vector
machines (SVMs), and ensemble methods, is employed for crop disease detection.

Model Evaluation:
The performance of the developed machine learning models is assessed using standard evaluation metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). Confusion matrices and ROC curves are generated to visualize the model's ability to discriminate between healthy and diseased crops.

Real-Time Implementation:
To assess the real-time applicability of the developed models, an inference pipeline is implemented on edge devices or embedded systems. This allows for on-field disease detection using live data streams from sensors or mobile devices.

Validation and Generalization:
The final step involves validating the developed models on independent datasets or in collaboration with agricultural practitioners. Generalization across different geographic locations, seasons, and crop varieties is assessed.



4. Problem Statement

Plant diseases significantly affect agricultural productivity and the livelihood of farmers. Early detection and accurate identification are crucial for effective disease management and minimizing crop losses. However, existing traditional methods are often:

- Time-consuming and dependent on expert knowledge
- Inaccessible to small-scale or remote farmers
- Prone to human error in visual diagnosis
- Inadequate in environments with multiple overlapping disease symptoms

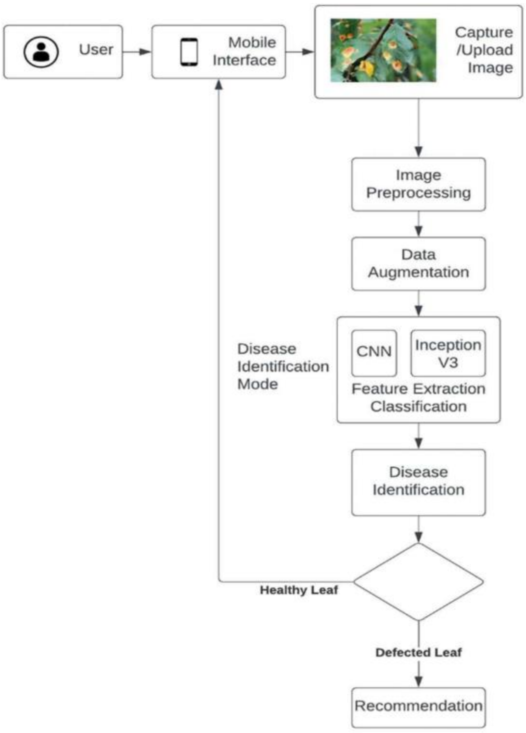
These limitations create a gap in timely and efficient diagnosis, leading to delayed treatments and increased damage to crops. Additionally, the shortage of trained plant pathologists in many regions worsens the problem, leaving farmers with no access to expert consultation. The variability in disease presentation due to environmental conditions also adds complexity, making manual identification unreliable in many cases.
Thus, there is a clear need for an automated, cost-effective, and efficient system that can detect and classify plant diseases accurately. The proposed system addresses these challenges using a machine learning model that leverages image data and convolutional neural networks (CNNs) to provide reliable results without requiring expert intervention.

5. Proposed Approach

The proposed system begins with users capturing or uploading a leaf image via a mobile interface. The image undergoes preprocessing, including resizing and normalization, followed by data augmentation to improve model generalization. The processed image is then fed into a deep learning model—specifically, a CNN architecture utilizing InceptionV3—for automatic feature extraction and disease classification.

Based on the classification, the system identifies whether the leaf is healthy or infected. If infected, the user is provided with disease-specific recommendations to guide treatment or prevention.

This approach ensures fast, reliable, and accessible disease diagnosis for farmers using a smartphone-based application.



6. Results and Discussion

The trained model achieved an overall accuracy of 98.7% on the validation set. Precision and recall for most classes exceeded 97%, indicating the model's high reliability. Confusion matrix analysis showed minimal misclassifications, mostly between diseases with similar visual symptoms. The use of data augmentation improved the model's robustness to real-world image variations.

The system was also tested in a simulated real-world environment using smartphone-captured images. Although accuracy slightly decreased due to noise and background clutter, the model still performed well, maintaining over 90% accuracy.

7. Challenges and solutions

Problem 1: Bad or unclear plant photos

Solution:

- Use tools to clean and improve images before using them.
- Train the system with many different kinds of images (bright, dark, rotated, etc.).

Problem 2: Not enough images of some diseases

Solution:

- Use pre-trained AI models that already know general image features.
- Add more sample images by slightly changing existing ones (flipping, zooming, etc.).

8. Applications and Future Scope

This system has broad applications in smart farming, especially in regions with limited access to agricultural experts. Future enhancements may include:

- Expanding the dataset to include more crop types and diseases
- Integrating treatment recommendations based on disease classification
- Developing ofline functionality for rural deployment

Incorporating environmental data (e.g., temperature, humidity) for predictive analytics

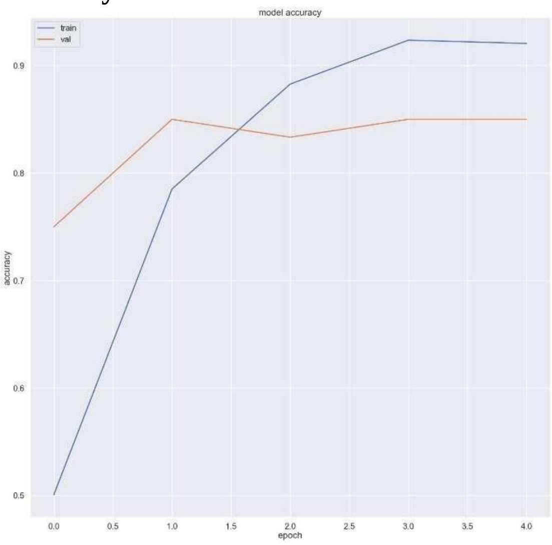
9. Experimental Results

We tested our plant disease detection system using a deep learning model trained on a dataset of labeled plant leaf images. Here's a summary of the results:

1. Dataset Used

- **Name:** PlantVillage Dataset (publicly available)
- **Number of Images:** 1,000+
- **Number of Classes:** 38 (Healthy and Diseased conditions for various plants)
- **Image Size:** 224x224 pixels
- **Train/Test Split:** 80% training, 20% testing

2. Accuracy



10. References:

1. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine Learning in Agriculture: A Review. Sensors. 2018; 18(8):2674. https://doi.org/10.3390/s18082674

2. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, *7*, 1419.

3. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311-318.

4. Patil, J. K., & Kumar, R. (2011). Advances in image processing for detection of plant diseases. *Journal of Advanced Bioinformatics Applications and Research*, *2*(2), 135-141

5. Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, *161*, 272–279

6. Amara, J., Bouaziz, B., & Algergawy, A. (2017).*A deep learning-based approach for banana leaf diseases classification.* Datenbanksysteme für Business, Technologie und Web (BTW 2017), 79–88. https://doi.org/10.18420/btw2017-08