Plant Disease Detection using AI

Sibsankar Maity

1 Abstract

Plant disease detection is a critical aspect of modern agriculture, aiming to improve crop health, yield, and food security. Traditional methods of disease detection often rely on manual inspection by experts, which can be time-consuming, labor-intensive, and prone to human error. With the advent of advanced machine learning techniques and the increasing availability of image data, automated plant disease detection has emerged as a promising solution.

This project focuses on developing a system that utilizes image processing and deep learning to identify and classify plant diseases from images. The system is designed to take input as a picture of a plant, preprocess the image, and apply a trained convolutional neural network (CNN) to detect the presence and type of disease. The model leverages large datasets of labeled plant images to achieve high accuracy in disease classification.

The proposed approach offers several advantages over traditional methods, including faster diagnosis, scalability, and the ability to process images in real-time. By integrating this system into mobile applications or agricultural management platforms, farmers can quickly and accurately identify diseases in their crops, enabling timely interventions and reducing the spread of disease.

This work contributes to the broader field of precision agriculture, providing a tool that can help farmers improve crop management, reduce losses, and ultimately enhance food production efficiency. Future research will focus on expanding the system to include a wider range of crops and diseases, improving model accuracy under diverse environmental conditions, and exploring the integration of other data sources, such as environmental sensors, to enhance detection capabilities.

2 Problem Statement

Plant diseases are a major challenge in agriculture, leading to reduced crop yields, economic losses, and threats to food security. Traditional methods of disease detection typically rely on manual inspection by experts, which can be inefficient, time-consuming, and often inaccessible to many farmers, especially those in remote or resource-limited areas. These methods are also prone to human error, and the lack of early detection can result in the rapid spread of diseases, exacerbating their impact. To address these issues, there is a critical need for an automated, accurate, and scalable system that can detect plant diseases from images captured by cameras or mobile devices. Such a system would leverage machine learning to identify and classify plant diseases in real time, providing farmers with immediate, actionable information. The goal is to reduce the dependency on expert inspection, making disease detection more accessible and efficient. By developing a user-friendly interface, possibly in the form of a mobile application, the system would empower farmers to diagnose and manage plant diseases promptly, thereby improving crop health, increasing productivity, and contributing to more sustainable agricultural practices.

3 Literature Review

Machine learning (ML) and deep learning (DL) have become increasingly popular in plant disease detection, offering new ways to identify diseases from plant images. Traditionally, ML methods involved manually extracting features like color, texture, and shape from images to train models that could distinguish between healthy and diseased plants. These methods have been used successfully in detecting diseases like leaf blotch and powdery mildew. However, they often struggle with early disease detection and processing complex images.

To address these limitations, DL techniques, especially Convolutional Neural Networks (CNNs), have been introduced. Unlike traditional methods, CNNs can automatically learn important features from images, making them more effective at detecting subtle symptoms. They also work well with high-resolution images, which is crucial for detailed analysis. However, DL models require large amounts of labeled data for training and are computationally demanding, making them challenging to implement in some settings.

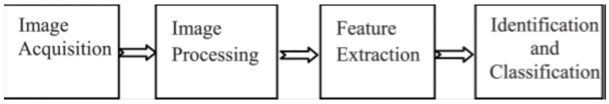
Researchers have tried to overcome these challenges by using techniques like data augmentation, which increases the diversity of training data, and transfer learning, which allows models to be fine-tuned on new datasets. These approaches have shown promise in making DL models more robust and generalizable. However, most research has focused on specific plants or diseases, limiting the broader applicability of these models.

Recent studies emphasize the need for models that can detect a wide range of diseases across different plant species. Ensemble methods, which combine multiple models, are also being explored to improve accuracy. While progress has been made, there are still challenges, such as the need for diverse datasets and the high computational requirements of DL models. Future research should focus on addressing these issues to develop more effective plant disease detection systems.

4 Introduction

The occurrence of plant diseases poses a significant threat to agricultural production, directly impacting food security worldwide. Timely detection of these diseases is crucial for effective prevention and control, serving as a cornerstone in the management and decision-making processes of agricultural practices. Traditionally, plant diseases have been identified through visual inspection by agricultural experts or farmers. However, this method is subjective, time-consuming, and prone to errors, particularly when conducted by less experienced individuals. These limitations can lead to misjudgments in disease management, resulting in unnecessary economic losses and environmental pollution.

Over the years, traditional image processing techniques have been developed to automate plant disease identification. These methods typically involve the extraction of fea-



tures such as color, texture, and shape, followed by the application of various classifiers like Support Vector Machines (SVM) and Principal Component Analysis (PCA). While these approaches have achieved commendable accuracy, they are often labor-intensive, highly subjective, and dependent on manual feature extraction. Furthermore, they struggle to perform effectively in complex environments, limiting their practical application on a broader scale.

In recent years, deep learning has emerged as a promising solution to these challenges, particularly through the use of Convolutional Neural Networks (CNNs). Unlike traditional methods, CNNs can automatically extract features from images and classify diseases with minimal human intervention. This has significantly reduced the dependency on manual feature extraction, making the process more efficient and accurate. However, the effectiveness of deep learning models is often constrained by the availability of large, diverse datasets, which are crucial for training robust models. To address this, transfer learning has become a popular technique, allowing pre-trained CNNs to be adapted for plant disease recognition using smaller, domain-specific datasets.

Despite the advancements in deep learning, there are still several gaps in the current research. Existing studies have largely overlooked recent developments in visualization techniques and the adaptation of deep learning models for early disease detection and classification based on small sample sizes. This project aims to address these gaps by reviewing the latest research in plant leaf disease recognition using a combination of image processing, hyper-spectral imaging, and deep learning techniques. The goal is to provide a comprehensive overview of the current state of the field and highlight potential areas for future research, particularly in the context of early detection and the challenges posed by limited datasets.

5 METHODOLOGY

Deep learning is a powerful approach in machine learning that has significantly alleviated the challenges of traditional feature engineering, eliminating the need for extensive domain expertise. This advancement is largely attributed to the use of deep learning techniques, which are built upon the foundation of artificial neural networks (ANNs). ANNs are mathematical models designed to mimic the general principles of brain function through interconnected neurons and synapses. A widely used library for implementing neural networks is TensorFlow, which provides comprehensive tools for building and training artificial neural networks. TensorFlow supports a variety of tasks, including classification of both text and images, making it a versatile tool for deploying deep learning models in various applications.

5.1 Convolution Neural Network (CNN)

Convolutional Neural Networks (CNNs) are widely used to detect diseases in plant leaves, offering superior performance compared to traditional Artificial Neural Networks (ANNs) when dealing with image data. This improvement is due to CNN's ability to recognize repeating patterns within images, which is essential for accurate disease detection. The two key functions of CNNs are convolution and pooling: convolution detects edges and patterns within an image, while pooling reduces the image size, preserving the most important features. Several CNN architectures have been applied to this problem, including Simple CNN, VGG, and InceptionV3. These models are typically trained using Jupyter Notebook and the Keras API within TensorFlow. Keras is a high-level API that simplifies the building and training of deep learning models, making it easier to implement complex neural network architectures.

5.2 Dataset Discussion

Two datasets are used to perform plant disease detection.The first dataset consists of 15 classes and the second one consists of 38 classes. Both databases have number of images of each plant. The first dataset has total 2952 images. The final findings of this work is on the PlantVillage dataset which contains 38 classes of different plants. It is also openly available on the internet. A description of these classes and dataset is given in the following Table- I (a) and (b)

Table- I (a): Dataset Description

Class	Plant Name	Healthy or Diseased	Disease Name	Images (Number)
C_0	Apple	Diseased	Apple_scab	2016
C_1	Apple	Diseased	Black_rot	1987
C_2	Apple	Diseased	Cedar_apple_rust	1760
C_3	Apple	Healthy	-	2008
C_4	Blueberry	Diseased	-	1816
C_5	Cherry_(including_sour)	Diseased	Powdery_mildew	1683
C_6	Cherry_(including_sour)	Healthy	-	1826
C_7	Corn_(maize)	Diseased	Cercospora_leaf_spotGray_leaf_spot	1642
C_8	Corn_(maize)	Diseased	Common_rust	1907
C_9	Corn_(maize)	Diseased	Northern_Leaf_Blight	1908
C_10	Corn_(maize)	Healthy	-	1859
C_11	Grape	Diseased	Black_rot	1888
C_12	Grape	Diseased	Esca_(Black_Measles)	1920
C_13	Grape	Diseased	Leaf_blight (Isariopsis_Leaf_Spot)	1722
C_14	Grape	Healthy	-	1692
C_15	Orange	Diseased	Haunglongbing (Citrus_greening)	2010
C_16	Peach	Diseased	Bacterial_spot	1838
C_17	Peach	Healthy	-	1728
C_18	Pepper_bell	Diseased	Bacterial_spot	1913
C_19	Pepper_bell	Healthy	-	1988
C_20	Potato	Diseased	Early_blight	1939
C_21	Potato	Diseased	Late_blight	1939
C_22	Potato	Healthy	-	1824
C_23	Raspberry	Healthy	-	1781
C_24	Soybean	Healthy	-	2022
C_25	Squash	Diseased	Powdery_mildew	1736
C_26	Strawberry	Diseased	Leaf_scorch	1774
C_27	Strawberry	Healthy	-	1824
C_28	Tomato	Diseased	Bacterial_spot	1702

Figure 1: Dataset Description

This table provides the number of images in each class, with approximately 2,000

Table- I (b): Dataset Description

Class	Plant Name	Healthy or Diseased	Disease Name	Images (Number)
C_29	Tomato	Diseased	Early_blight	1920
C_30	Tomato	Diseased	Late_blight	1851
C_31	Tomato	Diseased	Leaf_Mold	1882
C_32	Tomato	Diseased	Septoria_leaf_spot	1745
C_33	Tomato	Diseased	Spider_mites Two-spotted_spider_mite	1741
C_34	Tomato	Diseased	Target_Spot	1827
C_35	Tomato	Diseased	Tomato_Yellow_Leaf_Curl_Virus	1961
C_36	Tomato	Diseased	Tomato_mosaic_virus	1790
C_37	Tomato	Healthy	-	1926
Total				70295

Figure 2: Dataset Description

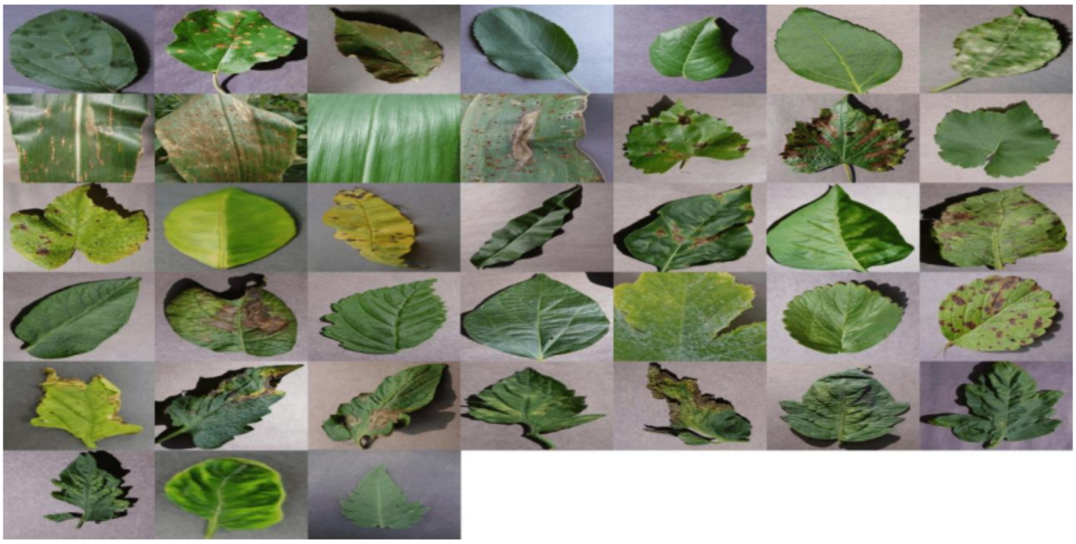


Fig. 1: Sample of Images from PlantVillage Dataset

Figure 3: Sample of Images from PlantVillage Dataset

images per class. The dataset features fourteen different plants, including both healthy and diseased leaf images. The majority of images are from Tomato and Apple plants, while Raspberry, Soybean, and Squash have the fewest images. The image below displays some examples of the various leaf types included in the dataset.

5.3 Model Description

First some preprocessing is applied on dataset in form of augmentation to increase size of dataset in order to achieve better accuracy. Then images size are reduced by 256x256 pixels. After that a convolution neural network based model will be created with multiple pooling and convolution layers and a dense layer for prediction. Five convolution layers with 3x3 filter are used and five MaxPooling2D layers with 2x2 filter. Batch Normalization is also used in this model. Batch normalization is used to scale data on particular scale but the difference is that it not just does it on input layer but it also do it at other hidden layers. At last model is trained on PlantVillage dataset.

Table- II: CNN Training Parameters

Parameter	Value
Epochs	25
Batch Size	32
Learning Rate	1e-1
Activation in middle layers	Relu
Activation in Final layer	Softmax

Figure 4: CNN Training Parameters

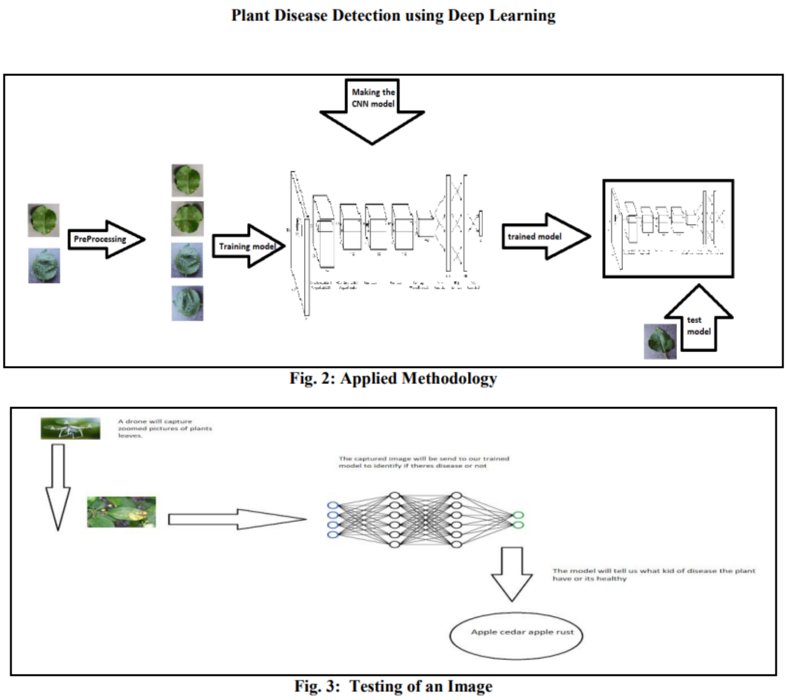


Figure 5: Caption

6 RESULT AND DISCUSSION

This study shows the importance of plant disease detection in these days. This model were developed using Deep Learning in python.20% (14,059) images from PlantVillage dataset were used to test the accuracy of this model. These images are from 38 different classes. 20% of each class randomly selected for testing. Some real time images were also used. Those images were captured from local environment. They do not belong to any class which are present in dataset. But model give us more than 95% accuracy on those images as well by telling either leaf is healthy of unhealthy. Total 100 images were used and 96 were classified correctly. Some images were captures at night with the help of flash light and some images have dirt upon it so that they were misclassified. Some of the images which we captured from local environment.

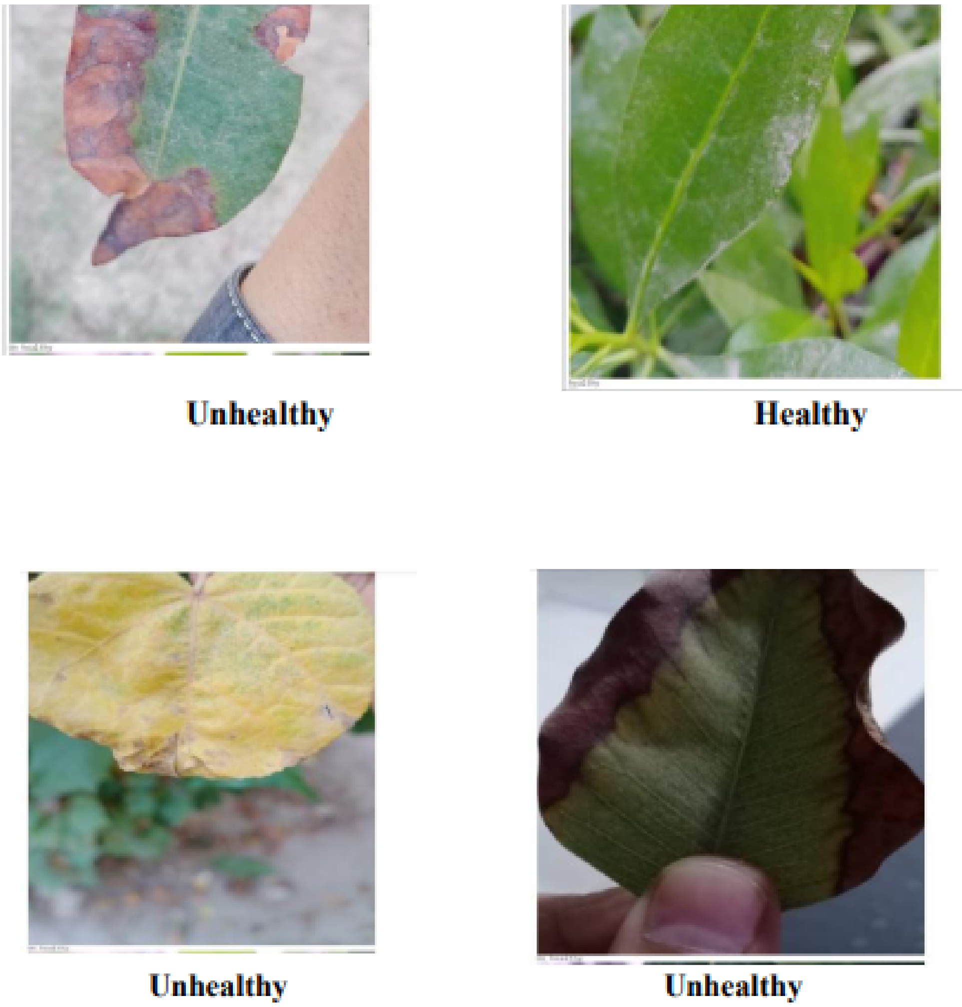


Fig. 4: Locally Captured Images

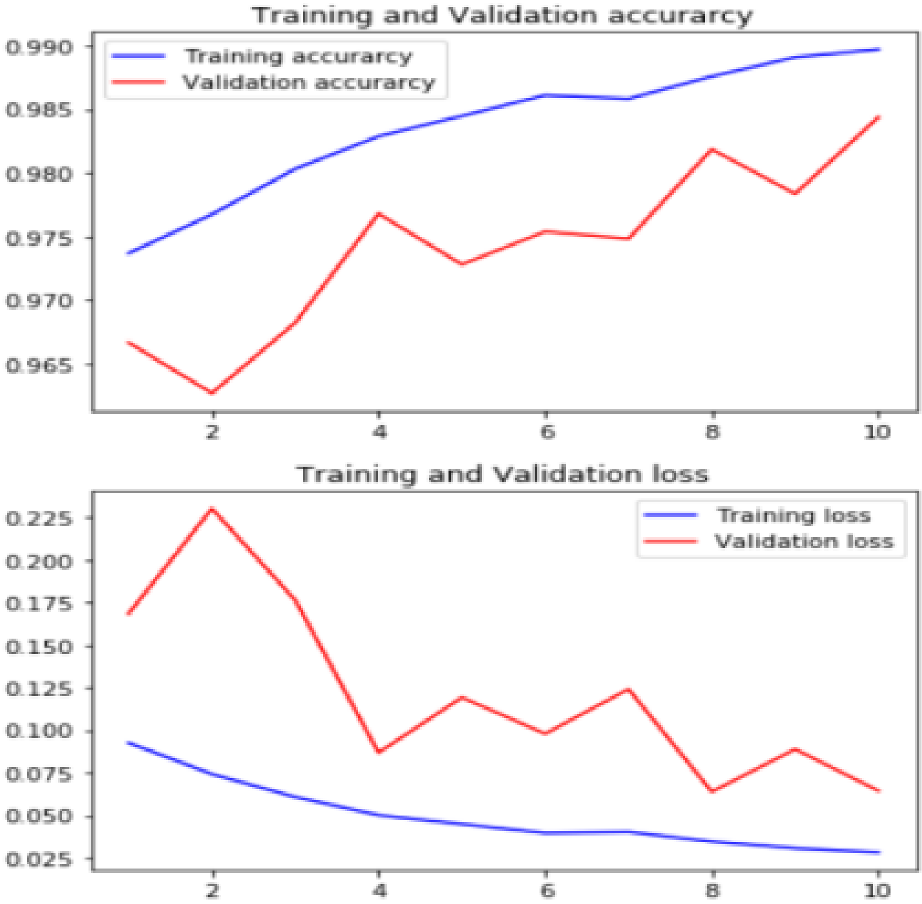


Figure 6: Training and Validation Accuracy on Testing Dataset

Classes like Corn_(maize) healthy, Tomato Tomato_mosaic_virus, Strawberry healthy, and Corn_(maize) Common_rust give approximately 100accuracy. Just 1 to 5 images misclassified from them. Apple Cedar_apple_rust,Cherry_(including_sour) healthy, Grape Black_rot, and Raspberry healthy classes give less accuracy than other classes. Below is the table of image classes which gives better accuracy while testing on our model.

Table- III: Image Classes with Better Accuracy

Class	Total	Correct Classified	Mis-classified
Corn_(maize)___healthy	186	186	0
Tomato___Tomato_mosaic_virus	179	176	3
Strawberry___healthy	182	175	7
Corn_(maize)___Common_rust_	191	188	3
Corn_(maize)___Northern_Leaf_Blight	191	175	16

Table- IV: Final Model Performance

Model	Dataset for Training	Dataset for Testing	Training Accuracy	Testing Accuracy
CNN	PlantVillage (80%)	PlantVillage (20%)	99%	98%+
CNN	PlantVillage (80%)	Actual Environment (100 Images)	99%	95%+

Figure 7: CONCLUSION AND FUTURE WORK

7 CONCLUSION AND FUTURE WORK

This study has utilized deep learning capabilities to achieve automatic plant disease detection system. This system is based on a simple classification mechanism which exploits the feature extraction functionalities of CNN. For prediction finally, the model utilizes the fully connected layers. The research was carried out using the publically accessible collection of 70295 images, and 100 images from experimental conditions and actual environment. The system has achieved an overall 98% testing accuracy on publically accessible dataset, and performed well on images of Sukkur IBA University plants. It is concluded from accuracy that CNN is highly suitable for automatic detection and diagnosis of plants. This system can be integrated into mini-drones to live detection of diseases from plants in cultivated areas. Though this system is trained on Plant Village dataset with only 38 classes it could tell if the plant has a disease or not as somehow symptoms are same in all kinds of plants. In addition, more actual environment images can be added to the dataset to improve the accuracy on real-condition images of leaves and classify more plant types as well as disease types. In the future, this system can also adopt 3 layer approach where the first layer detects if there's any plant in an image or not, second layer tells the plant type and the third layer tells if there is any disease or not and what type of disease is there if any.