

**INT 354 CA 1 PROJECT**

# On

**Plant Disease Detection Using Image Processing and Machine Learning**

Submitted by:

# NAME: SHAHBAZ AMAN O

# REG No: 12114313

**PROGRAM: B. Tech CSE**

**School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

# Abstract

One of the important and tedious tasks in agricultural practices is detection of disease on crops. It requires huge time as well as skilled labour. This paper proposes a smart and efficient technique for detection of crop disease which uses computer vision and machine learning techniques.

Keywords: Digital image processing, Machine learning, Plant disease detection.

# Introduction

In recent years, agriculture has undergone a transformative evolution, integrating cutting-edge technologies to address age-old challenges. Among these challenges, the effective management of plant diseases stands out as a critical factor influencing crop yield, quality, and overall agricultural sustainability. The presence of leaf diseases not only threatens the economic viability of farming but also jeopardizes food security on a global scale. Addressing this pressing issue demands proactive measures and innovative solutions.

In response to this imperative, the convergence of machine learning and agricultural science offers a promising avenue for revolutionizing disease management practices. Leveraging the vast potential of machine learning algorithms, researchers and practitioners can now develop sophisticated models capable of predicting leaf diseases with unprecedented accuracy and efficiency. By harnessing the power of data-driven insights, stakeholders in the agricultural sector can implement targeted interventions, optimize resource allocation, and mitigate the adverse effects of plant diseases on crop productivity.

This project report presents a comprehensive exploration into the realm of leaf disease prediction using machine learning techniques. By amalgamating principles from computer science, agronomy, and statistical modeling, our research endeavors to develop a robust predictive framework tailored to the unique challenges of disease identification and classification in agricultural settings. Through a meticulous analysis of diverse datasets encompassing factors such as environmental conditions, plant physiology, and disease incidence, we aim to unveil the underlying patterns and correlations that govern the dynamics of leaf diseases.

Furthermore, this report seeks to elucidate the practical implications of our predictive model within the context of real-world agricultural practices. By elucidating the potential benefits of early disease detection, timely intervention strategies, and adaptive management approaches, we strive to demonstrate how our machine learning-based solution can empower farmers, agronomists, and policymakers to make informed decisions and safeguard agricultural productivity in the face of evolving disease threats.

As we embark on this journey at the intersection of technology and agriculture, our endeavor transcends the mere development of predictive algorithms; it embodies a collective commitment to fostering sustainable agricultural systems, enhancing food security, and nurturing a resilient future for generations to come. Through collaborative efforts and continuous innovation, we aspire to chart a course towards a world where the specter of leaf diseases no longer looms large over the fields of our prosperity.

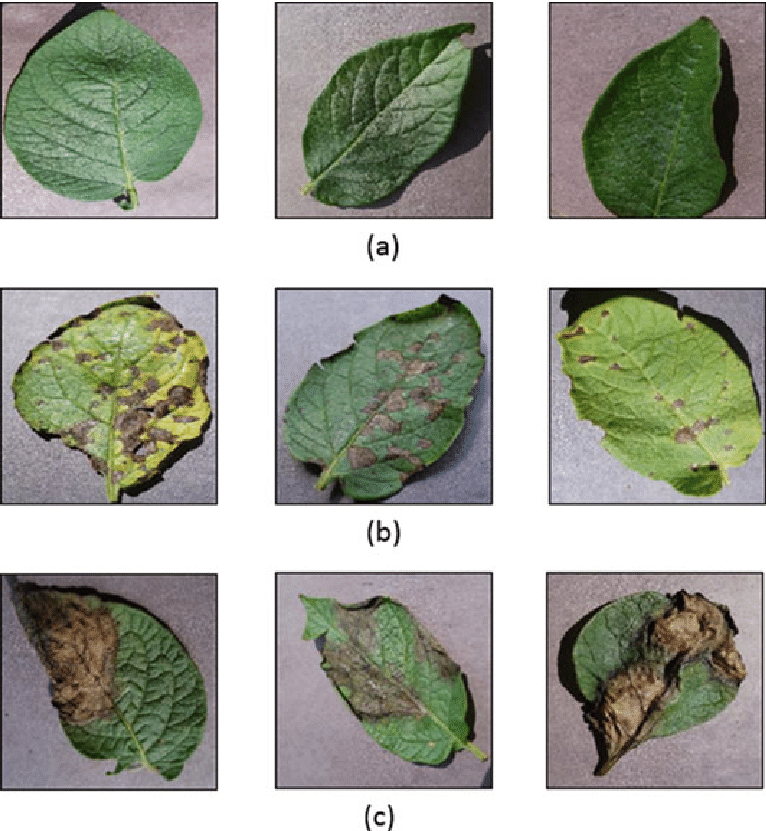
**Methodology**

## Dataset

For this project we have used public dataset for plant leaf disease detection called potato disease. This dataset consists of 1500 image files of 3 different classes, namely early blight, late blight, and healthy. This dataset includes images of healthy potato leaves and leaves infected with three common potato diseases:

* + - **Early Blight**
    - **Late Blight**
    - **Leaf Curl**
  + It contains a total of **600 images**, with **200 images** for each class.

Some samples from the dataset are shown in Fig. 1.



## Data preprocessing and feature extraction

Data preprocessing is important task in any computer vision-based system. Fig. 2 illustrates the preprocessing steps for each image. To get precise results, some background noise should be removed before extraction of features. Potato Leaf Disease Dataset available on Kaggle contains images of both healthy and infected potato leaves. These images are used for training deep learning models to classify the disease status of potato plants based on their leaves. Let’s delve into the details:

**Classes**:

The dataset categorizes leaves into three classes (test, train, valid)

**Healthy Potato Leaves**: Images of healthy, disease-free potato leaves.

**Infected Potato Leaves (Two Classes)**:

* **Early Blight**: Leaves affected by early blight disease.
* [**Late Blight**: Leaves affected by late blight disease](https://www.kaggle.com/datasets/muhammadardiputra/potato-leaf-disease-dataset).

## Feature selection

Feature selection is a crucial step in building effective machine learning models using the Kaggle Potato Leaf Disease Dataset. In this context, feature selection refers to the process of **choosing relevant attributes (features)** from the dataset to use as input for training the model. Here are some key points related to feature selection:

**Relevance**: Not all features contribute equally to the model’s performance. By selecting relevant features, we aim to improve the model’s accuracy and reduce overfitting. For instance, features related to leaf color, texture, shape, and vein patterns could be informative for disease classification.

**Dimensionality Reduction**: Feature selection helps manage the **curse of dimensionality**. With many features, models may become computationally expensive and prone to overfitting. Selecting a subset of relevant features reduces the dimensionality of the problem.

**Domain Knowledge**: Understanding the domain (in this case, plant pathology) is essential. Experts can guide the selection process by identifying features that directly impact disease detection. For potato leaves, features like lesion size, color variations, and leaf area could be relevant.

**Exploratory Data Analysis (EDA)**: EDA helps identify features with high variance, strong correlations, or discriminatory power. Techniques like scatter plots, histograms, and correlation matrices assist in making informed decisions about feature inclusion.

**Model Performance**: Iteratively selecting and evaluating features can lead to better model performance. Techniques like **feature importance scores** from tree-based models guide the selection process.

In summary, thoughtful feature selection ensures that our model focuses on the most informative aspects of the potato leaf images, leading to accurate disease classification and effective management. After feature selection, the data is now parsed to machine learning classifiers to find the patterns in the data.

## Classification Algorithm

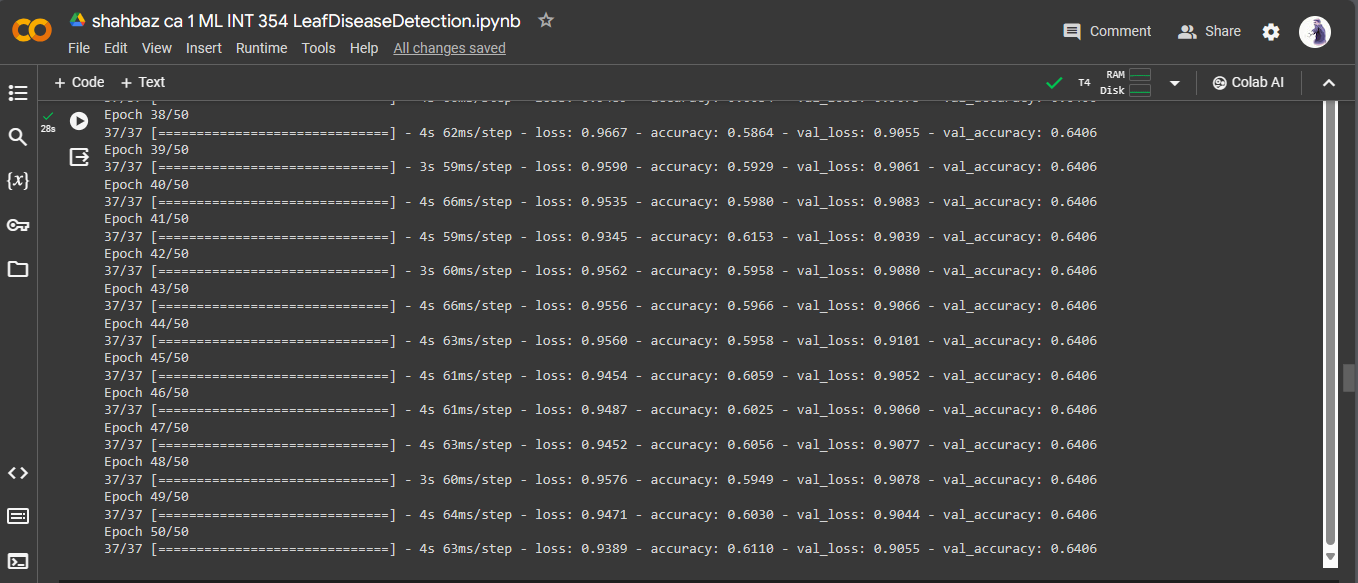
Random forest classifier has been used for classification or detection task. It is the part of ensemble learning, where the output is predicted from multiple base estimators. To achieve higher accuracies, decision trees are used. But they are prone to overfitting problems. So, to overcome this issue, random forest classifier is used which is a combination of multiple decision trees. Each tree is trained by using different subsets of the whole dataset, this can reduce the overfitting and improves the accuracy of the classifier. We have splatted the dataset into train set (80%) for fitting the model and test set (20%) for validation.

**Model Training**:

Deep learning models, particularly Convolutional Neural Networks (CNNs), are trained on this dataset. [The goal is to accurately classify potato leaves into the specified classes, aiding in disease detection and management](https://link.springer.com/article/10.1007/s11042-023-16993-4). After extracting all the features from all the images in the dataset, feature selection task is performed.

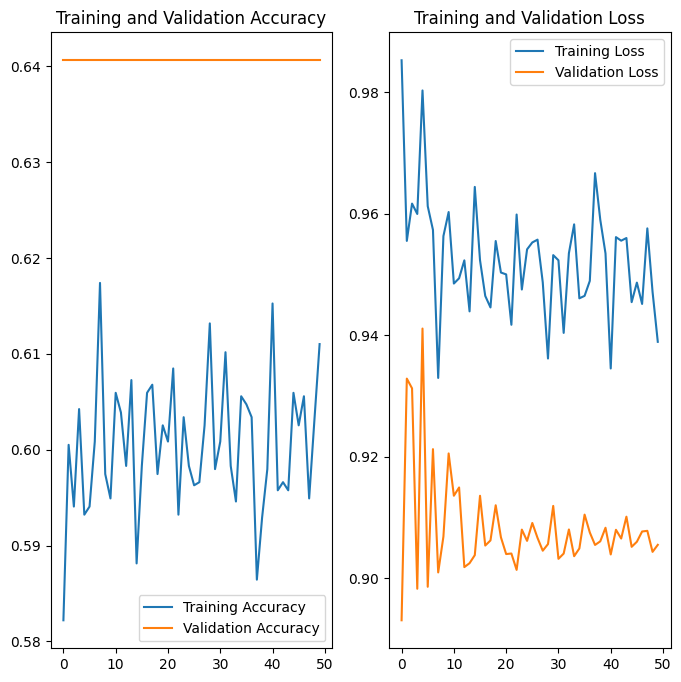
# Results and discussion

Table 2 shows the performance matrices for each model developed for each of the plant. We can observe that the accuracy scores are equal to f1 scores. This is because of balanced number of false negative and false positive predictions. This is considered as best case for any machine learning algorithm. The average accuracy was 60%.



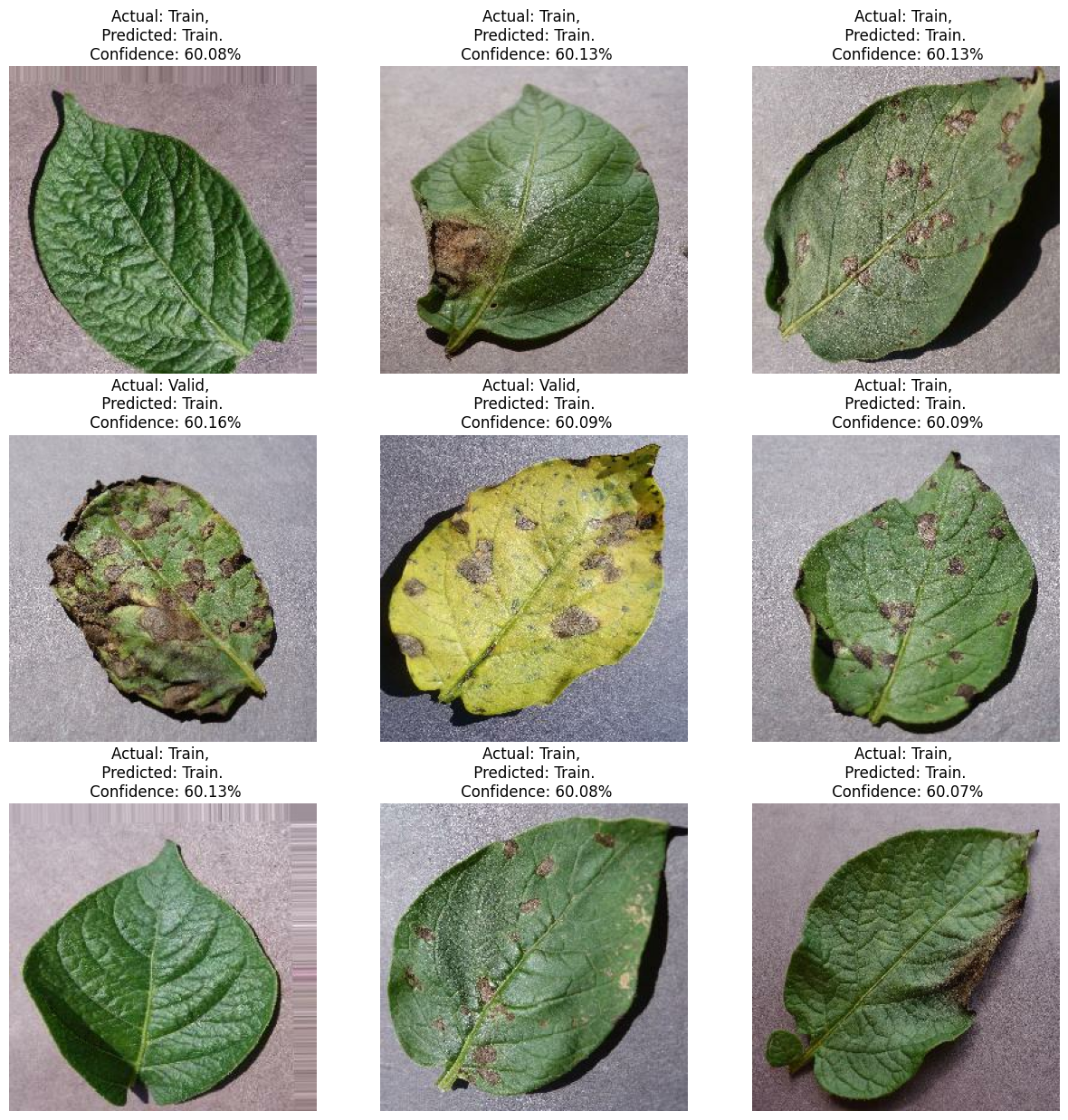
The system successfully detected the disease of leaf. However, we can deploy an intelligent robot vehicle with high end processor attached to it for real time plant disease detection. This system can detect the diseased plants in the agricultural site. Even we can automate the process of spreading the fertilizers by using such robots. Our proposed algorithm is computationally inexpensive, so it can detect the plant disease in efficient manner. Also, sometimes it happens that the farmer also could not identify the disease of the plant. So, they need an expert advice. So, we can deploy a website which can detect the plant disease based on images captured and uploaded by farmer and can give suggestions or can suggest some fertilizers based on detected disease.

**5 Plots**



# 6 Conclusion

We have successfully developed a computer vision-based system for plant disease detection with average 60% accuracy. Also, the proposed system is computationally efficient because of the use of statistical image processing and machine learning model.



We can observe that our technique is accurate and efficient compared with other systems.

Also, it will not require a specialized hardware, makes it cost effective solution.

# 7 References

1. Wikipedia
2. Kaggle
3. Stack Overflow
4. GitHub