Tomato Plant Disease Detection and Classification by Leaf Image Analysis

Concept Note

✅ Background

Plant diseases pose a significant threat to agricultural productivity, leading to substantial crop losses and impacting the livelihoods of farmers. According to a 2014 study by the Associated Chambers of Commerce and Industry of India, annual crop losses due to pests and diseases amount to Rs.50,000 crore[1]. Worldwide crop loss due to plant disease is estimated annually to be $220 billion USD or 14.1%[2].

Tomato plants, being one of the staple crops in India, are particularly vulnerable to a wide variety of diseases, some common examples being late blight, bacterial spot and Tomato mosaic virus. These diseases spread rapidly and can destroy entire harvests if not addressed at the right time.

✅ Problem Statement

Currently, the manual identification of plant diseases is time-consuming, inaccessible, and prone to human error, leading to delayed interventions and potential economic losses for farmers. A major factor contributing to the perpetuation of these diseases is the lack of affordable access to professionals who can diagnose and treat them in rural areas.

Semi-automating this sector using Machine Learning and Artificial Intelligence will allow farmers to improve their crop management practices, optimise resource allocation, and make data-driven decisions for disease prevention and treatment.

✅ Objective

- This project aims to help farmers detect plant diseases at an early stage.

---

[1] https://economictimes.indiatimes.com/news/economy/agriculture/crops-worth-rs-50000-crore-are-lost-a-year-due-to-pest-disease-study/articleshow/30345409.cms

[2] https://www.fao.org/news/story/en/item/1402920/

- Timely detection will allow for corrective measures such as targeted treatments or adjustments to farming practices, which can prevent the further spread of diseases and minimise crop losses.
- This will help boost crop yields and positively impact both the farmer's income and the nation's food security.

✅ Data Sources Used

Dataset of Tomato Leaves - https://data.mendeley.com/datasets/ngdgg79rzb/1

✅ Data Analytics Software/Platforms Used

Python, Google Colab

✅ Python Libraries Used

- Numpy and Pandas - loading and preprocessing data
- Matplotlib and Seaborn - visualisation
- OpenCV and PIL (Pillow) - image processing
- Tensorflow and Keras - loading and training CNN image classification model

---

✅ Probable Visualisations

- Distribution of different disease categories in data set - pie chart
- Example specimens from each category - image grid
- Different dataset configurations used (colour, grayscale, segmented) - image grid
- Prediction accuracy by used model, dataset configuration & disease category - bar chart

---

✅ Methodology

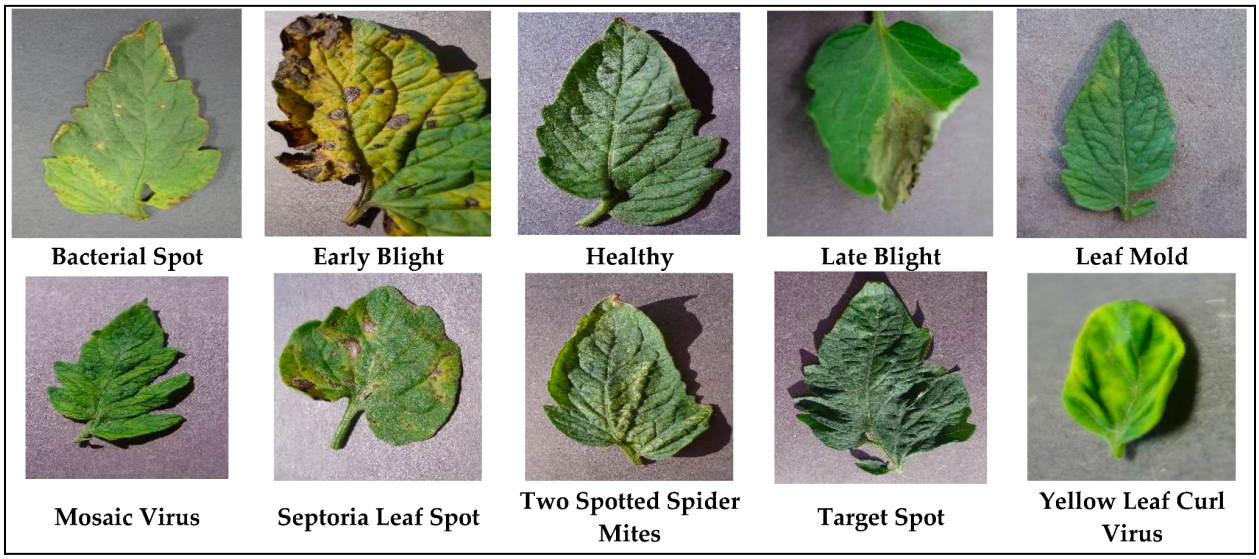
We aim to develop a machine learning model that accurately diagnoses tomato plant diseases using leaf images.

We will start by training a machine learning model on a dataset of more than ten thousand tomato leaf images using a Convolutional Neural Network (CNN), a model type suitable for image classification tasks. The dataset used for this project is "Dataset of Tomato Leaves" by

Mei-Ling Huang and Ya-Han Chang obtained from the Mendeley Data repository. The images in this dataset are sourced from the PlantVillage dataset which consists of 54303 healthy and unhealthy leaf images divided into 38 categories by species and disease. It is available for open access under the creative Commons 3.0 licence.

To avoid overfitting, we will be using the 5-fold cross validation technique for training the model. This technique involves splitting the dataset into five subsets or folds and iteratively training the model on four folds and evaluating it on the remaining fold. This process is repeated five times, with each fold acting as the test set once. By averaging the performance across these five iterations, we can obtain a robust estimation of our model's classification capabilities.

Python libraries like TensorFlow and Keras and cloud based platforms like Google Colab will be used for training. Multiple models will be trained and their accuracy will be compared across multiple relevant metrics like precision and recall.



Images will be classified into the following categories:

1. Bacterial spot
2. Early blight
3. Healthy
4. Late blight
5. Leaf mould
6. Septoria leaf spot
7. Target spot
8. Tomato mosaic virus
9. Tomato yellow leaf curl virus
10. Two-spotted spider mite

The most suitable model will be used to further develop a web app where farmers can upload pictures of their tomato plant leaves and get an instant diagnosis. The data obtained through this platform will be used to further train our model and improve its accuracy.

We will also be exploring lightweight models such as MobileNet to develop an offline app so that the logistical limitations of typical farm locations like limited network connectivity do not limit accessibility.

More plants and diseases will be added if this project achieves initial success.

✅ Probable Outcome

The outcome we expect from this project is a machine learning model that can accurately detect tomato plant diseases by analysing images of leaves. This model will be made publicly available through a user-friendly web application and/or android application allowing farmers to upload pictures of tomato plant leaves and instantly receive a diagnosis of any diseases present.

Team Name - Quantitative Conquerors
Members - Pranav Mittal, Dinesh Thakur, Sarthak Goel, Anupam Kumar, Mohit