

Project Problem Statement

Agriculture plays a vital role in sustaining the global economy and ensuring food security. However, one of the most significant threats faced by farmers worldwide is plant disease, which can drastically reduce crop yields and quality. Traditional methods of disease identification rely on manual observation by agricultural experts, which is often time-consuming, error-prone, and inaccessible to small-scale farmers.

With the rise of Artificial Intelligence (AI) and Deep Learning, it is now possible to create intelligent systems capable of identifying plant diseases automatically from images. Using Convolutional Neural Networks (CNNs), these systems can analyze leaf images, detect disease patterns, and predict the type of infection accurately.

To address this problem, there is a need for an intelligent, automated system that can classify plant leaf images into healthy or diseased categories. Such a system will support farmers in early detection, reduce dependence on manual inspection, and promote sustainable agriculture by minimizing pesticide misuse and improving crop health management.

DataSet

Dataset Name:

PlantVillage Dataset – A Comprehensive Image Dataset for Plant Disease Classification

About Dataset:

This dataset contains **over 50,000 high-quality images** of plant leaves collected under controlled conditions. The images represent a variety of **plants and diseases**, making it ideal for building **machine learning and deep learning models** focused on plant disease identification.

Key Details:

- **Number of Classes:** 38 classes (covering healthy and diseased leaves across multiple plant species)
- **Image Type:** RGB leaf images
- **Plants Included:** Apple, Tomato, Corn, Grape, Potato, etc.
- **Categories:** Each plant species has “healthy” and multiple “disease” subclasses (e.g., *Tomato Early Blight*, *Tomato Leaf Mold*).
- **Image Format:** JPEG / PNG
- **Image Size:** Approximately 256×256 pixels

Source:

[Kaggle – PlantVillage Dataset](#)

Next Steps

- 1) **Collect & Prepare Dataset** – Download and organize images into training, validation, and test folders.
- 2) **Train the CNN Model** – Use **Google Teachable Machine** or **TensorFlow/Keras** to train a deep learning model for disease classification.
- 3) **Evaluate & Test the Model** – Measure model accuracy, precision, recall, and confusion matrix performance.
- 4) **Build the Web Interface** – Develop a simple web app where users can upload a leaf image and get the predicted disease label.
- 5) **Test the Complete Application** – Ensure that the interface and model work seamlessly together.
- 6) **Deploy & Document** – Deploy on a cloud platform (e.g., Streamlit, Flask, or Heroku) and document the results.