Data Collection and Preprocessing Phase

Date	6 July 2024
Team ID	SWTID1720076593
Project Title	Visual Diagnostics: Detecting Tomato Plant Diseases through Leaf Image Analysis
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Detecting and diagnosing tomato plant diseases at an early stage is crucial for effective crop management and yield optimization. This project aims to develop a deep learning model to accurately identify various diseases in tomato plants using leaf images. Leveraging a comprehensive dataset from Kaggle, this project will ensure meticulous data curation and integrity for informed decision-making.

Data Collection Plan Template

Section	Description
Project Overview	By using deep learning, the project can automatically tell if a plant is sick from a picture of its leaf. This can help farmers and plant experts find problems faster and easier, so they can keep their tomato plants healthy. **Objectives:** • Develop a robust deep learning model to classify tomato leaf images into different disease categories. • Enhance the dataset with preprocessing techniques to improve model performance. • Evaluate the model's accuracy and reliability using metrics. • Provide actionable insights to farmers and agronomists for better crop management.

Data Collection Plan	**Source:** Kaggle (Tomato Diseases Dataset) • **Content:** Annotated images of tomato leaves, categorized into healthy and various diseased classes. • **Classes:** Includes bacterial spot, early blight, late blight, leaf mold, septoria leaf spot, spider mites, target spot, mosaic virus, and yellow leaf curl virus. • **Number of Images:** Approximately 18,000 images. • **Image Size:** Typically around 256x256 pixels.
Raw Data Sources Identified	**Kaggle Dataset:** • The primary source of images for this project. • Provides labeled images for various tomato plant diseases.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset 1	The Kaggle Tomato Diseases Dataset provides a comprehensive collection of images depicting various conditions of tomato leaves, including healthy leaves and those affected by different diseases. This dataset is instrumental in developing and training deep learning models for disease detection and classification.	https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf	Image	188 MB	Public
Kaggle Dataset 2	This collection of pictures on Kaggle shows tomato leaves. Some leaves are healthy, but others have different diseases. You can use these pictures to train computers to identify these diseases, like spotting early blight or something called septoria leaf spot. This helps people who study plants and farmers	https://www.kaggle.com/datasets/ashishmotwani/tomato	Image	1 GB	Public