Project Title:

Leaf disease prediction (Tomato)

Full name: Catarina Ieze Vuzi

Report Submission Date: 10/18/2024

1. Introduction

Over the past two weeks, I have made significant progress on my project focused on Leaf disease Prediction using deep learning. The primary objective was to explore the data to gain a deeper understanding of the project and extract useful insights. I successfully navigated through the project folders, and observed its structure. The dataset is splitted into test, train and validation. Inside of each folder there are sub folders with images of different plant leaves. For this project I will be focusing only on tomato leaves . This report provides a detailed overview of the progress made on the project, including any challenges faced and a step-by-step account of the actions taken.

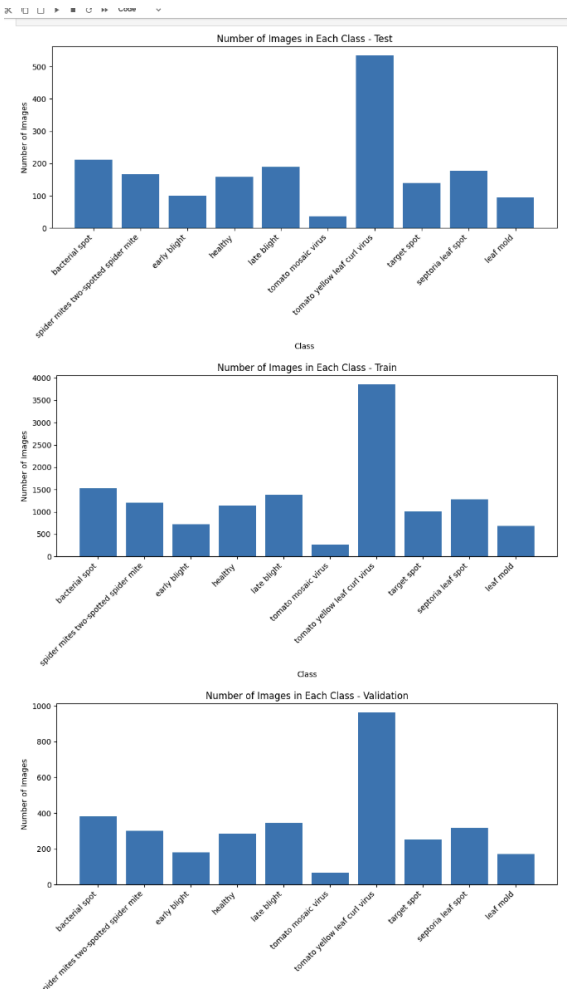
2. Summary

- **Summary of work Done**

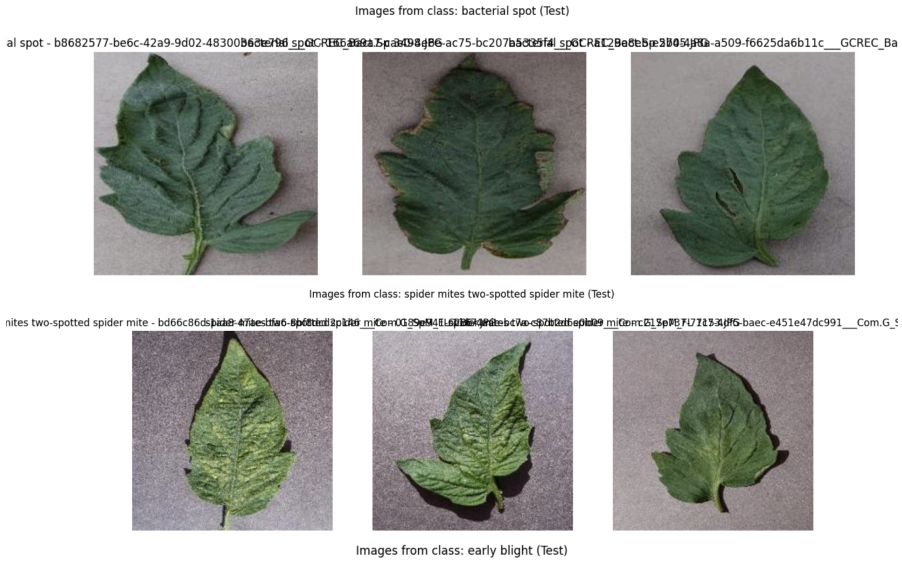
The current completed tasks in my project include data exploration and initial image data preprocessing. For data exploration, I began by examining the entire folder structure, as mentioned in the above paragraph the folder is splitted into test, train and validation. Additionally, I created a barplot to visualize the number of images in each class. I noticed that there is an imbalance between classes.The dataset does not need too much preprocessing since

the images are noiseless and clean. The images had all the same sizes, but still I decided to resize it to 224, 224 and normalize the pixel to [0,1].

- **Graphs and Visualization**



- **Some of the images**



3. Progress and Milestone

The completed tasks in the project include data exploration and partial data preprocessing. I examined the folder structure, created visualization for a better understanding of the dataset . For data preprocessing, I executed image resizing and normalization. The next pending task is to deal with imbalance and create a base model with the imbalance data and without data augmentation.

4. Problem-Solving and Challenges

The project challenge is working with an imbalance dataset. There are more images of tomato leaf curl virus than the rest of leaf diseases and the healthy leaf images. This can cause bias during the model phase. This problem will be solved by applying techniques like oversampling or undersampling to handle the imbalance.

5. Technical depth

The images were resized and normalized using a function. The function resizes all images within a specified folder to a new size. For each image, it opens the file using the Python Imaging Library (PIL), resizes it to the specified dimensions (defaulting to 224x224 pixels), and then saves the resized image, overwriting the original file. This approach ensures that all images in the folder are uniformly resized, which can be particularly useful for preparing datasets for the deep learning part.

6. Future Plans Goals

The next steps will be focused on creating a benchmark model with the imbalance data a and without data augmentation to evaluate its performance. After that, handling an imbalanced dataset using different techniques. Finally, continue to learn more about the dataset for a better understanding of the project.