Capstone Three - Project Proposal

Plant Disease Recognition Using Deep Learning

Problem statement formation

The goal of this project is to develop algorithms that can accurately diagnose a plant disease based on an image.We will use deep learning for disease recognition on the plant dataset using deep learning and PyTorch.

Context

Plant diseases, a deterioration in the original condition of the plant that upsets and changes its vital functions, are common problems that farmers and agronomists face from season to season.Developing algorithms that can accurately diagnose a disease based on an image is the next big disease diagnostics tool which assists in turning a vision of productive agriculture into reality. The aim of this project is to create a framework that will detect and classify plant diseases.We are using augmented dataset available on kaggle.

The dataset contains 38 classes of crop disease pairs.

14 unique varieties of plants.

There are 70295 images for training.

Criteria for success

The success of the project is predicated on the accuracy of the classification results and the extent of analysis conducted.

Scope of solution space

The focus of the project will be to identify a model that can best differentiate infected vs healthy plants based on leaf pictures.

Constraints

We are currently constrained to the classification of single leaves, facing up, on a homogeneous background. While these are straightforward conditions, a real world application should be able to classify images of a disease as it presents itself directly on the plant. Indeed, many diseases don’t present themselves on the upper side of leaves only (or at all), but on many different parts of the plant. Thus, new image

collection efforts should try to obtain images from many different perspectives, and ideally from settings that are as realistic as possible.

Data sources

https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset

This dataset is recreated using offline augmentation from the original dataset. The original dataset can be found on this github repo. This dataset consists of about 87K rgb images of healthy and diseased crop leaves which is categorised into 38 different classes. The total dataset is divided into 80/20 ratio of training and validation set preserving the directory structure. A new directory containing 33 test images is created later for prediction purpose.

Problem approach

1. Importing the data set
    - Have a General Overview of the data

2. Exploratory Data Analysis

    - Perform Exploratory Data Analysis(EDA) to gain more clear insights of the data

3. Data Preparation for training

    - Normalize the image data
    - Helper functions to use GPU

4.Building model architecture

    - Perform classification using Residual Network Resnet
    - Hyperparameter tuning
    - Evaluate model using Accuracy function which calculates the overall accuracy of the model on an entire batch of outputs,also calculating validation training losses.

6. Results

    - Evaluate predictions on test data

Deliverables

- A GitHub repo containing the work completed for each step of the project.
    a. A slide deck
    b. A project report