AI 539 Final Project:
Botanist

Quick Overview

Each team can have a maximum number of 2 members, and would need to perform both projects. .

In this project you would be dealing with image data. This is a classification problem, where the ultimate goal is to train a model which is fed with an image of a leaf, and predicts a label corresponding to the type and decease of the plant. There are 38 classes to predict from.

Problem Setup

Many of you who do your own landscaping and gardening may have thought about this before. Every time you see a weird spot on your favourite plant, you get concerned whether this could be a serious issue or something that goes away with time. In this project, you would be designing a prototype system, which does a botanist's job. It simply takes an image of the plant (specifically the leaf), and predicts the class of problem going on with the plant.

In this project, we are planning to come up with such prototype system. Due to the limited computational resources, we would only explore 38 class labels, although upon the

availability of more data the approach can be extended to a more commercialized product. We would like to note that the 38 labels considered in this problem do not all correspond to deceases. For example, a label 4 corresponds to a healthy apple, and label 22 indicates that we have a healthy potato. In case it helps with your classification, the list below summarizes what plants are the subject of which labels:
Apple: 1, 2, 3, 4
Blueberry: 5
Cherry: 6, 7
Corn: 8, 9, 10,11
Grape: 12, 13, 14, 15
Orange: 16
Peach: 17, 18
Pepper: 19, 20
Potato: 21, 22, 23
Raspberry: 24
Soybean: 25
Squash: 26
Strawberry: 27, 28
Tomato: 29, 30, 31, 32, 33, 34, 35, 36, 37, 38

For technical reasons related to the competition, no other information about the labels is available.

Modeling Instructions

You would need to use the shared images to train a classification model that accepts a leaf image and predicts the corresponding label (an integer between 1 to 38). You are only allowed to use the dataset provided and augmenting the data from external resources is not acceptable. The training labels and the file names are available in `Botanist_Training_Set.csv`.

Please make sure to communicate with the instructor and Piazza about any potential questions related to the data.