DS 4002 Case Study Spring 2025

Tomato Trouble: Using Machine Learning to Combat Crop Disease

In a world where global food systems are becoming increasingly fragile, and climate change is accelerating the spread of crop diseases, the need for fast, accessible agricultural technology has never been greater. In many regions, small-scale farmers do not have access to expensive diagnostic tools, instead relying on visual cues and local knowledge to identify problems in their crop. Yet early detection of crop diseases can be the difference between a successful harvest and devastating loss.

This is where you come in.

You are part of a growing network of data scientists working to bridge the gap between advanced research and practical, real-world solutions. Your challenge is to design a robust, image-based machine learning system that can detect diseases in tomato plants, one of the world's most vital crops. Using the PlantVillage dataset, which contains thousands of labeled leaf images, you will harness computer vision to recognize and classify several types of tomato leaf disease.

As part of your work, you will build multiple models. This includes three models that apply transfer learning using pre-trained convolutional neural networks (ResNet50, VGG19, and InceptionV3), as well as a support vector classifier (SVC) for comparison. Through experimentation, visualization, and performance analysis, you will examine how these different architectures handle complex image data and explore the strengths and limitations of each approach.

The Deliverable:

You have been asked by a nonprofit agricultural tech startup to deliver a working prototype of a tomato disease detection tool. Your submission will include multiple models, along with the code, analysis, and visualizations needed to clearly communicate how each model performs. These visualizations should help explain your findings and provide insight into the strengths and weaknesses of your approach. While your focus is on tomatoes, your work could eventually be extended to other crops and broader agricultural challenges. This project is more than a classification task; it is a chance to create something meaningful, with the potential to support real farmers and influence the future of agriculture through machine learning.

GitHub Starter Repo: https://github.com/ytkidanu/CS2-DS4002-SPR25