📖 **write-up**

Farmer Crop Disease Aid

What is my research question?

How can we better aid farmers detect the early signs of crop disease? Can we make their job more automated in someway related to finding diseases earlier, possibly reducing labor costs and/or loss of crops?

What is and where to collect you data?

The dataset will be downloaded through AWS.

The instructions for this can be found here.

This dataset is a collection of crop images used for a competition. Called 'Agriculture-Vision'. The challenge is basically what I am trying to do. I find it extremely interesting.

How does machine learning and data mining help my project?

Training

Training these models will be very computationally expensive. We will Hadoop's map reduce functionality to execute distributed programming letting us create our models significantly faster.

Incorporating a CNN

I plan to use CNN's to help me achieve the goal of the classification. I want to use K-means clustering based on RGB values. To enhance this, maybe I can apply a filter that ignores the blue value. I believe this could work since the healthy crops would have a large amounts of green because of their chlorophyll. I need to do more research on this, but operations like this could be of interest. I would possible narrow the scope of this

project to the effects of a particular disease, but would have to carefully inspect the dataset to see what the data annotation support. Depending on that, I can warp the project accordingly.

Looking at the dataset, we would need to apply filters to make sure we account for the time of day and segmentation to avoid noise like random roadways.

Incorporating a LLM

A possible way to incorporate an LLM would be to pass the results of the CNN into a LLM, and have the LLM give a more natural language response. For example, when we get a huge list of results giving us the probability of each photo, we can have the LLM tell us which photos to look at, and maybe even which fields they are associated with. This will make the results from the CNN much more understandable to someone who maybe is not as tech-savvy. I could not find a dataset that would support this kind of training, but maybe we can train it off the data the actual users provide over time. Maybe a year after deployment we can use the training set to train a LLM.

Overtime we can have the user tag the photos (which would have to be geotagged) manually. After sufficient data is collected, we can use that data to train out model. An OpenAI wrapper would seem appropriate for this.

What is the expected outcome of your project?

The outcome will be binary output for each photo provided. Since this will be my first CNN project, I want to keep it simple to start out. I want it to take in a photo, then output binary. The 0 class will be the normal class, healthy. The one will be the anomaly class, unhealthy in some capacity. Signaling to the farmer that this area of crops needs to be looked at. This will make the maintenance of keeping the crops healthy much easier. Benefits from this project will include:

1. Drone support for analysis - birds eye view
2. Allow RGB farmer to spend either less time, or money on manual labor looking through the crops. Even searching through the photos, the model will do this
3. More accurate predictions that humans, as the RGB values will be judged more objectively

After the CNN portion is completed, such as collected photos for the day and such, an LLM will look at the results and the geotags of the photos thus far, and give the farmer a

GPS coordinate to look at. This coordinate will be the cluster most likely to have some sort of disease, which hopefully the LLM will be able to explain.