Proposal

ML 2 Project Proposal

Authors: Nigel Martis, Rohan Paul, Pranay Bhakthula

• What problem did you select and why did you select it?

The problem we selected for this project is a multiclass Cotton Disease Prediction Task. There is a need for a system which can automatically detect the diseases as it can bring revolution in monitoring large fields of crops and then the leaves of these plants can be taken care of as soon as possible. This can reduce the cost which is incurred while agricultural specialists keep track of this disease.

• What database/dataset will you use?

The dataset that we will use is from Kaggle. It is split into train, test and validation sets with ~ 4000 images and 4 classes. This is also an imbalanced data with one class being the minority class. The 4 classes are diseased cotton leaf, diseased cotton plan, fresh cotton leaf and fresh cotton plant.

• What methods will you pick from the concept list? Will it be a classical model or will you have to customize it?

We are thinking of using at least 3-4 pre-trained models along with creating a CNN model from scratch and then we might ensemble them and check for performance.

• What packages are you planning to use? Why?

The packages we will use are NumPy, Pandas, Matplotlib, and PyTorch. NumPy and Pandas will be used for data cleaning and preprocessing, Matplotlib will be used for visualizing our findings, PyTorch used to implement, train, and test our models.

• What tasks will you work on?

The reference materials we will use to obtain sufficient background on applying the chosen network to our vehicle image classification are the NLP course materials, the official documentation websites for the packages we will be using, and various websites such as medium.com.

• How will you judge the performance of the network? What metrics will you use?

We will judge the performance of our network by checking the accuracy score of our test dataset. The metrics we will calculate are precision, recall, F1. We may use accuracy as a metric, dependent on how balanced the dataset is. This will be examined during exploratory data analysis.

• Provide a rough schedule for completing the project

The table below outlines the progress and deliverables that we plan to achieve at each date.

Proposed Date of Completion	Actions/Deliverables
Tuesday, November 2	- Decide on a dataset - Create GitHub repository
Tuesday, November 9	- Submit group proposal - Begin exploratory data analysis
Tuesday, November 16	Clean and process the data for train-test and apply augmentation.
Tuesday, November 23	- Evaluate the results of the models - Refine models if needed - Begin writing the final report
Tuesday, November 30	- Create GUI - Create presentation

Sunday, December 2	● Finalize final report ● Individual final reports
Monday, November 6	● Final Project