PROJECT PROPOSAL

Ridhi Bandaru, Adam Abdelrehim
CSE 5717

Motivation: In developing nations, agriculture confronts hurdles like knowledge gaps and pest issues. This project deploys AI for transformative solutions. The problem addressed by this project lies in the multifaceted challenges faced by the agriculture sector in developing countries. Key issues include a significant knowledge gap between farmers and technology, prevalent disease and pest infestations, and inadequate storage facilities. These challenges hinder agricultural productivity and sustainability.

Goal: The overarching goal of the project is to leverage AI to mitigate these problems by controlling crop pests and diseases, reducing costs, and enhancing crop yields. By bridging the knowledge gap and providing technological solutions, the project aims to contribute to the overall improvement of agriculture in developing regions.

Dataset: The dataset we use is bifurcated into raw and augmented images, totaling 127,857 samples encompassing cashew, cassava, maize, and tomato. Raw images comprise 24,881 instances, while augmented images are categorized into 22 classes, summing up to 102,976 images. Validated by plant virologists, these de-identified images from the University of Energy and Natural Resources in Ghana are openly accessible, facilitating research within the scientific community. Cashew is categorized into five classes, including anthracnose, gummosis, healthy, leaf miner, and red rust. Cassava exhibits five classes: bacterial blight, brown spot, green mite, healthy, and mosaic. Maize is divided into seven classes, encompassing fall armyworm, grasshopper, healthy, leaf beetle, leaf blight, leaf spot, and streak virus. Tomato presents five classes: healthy, leaf blight, leaf curl, septoria leaf spot, and verticillium wilt. The images were captured in diverse conditions and backgrounds, ranging from white and dark to illuminated and real settings.

Methods: We propose to experiment with Support Vector Machines (SVM) for multi-class classification, as SVM works well with high-dimensional data. We might also use Random Forest for handling complex relationships in data, especially with a large number of features. Next, Decision Trees: simple and interpretable, decision trees can be useful for understanding feature importance. Lastly, we will use Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN).

Performance Measures: We will use accuracy, precision, recall, F1 score, and create confusion matrices for each crop. We will also measure the area under the ROC curve.