Machine Learning Engineer Nanodegree

Capstone Project

Plant Disease Detection

Hesham Abdallah

October 2019

Domain Background

Food losses due to crop infections and pests are persistent issues in agriculture for centuries across the globe. In order to minimize the disease induced damage in crops during growth, harvest and postharvest processing, as well as to maximize productivity and ensure agricultural sustainability. Globally, over one billion people are suffering from different situations of malnutrition due to lack of food supply and approximately twice that population do not have access to sufficient nutrients or vitamins to meet their daily nutrition needs. To cut down the wasted agriculture crops due to pests and disease, a plant disease detection can help with this task. The main goal for this model is early detection of the common plant disease for contamination and reduce the waste.

For instance, in my homeland Egypt agriculture is one of the main pillars of the Egyptian economy; in 2012, agriculture contributed with 13.4% of the national gross domestic production and employed 27.1% of the total Egyptian workforce. Such an application will have a major impact on both of food security side and the economic side.

Problem Statement

It is estimated that the agriculture waste due to plant diseases and pests is ranging between 20 and 40% of global agricultural productivity. Cutting down this waste will be helpful as the demand for food will continue to increase as the population increasing rapidly. The projections indicate that an additional 70% of food production is required by 2050 to meet the needs.

The challenge is to predict whether the captured pictures of the plants is healthy or infected, this is done by detecting the infection and bacterial patterns that the model will learn. This is a classification problem with supervised learning technique.

Datasets and Inputs

After a long search, I found an opensource dataset called" Plant Village Dataset "on Kaggle. It contains thousands of samples of healthy plants and infected plants. The dataset contains accurate disease classification for each plant species as shown in the below samples. The available species in the dataset is the widely cultivated and consumed around the world.

The dataset consists of 18335 images of 256*256 pixels each. due to the high difference of the number of samples of each class and to reduce the class imbalance, I will choose a fixed number of images from each class. The dataset will still have a minor class imbalance, but it will not be a huge problem as we are actually focusing on the majority class (the infected class). The total size of the dataset is 326 MB.

Figure 1 Healthy Plant

Figure 2 Inffected plant

Solution Statement

the main goal here is to detect Crop diseases using CNN model that is capable of detecting the infected plant and the healthy plant with decent accuracy. Early detection will be crucial to minimize the wasted production. For broader access, the model can be deployed as a mobile application to be used by the farmers around the world.

Benchmark Model

The benchmark model will be a CNN created from scratch. The structure will the following:

- 3 Convolutional layers separated by MaxPooling and Dropout layers with RELU activation functions
- 2 Fully connected layers. The final one will have a dense equal to the number of classes with SoftMax activation function

The performance of the model will be evaluated by the following metric.

Evaluation Metrics

Accuracy will be the evaluation metric. By plotting the accuracy through each epoch we will have a clear view of the model's performance.

Project Design

The model development steps will be the following:

-**Data Acquisition**: Download the dataset.

-Data preprocessing : reshaping the images and other preprocessing steps that help with the training like one-hot encoding and normalization

-**Data Split**: Split the data into training, testing and validation sets.

-**Build the model architecture:** using layers of Conv2D, MaxPooling2D, Dropout and Dense layers.

-**Training**: Training model on training set and then evaluate its predictions using loss and accuracy scores.
-Transfer Learning models: we create other model(s) with transfer learning approach using the pretrained models like Inception-V3, RESNET 50, VGG-16, VGG-19.
-**Comparing**: Compare the results with the result of benchmark model and decide whether our model made improvements or not.

References:

- https://www.researchgate.net/publication/257788783_Crop_losses_due_to_diseases_and_their_implications_for_global_food_production_losses_and_food_security
- https://www.bibalex.org/SCIplanet/en/Article/Details?id=5181
- Dataset: https://www.kaggle.com/emmarex/plantdisease