CV PROJECT REPORT

TOPIC: Plant Disease Detection

GROUP MEMBERS:

Haris Jafar i191690

Muhammad Millwala i191735

1. ABSTRACT:

Problem Introduction:

Our problem is based on detecting diseases in plants e.g Apple trees. Through our computer vision model, we will analyse the health of an apple tree leaf- such as if it is healthy or has diseases like scab, rust, or having more than 1 diseases.

Problem Statement:

We all know how vital a booming agriculture infrastructure is for any country's economy. Having diseases in certain plants and their leaves is a common finding, due to which crop and yield productivity begins to diminish, causing a major loss. Therefore, it is important to stop the spread of such a disease in its track, rather than facing such a loss later. Hence, there's big need in the market for an automatic disease detection system. So, we have proposed an automatic detection and diagnosis algorithm using deep learning classification, which will help stop infection and the further use of chemicals/pesticides on the plants.

Contribution and Results:

We used Plant Pathology Apple Tree Disease Dataset from Kaggle. This dataset contains about 3800 train images and test images combined. With each image containing either disease such as scab, rust, more than 1 disease or being healthy and containing no disease. We used Adam optimiser and ran 30 epochs with accuracy over 90 percent, successfully classifying diseases and its health.

2. INTRODUCTION AND BACKGROUND:

As said previously, our problem is based on detecting diseases in plants e.g. Apple trees. Through our computer vision model, we will analyse the health of an apple tree leaf- such as if it is healthy or has diseases like scab, rust, or having more than 1 diseases.

Moreover, over the years there has been work done along this and in this domain as well. Recently, there has been a major break - through in classifying diseases in plants using strong image processing convolution neural networks. This work includes classification, detection, and segmentation of the infections. In summary, in the field of plant diseases and pests detection which emphasizes detection accuracy at this stage, more models based on two-stage CNN is used where we first obtain the feature map of our image through backbone network, then calculate the anchor box confidence using RPN to get the proposal, and in the field of plant diseases and pests detection which pursue detection speed more models based on one-stage are used, where region proposal is removed by directly add detection to our backbone network to classify etc. Finally, through our contribution, we have successfully and accurately diagnosed and classified whether an apple leaf is healthy, has diseases, and what type of disease to be exact. Our customised model has a fast convergence rate, and comparatively small number of parameters. We believe our project has done fairness in detecting the health of an apple tree leaf, and can be potentially used in apple tree cultivation smart systems.

3. Motivation

When it comes to accurately classifying human diseases, a great deal of research and development has been done; nevertheless, when it comes to classifying diseases in plants, this field is not given much priority. Misdiagnosis of the various illnesses that affect agricultural crops can result in the abuse of pesticides, the creation of pathogen strains that are resistant to those chemicals, increased input costs, and additional outbreaks that have serious negative effects on the economy and the environment. Although computer-vision based models have the potential to increase efficiency, the current disease diagnosis method relies on human scouting, which is time-consuming and expensive. The great variation in symptoms caused by the age of infected tissues, genetic variations, and light conditions within trees reduces the accuracy of detection making it far more susceptible to classify the disease.

4. Results

We created and trained our model using Python3. We used Python as well as the features of several frameworks and libraries, including Keras, Tensorflow, Pandas, and Numpy, to achieve our goals. We used Google Colab, which provides limited free access to GPU. Google created Tensor Processing Units, which are specially created application-specific integrated circuits (ASICs) to speed up tasks related to deep learning and machine learning. The dataset used to train our model was taken from Kaggle.

The pilot data set consists of 3800 high-quality annotated RGB photos of apple leaves, some of which demonstrate complex disease symptoms (i.e., more than one illness on the same leaf), such as cedar apple rust and apple scab, as well as healthy apple leaves. Images were captured in a variety of lighting, angle, surface, and noise circumstances and depict actual field scenarios.

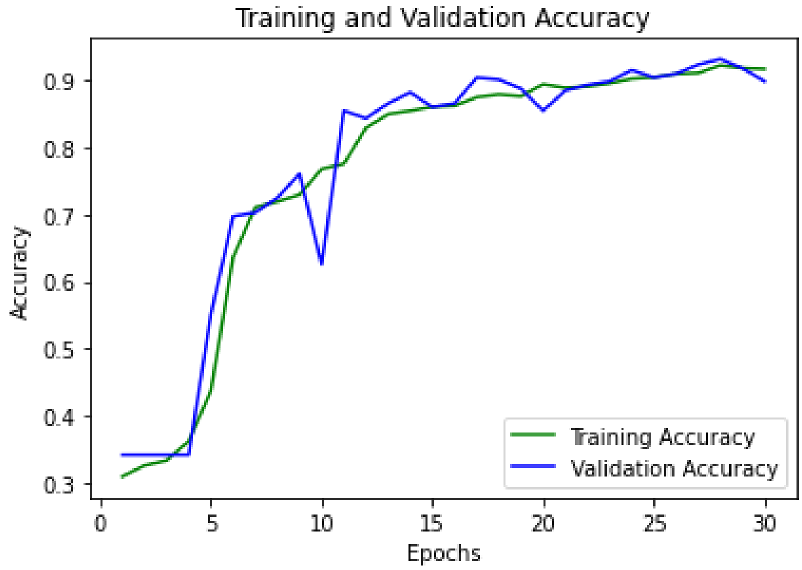
The fact that the data set was very small could have contributed to the models' high accuracy ratings. The images of the foliar symptoms in the data set, on the other hand, were taken over the course of four months from a large number of apple cultivars under various lighting conditions. They represented visually diverse symptoms, from early to late stage of infection on young and old leaves in each disease category. Additionally, in order to lessen the bias toward elements like identical cultivars, leaf kinds, and imaging settings, 80% and 20% of the photos, respectively, were randomly assigned to the training and test data sets.

In order to classify apple scab, cedar apple rust, complicated disease signs (leaves with multiple diseases in the same leaf), and healthy leaves, we trained an off-the-shelf CNN on this pilot data set. We specifically used our annotated disease data set to fine-tune the network weights on a custom network which contained 5 Convolution layers and 5 Max pool layers. The activation function used was Relu and in the end Softmax was used to normalise the probabilities of what class a certain leaf belonged to for far accurate results.

We loaded our model using Keras' sequential class and set image net weights. Functional calls to the necessary metrics, such as accuracy, F1, etc., were made while the model was being loaded. We used a batch size of 16 and ran about 30 epochs with 30 steps each. The findings were obtained using the history object that was retrieved after our models had been fitted. The total number of parameters were 821,700.

Our custom built CNN achieved an overall test accuracy of 97% (i.e., 97% of test images were properly classified), with the network making high accuracy predictions on most categories. As can be seen in the figure below:

The accuracy increases exponentially as the epochs are increased from 5 – 15 and then accuracy reaches more than 90% from epoch 25-30.



If we take a look at our results and perform an error analysis, our findings were as follows: Majority of the data points/images were classified accurately only a few were labelled incorrectly or were not able to detect multiple diseases in one leaf image. This resulted the accuracy to lie between 85% - 90%. Further increasing epochs decreased the overall loss and hence gave rise to far accurate results. The dataset contained high quality images and was uniformly oriented therefore decreased the chances of error.