***Abstract***

**Deep Learning Project in Agricultural domain.**

**It includes Data collection and Model building.**

**The common disease of Bell\_Pepper plant is Bell\_Pepper leaf with Bacterial spot. If a farmer can detect these diseases early and apply appropriate treatment, it can save lot of waste and prevent the economic loss.**

**Atliq Agriculture is an AI company that happen to solve problems in Agricultural domain. It uses Deep Learning(DL) and Convolutional Neural Network(CNN).**

**Any supervised Machine learning project starts with Data collection. The data here we use as a Training Dataset.**

**We have collected the images of 1).healthy Bell\_pepper leaf ,and 2).Bell\_pepper leaf with Bacterial spot disease.**

**It is followed by Data cleaning and Preprocessing step, for which we will be using tf dataset and Data Augmentation.**

**Model building is done using Convolutional Neural Network. CNN is a standard way of doing image classification.**

**Data Collection, Pre-processing :**

**-In this project we have used ready-made data from Kaggle for training.**

**-The dataset was downloaded and the directory was named as "PlantVillage".**

**-We launched GitBash which allows us to run all the unix command, which inturn launched Jupyter notebook.**

**-To download the dataset into tf dataset TF data input pipeline, followed by data cleaning and we made our dataset ready for Model Training.**

**-Essential modules were downloaded.**

**-We used Tensorflow dataset to download all the images into tf.data.dataset .**

**-We created couple of constants such as Image\_size, Batch\_size, Channels and EPOCHS .**

**-We split our data into Train and Test.**

**-80% of the data was kept as Training Data.**

**-20% of the data was split into 10% Validation and 10% Test.**

**-Here 10% of the data used for test is used to check the accuracy of our model.**

**-We let Tensorflow determine how many batches to load while GPU is training.**

**-Dataset were optimised for Training performance such that training will run fast.**

**-We supplied preprocessing pipeline using an API which would scale the image.**

**-All the image dimensions were resized to 256 by 256, which is followed by Data augmentation to make our model robust.**

**Model Building:**

**-We trained Convolutional Neural Network in Tensorflow using bell\_pepper leaf images. The goal of this model would be to classify these images either healthy or Bacterial spot disease.**

**-We built a trained CNN model.**

**-Plotted training history on the graph.**

**-Made predictions/inference on sample images.**

1. ***Introduction***

Plants are a significant source of energy. Crop protection and production can be increased using the early and accurate diagnosis of plant diseases. In the old-fashioned environment, the identification is processed whether the plant leaf is healthy or infected either by visual observation or by testing leaves in the laboratory. The visual identification is done by the experts of the plant domain but opinion may vary from expert to expert. Testing of the plant leaf in the laboratory is a very time consuming and strenuous process and hence results may not come on time.

In today's Agriculture, leaf disease is a major issue. It hinders the natural growth of the plant. It will lowers the quality of agricultural products. Leaf disease can be developd as a result of bacterial, fungal, or other causes. Finding and detecting sick plants in the open eye takes a long time.

Capsicum (Capsicum annum) is the most used food crop in the world. Bell pepper bacterial spot (a fungal disease in capsicum) has caused a momentous economic and commercial loss and just by eliminating 20% of this bacterial infection, the farmers may benefit from an extraordinary profit. Therefore, early detection and identification of capsicum bacterial diseases play the utmost important role to take timely measures for the quality of the plant.

There are numerous ways to perceive plant pathologies. Some plants have no visible symptoms of diseases associated with them or diseases may appear only when too late to cure. So it is necessary to perform a classy analysis of the plant diseases in the laboratories by the experts using powerful microscopes or employing different electromagnetic spectrum that is not visible to humans.

The population is increasing day by day and the importance of agriculture is also increasing in parallel. There is an attempt to apply machine machine learning and deep learning methods in agriculture under the name of smart farm technologies both to increasing agricultural production solve some challenges of agriculture. Some of the machine learning challenges include Data availability and quality. Agricultural data can be complex, heterogenous and difficlut to collect. ML algorithms require large amount of high quality data to produce accurate predictions, which can be a challenge in agriculture, particularly for small scale farmers.

Computer vision, artificial intelligence, machine learning and deep learning techniques are more popular research areas for object detection and classification from images, text and videos. Digital image processing technique minimized the inaccurate manual disease detection and improve the accuracy, feasibility and efficiency to predict the disease on a time from a plant.

The first aim of the current project was the collection of a sufficient capsicum image dataset from the field and then classifies the images into two categories (Bacterial spot images and healthy images).

* 1. ***Initial Hypothesis***

The common disease of Bell\_Pepper Plant is Bell\_Pepper Leaf with Bacterial Spot. If a farmer can detect these diseases early and apply appropriate treatment, it can save lot of waste and prevent the economic loss.

The Decision on whether the image is Diseased or not is done by considering the CNN Algorithm’s Method known as Grid of Numbers.

* 1. ***Data Collection Approach***

Data Collection:-

-In this project we have used ready-made data from Kaggle for training.

-The dataset was downloaded and the directory was named as "PlantVillage".

-We launched GitBash which allows us to run all the unix command, which inturn launched Jupyter notebook.

-To download the dataset into tf dataset TF data input pipeline, followed by data cleaning and we made our dataset ready for Model Training.

-Essential modules were downloaded.

-We used Tensorflow dataset to download all the images into tf.data.dataset .

-We created couple of constants such as Image\_size, Batch\_size, Channels and EPOCHS .

-We split our data into Train and Test.

-80% of the data was kept as Training Data.

-20% of the data was split into 10% Validation and 10% Test.

-Here 10% of the data used for test is used to check the accuracy of our model.

-We let Tensorflow determine how many batches to load while GPU is training.

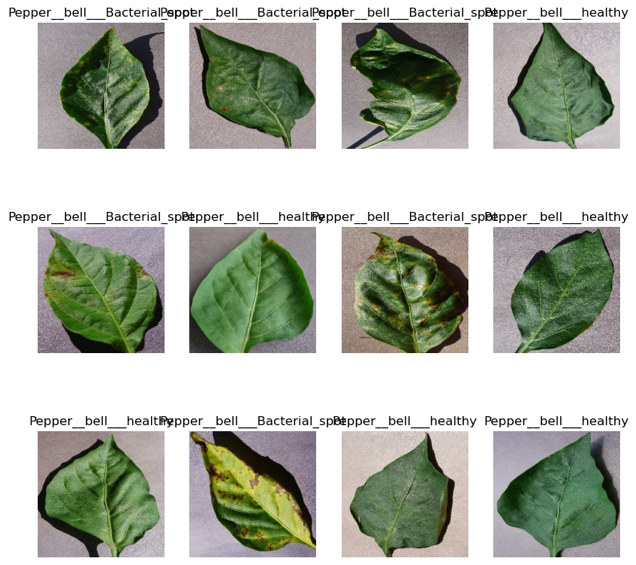
-Dataset were optimised for Training performance such that training will run fast.

-We supplied preprocessing pipeline using an API which would scale the image.

-All the image dimensions were resized to 256 by 256, which is followed by Data augmentation to make our model robust.

- Visualisation

-We Visualised some of the images which majorly divided into pepper leaf Diseased and pepper leaf Healthy.



* 1. ***Machine Learning Approach***

Based on the problem statement, we just analysed that we should work on the classification algorithms.

As it is a Dataset which contains images, we should deal with Deep learning model which uses a very complex Algorithms. So we use the CNN Algorithm Here.

We trained Convolutional Neural Network in Tensor Flow using Bell Pepper Leaf images. The goal of this model would be to classify these images either healthy or Bacterial spot disease.

Creating a Layer for Resizing and Normalisation :-

-Convolutional Neural Network has 2 parts.

1). Convolutional layers

2). Pooling layers

- Usually the model contains set of layers, wherein our model contains 32 layers.

-It includes filters and kernels.

-The popular activation layer for the hidden layers is always "Relu", since it is very fast to compute.

-After convolutional layer we have Pooling layer. The type of pooling we use in our model is Max Pooling.

-We used 'Soft max' activation function, which normalises the probability of classes.

Data Augmentation:-

- The Designed model had applied augmentation on the training dataset to improve efficiency and accuracy.

Building a Trained CNN Model:-

-In Deep Learning first we define the Neural Network architecture which is followed by compiling using optimizer.

-We compiled using a famous optimizer named "Adam".

-Then we defined Loss function and metric in each epoch.

-"Accuracy" is the kind of metric we used to track the training process, which is followed by training the network at model.fit() .

-We will record the history of every epochs in the history parameter such that we can plot some charts later on.

-Our validation accuracy was close to 1 , which is pretty good.

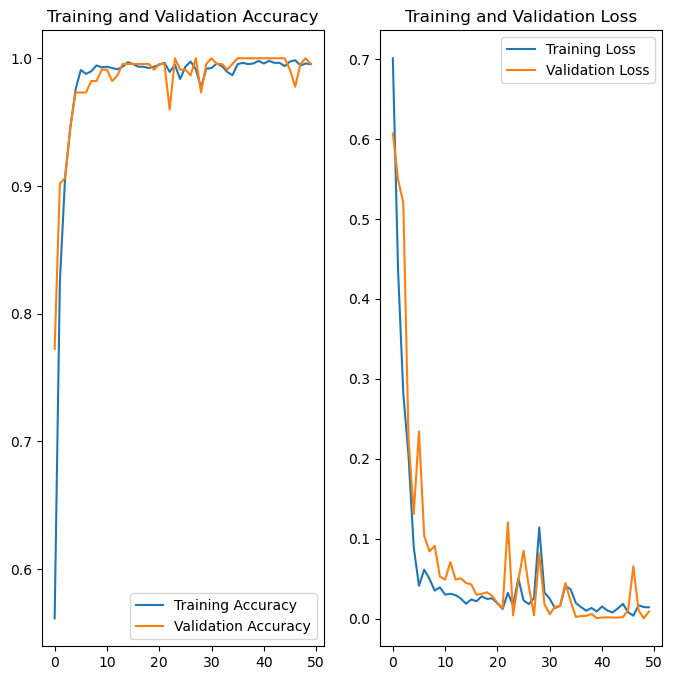
-We checked the accuracy of our model by using test set and the accuracy was found to be 98.96%.

-History object has an element called history.key, which contains 4 parameters. Those are Loss, Accuracy, Validation loss, Validation accuracy.

Evaluating The Model:-

-Plotted training history on the graph.

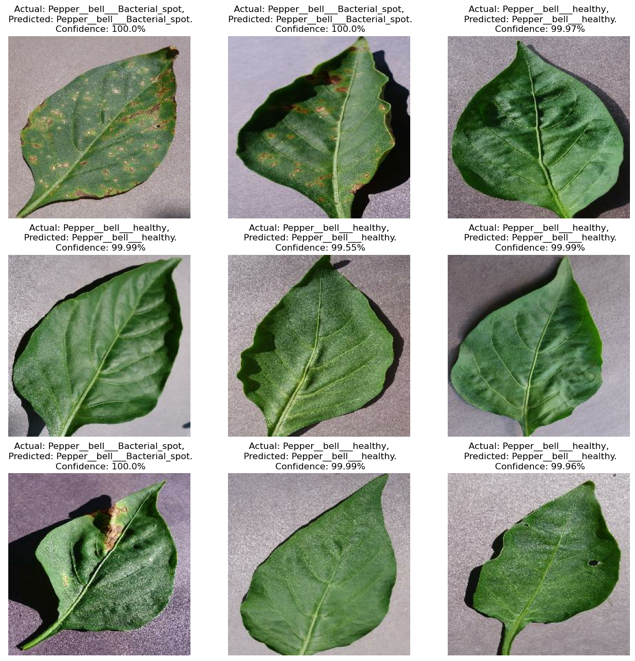
-We plotted Training accuracy versus Validation accuracy and Training loss versus Validation loss.



Making Predictions / inferences on simple Images:-

-Intially we made predictions using first image and then we made predictions for 32 images.

-We checked the Actual class, Predicted class and Confidence % and we found that actual class was as same as predicted class and the confidence was found to be 100%.



Conclusion :-

We can conclude the actual class is same as predicted class and the confidence is 100%. The proposed model was to be found 100% accurate.