**Industrial Internship Report on**

**”** Plant Disease Detection Machine Learning**”**

**Prepared by**

**SANIKA SATYAWAN TOTRE**

**TABLE OF CONTENTS**

[1 Preface 3](#_Toc139702806)

[2 Introduction 4](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 4](#_Toc139702808)

[2.2 About upskill Campus 8](#_Toc139702809)

[2.3 Objective 9](#_Toc139702810)

[2.4 Reference 9](#_Toc139702811)

[2.5 Glossary 10](#_Toc139702812)

[3 Problem Statement 11](#_Toc139702813)

[4 Existing and Proposed solution 12](#_Toc139702814)

[5 Proposed Design/ Model 13](#_Toc139702815)

[5.1 High Level Diagram (if applicable) 13](#_Toc139702816)

[5.2 Low Level Diagram (if applicable) 13](#_Toc139702817)

[5.3 Interfaces (if applicable) 13](#_Toc139702818)

[6 Performance Test 14](#_Toc139702819)

[6.1 Test Plan/ Test Cases 14](#_Toc139702820)

[6.2 Test Procedure 14](#_Toc139702821)

[6.3 Performance Outcome 14](#_Toc139702822)

[7 My learnings 15](#_Toc139702823)

[8 Future work scope 16](#_Toc139702824)

# Preface

Summary of the whole 6 weeks’ work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to all (with names), who have helped you directly or indirectly.

Your message to your juniors and peers.

# Introduction

## About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various**Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end**etc.



1. UCT IoT Platform **(****)**

**UCT Insight** is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

* It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
* It supports both cloud and on-premises deployments.

It has features to  
• Build Your own dashboard  
• Analytics and Reporting  
• Alert and Notification  
• Integration with third party application(Power BI, SAP, ERP)  
• Rule Engine

1. **Smart Factory Platform (****)**

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

* with a scalable solution for their Production and asset monitoring
* OEE and predictive maintenance solution scaling up to digital twin for your assets.
* to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
* A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

1.  based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

1. Predictive Maintenance

UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



## About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

<https://www.upskillcampus.com/>

upSkill Campus aiming to upskill 1 million learners in next 5 year



## The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

## Objectives of this Internship program

The objective for this internship program was to

 ☛ get practical experience of working in the industry.

 ☛ to solve real world problems.

 ☛ to have improved job prospects.

 ☛ to have Improved understanding of our field and its applications.

 ☛ to have Personal growth like better communication and problem solving.

## Reference

[1] Google

[2] Youtube

[3] Chatgpt

# Problem Statement

In the assigned problem statement

[Plant Disease Detection Using Image Processing and Machine Learning]

# Existing and Proposed solution

In India about 70% of the populace relies on agriculture. Identification of the plant diseases is important in order to prevent the losses within the yield. It's terribly troublesome to observe the plant diseases manually. It needs tremendous quantity of labor, expertize within the plant diseases, and conjointly need the excessive time interval. Hence, image processing and machine learning models can be employed for the detection of plant diseases. In this project, we have described the technique for the detection of plant diseases with the help of their leaves pictures. Image processing is a branch of signal processing which can extract the image properties or useful information from the image. Machine learning is a sub part of artificial intelligence which works automatically or give instructions to do a particular task. The main aim of machine learning is to understand the training data and fit that training data into models that should be useful to the people. So it can assist in good decisions making and predicting the correct output using the large amount of training data. The color of leaves, amount of damage to leaves, area of the leaf, texture parameters are used for classification. In this project we have analyzed different image parameters or features to identifying different plant leaves diseases to achieve the best accuracy. Previously plant disease detection is done by visual inspection of the leaves or some chemical processes by experts. For doing so, a large team of experts as well as continuous observation of plant is needed, which costs high when we do with large farms. In such conditions, the recommended system proves to be helpful in monitoring large fields of crops. Automatic detection of the diseases by simply seeing the symptoms on the plant leaves makes it easier as well as cheaper

## Code submission (Github link)

## Report submission (Github link) :

# Proposed Design/ Model

Given more details about design flow of your solution. This is applicable for all domains. DS/ML Students can cover it after they have their algorithm implementation. There is always a start, intermediate stages and then final outcome.

## High Level Diagram ()

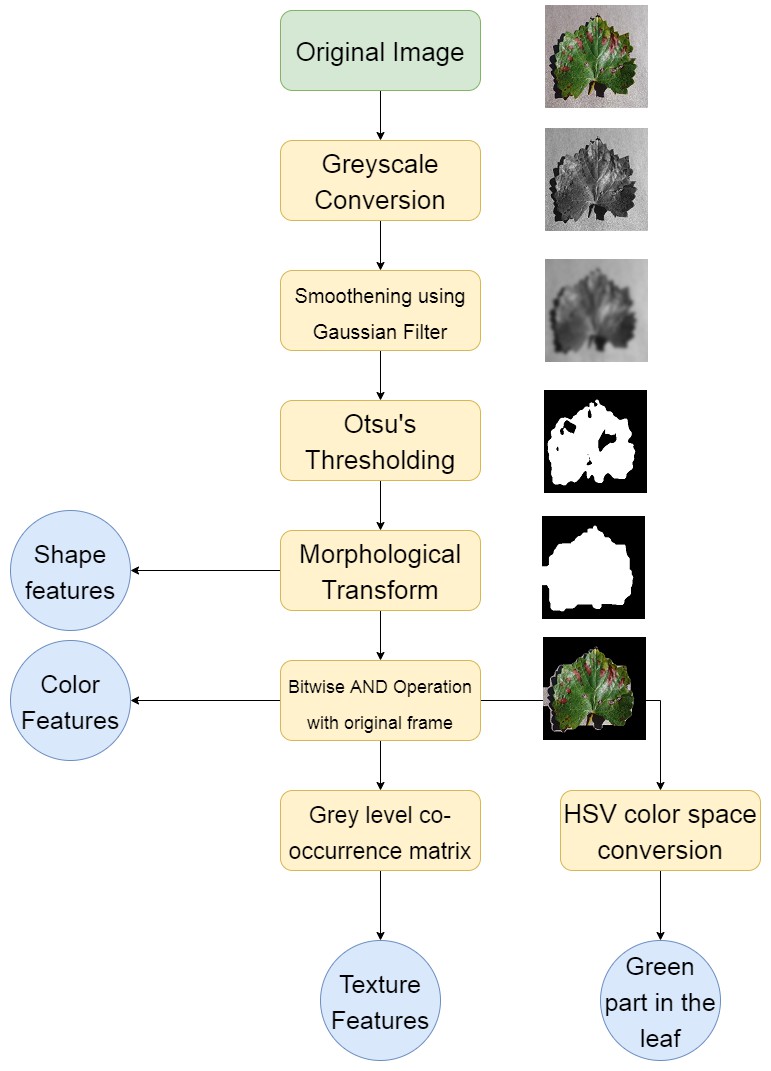
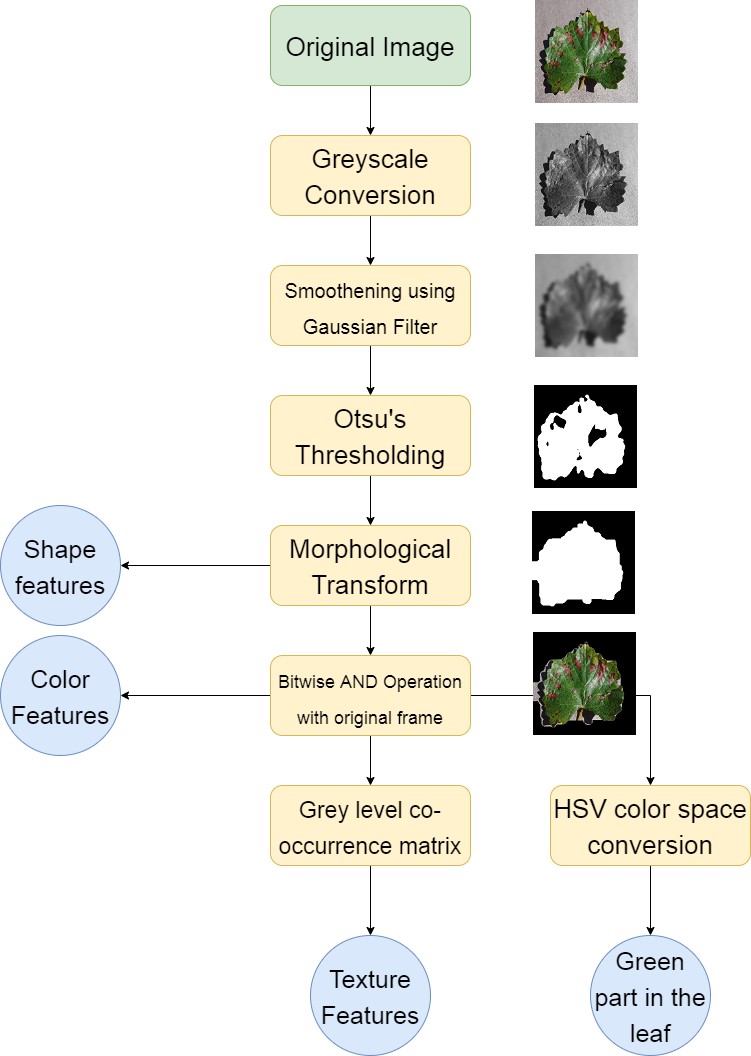


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

## Low Level Diagram (if applicable)

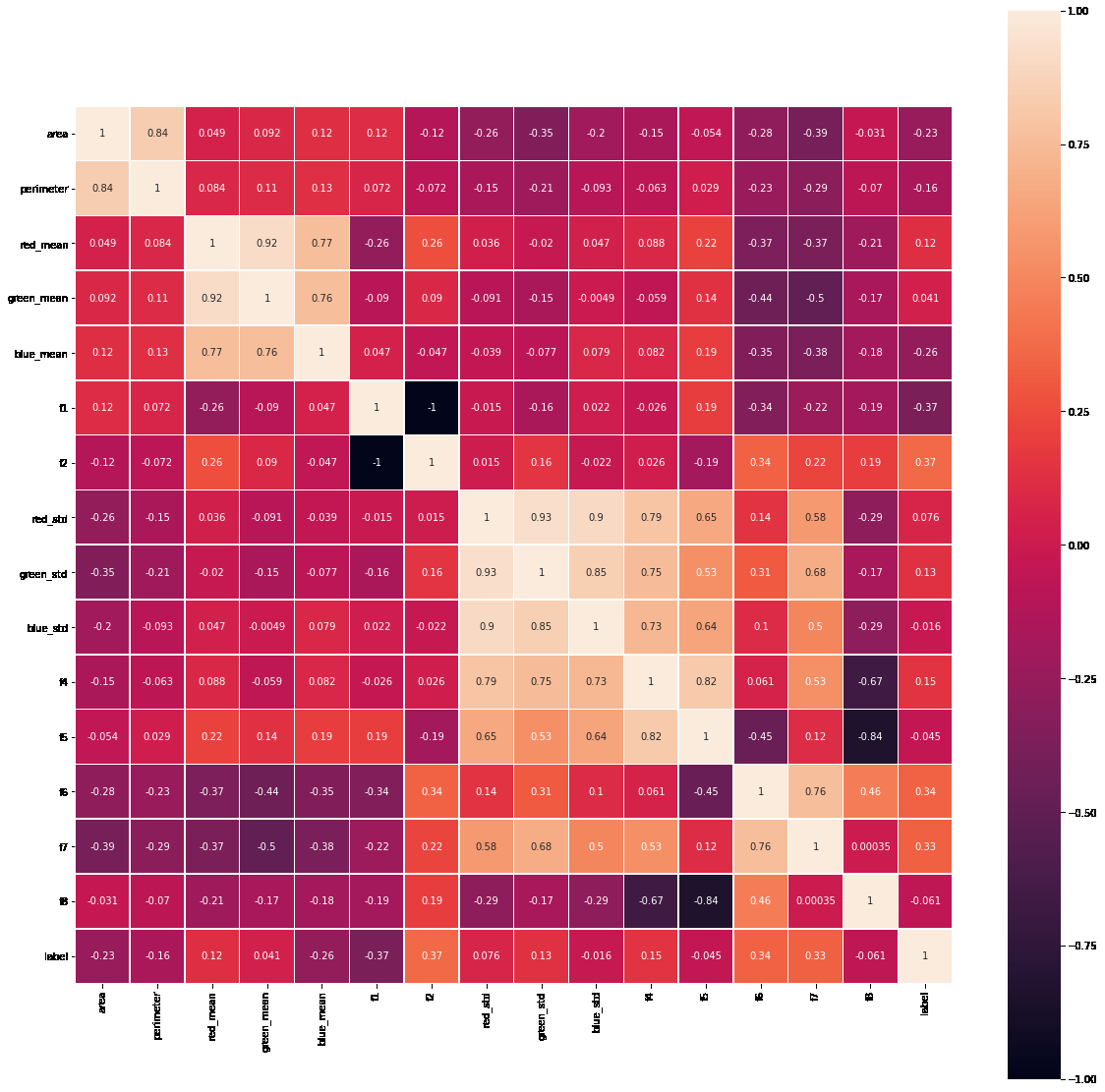
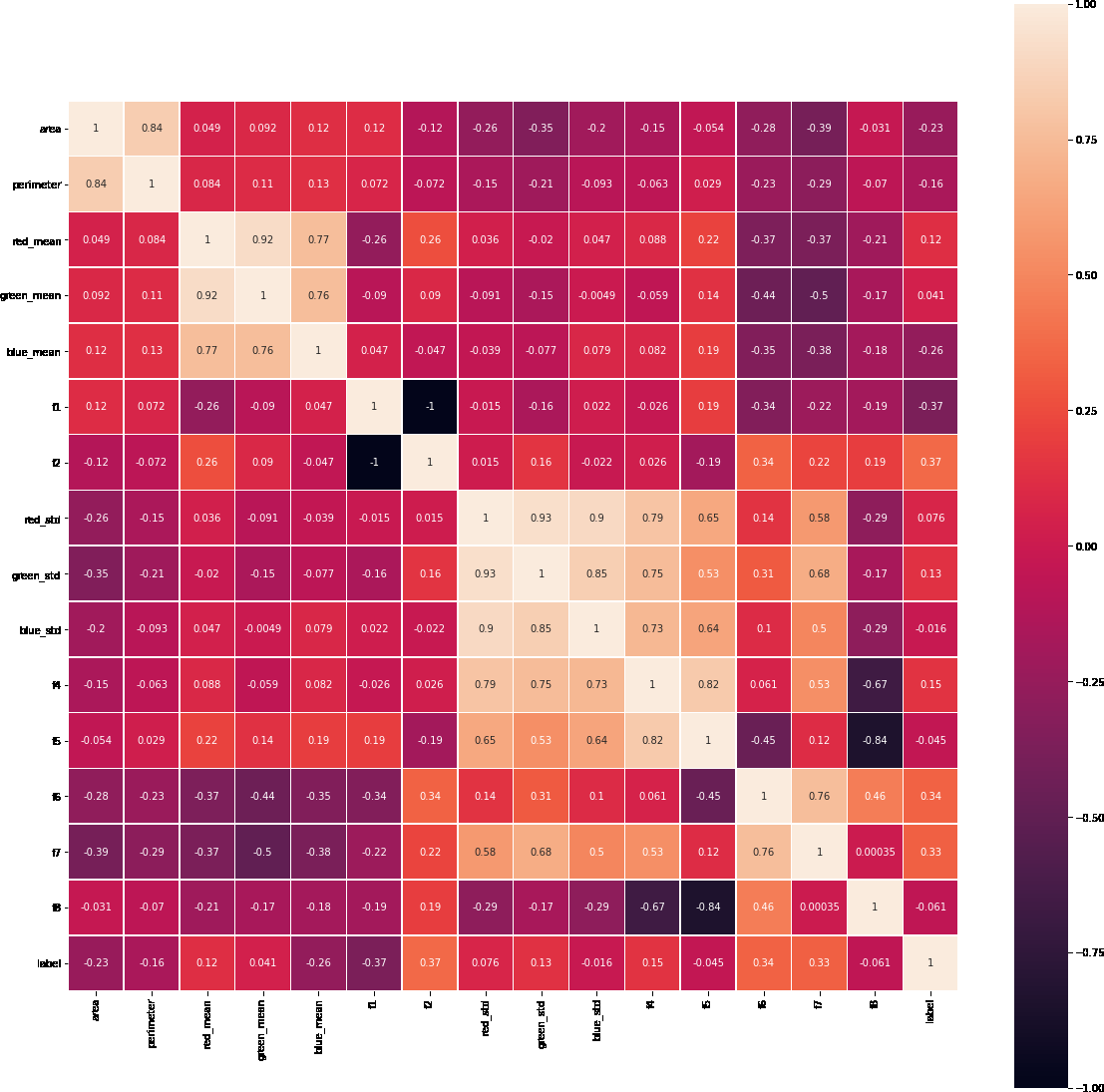
## Interfaces (if applicable)

Update with Block Diagrams, Data flow, protocols, FLOW Charts, State Machines, Memory Buffer Management.

# Performance Test

## Test Plan/ Test Cases

Feature selection is an important step in all machine learning problems. In this project we are selecting the features on the basis of correlation of variables with target variable. Fig. 3 shows the correlation of each variable with each other for apple dataset. The correlation of feature green part of leaf (F1) and green part of leaf (F2) is very high (1) which means both variables are dependent on each other. So we have dropped one of them (F2). Now for apple disease prediction, less correlated features such as green channel mean, red channel standard deviation, blue channel standard deviation, dissimilarity (f5) and correlation (f8) will not contribute too much in model development. So we have dropped these variables also. After feature selection, the data is now parsed to machine learning classifiers to find the patterns in the datA



**Fig. 3.** Correlation plot for Apple dataset.

## Test Procedure

Random forest classifier has been used for classification or detection task. It is the part of ensemble learning, where the output is predicted from multiple base estimators [8]. Generally, to achieve higher accuracies, decision trees are used. But they are prone to overfitting problems. So to overcome this issue, random forest classifier is used which is a combination of multiple decision trees. Each tree is trained by using different subsets of the whole dataset, this can reduce the overfitting and improves the accuracy of the classifier. We have splitted the dataset into train set (80%) for fitting the model and test set (20%) for validation. K-fold cross validation technique is implemented to find the accuracy score. This method can find the accuracy on whole dataset without any bias. After fitting the data, f1 score, precision, recall, accuracy has been calcula

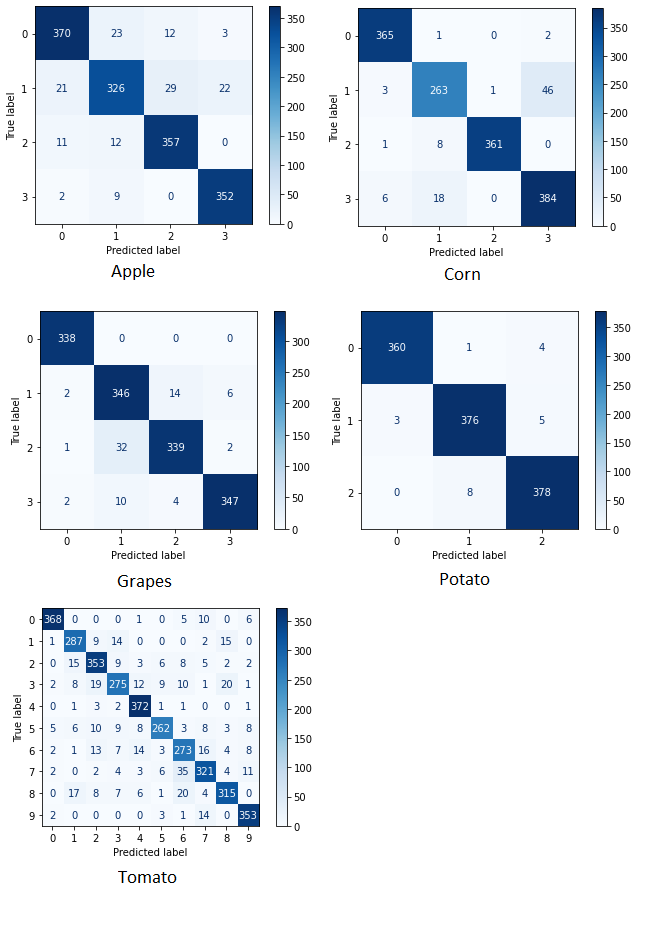
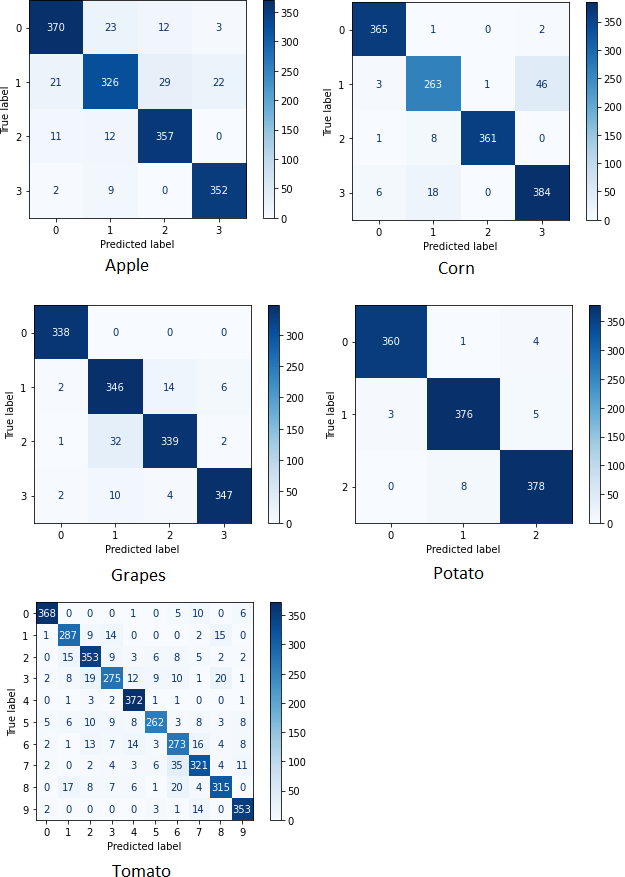
## Performance Outcome

Table 2 shows the performance matrices for each model developed for each of the plant. We can observe that the accuracy scores are nearly equal to f1 scores. This is because of balanced number of false negative and false positive predictions. This is considered as best case for any machine learning algorithm. The average accuracy was 93%.

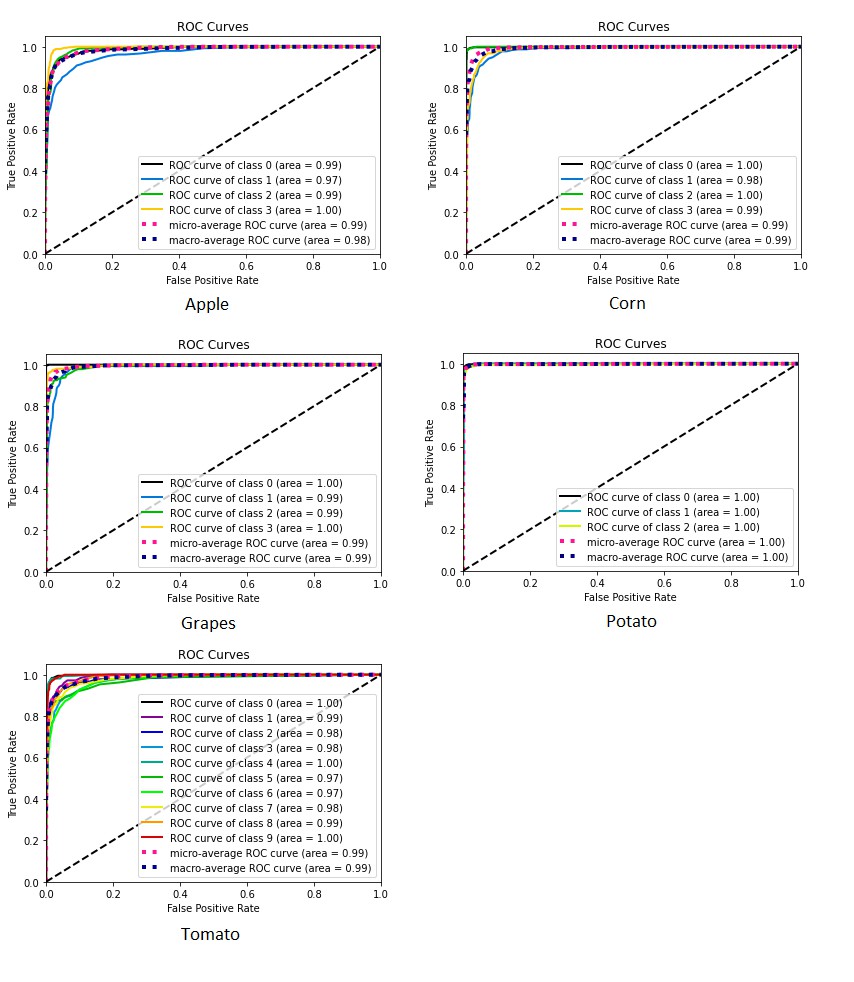
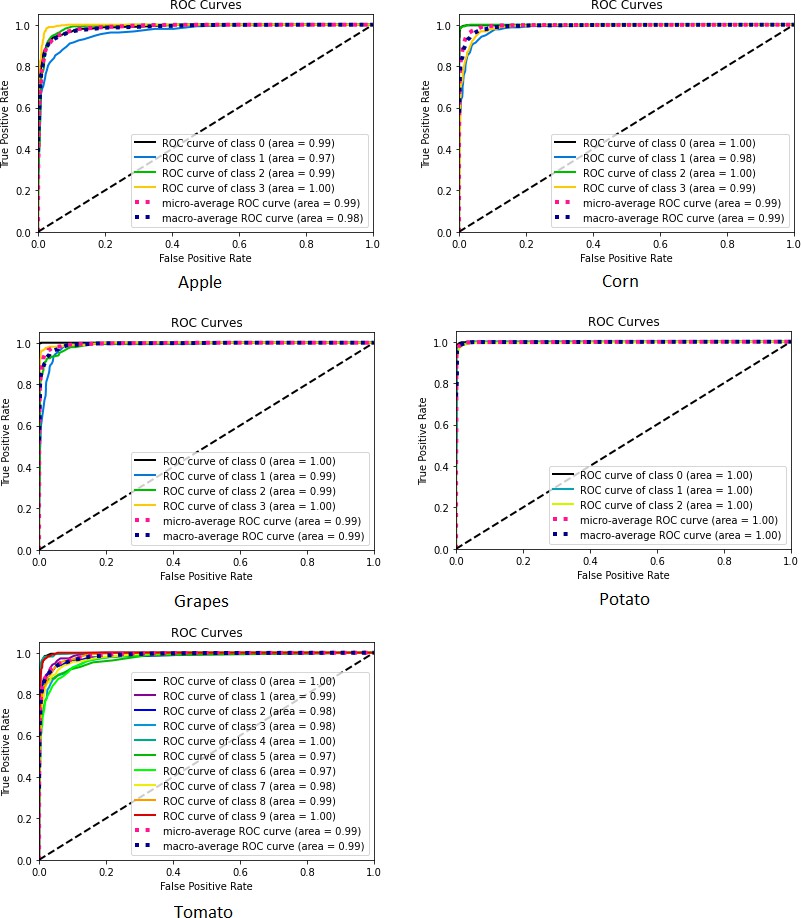
**Table 2.** Performance matric for all models.

|  |  |  |
| --- | --- | --- |
| Plant | Accuracy | F1 Score |
| Apple Corn Grapes Potato | 0.91  0.94  0.95  0.98 | 0.91  0.94  0.95  0.98 |
| Tomato | 0.87 | 0.87 |

Fig. 4 shows the confusion matrices for each of the model. With the help of confusion matrices, number of false negatives, false positives, true predictions can be analyzed. Fig. 5 shows the receiver operating characteristic (ROC) curve for each of the model. An ROC curve is a graph showing the performance of a classification model at all classification thresholds. It depends upon two parameters, true positive rate and false positive rate.

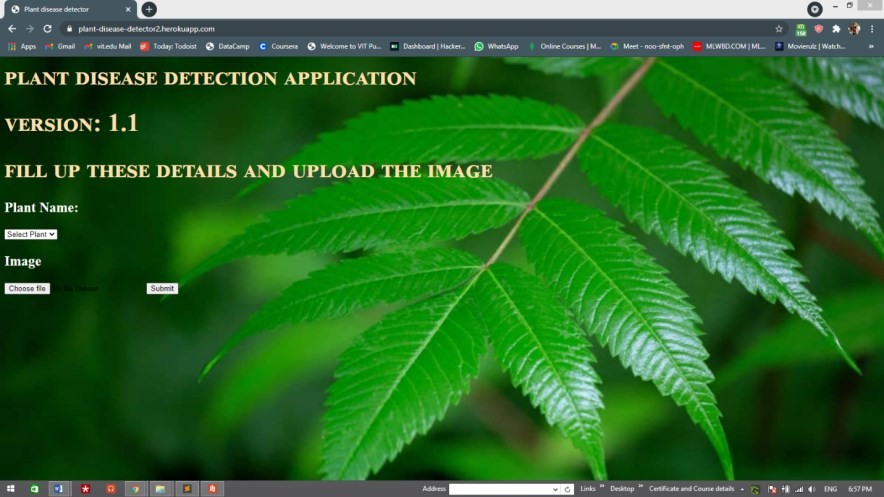
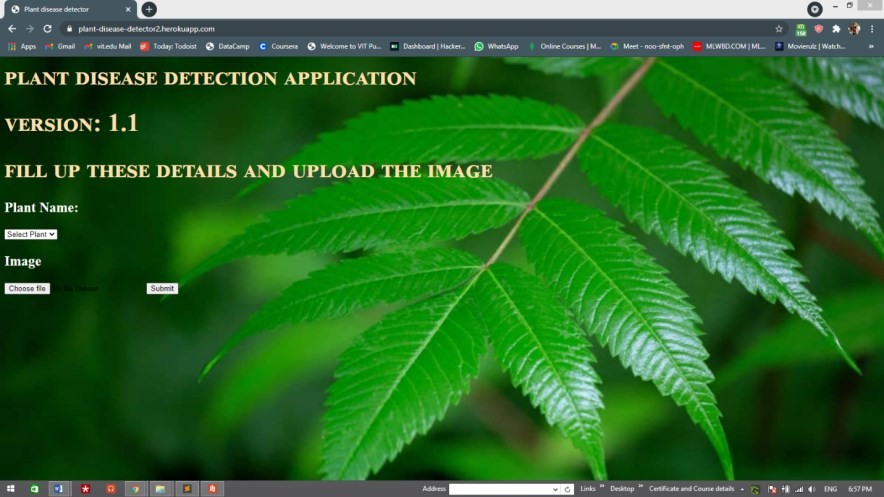


**Fig. 4.** Confusion matrices for all the models.



**Fig. 5.** ROC curves for all the models.

We have developed a flask based web application for detecting the plant disease and deployed it on heroku (free cloud hosting server). Fig 6 shows the homepage of deployed web application and Fig 7 shows the input images and their corresponding predictions made by our system. It shows that the system successfully detected the disease of leaf.



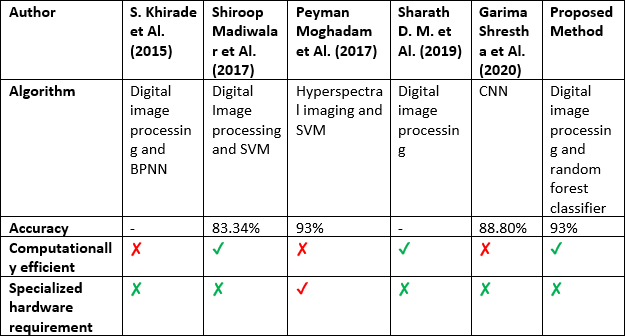
However, we can deploy an intelligent robot vehicle with high end processor attached to it for real time plant disease detection. This system can detect the diseased plants in the agricultural site. Even we can automate the process of spreading the fertilizers by using such robots. Our proposed algorithm is computationally inexpensive, so it can detect the plant disease in efficient manner. Also sometimes it happens that the farmer also could not identify the disease of the plant. So they need an expert advice. So we can deploy a website which can detect the plant disease based on images captured and uploaded by farmer and can give suggestions or can suggest some fertilizers based on detected disease.

# 

# My learnings

We have successfully developed a computer vision based system for plant disease detection with average 93% accuracy and 0.93 F1 score. Also the proposed system is computationally efficient because of the use of statistical image processing and machine learning model. Table 3 illustrates the overall benefits of our system over the other approaches.

**Table 3.** Comparison of proposed system with other existing systems.



We can observe that our technique is accurate and efficient compared with other systems. Also it won't require a specialized hardware, makes it cost effective solution.

# Future work scope

* Initial detection of pests and pathogens in standing crops is a big challenge.

Remote sensing and image analysis-based technologies have been used to detect the occurrence of pests and pathogens.

Cutting-edge technologies could be used to ensure sustainable and safe crop production.

Different sensors, algorithms, and computational skills can be utilized in efficient disease detection.