# Classification and Detection of Fruit Diseases Using Image Processing

**K M Rayudu[1], K Manikanta[2], G Tharun[3], N Mahesh Babu[4]**

[1]Professor, [2,3,4] Scholars

Department of Computer Science and Engineering

Qis College of Engineering and Technology

**ABSTRACT**-Fruit disease detection is vital at early stage since it will affect the agricultural field. In this paper, mainly consider the detection and analysis of fruit infections which is available in the plant areas and storage of data about the agricultural filed and details of farmers in database and recovering the data using Cloud computing. There are more fruit diseases which occur due to the surrounding conditions, mineral levels, insects in the farm area and other factors. The detected data from the plant area is determined by image processing and stored in the database.

Keywords-K-Means Clustering; Support vector Machine;

## 1. INTRODUCTION

### 1.1 Background:

Agriculture has been the base for every people. It is most important that more than 70% of the people depend on agriculture for their livelihood in India. Nowadays the growth of productivity of plants, crops and fruits are normally affected by the diseases. The disease is a major problem arising in an agricultural field. In plants, most of the leaves and fruits are affected by diseases due to bacteria and virus. This technique is used to determine the infection on leaves, fruits and stem of the plants. In order to generate an automated database to examine the infections using proposed method. The database consists of data related to plant leaves, fruit conditions and the symptoms of disease to be affected.

The fruit details and the identification of disease from the feature extraction are stored in the database. The entire database is viewed and compared with the captured image. The mobile application is developed for processing the data and providing intimation to the farmers. Thus the variation in

image from the database and also indicates the disease in the fruits.

### 1.2 Trust Based Recommendation System:

Agribusiness has been the base for each individual. It is most significant that over 70% individuals rely upon horticulture for their business in India. These days the development of efficiency of plants, yields and organic products are regularly influenced by the infections. The illness is a significant issue emerging in an agrarian field. In plants, the greater part of the leaves and natural products are influenced by infections because of microscopic organisms and infection. This method is utilized to decide the disease on leaves, products of the soil of the plants. To produce a mechanized data set to analyze the contaminations utilizing proposed strategy. The data set comprises of information identified with plant leaves, organic product conditions and the manifestations of illness to be influenced. The organic product subtleties and the recognizable proof of illness from the element extraction are

put away in the data set. The whole

data set is seen and contrasted and the caught picture. The portable application is created for preparing the information and giving insinuation to the ranchers. Accordingly, the variety in picture from the information base and furthermore demonstrates the sickness in the organic products.

## 1.3 Motivation of the Project:

Agriculture has been the base for every people. It is most important that more than 70% of the people depend on agriculture for their livelihood in India. Nowadays the growth of productivity of plants, crops and fruits are normally affected by the diseases. The disease is a major problem arising in an agricultural field.

## 1.4 Problem statement:

Agriculture has been the base for every people. It is most important that more than 70% of the people depend on agriculture for their livelihood in India. Nowadays the growth of productivity of plants, crops and fruits are normally affected by the diseases. The disease is a major problem arising in an agricultural field.

### 1.5 Scope of the Project:

Agriculture has been the base for every people. It is most important that more than 70% of the people depend on agriculture for their livelihood in India. Nowadays the growth of productivity of plants, crops and fruits are normally affected by the diseases. The disease is a major problem arising in an agricultural field. In plants, most of the leaves and fruits are affected by diseases due to bacteria and virus. This technique is used to determine the infection on leaves, fruits and stem of the plants. In order to generate an automated database to examine the infections using proposed method. The database consists of data related to plant leaves, fruit conditions and the symptoms of disease to be affected.

## 2. LITURATURE SURVEY

[1] Asha R.PatilVarshaI.Patil, B.S.Panchbhai, "Detection of Plant Diseases Using Image Processing Tools". Asha R. PatilVarshaI.Patil. Int. Journal of Engineering Research and Application ISSN : 2248-9622, Vol. 7, Issue 4, ( Part -2) April 2017, pp.44-45.

[2] Pratik Agarwal, "AGROCLOUD-Open surveillance of Indian Agriculture via cloud" 2016.International Conference on Information Technology (InCITe)-The Next Generation IT Summit

[3] Athmaja S1, Hanumanthappa M2, "Applications of Mobile Cloud Computing and Big Data Analytics in Agriculture Sector- A Survey". International Journal of Advanced Research in Computer and Communication Engineering ICRITCSA M S Ramaiah Institute of Technology, Bangalore Vol. 5, Special Issue 2, October 2016.

[4] Wang-Su Jeon1 and Sang-Yong Rhee,"Plant Leaf Recognition Using a Convolution Neural Network "International Journal of Fuzzy Logic and Intelligent Systems Vol. 17, No. 1, March 2017, pp. 26-34

[5] Li Tan1,2†, Hong Fei Hou1,Qin Zhang2."An Extensible Software Platform for Cloud-based Decision Support and Automation in Precision Agriculture".2016 IEEE 17th International Conference on Information Reuse and Integration.

Pratik Agarwal, "AGROCLOUD-Open surveillance of Indian Agriculture via cloud" 2016.International Conference on Information Technology (InCITe)-The Next Generation IT Summit.

Athmaja S1, Hanumanthappa M2, "Applications of Mobile Cloud Computing and Big Data Analytics in Agriculture Sector- A Survey". International Journal of Advanced Research in Computer and Communication Engineering ICRITCSA M S Ramaiah Institute of Technology, Bangalore Vol. 5, Special Issue 2, October 201

## 3. PROPOSED SYSTEM

Agriculture has been the base for every people. It is most important that more than 70% of the people depend on agriculture for their livelihood in India. Nowadays the growth of productivity of plants, crops and fruits are normally affected by the diseases. The disease is a major problem arising in an agricultural field. In plants, most of the leaves and fruits are affected by diseases due to bacteria and virus. This technique is used to determine the infection on leaves, fruits and stem of the plants. In order to generate an automated database to examine the infections using proposed method. The database consists of data related to plant leaves, fruit conditions and the symptoms of disease to be affected. The fruit details and the identification of disease from the feature extraction are stored in the database. The

entire database is viewed and compared with the captured image. The mobile application is developed for processing the data and providing intimation to the farmers. Thus the variation in image from the database and also indicates the disease in the fruits

## 3.1 Advantages:

Agriculture is becoming digital, AI in agriculture is emerging in three major categories which are agricultural robotics, soil & crop monitoring, and predictive analytics.

## 3.2 System specifications:

### 3.2.1 Software Requirements
Functional requirements for a secure cloud storage service are straightforward:

1. The service should be able to store the user'sdata;
2. The data should be accessible through any devices connected to the Internet;
3. The service should be capable to synchronize the user's data between multiple devices (notebooks, smart phones, etc.);
4. The service should preserve all historical changes(versioning);
5. Data should be shareable with other users;
6. The service should support SSO;and
7. The service should be interoperable with other cloud storage services, enabling data migration from one CSP to another.

• Operating System**:** Windows

• Coding Language: Python 3.7

• Script

• Database

### 3.2.2 Hardware Requirements:

**Processor** - Pentium–III

**Speed** – 2.4GHz

**RAM** - 512 MB (min)

**Hard Disk** - 20 GB

**Floppy Drive** - 1.44MB

**Key Board** - Standard Keyboard

**Monitor** – 15 VGAColour

Cloud computing has three fundamental models
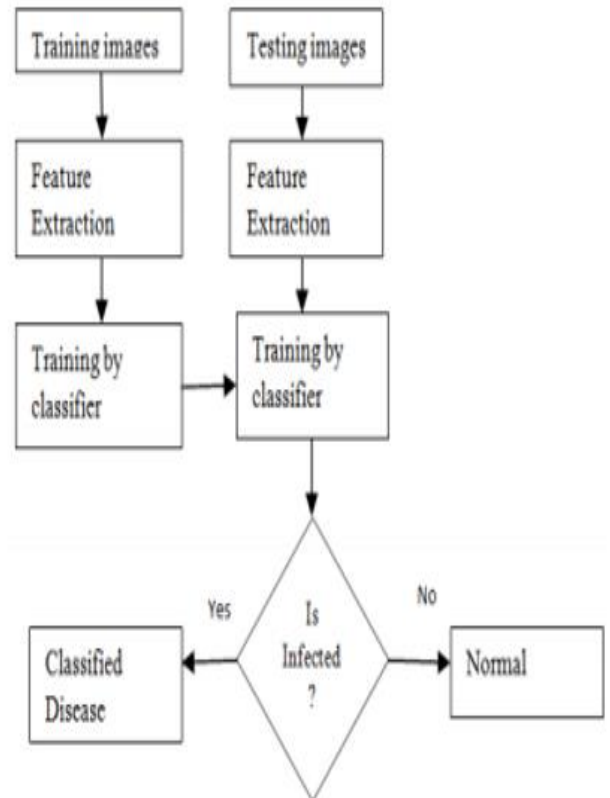
## 4. System Design
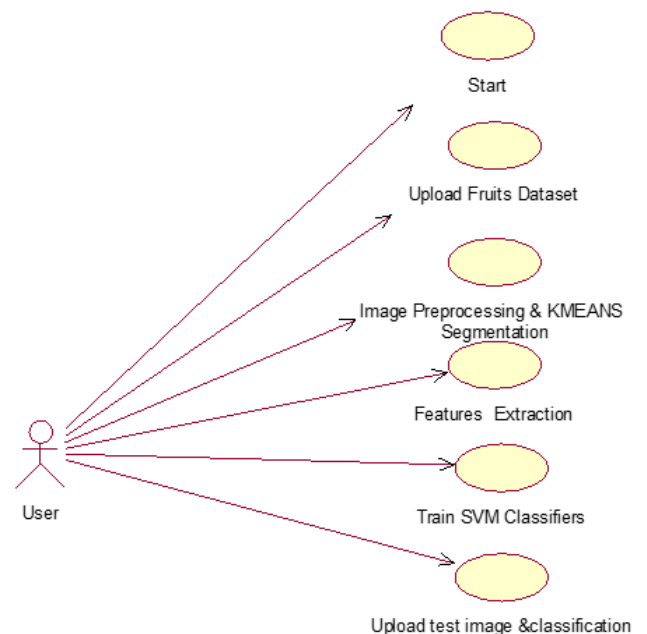
**System Architecture:**
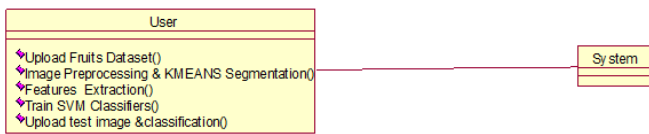


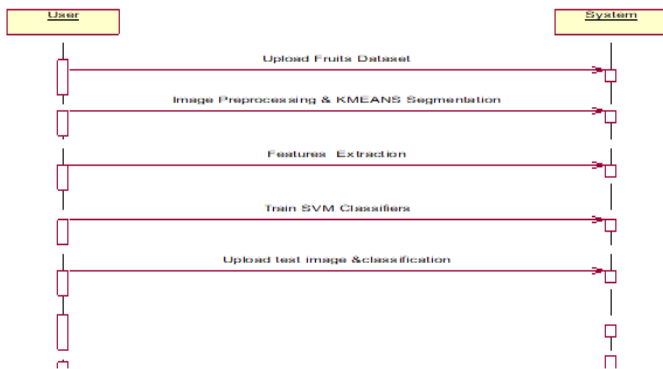Fig 1: System Architecture



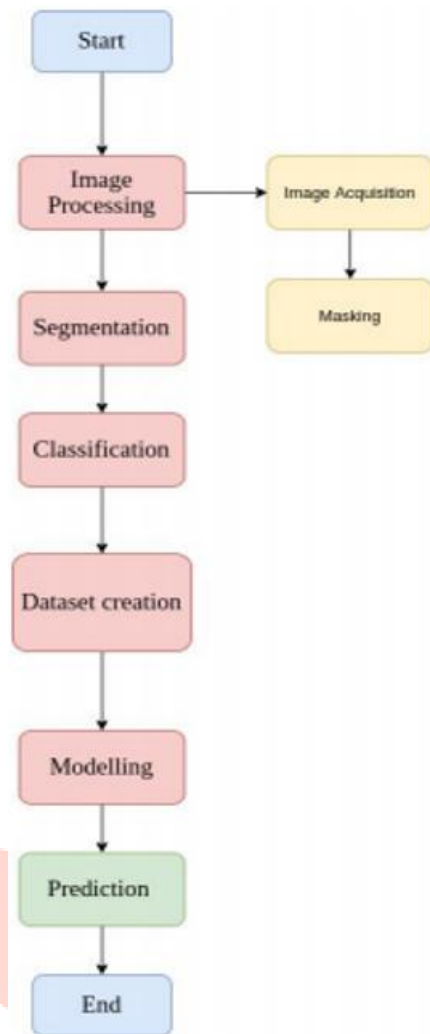Fig 2: Use case diagram
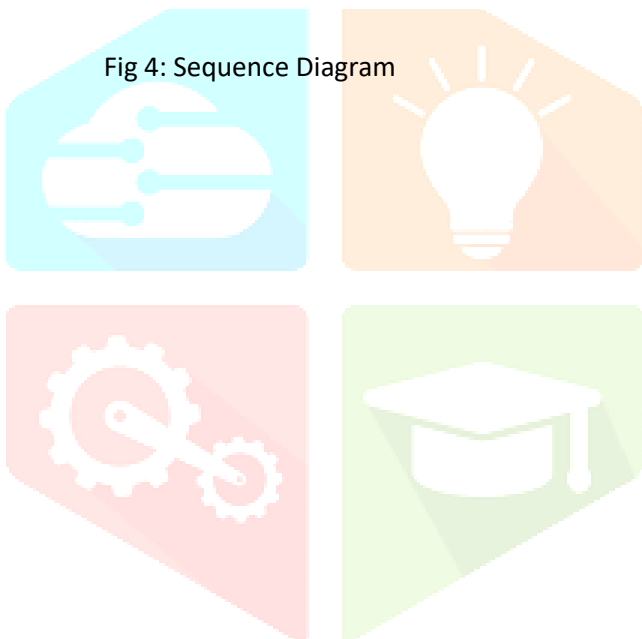
Fig 3: Class Diagram



Fig 4: Sequence Diagram



Fig 5: Data Flow Diagram

## 5. IMPLEMENTATION

### 5.1 Software Environment:

### 5.1.1 Python Technology

What is Python:-

Below are some facts about Python.

Python is currently the most widely used multi-purpose, high-level programming language.

Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java.

Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.

Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber… etc.

The biggest strength of Python is huge collection of standard library which can be used for the following

- Machine Learning
- GUI Applications (like Kivy, Tkinter, PyQt etc.)
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like OpenCV, Pillow)
- Web scraping (like Scrapy, BeautifulSoup, Selenium)
- Test frameworks
- Multimedia

## What is Machine Learning: -

Before we take a look at the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of *building models of data*.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models *tunable parameters* that can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain. Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here.

## Applications of Machines Learning: -

Machine Learning is the most rapidly growing technology and according to researchers we are in the golden year of AI and ML. It is used to solve many real-world complex problems which cannot be solved with traditional approach. Following are some real-world applications of ML –

- Emotion analysis
- Sentiment analysis
- Error detection and prevention
- Weather forecasting and prediction
- Stock market analysis and forecasting

- Speech synthesis
- Speech recognition
- Customer segmentation
- Object recognition
- Fraud detection
- Fraud prevention
- Recommendation of products to customer in online shopping

## Modules Used in Project:-

### TensorFlow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

### NumPy

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using NumPy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

### Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution

towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

## Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and I Python shells, the Jupiter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with I Python. For the power user, you have full control of line styles, font properties, axes properties, etc., via an object-oriented interface or via a set of functions familiar to MATLAB users.

## Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. **Python**

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

- Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

- Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

## 6. System Testing

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

## 6.1 Types of Testing
## Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application.it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

# Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

# Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input: identified classes of valid input must be accepted.

Invalid Input: identified classes of invalid input must be rejected.

Functions: identified functions must be exercised.

Output: identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

## System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process

links and integration points.

## White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

## Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated,

asablackbox.youcannot−seeintoit.Thetestprovidesinputsa ndrespondsto outputs without considering how the software works.

### 6.1.1 Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.

- Pages must be activated from the identified link.

- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format

- No duplicate entries should be allowed

- All links should take the user to the correct page.

## 6.1.2 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.



**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

## 6.1.3 Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

Output Screenshots:
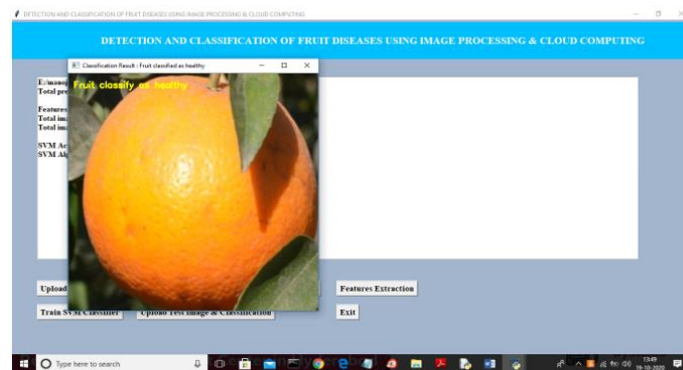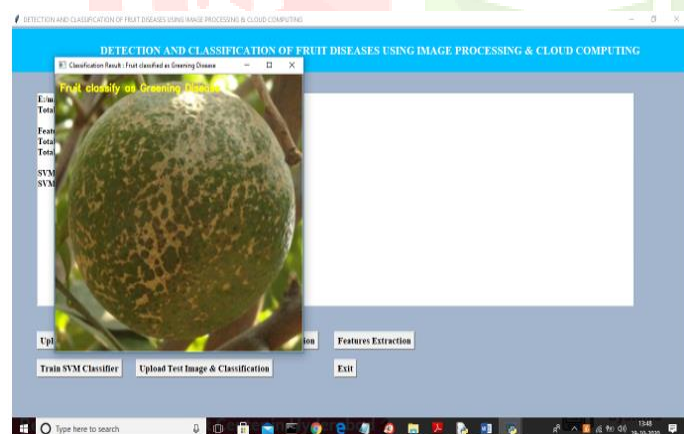


## 8.Conclusion

The development of cloud based scheme for helping Indian farmers and agriculture, helps to analyze the agriculture data in a better way to reduce the hoardings and in bringing up a prosperous safe and peaceful farmer society in India. The classification and segmentation of fruit images were performed using K-Means Algorithm and SVM technique. The various features of few fruits were initially extracted and segment the respective images. After comparison with feature values, the various disease names are analyzed and the optimal disease for the image is identified and the disease is indicated by an alert box and can be provided as the message through mobile application. The total number of samples provided, the true and false positions, the true and false negativities, the accuracy and the specificity are also indicated in an alert box.

## 9.References

[1] Asha R.PatilVarshaI.Patil, B.S.Panchbhai, "Detection of Plant Diseases Using Image Processing Tools". Asha R. PatilVarshaI.Patil. Int. Journal of Engineering Research and Application ISSN: 2248-9622, Vol. 7, Issue 4, ( Part -2) April 2017, pp.44-45.

[2] Pratik Agarwal, "AGROCLOUD-Open surveillance of Indian Agriculture via cloud" 2016.International Conference on Information Technology (InCITe)-The Next Generation IT Summit.

[3] Athmaja S1, Hanumanthappa M2, "Applications of Mobile Cloud Computing and Big Data Analytics in Agriculture Sector- A Survey". International Journal of Advanced Research in Computer and Communication Engineering ICRITCSA M S Ramaiah Institute of Technology, Bangalore Vol. 5, Special Issue 2, October 2016.

[4] Wang-Su Jeon1 and Sang-Yong Rhee, "Plant Leaf Recognition Using a Convolution Neural Network "International Journal of Fuzzy Logic and Intelligent Systems Vol. 17, No. 1, March 2017, pp. 26-34

[5] Li Tan1,2†,Hongfei Hou1,Qin Zhang2."An Extensible Software Platform for Cloud-based Decision Support and Automation in Precision Agriculture".2016 IEEE 17th International Conference on Information Reuse and Integration.

[6] P. Deepan "Detection and Classification of Plant Leaf Diseases by using Deep Learning Algorithm "International Journal of Engineering Research & Technology (IJERT) ISSN: 2278- 0181Published by, www.ijert.org,ICONNECT - 2k18