

**PLANT DISEASE DETECTION USING CLUSTERING –BASED  
SEGMENTATION AND NEURAL NETWORKS**

*Project report submitted in partial fulfillment of the requirements for the degree of*

**BACHELOR OF TECHNOLOGY**

**IN**

**ELECTRONICS AND COMMUNICATION ENGINEERING**

**By**

**Aditya Mohan (161035)**

**Arjun Sharma (161040)**

**Utkarsh Sharma (161097)**

**UNDER THE GUIDANCE OF**

**Dr. NafisUddin Khan**



**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT**

**May, 2020**

# TABLE OF CONTENT

<b>TABLE OF CONTENT</b> .....	<b>ii</b>
<b>DECLARATION</b> .....	<b>iv</b>
<b>ACKNOWLEDGEMENT</b> .....	<b>v</b>
<b>CERTIFICATE</b> .....	<b>vi</b>
<b>TERMINOLOGIES</b> .....	<b>vii</b>
<b>LIST OF FIGURES</b> .....	<b>viii</b>
<b>ABSTRACT</b> .....	<b>x</b>
<b>CHAPTER 1</b> .....	<b>11</b>
<b>PROJECT INTRODUCTION</b> .....	<b>11</b>
1.1 Introduction.....	11
1.2 PLANT DISEASES MANAGEMENT.....	12
1.3 Cross-Validation.....	15
1.3.1 Methods of Cross Validation.....	15
1.4 Contextual segmentation: Region growing.....	18
1.4.1 Pixel connectivity.....	19
1.4.2 Region similarity.....	20
1.4.3 Region growing.....	21
1.4.4 Split-and-merges.....	21
1.4.5 Texture segmentation: Spectral features.....	21
1.5 Feature extraction and selection in image processing.....	22
1.5.1 ORB.....	22
1.5.2 Vantage Point Tree.....	22
1.5.3 KAZE.....	22
1.6 Image Segmentation techniques.....	23
1.6.1 Types of Image Segmentation.....	23
1.6.2 Image Segmentation based on Clustering.....	25
1.7 Color Spaces.....	25
1.7.1 SATURATION.....	27
1.7.2 HUE.....	28
1.7.3 RGB color space.....	28
1.8 SOFTWARE USED.....	29
1.8.1 Open CV.....	29
1.8.2 Scikit-learn.....	29
1.8.3 Keras.....	30
1.8.4 Matplotlib.....	32
1.9 Clustering Methods.....	35
1.9.1 Comparing Clustering Algorithms.....	36
<b>CHAPTER 2: LITERATURE REVIEW</b> .....	<b>37</b>

**CHAPTER-3: CLASSIFICATION ALGORITHMS AND NEURAL NETWORKS, EVALUATION METRIC ..... 39**

- 3.1 Feed forwards Perceptrons ..... 39
- 3.2 Recurrent neural networks ..... 41
- 3.3 LSTM..... 42
- 3.4 DIFFFERENT EVALUATION METRICS ..... 43
- 3.5 Motivation ..... 46
- 3.6 Objective ..... 46

**CHAPTER 4 ..... 47**

**PROGRESS AND RESULTS ..... 48**

**REFERENCES ..... 60**

**PLAGIARISM REPORT ..... 61**

## DECLARATION

We hereby declare that the work reported in the B-Tech thesis entitled “Plant Disease Detection Using Clustering -Based Segmentation and Neural Networks” submitted at Jaypee University of Information Technology, Waknaghat India, is an authentic record of our work carried out under the supervision of Dr. NafisUddin Khan. We have not submitted this work elsewhere for any other degree or diploma.



Aditya Mohan (161035)



Arjun Sharma (161040)



Utkarsh Sharma (161097)

Department of Electronics and Communication Engineering  
Jaypee University of Information Technology, Waknaghat , India

## **ACKNOWLEDGEMENT**

We would like to thank Prof. Dr. M.J. Nigam Head of Dept., Electronics and Communication, for providing us the opportunity of working on a project.

We are highly indebted to Dr. NafisUddin Khan for motivating and enlightening us for our project work. We thank you for being a constant support throughout, without whose valuable guidance and insights, this project would not be a complete one. We offer you our sincere gratitude to you for instructing and directing us through thick and thin.

We would also like to acknowledge Mr. Abhishek Ray for helping us in the labs. Thank you for being there and helping us in and out.

## CERTIFICATE

This is to certify that the work reported in the B-Tech. thesis entitled “Plant Disease Detection Using Clustering -Based Segmentation and Neural Networks”, submitted by Aditya Mohan (161035), Arjun Sharma (161040), Utkarsh Sharma (161111) at Jaypee University of Information Technology, Wanknaghat, India, is a bonafide record of their original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.



(Signature of Supervisor)

Dr. NafisUddin Khan

Dept. of ECE, JUIT

26<sup>th</sup>May, 2020.

## **TERMINOLOGIES**

1. Parasite- organism that depend upon other plants
2. Life cycle- the growth and development of a pathogen.
3. Vector- help in spread/movement of the disease agent
4. Non-host - Plant thatdoes not have any effect on host plant
5. Host range- species of plant that have a probability to be infected by some insects.
6. Susceptibility- Inability of a number plant to resist disease.
7. Symptom- change in the appearance of a plant.
8. Sign- Physical change due to the presence of a pathogen
9. Pathogen-the power of adapting disease.
10. Host - Plant with the ability to develop a disease caused by any pathogen.
11. Resistance- the strength of a number plant to resist any disease.

# LIST OF FIGURES

FIGURE 1.1: LEARNING PHASE.....	13
FIGURE 1.2: TESTING PHASE .....	14
FIGURE 1.3: GREYSCALE IMAGE .....	16
FIGURE 1.4: LEVEL HISTOGRAM.....	17
FIGURE 1.5: BINARY REGIONS FOR A LOWER THRESHOLD.....	17
FIGURE 1.6: BINARY REGIONS FOR A HIGHER THRESHOLD .....	17
FIGURE 1.7: GREYSCALE VS COLOUR THRESHOLDING.....	18
FIGURE 1.8: PIXEL CONNECTIVITY .....	19
FIGURE 1.9: LABELLING OF REGION.....	20
FIGURE 1.10: DIFFERENT SEGMENTATION RESULTS.....	20
FIGURE 1.11: EXAMPLE OF INSTANCE SEGMENTATION.....	24
FIGURE 1.12: RGB .....	26
FIGURE 1.13: HSV .....	27
FIGURE 1.14: RELATIONSHIP OF KERAS.....	31
FIGURE 1.15: GENERATED OUTPUT.....	33
FIG 1.16: GENERATED OUTPUT .....	34
FIGURE 1.17: ALGORITHM COMPARISON.....	36
FIGURE 3.1: FEED FORWARDS PERCEPTRONS .....	39
FIGURE 3.2: LOCAL RESPECTIVE FIELDS WITH INPUT NEURONS .....	40
FIGURE 3.3: SCHEMATIC APPROACH OF REPRESENTING RNN .....	42
FIGURE 3.4: LONG SHORT-TERM MEMORY MODEL.....	42
FIGURE 3.5: EXAMPLE OF CONFUSION MATRIX.....	44
FIGURE 4.1: EXAMPLE .....	48
FIGURE 4.2: CLUSTER DETECTED.....	49
FIGURE 4.3: CLUSTER 1.....	49
FIGURE 4.4: CLUSTER 2.....	50
FIGURE 4.5: CLUSTER 3.....	50
FIGURE 4.6: CLUSTER SELECTION.....	51
FIGURE 4.7: SNAPSHOT OF DATASET PREPARED .....	51
FIGURE 4.8: ARCHITURE OF KERAS MODEL.....	52
FIGURE 4.9: ACCURACY OF KERAS MODEL .....	52
FIGURE 4.10: RESULT OF DECISION TREE CLASSIFIER.....	53
FIGURE 4.11: RESULT OF BERNOULLI NAÏVE BAYES CLASSIFIER.....	53
FIGURE 4.12: ADABOOST CLASSIFIER USING HEAT MAP .....	54



FIGURE 4.13: BERNAULLI NAÏVE BAYES CLASSIFIER ..... 55  
FIGURE 4.14: DECISION TREE CLASSIFIER..... 56  
FIGURE 4.15: GAUSSIAN NAÏVE BAYES CLASSIFIER..... 57  
FIGURE 4.16: K-NEAREST NEIGHBOR..... 58  
FIGURE 4.17: MULTINOMIAL NAÏVE BAYES CLASSIFIER ..... 59

## **ABSTRACT**

Plant infection, there are some typical condition of a plant that is unchanged. All plants and wild developed the same, are dependent upon sickness. Albeit that all specie is helpless to trademark infections, these are, for each situation, moderately very few. The event and predominance of plant ailments differ from season to season, contingent upon the nearness of the pathogen, natural conditions, and the yields and assortments developed. Some plant assortments are especially dependent upon episodes of how sick they are while others are increase problem to them.

Loss on crops due plant disease may result in abnormal condition and starvation in less developed area in all these places disease control method is limited and annual loss of 25 % to 40 % of crop.

Plant disease a normal part of nature and usually it's not easy for farmer to find the type of disease. The main motivation of project detects the disease in plant by using semantic segmentation and neural networks.

# CHAPTER 1

## PROJECT INTRODUCTION

### 1.1 Introduction

Farmer suicides include 1.55 out of 10.3 in suicide rate per hundred thousand people in India, mostly due to failed crop.

How to stop or reduce crop failure?

Identifying diseased crops and working towards a solution.

Due to this, researchers have proposed several methods aimed at detecting diseased crops using various software algorithms.

These methods include subjective, evolutionary, scientific, and area-specific approaches for working on various crops.

- Physical approaches

On-field identification of disease types by experts and estimation.

- Communication-based methods

Sending crop images to experts via mobile systems and getting evaluated data back.

- Software Methods

Software to identify disease type and level from photo.

### **Plant pathogens are classified into**

Fungi and Bacteria Plant pathogenic bacteria causes different kind symptoms such as galls and overgrowths, leaf spots, sand blights, soft rots, scabs and cankers.

## 1.2 PLANT DISEASES MANAGEMENT

These are methods of process

1. Disease
2. Diagnosis
3. Management:

Use Integrated Pest Management (IPM)

Using different type of management practices like cultural, physical, and biological help in reducing the amount of cause in a suchway that is efficient, and eco safe.

Basic Physical approaches

### Chemicals

Chemicals are used in plants to prevent them from insects and kill weeds to have control growth. when are used fertilizers substances that help in dispersing and have cleansing properties like detergents are main used cause damage . Sulfur dioxide causes from the burning of oil causes toxicity for plants and have concentrations of as 0.15ppm in air.

Proposed and analyzed approach disease identification has these steps

- image loading
- Image preprocessing,
- image segmentation
- feature extraction
- classification.

Image processing technique can be applied on various applications:

- To detect the plant leaf, stem and fruit disease.
- Finding the boundaries of effected area of plant.
- Determine the size and type of fruit or plant.

## Image Segmentation using K-Means clustering

Image segmentation is used to separate objects and use each object individually for it is. Image segmentation. the data before pattern recognition is used in feature extraction, and identify of the image.

Segmentation of image is use to classify image into different groups. different methods are used by methods is K-Means is used most.

The complete procedure for detecting diseases in a given close-up image is primarily divided into three phases:

- Feature extraction phase
- Training and learning phase
- Testing phase

### 1) Feature Extraction

This is used to reduce the undesired content by comparing the data. now the new data set is features less and is based on component comparing data by extracting the image data set useful to the user and segmentation is done.

### 2) Training and learning phase

This means repeats. The process is updated by the weights this is called one training step.

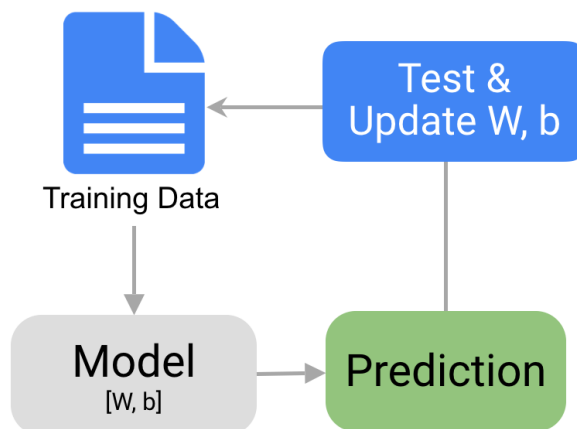
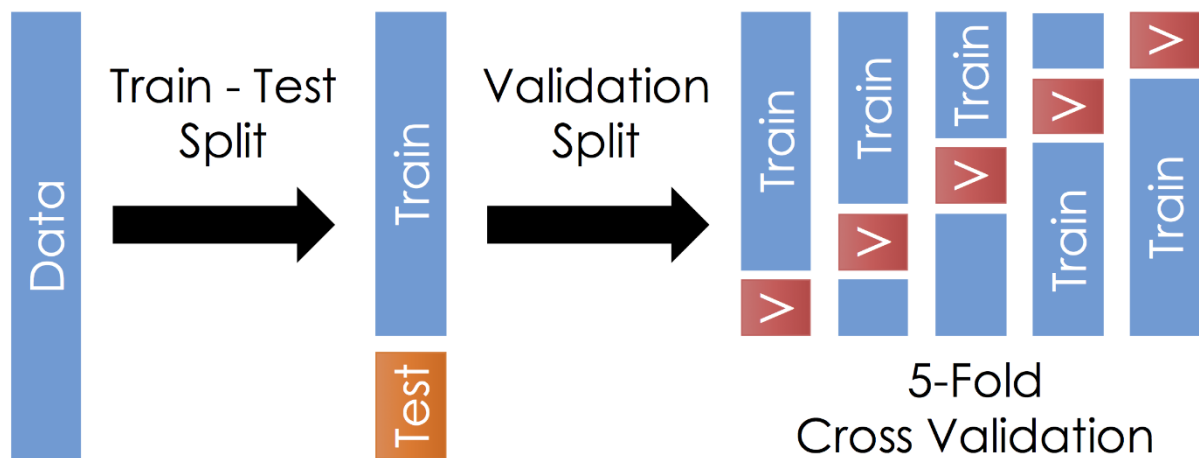


Figure 1.1: Learning Phase

The AI procedure appears, that you start with a preparation stage. A stage where you are essentially preparing the calculations to make the correct yield. In the in the learning stage you are having the info parameters. Fundamentally the arrangement of the model and you have the info information. What you're doing is you are preparing the calculation. While preparing, the calculation changes the preparation parameters. It additionally changes the pre-owned information and afterward you are getting to a yield.

### 3) Testing phase

**Test set:** using unseen data from the data set only to enhanced the performance of a full classifier.



**Figure 1.2: Testing Phase**

The data is separated into 3 segments and start the training process.

A training set of implementing model, on other hand a test set by validating the model to be from open source built. All the points excluded from the test set during the training of data. The data set is used into a training set as well as validation set in some iteration and training set is divided into the validation of these set are test of the set each iteration. Model are usually uses any process of 3 step and 3 point. Once the model is prepared we are using the trained model as a predicting channel for using testing data. Once it is done confusion matrix is developed, this help is how well our model is trained. A confusion matrix is classified into four paraphrases are used, True and False. We want to get all values in the True and negatives or true and positives to get a more accurate model. the Confusion matrix is classified into different classes.

**Training data set** — help in training the model, it can vary from 40% due to availability data on training time.

**Validation data set** — the selected model used all the training data as well as the running modelling on useable data subset of the data has ranges from 15% to 25%. Selected data helps by providing the discrete of the model. If the error exceeds validation dataset increases, then we have an overfitting model.

**Test dataset** — it is classified as hold to our data set. It consists of data that has never been able to operate on training. Test data set is helping with modelling final evaluation and is around 7% to 24% of the database.

No validation is used on test data set.

As the *sample is varied between data training and data set* continue model shows better on results data it unable to generate dataset. This results in low training and high test error rate.

The dataset is used to split training, validation and data set, subset of data is used and as a result training is usually on fewer observations the model will not perform well and the test error rate for the model to use on the complete dataset

## 1.3 Cross-Validation

It is technique which splits the data in one or more minor subsets then trains the data on different subset and using it in evaluating the performance of the models.

To decrease variability of many rounds of cross-validation are performed with many sets from the same data. But on the results from many rounds it comes up with an estimate of the modelling that predictive nature.

### 1.3.1 Methods of Cross Validation

Method involve in performing training data with 50% and rest 50% is used for the purpose of testing. One of disadvantage of this method is training is being performed on only on half of the data, it is possible that the other half of the data may have some information which can be unused while training our model i.e. higher bias.

## Leave One Out Cross Validation

In this method is used to train on whole data but leaves only one to many data-point because of unviabilityof dataset andpoint.An advantage is that by using this kind methodis all data points arebias.one and disadvantage it results in variation that is high in number in the testing model aswell as testing many to one data point. If the data points are used in outlying it can predict higher variation. Another disadvantage is that the dataset uses a lot of time to execute the data.

## Validationof k-Fold Cross

This process is used to spate the the data inmany small and big folds as used in final data output and then perform date set is used for training the uniform subsets but it leavesthe till last subset because the evaluation depend upon quality of the data. the mode of iterateis M times is differentfor all subset but is reserved for purpose of testing.

Segmentation Techniques: partitions of this image is done into different regions with all pixel with uniform nature. It is used for image segmenting the data the regions showthe objectsthat areusing. segmentationfor classifying some step all levelsof image process to transform a grey and then imageof different colours to all function of image separation in terms of features, objects, and scenes. All used imageare processed and it depend on the realtime andseparation,all uniformedpositions of an image have results inunsavoury.

## Simple thresholding

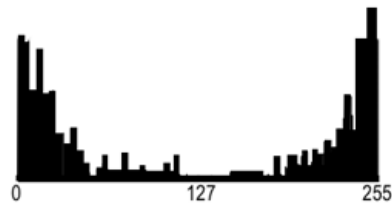
Pixel grey level is the most common of all that is

$g(x,y) = 0$  if  $f(x,y) < T$  and  $g(x,y) = 1$  if  $f(x,y) \geq T$ , where  $T$  is the threshold. The 2 thresholds,  $T1$  and  $T1$ , region 1 can be defined:  $g(x,y) = 0$  if  $f(x,y) < T1$  OR  $f(x,y) > T2$  and  $g(x,y) = 1$  if  $T1 \leq f(x,y) \leq T2$ .



Figure 1.3: Image using greyscale





**Figure 1.4: all level mostly grey**



**Figure 1.5: regions for a lower threshold**



**Figure 1.6: Binary regions for a higher threshold**

### **Adaptive thresholding**

It is used to separate dataset of the object from the background and it is used for adjust background.looking into errors, they are used for assigning different path sedation and pixel path transmission it is using all the defined path as told as threshold.

It is based on reefing of the data is use in data set, on at estimate peak time it is using the grey level for the pixel'sadjustment and is relating the

Its peak value due to this this the pixel probability decrease .at background the pixel is done at at region by define the grey  $f(x, y)$  and define its output as to by the grey mean level that found the object and background at high levels

### Thresholding color

It is a separation that is used to using colors by comparing the grey level and red green blue, it shows that color that are alteration have strong color component and HIS component too. It is used to have the stability and enhance the performance to respective color by using more stable saturation methods .it focuses on grey scaling of more and more color segmentation.

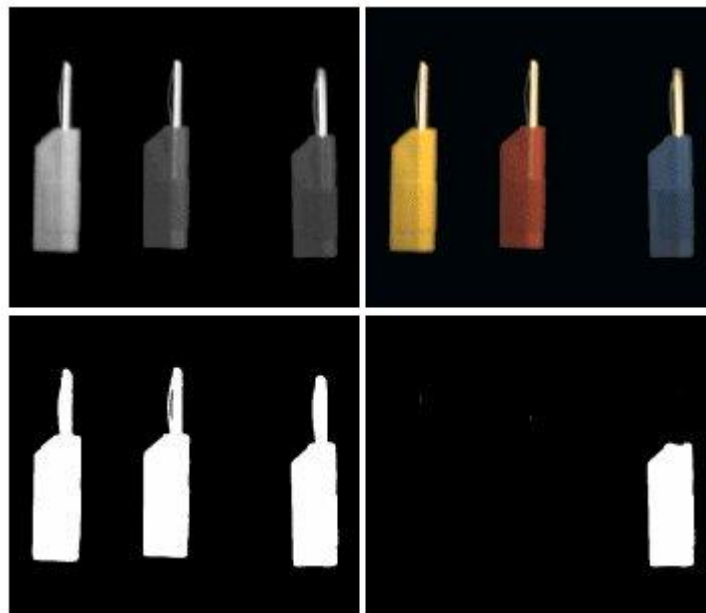


Figure 1.7:fig showingGreyscalevscolorthresholding

### 1.4 Contextual segmentation: Region growing

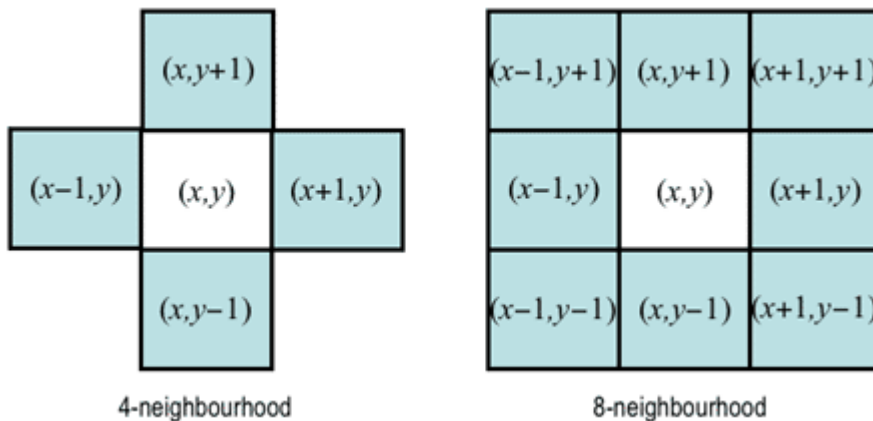
This pixilation is Non-contextual kind of thresholding that has are not using any locations in the image graph. It canhave used for separating all the objects from using that close to pixel label that operate on single term, it has only to contextual segmentation that are based on Discontinuity .it is used to find the complete boundaries by using the data and in uniform format by assuming not even a single change

inboundary condition .it is used to group the two segmentation by by using the criteria and other split between the regions.

### 1.4.1 Pixel connectivity

It is defined as the peeling the adjutant neighbor of the data in single format that is uniform in nature and predicts that higher value if the data as hoe in and illustrated in the figures.

- All the Alternate measurement  $\{(x-1, y), (x, y+1), (x+1, y), (x, y-1)\}$
- A 8 interchanges are adding to the 4 substitute • four corner to corner  $\{(x-1, y-1), (x-1, y), (x-1, y+1), (x, y+1), (x+1, y+1), (x+1, y), (x+1, y-1), (x, y-1)\}$ .



**Figure 1.8: Pixel Connectivity**

All four alternate paths starting  $p_1$  and to next pixel  $p_n$  are defined as  $(p_1, p_2, \dots, p_n)$  and  $p_{i+1}$  is a that is extended till  $n-1$ . In I The path is 8-connected pixel are 8-neighbour of  $p_i$ . Pixels all the region are showing and illustrating that the data as the subset go higher valuing of the data for higher pixel value.

labeling region to show pixel as the data set that have stars:

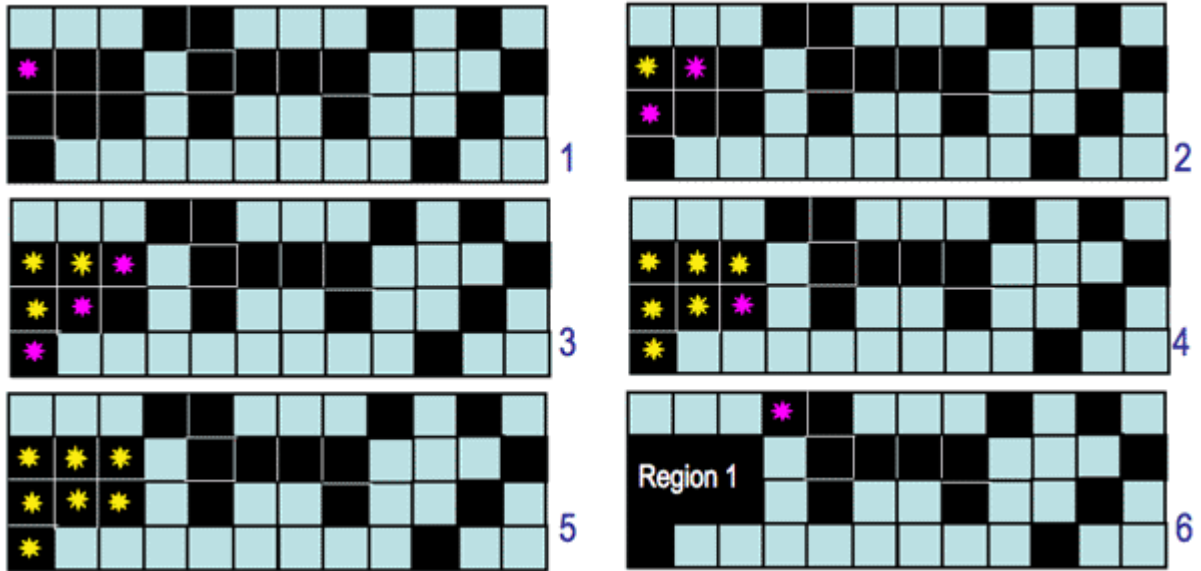


Figure 1.9: Labelling of region

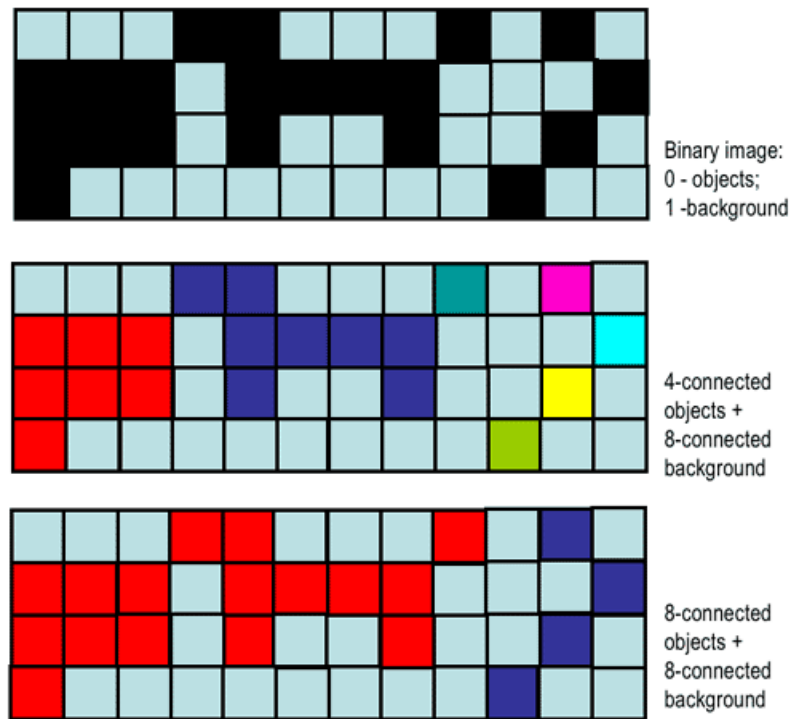


Figure 1.10: Different segmentation results.

### 1.4.2 Region similarity

The uniform usage of the data as the set of using an integrating the sub data set involves illogical statement and condition value of the data that show the pixel in the region and it is used to predict the data with the signal variation in the data subset and having different property such as color using grey

green red black and white sometime for showing pixel data in more effective manner as illustrate in pictures the variation on small size of pixel can result in more and more effect in final color picture.

### **1.4.3 Region growing**

This is the kind of segmentation in which the user that data without changing pixilation value to higher or lower segmentation for alteration of the data to show the data segmentation this is uniform in nature will sow the uniform data to improve the user experience. this kind of segmentation is highly unstable and small change in segmentation can result in large change in region of data maps are used into different segmentation of the data for showing eachregion, different seed value and pixel connectivity.

### **1.4.4 Split-and-merges**

It is kind of segmentation that uses that data set that considered the entire picture in one region and splits the data into many region and help to identify the the set of data and additional data is added intouniformly by adding the data for growing it show the desired number of establishment.

It usually split the data in the M number of the data segmentation, the uniform P<sup>®</sup> are used as true and p<sub>9imsge</sub>) is false of the dataset and p (quadrant) has all the value of the data set and all the other values are true in nature.

### **1.4.5 Texture segmentation: Spectral features**

It is characterized in special property of the region, all the local use the text of computedneigh our such as simplest statics example to show the level of square that has nan in neighborhoodcentered of the pixel.

It helps showing the radical Spectral segmentation techniques angularly integrated power spectrum of all thedifferent region go the data segmentation. Radial parameters

Ring of radius r

An orientation,  $\theta$

An angular width,  $\Delta\theta$ .

## **1.5 Feature extraction and selection in image processing**

There are different kinds of feature extraction techniques; many of them work on a single data set while some work on different sets of data. In this kind of technique, the algorithm chooses to travel all data and at the end it chooses to filter in interested data information to process this is called the feature extraction of the image.

### **1.5.1 ORB**

ORB (Oriented FAST and Rotated BRIEF) algorithm is the most famous of all as it acts as the combination of FAST and BRIEF data set.

It works as it finds the edge of the image and fast component help in identifies features as areas of the image on bases of color and brightness. More than 8 mean it has mean flag to process the desired data as fast technique, once detecting the entire data set it keep working on the data as its first priority to process the data in different segmentation the usually strings are of the range 150 to 300 ppm, CVS help in comparing the frames in dataset, this method is great for any of feature extraction in data.

### **Color gradient Histogram**

This kind of method used to frame different set of instruction on domain of the color like green red grey white etc. it is used to predict the prime value of color grid in data set of picture.

### **1.5.2 Vantage Point Tree**

It works on fundamental principle that is copy the preferred information from dataset, the algorithm helps in find the image data and alternate data in or near identical image of data it comprise all the data information that is 32 bits and keep moving. the image that have the similar composition of the data as using the similar composition of the data and the data would be neighbor base on the data, this is called the hashing of data.

### **1.5.3 KAZE**

This algorithm work on principle of the wind that is nonlinear in nature. the alga is processing the data on basics of the data that usually take the principle working of the nonlinear data set in the image domain.

This is puller kind of operation that work manually by user some of them are shown below:

- Local Binary Patterns
- Speeded up Robust Features
- Histogram of oriented gradients
- Heat wavelets
- Color histograms

## **1.6 Image Segmentation techniques**

It is used to divide image in different segmentation and it will be easier for u to differ ate the image on different segments about the region do not contain any image. image segmentation is dived into various operation as shown

Images collect all data set and considered the image that are processing the data that has attributes using image segmentation.

Image segmentation is created using a pixel oriented work for each data set that image and the technique help in building the understanding the different object during in the image.

Image segmentation have different kind of application.

- AI and machine learning segmentation
- Self driving cars
- Satellite image processing

### **1.6.1Types of Image Segmentation**

#### **Semantic Segmentation**

It is kind of the process of that linking different pixel in an image of different classes to complete a class label. These labels can anelephantand, rat, car, homeetc.

It has various applications like autonomous self-driving cars and robotics satellite etc.because the models need areknown for its real time data output in in real environment during the operations.

## Instance Segmentation

Semantic segmentation can be used as image classification at a pixel level. An image that has many cars and trees but by separating the class of models known as instance segmentation we are able to label the separate instances where an object appears in an image. This kind of segmentation can be used in applications that have a number of objects, such as counting the amount of trees in a park.

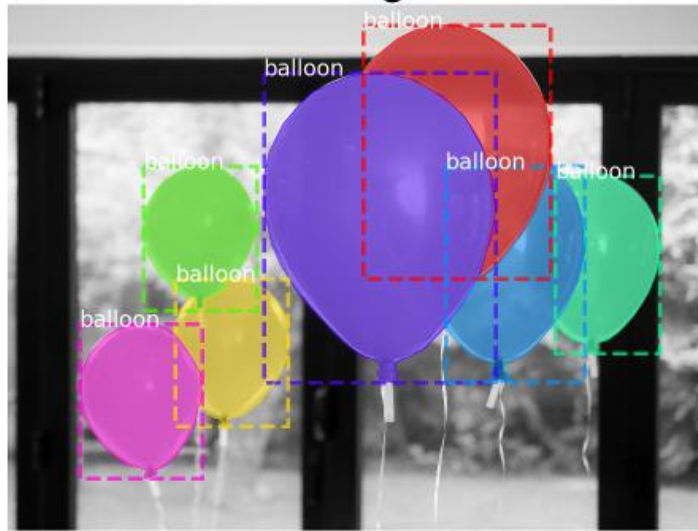


Figure 1.11: Example of instance segmentation

## Region-based Segmentation

It might be used as segment alignment project to identify different work in data set and therefore image background is working on its own data set.

In some cases, threshold value is fixed to data. The pixel values that are below or above that threshold shows problem because objects do not align with the background. This is known as

It can be shown as

## Edge Detection Segmentation

When can divide the data have these are 2 different region grayscale values and pixel value. The edges are often considered because of the discontinuous local features of a picture.

Step to show the data process

- Choose the weight matrix
- Then take the image
- Perform operation such as multiplication and to show the output
- Move the weight matrix from dataset.
- Convolve all the pixels of the input for complete usage.



The output of this process is convention in nature of data set and always remain into segmented parts Clustering is the process of dividing the different data set in in to groups and and these groups are known as clusters.

### **1.6.2 Image Segmentation based on Clustering**

By image clustering we use the data set to read and elite new algorithms of its choice as the output. Such matrix is sober in nature and the data appoint are uniform at the output to show uniform type of results.

- Take the weight matrix
- Put it on top of the image
- Perform element-wise multiplication and get the output
- Move the weight matrix as per the stride chosen
- Convolve until all the pixels of the input are used

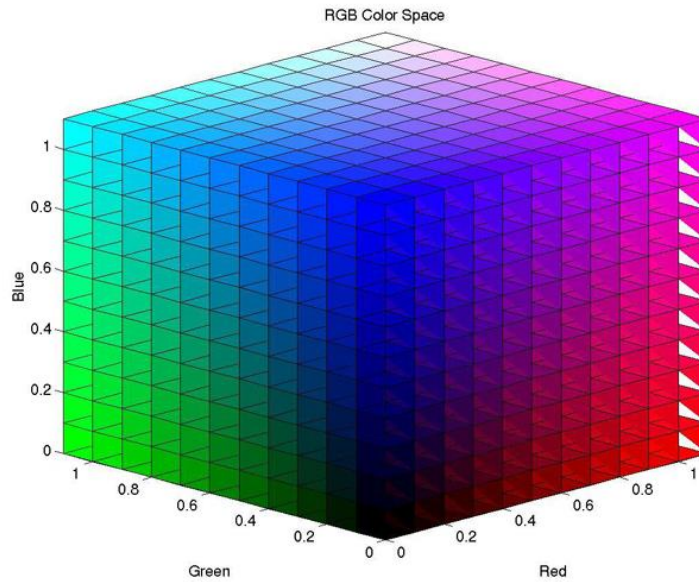
Convolution is backend of the process the entire data set depend upon the convlutin as main stream guideline It also helps in detecting all edges horizontal and vertical .

One such weight matrix is the sober operator. It is typically used to detect edges. The sober operator has two weight matrices – one for detecting horizontal edges and the other for detecting vertical edges.

## **1.7 Color Spaces**

A range of colors are often created by first color that we choose in the segment to show the pigment and the colors are defining a color space.

Color space depend upon the color model system that help in the abstract rationof mathematical model which has been describes from the the range of templates in color and numbers or color components RGB. Spacing is used in the frame of reference and data sub. singledot represent the color validation of the data set in image.



**Figure 1.12: RGB color space.**

L\*A\*B color space

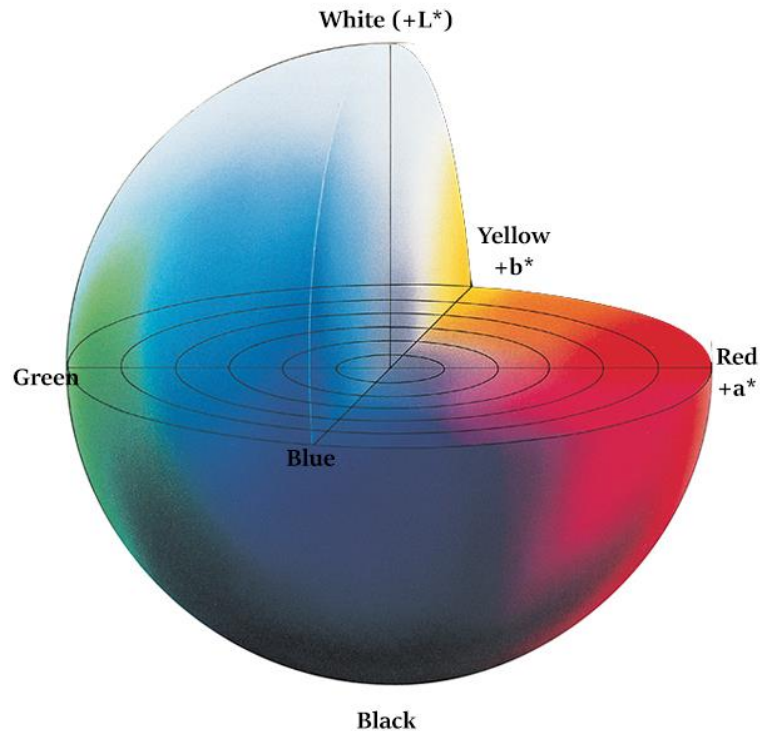
L\*a\*b\* color space is the CIELAB and it mostly used foremost popular color spaces for calculating data colors. It is used in color management and is widely adapted by organizations ..

In the L\*a\*b\* color space mean L\* shows lightness a\* and b\* shows the coordinates. a\* and b\* are color directions: +a\* is that the red axis, I -a' is that the green axis, +b\* is that the yellow axis -b\* is that the blue axis.

Plotting all the variable on the color space predicts that data as an apple and the L\* shows lightwith some property of apple.

The L\*a\*b\* shows the spacing between all selected segmentation and dataset with management to show the excite products are list color lab

Color are different because it calculates it with the L\*a\*b\* to add the values for major and data show aquantity check.



**Figure 1.13: HSV color space**

Otherwise both RGB and CMYK, are used for choosing the colors, HSV is operated humans on human color interaction and has different components:

- hue,
- saturation,
- value.

This color space is used to predict the terms and shade like saturation gray and also the brightness value. Various color is picked due to inflection the value for HSV and HSB work frame and is also work on color modelling.

The HSV colors are used to predict in circle cone or cylinder, but shows three components

### 1.7.1 SATURATION

It shows the quantity of gray of different color 0 to one hundred pc. Reducing this component toward zero it will turn into more grey in nature .

## **1.7.2 HUE**

Hue is varied from 0 to 360 degrees. It is a color segmentation modelling:

Red falls somewhere in the range of 0 and 60 degrees. Yellow falls somewhere in the range of 61 and 120 degrees. Green falls between 121-180 degrees. Cyan falls between 181-240 degrees. Blue falls between 241-300 degrees. Fuchsia falls between 301-360 degrees. Tone has shading part of the model, communicated as assortment from 0 to 360 degrees: Red falls somewhere in the range of 0 and 60 degrees. Yellow falls somewhere in the range of 61 and 120 degrees. Green falls between 121-180 degrees. Cyan falls between 181-240 degrees. Blue falls between 241-300 degrees. Fuchsia falls between 301-360 degrees.

## **VALUE (OR BRIGHTNESS)**

It depends on the Value like saturation show the intensity of the combination of colors, range of the 0-100 percent, shows the 0 is marked as black where 100 is marked same all colored.

## **Uses of HSV**

It is used as a color model as selecting colors with paint and HSV better shows better results in framework.

This shows the main contribution of high-quality interface of the graphics. But less use of the RGB and CMYK, the HSV is better than other software algorithms.

HSV shows the values and change in the figure of desired output because of change in graphics and color content.

## **1.7.3 RGB color space**

This kind of algorithm is used for spacing the color content for the employed by graphics cards and monitors and on LED and LCD

It consists of many components like components, green red black and there is the color the so-called base coloring. Depending on what proportion you're taking from each base color,

We can create some of the effect using some parameter as discussed in following -.

In the event that we utilize none of the three base hues the outcome's dark. On the off chance that we blend red and green, you'll get yellow. In the event that we blend red and blue, you'll get red. By blending each of the three base hues in with full force, we'll get white. So by blending the base hues in a few forces, you'll make any necessary shading.

Since every one of the three base shading segments are equivalent in significance, a 3 dimensional solid shape are frequently drawn, as demonstrated to contents called a Color space.

## **1.8 SOFTWARE USED**

### **1.8.1 Open CV**

Open CV may be a cross-platform library using which we will develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection.

It are often defined as a discipline that explains the way to reconstruct, interpret, and understand a 3D scene from its 2D images, in terms of the properties of the structure present in the scene. It deals with modeling and replicating human vision using computer software and hardware.

Computer Vision overlaps significantly with the subsequent fields –

Image Processing – It focuses on image manipulation.

Photogrammetry – It is concerned with obtaining accurate measurements from images.

Computer Vision Vs Image Processing

Image processing deals with image-to-image transformation.

Computer vision is used for classify of physical objects from their image. The output of computer vision is used for the description or an interpretation of structures in 3D structure.

### **1.8.2 Scikit-learn**

It is used for uploading the robust library with machine learning as Python at its backend. It shows the usage by selecting all efficient tools which are embedded in machine learning and statistical modeling that have clustering and have interface such Python. This kind of operation are essential for data that operate on Python, having libraries as NumPy and Scipy.

Scikit-learn focused on datamodeling such as jumpySclera are as follows –

Unsupervised Learning algorithms – On the other hand, it also has all the popular unsupervised learning algorithms that operate on clustering as well as factor Principal Component Analysis to all CNN networks.

Clustering – It is model is used for grouping unlabeled data.

Dimensionality Reduction – It is used to reduce the number of dataset that are unused which can be further used for summarization, visualization and feature selection.

Ensemble methods – As name suggest, it is used for combining the predictions of multiple supervised models.

Feature extraction – It help to extract the features from data to define as text datainterpretation .

### **1.8.3 Keras**

Keras is an opensource software that has deep learning framework that operate on python. Keras models, Keras layers, Keras modules and eventually conclude with some real-time applications.

Keras is open source and run machine libraries Tensor Flow, Thea no or Cognitive Toolkit (CNTK). Thea no may be a python library used for fast numerical computation tasks. Tensor Flow is that the most famous symbolic math library used for creating neural networks and deep learning models. Tensor Flow is very flexible and the primary benefit is distributed computing. CNTK is deep learning framework developed by Microsoft. It uses libraries like Python, C#, C++ or standalone machine learning toolkits. Thea no and Tensor Flow are very powerful libraries but difficult to understand for creating neural networks.

### **Features**

Keras leverages various optimization techniques to form high level neural network API easier and more performant. It supports the following features –

- Consistent, simple and extensible API.
- It supports multiple platforms and backbends
- Easy to test.

- Keras supports both convolution and recurrent networks.
- Deep learning models have discrete components that you can use into many ways.

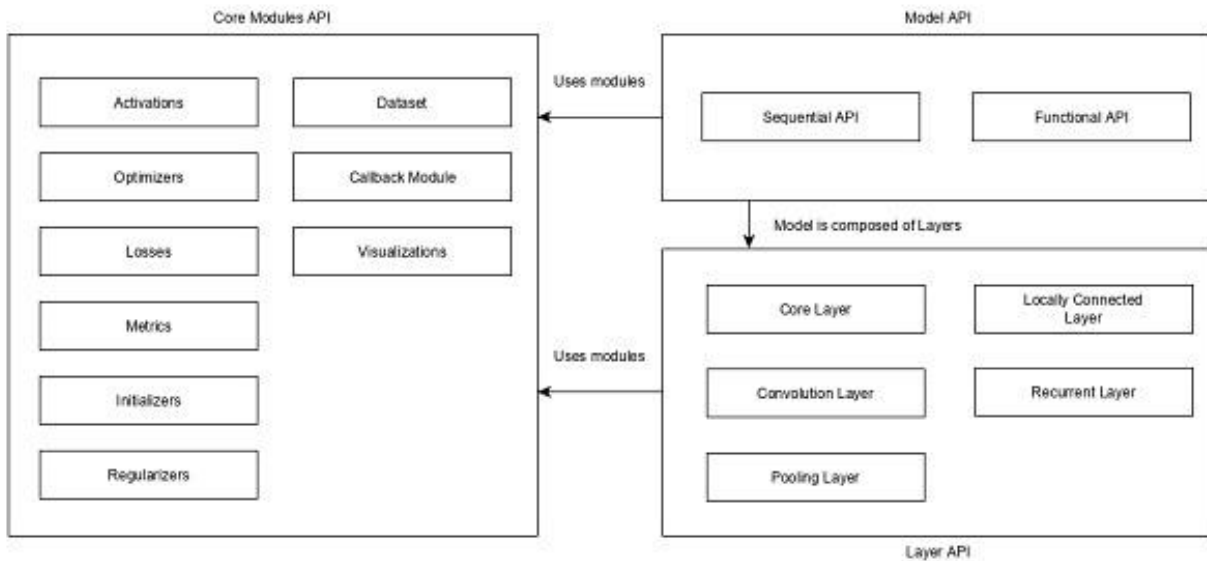
## Architecture of Keras

Keras API can be divided into three main categories –

- Model
- Layer
- Core Modules

In Keras all ANN are used by Keras Models. Keras Layers and represents ANN layers like input, hidden layer, output layers and pooling layer, Keras model and layer access Keras modules for activation function, loss function, regularization function, etc., Using Keras model, Keras Layer, and Keras modules, any ANN algorithm (CNN, RNN, etc.,) are often represented during a simple and efficient manner.

The following diagram depicts the connection between model, layer and core modules



**Figure 1.14: Relationship of Keras**

## **Keras layer**

A Keras layer requires shape of the input (input shape) to know the structure of the input file, initializer to line the load for every input and eventually activators to rework the output to make it non-linear. Keras layer requires below minimum details to make an entire layer.

- Quality of input data
- Neurons that are used / units using in layer
- Initializers
- Regularizes
- Constraints
- Activations

### **1.8.4 Matplotlib**

It is the software that help in Python packages library that is used for data interpretation. It helps in cross-platform library for 2D plotting and data to show arrays. Matplotlib has a backend of Python and NumPy.

It has library that that is called Pilaf that help us in designing the tooling such as MATLAB, a programming platform such as Math Works. It works with the NumPy that its usage is considered as value base as web application in developing industry

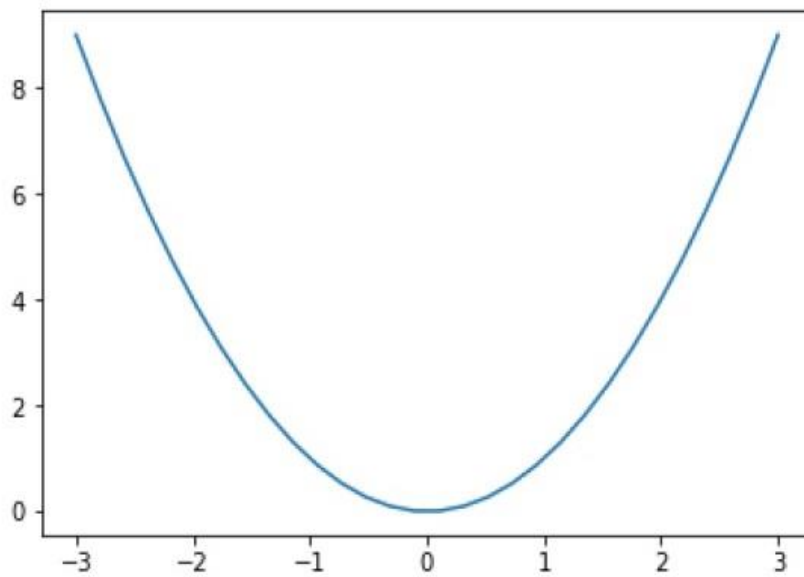
#### **Matplotlib - PyLab module**

PyLab is used in Matplotlib as library that is object-oriented and help in plotting the data in significant graphs.

- It is considered to be in module frame work but act as a bulk to imports pilot that is used for plotting
  - NumPy that is used for Math's and array operation in algorithm.
- a) Plot command is used for plotting of the graph in bases of data viability. It considers a good length of operation in plotting the graph that is used for the plotting the desired data set



```
from numpy import *  
from pylab import *  
  
x = linspace(-3, 3, 30)  
  
y = x**2  
  
plot(x, y)  
  
show()
```



**Figure 1.15:Generated Output**

### **Matplotlib - 3D Wireframe plot**

It helps you in plotting the grid values in 3-D surface, and this also help plotting the. The plot wireframe() function is used for the purpose –

```

from mpl_toolkits import mplot3d

import numpy as np

import matplotlib.pyplot as plt

def f(x, y):

    return np.sin(np.sqrt(x ** 2 + y ** 2))

x = np.linspace(-6, 6, 38)

y = np.linspace(-6, 6, 38)

X, Y = np.meshgrid(x, y)

Z = f(X, Y)

fig = plt.figure()

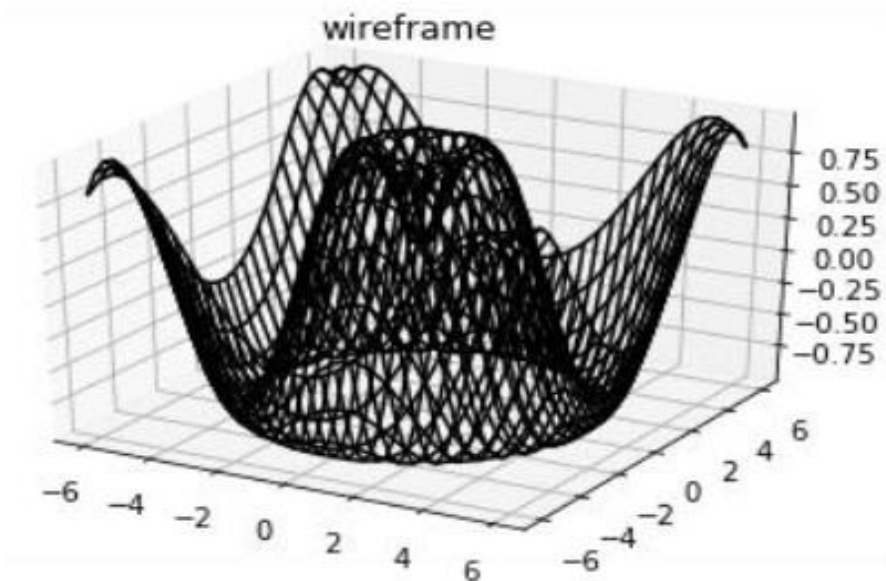
ax = fig.add_subplot(1, 1, 1, projection='3d')

ax.plot_wireframe(X, Y, Z, color='black')

ax.set_title('wireframe')

plt.show()

```



**Fig 1.16: Generated Output**

## 1.9 Clustering Methods

### 1) Scikit Learn

It most power full machine learning methods as it works on finding the similarity & relationship patterns among data sets and the cluster that is using the samples from the group have similarity support. Clustering shows the grouping to all the present unlabeled data, that is why it is important. The Scikit-learn library have cluster to perform clustering of unlabeled data. Under this module scikit-learn have the subsequent clustering methods –

### 2) KMeans

This algorithm helps in changing the requires and clusters to be specified that's why it assumes that they are already known. The main logic of this algorithm is to cluster the data separating samples in n number of groups of equal variances by minimizing the criteria known as the inertia. it is the number of clusters shown in the algorithm is by Kmean.

Scikit-learn have sklearn. cluster. Means module to perform K-Means clustering. While computing cluster centers and value of inertia, the parameter named sample weight allows sklearn. cluster. Means module to assign more weight to some samples.

### 3) Affinity Propagation

This algorithm is used in the concerting of between different pairs of samples until convergence. It does not require many clusters seems to be uniform in nature before running the algorithm. The algorithm shows the time complexity of different order ( $N^2T$ ), which act as the biggest disadvantage of it.

Scikit-learn have sklearn. cluster Affinity Propagation module to perform Affinity Propagation clustering.

### 4) Mean Shift

This algorithm mainly discovers blobs during a smooth density of samples. It is using the the data points to assign the clusters on bases of shifting points with density that is higher than the data points.

The parameter such as bandwidth are operating on size of the dataset and is automatically set the quantity of clusters.

### 5) Spectral Clustering

It is matrix that perform data by reducing it into fewer dimensions. The use of this algorithm is not advisable when there are large number of clusters. Scikit-learn have sklearn. cluster. Spectral Clustering module to perform Spectral clustering.

#### 1.9.1 Comparing Clustering Algorithms

Following table will give a comparison (based on parameters, scalability and metric) of the clustering algorithms in scikit-learn.

Sr.No	Algorithm Name	Parameters	Scalability	Metric Used
1	K-Means	No. of clusters	Very large n_samples	The distance between points.
2	Affinity Propagation	Damping	It's not scalable with n_samples	Graph Distance
3	Mean-Shift	Bandwidth	It's not scalable with n_samples.	The distance between points.
4	Spectral Clustering	No.of clusters	Medium level of scalability with n_samples. Small level of scalability with n_clusters.	Graph Distance
5	Hierarchical Clustering	Distance threshold or No.of clusters	Large n_samples Large n_clusters	The distance between points.
6	DBSCAN	Size of neighborhood	Very large n_samples and medium n_clusters.	Nearest point distance
7	OPTICS	Minimum cluster membership	Very large n_samples and large n_clusters.	The distance between points.

**Figure 1.17: Algorithm Comparison**

## CHAPTER 2: LITERATURE REVIEW

In this chapter different research papers are studied on the different topics related to our project.

- 1) Mohanty SP, Hughes DP and Salathé M (2016): In this paper they utilized the mix of expanding worldwide cell phone infiltration and late advances in computer vision by a collecting a public dataset of fifty four thousand unhealthy also, healthy plant leaves under uniform conditions and it is trained a deep convolutional neural network to spot 14 crop species and 26 diseases.
- 2) Sarah J. Pethybridge and Scot C. Nelson (2015): Here the analysts made an intuitive, iterative cell phone application on shading pictures to recognize sick from solid plant tissues and compute level of ailment seriousness. In this the client contacts the application's presentation screen to choose up to eight distinct hues that speak to solid tissues and he at that point moves a limit slider until just the suggestive tissues have been changed into a blue shade.
- 3) Bhimte, Namrata R., and V. R. Thool (2018): In this exploration paper the scientists introduced a framework utilizing basic picture preparing approach for programmed analysis of cotton leaf sicknesses where the characterization dependent on choosing suitable highlights, for example, shading, surface of pictures was finished by utilizing a help vector machine classifier alongside the use of different preprocessing strategies as separating, foundation expulsion, improvement.
- 4) Asheesh Kumar Singh, BaskarGanapathysubramanian ,SoumikSarkar, Arti Singh(2018):Here the specialists inspected the ongoing work where profound learning standards have been used for advanced picture based plant pressure. It give a similar appraisal of profound learning instruments against other existing methods.
- 5) Alexander Johannaesa, ArtzaiPiconb, Aitor Alvarez Gilab, JoneEchazarrab, Sergio Rodriguez Vaamondeb, Ana Díez, NavajascAmaia Ortiz Barredoc(2017): Here the analysts proposed a novel picture handling calculation dependent on up-and-comer problem area discovery in mix with measurable induction techniques to handle infection distinguishing proof in wild conditions by examining the presentation of early ID of three European endemic wheat maladies – septoria, rust and tan spot which was finished utilizing 7 cell phones and in excess of 3500 pictures caught in two pilot destinations in Spain and Germany during 2014, 2015 and 2016.

- 6) Islam, Monzurul, Islam, Monzurul, et.al. (2017): Here the scientists proposed a methodology that coordinates picture preparing and AI to permit diagnosing sicknesses from leaf pictures. Their technique groups maladies on potato plants from an openly accessible plant picture database called 'Plant Village'. Their use of help vector machine showed infection grouping more than 300 pictures with an exactness of 95%.

# CHAPTER-3: CLASSIFICATION ALGORITHMS AND NEURAL NETWORKS, EVALUATION METRIC

## 3.1 Feed forwards Perceptrons

This is a type of neural networks that is used to fake by associating the units that do not use shape for process. It is an essential kind of counterfeit neural system imagined and are less difficult than their partner, they're called feed forward

It is primarily used for supervised learning.

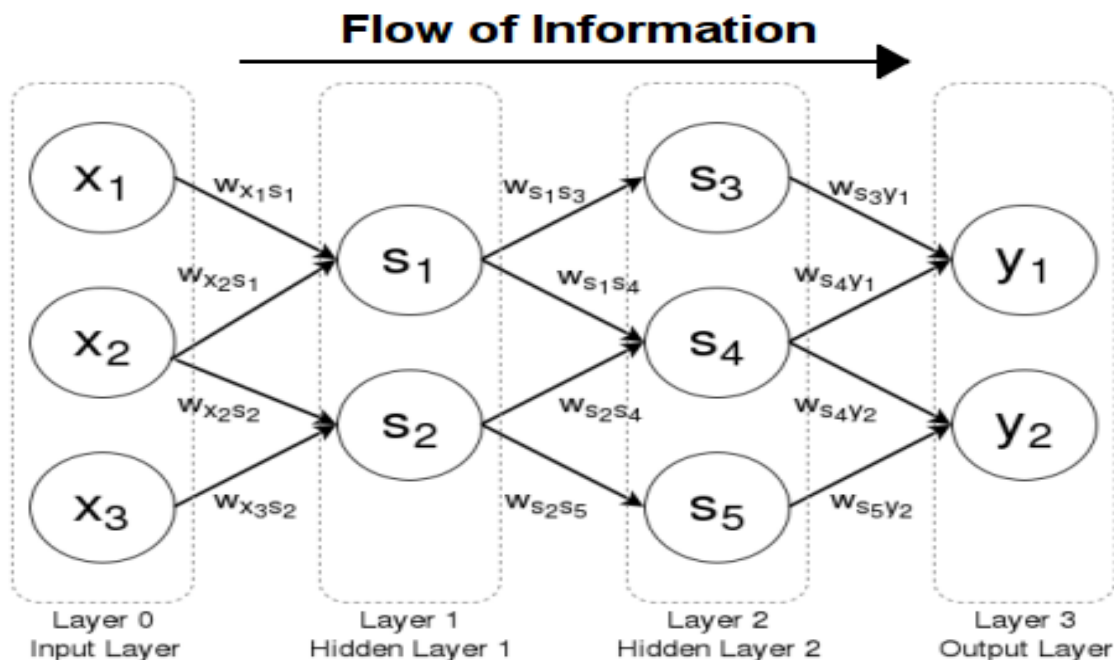


Figure 3.1: Feed forwards Perceptrons

## Convolutional Neural Networks

CNN is intended for handling the information through numerous layers of exhibits. This kind of neural systems is utilized in applications like picture acknowledgment or face acknowledgment. The fundamental contrast among CNN and the other normal neural system is that CNN has contribution as

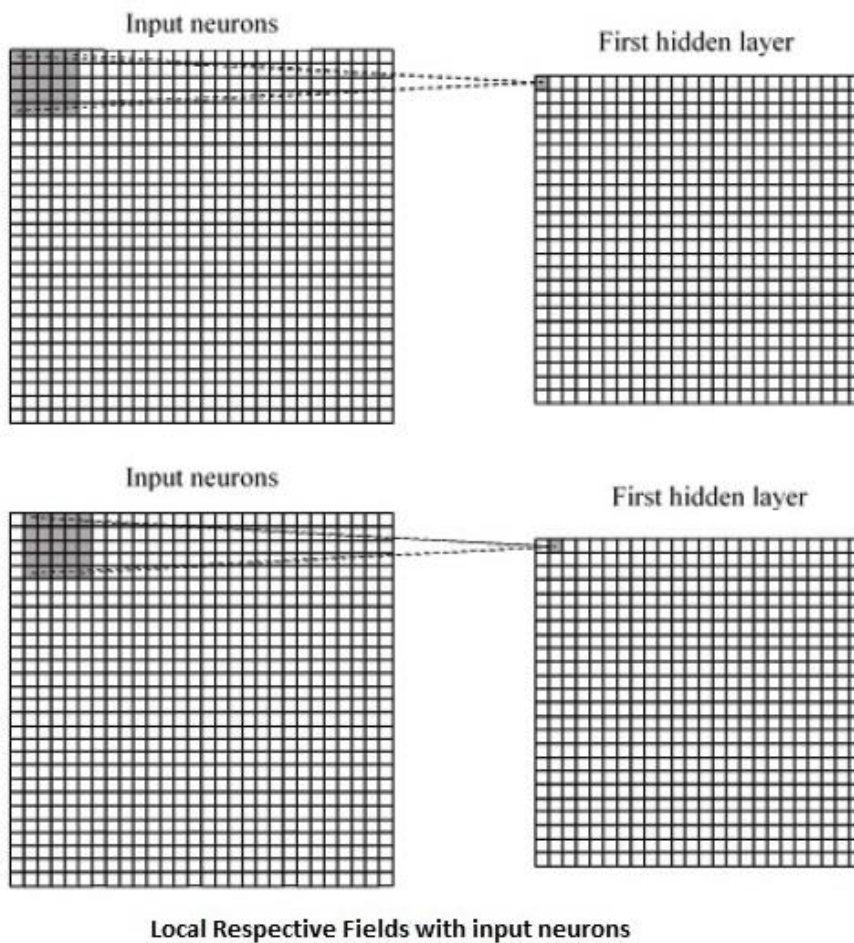
a two-dimensional cluster and works straightforwardly on the photos as opposed to concentrating on including extraction.

CNN classify the solutions for problems of recognition.

A convolutional neural network uses three basic ideas –

- Local respective fields
- Convolution
- Pooling

CNN uses spatial connections that exist inside the info document. Each simultaneous layer of a neural system associates some information neurons. This particular district is called neighborhood responsive field. Neighborhood open field focusses on the concealed neurons. The concealed neurons process the info document inside the referenced field not understanding the progressions outside the exact limit.



**Figure 3.2: Local Responsive fields with input neurons**



Every association learns a load of the shrouded neuron with a related reference to development from one layer to an alternate. Here, singular neurons play out a move now and again. This procedure is designated "convolution".

The mapping of associations from the info layer to the shrouded highlight map is characterized as "shared loads" and predisposition included is named "shared inclination".

CNN or convolutional neural systems use pooling layers, which are the layers, situated following CNN presentation. It takes the contribution from the client as an element map that comes out of convolutional arranges and readies a dense element map. Pooling of layer helps in making another layer with neurons of past layers.

### **3.2 Recurrent neural networks**

It is a kind of profound learning-focused calculation, which follows a consecutive methodology. In neural systems, we generally expect that each info and yield is free of every single other layer. These kinds of neural systems are called repetitive in light of the fact that they perform scientific calculations in successive way.

Consider the subsequent steps to coach a recurrent neural network –

Step 1 – Info a chose model from dataset.

Step 2 – System will take the information

Step 3 – An anticipated outcome is then figured.

Step 4 – The examination of real outcome created with the number-crunching mean will deliver a mix-up.

Step 5 – To follow the mistake, it is spread through same way where the factors are additionally balanced.

Step 6 – The steps from 1 to 5 are repeated until we are absolutely certain that the values declared for declaring the output are defined properly.

Step 7 – A precise expectation is made by applying these factors to get new concealed info.

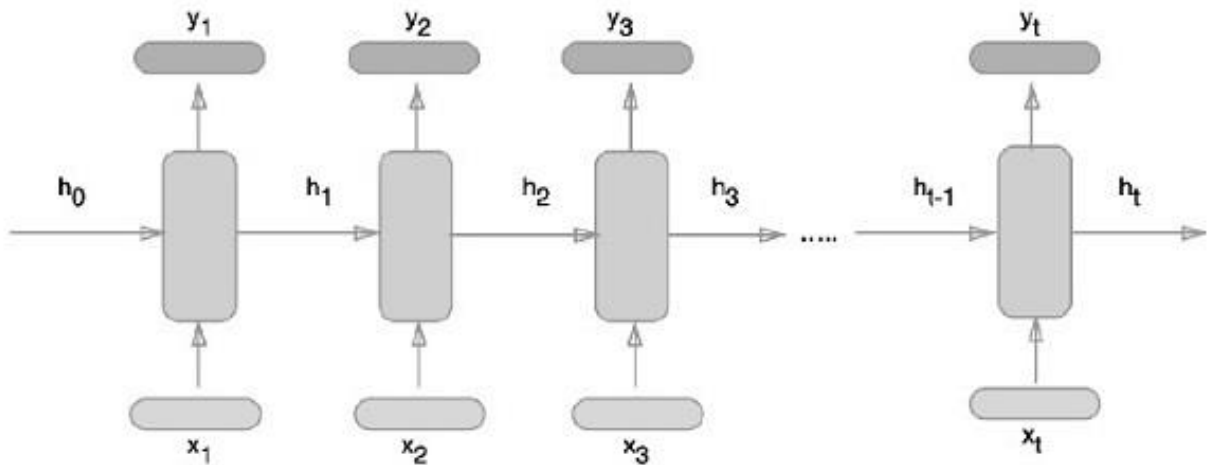


Figure 3.3: Schematic approach of representing recurrent neural networks

### 3.3 LSTM

This type of modelling is a class of RNN. So before we can jump to LSTM, it is essential to understand neural networks and recurrent neural networks.

It is a special kind of recurrent neural network that is fit for learning long haul conditions in data. This is achieved because the recurring module of the model has a combination of four layers interacting with each other.

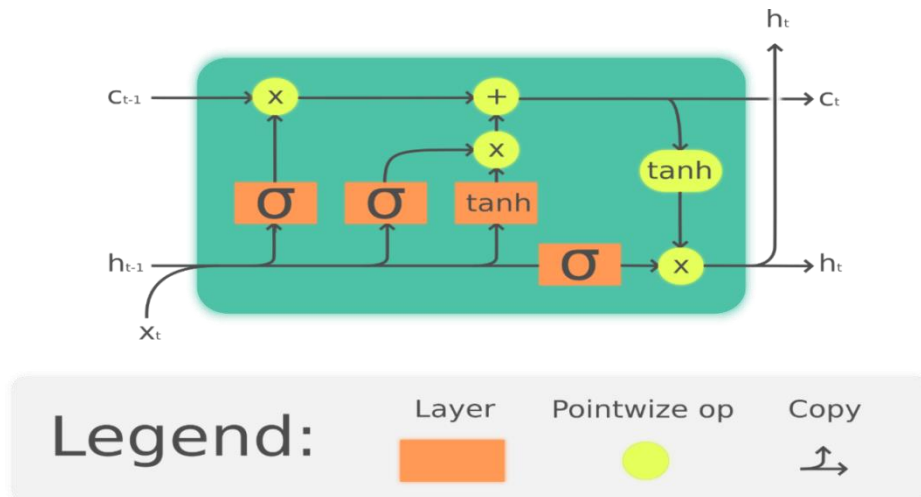


Figure 3.4: Long Short-term Memory model

### 3.4 DIFFERENT EVALUATION METRICS

Assessing your AI calculation is a significant a piece of any task. Your model may offer you fulfilling results when assessed. The vast majority of the days we use grouping exactness to live the exhibition of our model, anyway it's insufficient to truly pass judgment on our model.

- Classification Accuracy
- Logarithmic Loss
- Confusion Matrix
- Area under Curve
- F1 Score
- Mean Absolute Error
- Mean Squared Error

#### 1. Classification Accuracy

It is type of Accuracy that is used for the term exactness. And work on the principle of the proportion to number of right expectations for information tests.

**Accuracy = number of correct predictions/total number of predictions made**

It works well every class belong to the same number of the sample.

#### 2. Logarithmic Loss

It is kind of algorithms that work on principle of Logarithmic Loss or Log Loss, it has multi-class classification.

$$\text{Logarithmic loss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log p_i + (1 - y_i) \log (1 - p_i))$$

$y_{ij}$  indicates whether sample  $i$  belongs to class  $j$  or not and  $p_{ij}$  indicates the probability of sample  $i$  belonging to class  $j$

### 3. Confusion Matrix

Confusion Matrix act as matrixs table the shows us the us yield and portrays of whole execution for working on dataset of the program.

		Predicted: NO	Predicted: YES
n=165	Actual: NO	50	10
	Actual: YES	5	100

Figure 3.5: Example of Confusion Matrix

There are 4 important terms:

- True Positives : The cases in which it show YES and the output is YES.
- True Negatives : The cases in which it show NO and the output was NO.
- False Positives : The cases in which it show YES and the output was NO.
- False Negatives: The cases in which it shows and the actual output is YES.

Accuracy of the matrix are often calculated having the average of all the values to total number of sample .

$$\text{Accuracy} = (\text{True positives} + \text{False negatives}) / (\text{total number of samples})$$

#### Area Under Curve

Area Under Curve (AUC) is one of the most broadly utilized measurements for assessment. It is utilized for double grouping issue. AUC of a classifier is equivalent to the likelihood that the classifier will rank a haphazardly picked positive model higher than an arbitrarily picked negative.

True Positive Rate (Sensitivity): True Positive Rate is defined as  $TP / (FN + TP)$ . True Positive Rate compares to the extent of positive information focuses that are effectively considered as positive, w.r.t. every positive datum focuses.

$$\text{True positive rate} = (\text{True positives}) / (\text{false negative} + \text{true positive})$$

### F1 Score

It is used for having live a test accuracy of what number of examples it groups accurately, likewise as how powerful (it doesn't miss a major number of cases).

$$F1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

### Precision

Precision: it's the measure of right positive outcomes partitioned for the measurement of outcomes anticipated for classifying the dataset .

$$\text{Precision} = (\text{True positives}) / (\text{True positives} + \text{False positives})$$

### Recall

Recall is used measure of right normal outcomes separated by measuring every single significant example (all examples that should are distinguished).

$$\text{Recall} = (\text{True Positives}) / (\text{True Positives} + \text{False Negatives})$$

### Mean Absolute Error

It show the relation between the original and the predicted value and is used to furcating the different value on different time and is show mathematically in form of

$$\text{Mean absolute error} = \text{MSE}(\theta_{\text{obs}, i}) = E[(\theta_{\text{obs}, i} - \theta)]$$

### Mean Squared Error

Its is square mean of all the data point from 0 to n .It is represented as

$$\text{Mean squared error} = \text{MSE}(\theta_{\text{obs}, i}) = E[(\theta_{\text{obs}, i} - \theta)^2]$$

### **3.5 Motivation**

For the most part, it's evaluated that various pests (insects, weeds, nematodes, creatures, maladies) every year cause crop misfortunes around 40 % to 60%. All the more absolutely, there is a few information that keeps up that crop infections cause average yield misfortunes of 42% for the most significant food crops, now and again, crop diseases destroy the entire harvest creation.

Consequently, it's critical for ranchers to discover everything they can about the yield infections so they can oversee them appropriately.

Plant malady a typical piece of nature and normally it difficult for rancher to discover the kind of sickness. The primary inspiration of task identifies the illness in plant by utilizing semantic division and neural systems. While trying to arrive at significant returns and sound yields, ranchers all through the world battle to forestall and kill different infections from their harvests. Understanding the sickness and the development process is the initial move towards effective illness the board. There are a few special conditions that are helpful for disease improvement. First, each crop is defenseless to some sickness. Second, abiotic factors, to be specific climate, debilitate the plant altogether. Along these lines, the plant is vulnerable to the assault of the pathogen. When the entireties of the previously mentioned factors (vulnerable harvest, abiotic stress, pathogen assault) are available together and simultaneously, illness happens. That is called the Plant Disease Pyramid.

### **3.6 Objective**

Sparing the yields and protecting the crops from maladies is a burning issue for each rancher who wishes to see the best outcomes. From the earliest starting point, the first and urgent advance in crop security is knowledge. Each rancher ought to know of susceptibility of the yield to specific infections, just as the abiotic operators that are ideal for the event of the ailment.

Some extra thing that could make a critical difference in ailment ranch the executives. Ordinary observing of harvests and fields, just as on-time response, can be a genuine lifeline for each yield creation.

To identify Disease, the following steps are taken loading, Image preprocessing and segmentation extraction of various feature and at last using clustering to identify the disease

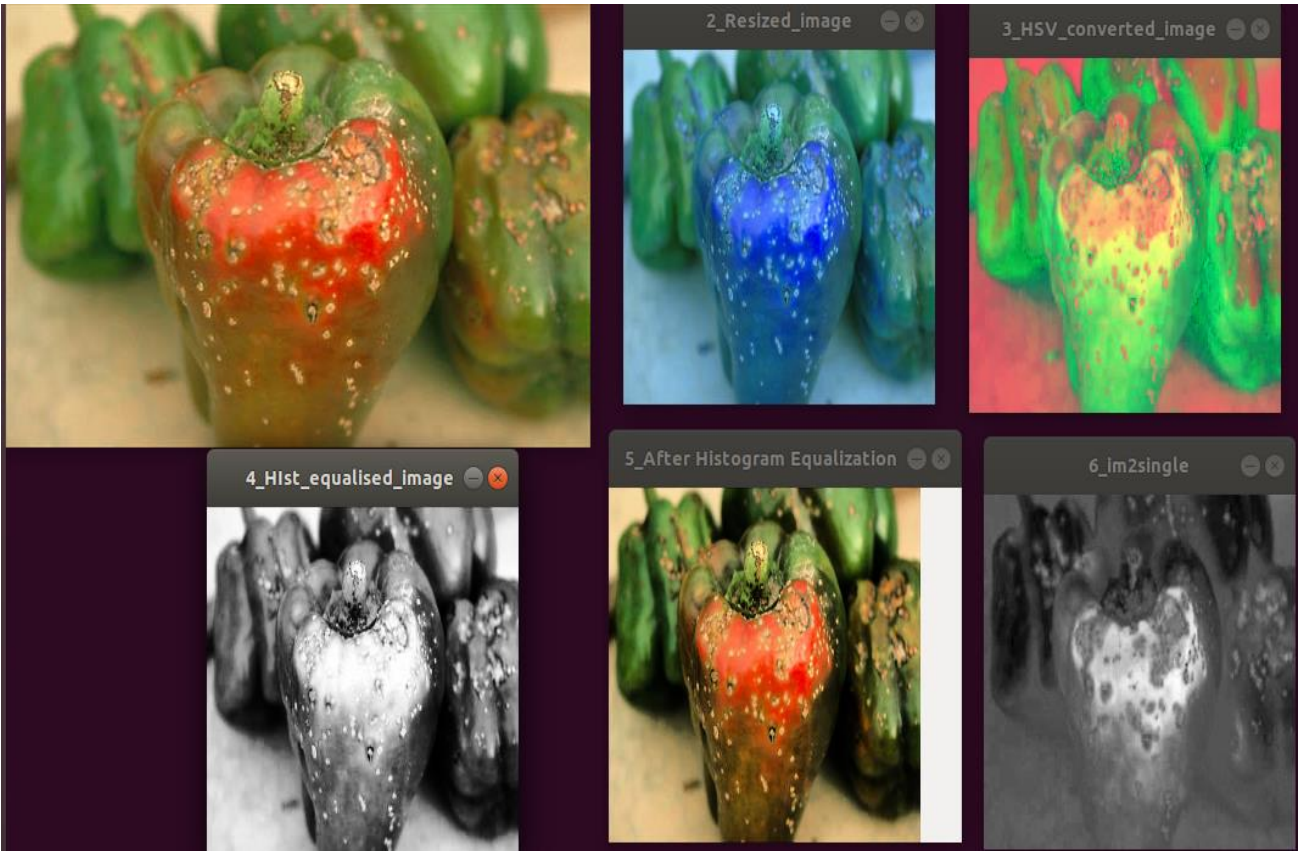
Image processing technique can be applied on various applications:

- To detect the plant leaf, steam and fruit disease.
- Finding the boundaries of effected area of plant.
- Determine the size and type of fruit or plant.

## CHAPTER 4

### PROGRESS AND RESULTS

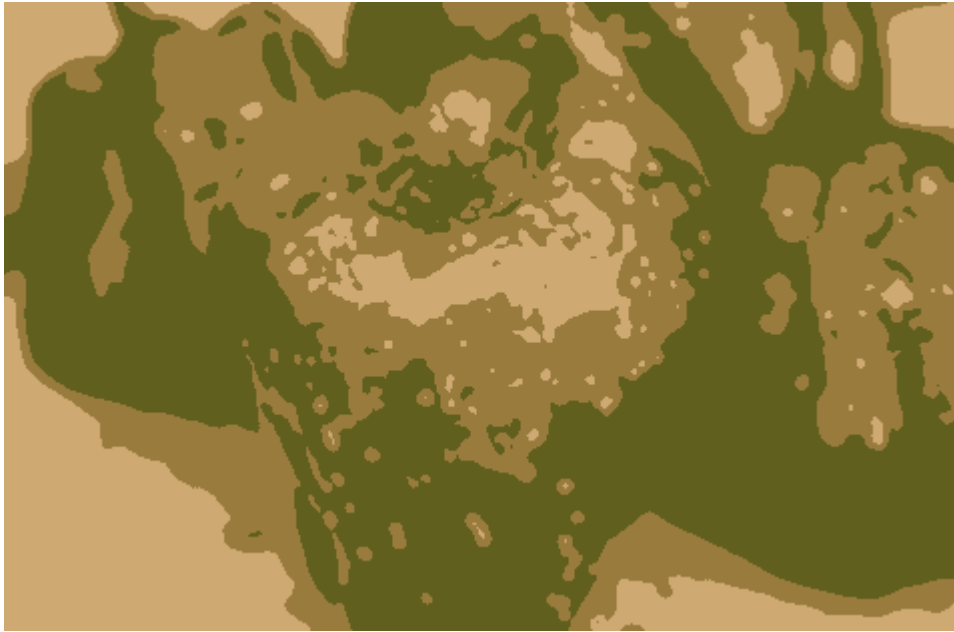
Application of various filters and changing into various color spaces on target image and getting it in proper place for performing feature extraction.



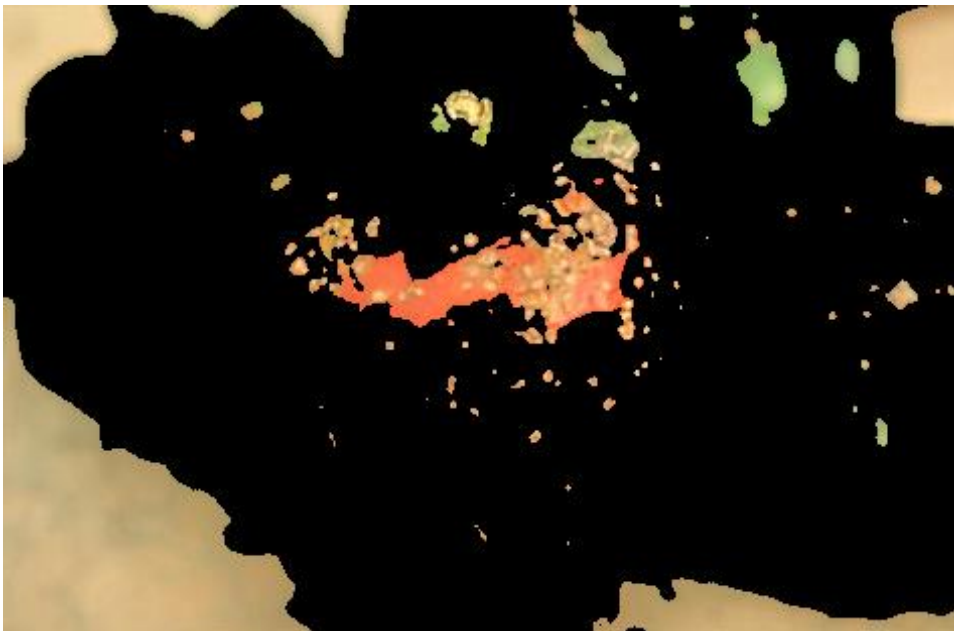
**Figure 4.1: Example**

Finding out various clusters in the target image using k-means clustering based segmentation. Here are the different clusters combined.

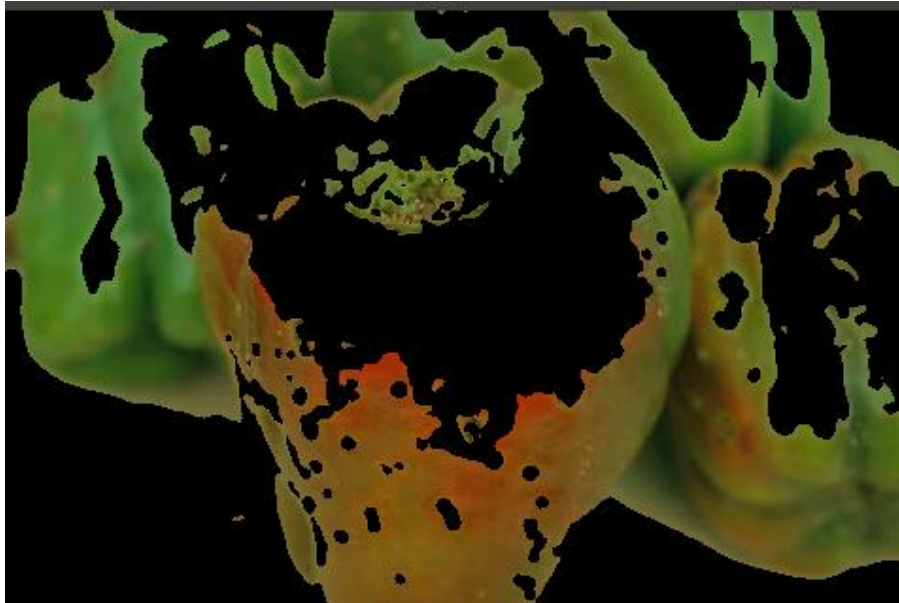




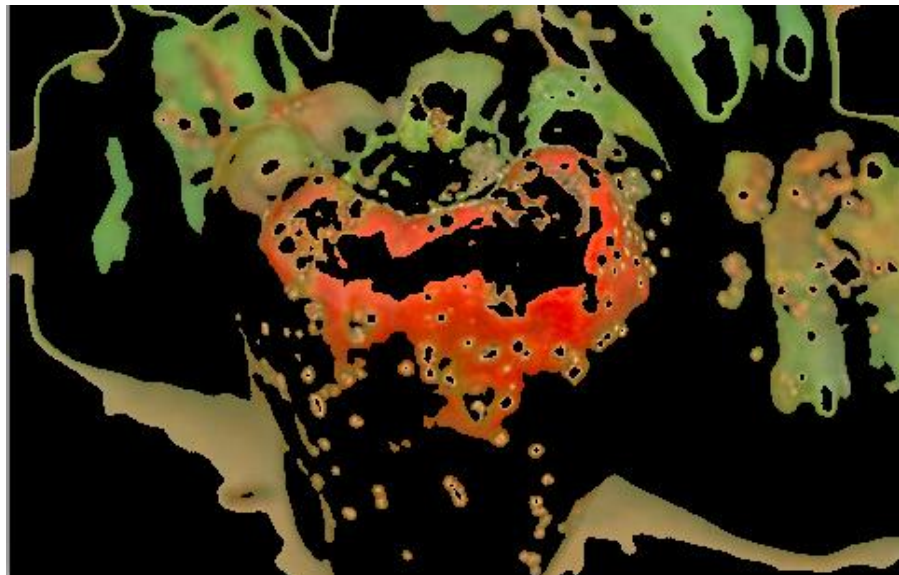
**Figure 4.2: Cluster Detected**



**Figure 4.3: Cluster 1**



**Figure 4.4: Cluster 2**



**Figure 4.5: Cluster 3**

- Selecting the cluster containing the diseased part and extracting features.

```

aaditya@Chaturvedi97:~/Downloads/final_year_project$ python3 final_program.py
Enter the cluster no. having diseased part- 1,2 or 3:2
Considered features are:
Homogeneity
0.8645
Energy
0.641
entropy:
21.7307725
correlation
0.9582
aaditya@Chaturvedi97:~/Downloads/final_year_project$

```

Figure 4.6: Cluster Selection

- Here is a screenshot of the dataset prepared.

1	Contrast	Correlation	Energy	Homogeneity	entropy	disease_name
2	0.888723748558223	0.9881	0.8719	0.9723	7411297.5	Alternaria_solani
3	0.82459027794914	0.8906	0.8566	0.9324	1617397.5	Alternaria_solani
4	0.547363503096818	0.9297	0.8942	0.9503	485678.88	Alternaria_solani
5	0.32017799256814	0.9433	0.6671	0.8862	2069492.9	Alternaria_solani
6	0.495179450833856	0.9445	0.9212	0.9659	488921.94	Alternaria_solani
7	0.372373818078736	0.9581	0.8327	0.9299	8276353	Alternaria_solani
8	0.63902629406853	0.9608	0.7375	0.8888	1624042.9	Alternaria_solani
9	0.645564596509017	0.981	0.7395	0.9112	4114036.2	Alternaria_solani
10	0.521477218106926	0.9308	0.7017	0.8855	154655.84	Alternaria_solani
11	0.861475011334904	0.9585	0.843	0.9341	8320437.5	Alternaria_solani
12	0.635640558595722	0.9451	0.8807	0.9527	319413.78	Alternaria_solani
13	0.186446159084269	0.9546	0.7131	0.8866	3832169.8	Alternaria_solani
14	0.786743536248218	0.9646	0.8169	0.9414	3680993.5	Alternaria_solani
15	0.890826037735352	0.9424	0.8249	0.9302	1805861.8	Alternaria_solani
16	0.227383120409514	0.9705	0.6868	0.918	6177666.5	Alternaria_solani
17	0.807323132165095	0.9519	0.8303	0.9185	397991.47	Tomato_anthracnose
18	0.808435043710728	0.9576	0.6682	0.8542	422950.88	Tomato_anthracnose
19	0.654648644224574	0.9617	0.8081	0.9573	1393754.1	Tomato_anthracnose
20	0.597578976474387	0.9868	0.8655	0.9495	17967824	Tomato_anthracnose
21	0.173200250100465	0.9864	0.7994	0.9528	8149867.5	Tomato_anthracnose
22	0.868051897703702	0.9606	0.7295	0.8948	1038096.1	Tomato_anthracnose
23	0.750822274360462	0.9936	0.8006	0.9513	8148805	Tomato_anthracnose
24	0.384413504265864	0.9351	0.9526	0.9788	4819898	Tomato_anthracnose
25	0.772824510879618	0.9733	0.8024	0.9339	686968	Tomato_anthracnose
26	0.819602252078201	0.9908	0.8834	0.9583	20699358	Tomato_anthracnose

Figure 4.7: Snapshot of Dataset prepared

- Specification and architecture of chosen Keras model.

Layer (type)	Output Shape	Param #
batch_normalization_1 (Batch Normalization)	(None, 5)	20
dense_1 (Dense)	(None, 32)	192
dropout_1 (Dropout)	(None, 32)	0
batch_normalization_2 (Batch Normalization)	(None, 32)	128
dense_2 (Dense)	(None, 32)	1056
dropout_2 (Dropout)	(None, 32)	0
batch_normalization_3 (Batch Normalization)	(None, 32)	128
dense_3 (Dense)	(None, 32)	1056
dropout_3 (Dropout)	(None, 32)	0
dense_4 (Dense)	(None, 4)	132
Total params: 2,712		
Trainable params: 2,574		
Non-trainable params: 138		

Figure 4.8: Architecture of Keras Model

- Training of Keras model on the final dataset. Accuracy using feed forwards perceptron coming out to be 80%.

```

44/44 [=====] - 0s 153us/step - loss: 0.3753 - accuracy: 0.8182 - val_loss: 0.4946 - val_accuracy: 0.7833
Epoch 184/200
44/44 [=====] - 0s 149us/step - loss: 0.3152 - accuracy: 0.8580 - val_loss: 0.4930 - val_accuracy: 0.7833
Epoch 185/200
44/44 [=====] - 0s 95us/step - loss: 0.3714 - accuracy: 0.8295 - val_loss: 0.4916 - val_accuracy: 0.7833
Epoch 186/200
44/44 [=====] - 0s 99us/step - loss: 0.3951 - accuracy: 0.8182 - val_loss: 0.4908 - val_accuracy: 0.7833
Epoch 187/200
44/44 [=====] - 0s 232us/step - loss: 0.4127 - accuracy: 0.7784 - val_loss: 0.4898 - val_accuracy: 0.7833
Epoch 188/200
44/44 [=====] - 0s 94us/step - loss: 0.3647 - accuracy: 0.8239 - val_loss: 0.4890 - val_accuracy: 0.7833
Epoch 189/200
44/44 [=====] - 0s 141us/step - loss: 0.4132 - accuracy: 0.7955 - val_loss: 0.4878 - val_accuracy: 0.7833
Epoch 190/200
44/44 [=====] - 0s 97us/step - loss: 0.3333 - accuracy: 0.8409 - val_loss: 0.4871 - val_accuracy: 0.7833
Epoch 191/200
44/44 [=====] - 0s 98us/step - loss: 0.3949 - accuracy: 0.8295 - val_loss: 0.4865 - val_accuracy: 0.7833
Epoch 192/200
44/44 [=====] - 0s 162us/step - loss: 0.3627 - accuracy: 0.8352 - val_loss: 0.4863 - val_accuracy: 0.7833
Epoch 193/200
44/44 [=====] - 0s 105us/step - loss: 0.3318 - accuracy: 0.8466 - val_loss: 0.4858 - val_accuracy: 0.7833
Epoch 194/200
44/44 [=====] - 0s 158us/step - loss: 0.3726 - accuracy: 0.8295 - val_loss: 0.4845 - val_accuracy: 0.7833
Epoch 195/200
44/44 [=====] - 0s 96us/step - loss: 0.3638 - accuracy: 0.8182 - val_loss: 0.4843 - val_accuracy: 0.7833
Epoch 196/200
44/44 [=====] - 0s 133us/step - loss: 0.3801 - accuracy: 0.8125 - val_loss: 0.4838 - val_accuracy: 0.7833
Epoch 197/200
44/44 [=====] - 0s 125us/step - loss: 0.3235 - accuracy: 0.8295 - val_loss: 0.4833 - val_accuracy: 0.7833
Epoch 198/200
44/44 [=====] - 0s 92us/step - loss: 0.4035 - accuracy: 0.7955 - val_loss: 0.4821 - val_accuracy: 0.7833
Epoch 199/200
44/44 [=====] - 0s 96us/step - loss: 0.4093 - accuracy: 0.7898 - val_loss: 0.4810 - val_accuracy: 0.7833
Epoch 200/200
44/44 [=====] - 0s 115us/step - loss: 0.3839 - accuracy: 0.8011 - val_loss: 0.4808 - val_accuracy: 0.7833
FOR PERCEPTRON, metrics f1_score:

```

Figure 4.9: Accuracy of Keras Model



- Evaluation and results of various algorithms apart from feed forward perceptrons.

```

metrics f1_score:
1.0
accuracy_score:
1.0
balanced_accuracy_score:
1.0

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4
1	1.00	1.00	1.00	2
2	1.00	1.00	1.00	7
3	1.00	1.00	1.00	2
accuracy			1.00	15
macro avg	1.00	1.00	1.00	15
weighted avg	1.00	1.00	1.00	15

Figure 4.10: Result of Decision Tree Classifier

```

precision recall f1-score support
0 0.00 0.00 0.00 4
1 0.13 1.00 0.24 2
2 0.00 0.00 0.00 7
3 0.00 0.00 0.00 2
accuracy 0.13 15
macro avg 0.03 0.25 0.06 15
weighted avg 0.02 0.13 0.03 15

```

```

[[0 4 0 0]
 [0 2 0 0]
 [0 7 0 0]
 [0 2 0 0]]

```

Figure 4.11: Result of Bernoulli Naïve Bayes Classifier

- Confusion matrix from various algorithms as mentioned shown.

### 1. AdaBoost classifier:

It helps in adding extra copies of the data to imply the perfect data set for the graphical interface to use the projected dataset for implying the various role in adaboost to classification

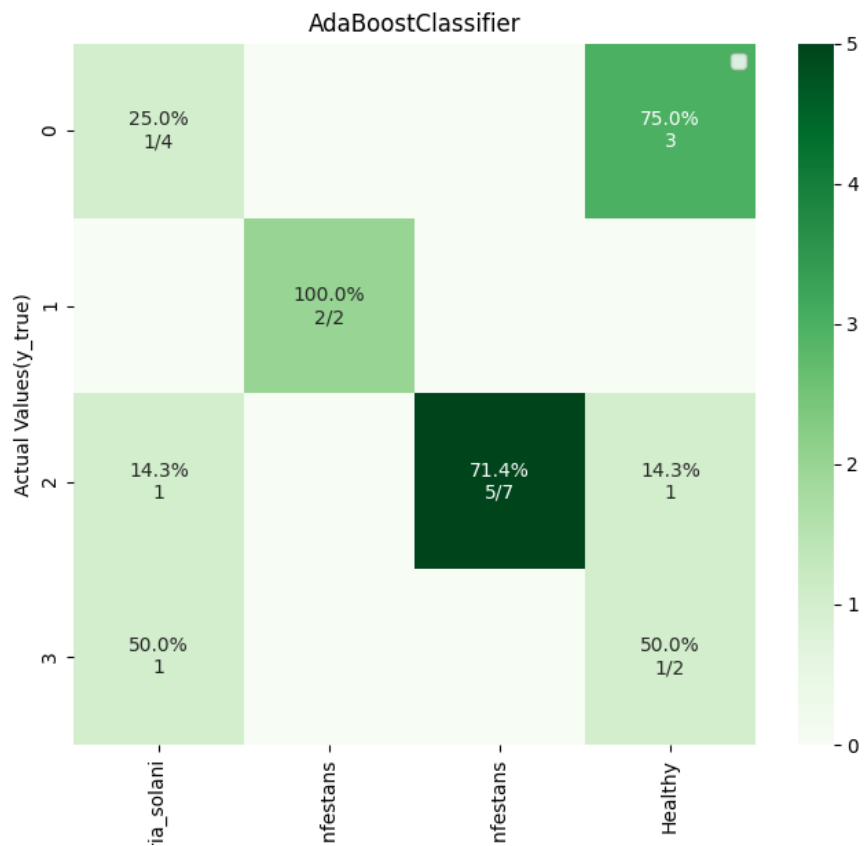


Figure 4.12: Heat Map of AdaBoost Classifier

## 2. Bernoulli Naïve Bayes' :

It is type of classifier that work on probability taking into account applying BNM is blameless and self-sufficient. They are among the least troublesome Bayesian framework models. In any case, they could be joined with Kernel thickness estimation and achieve higher precision levels.

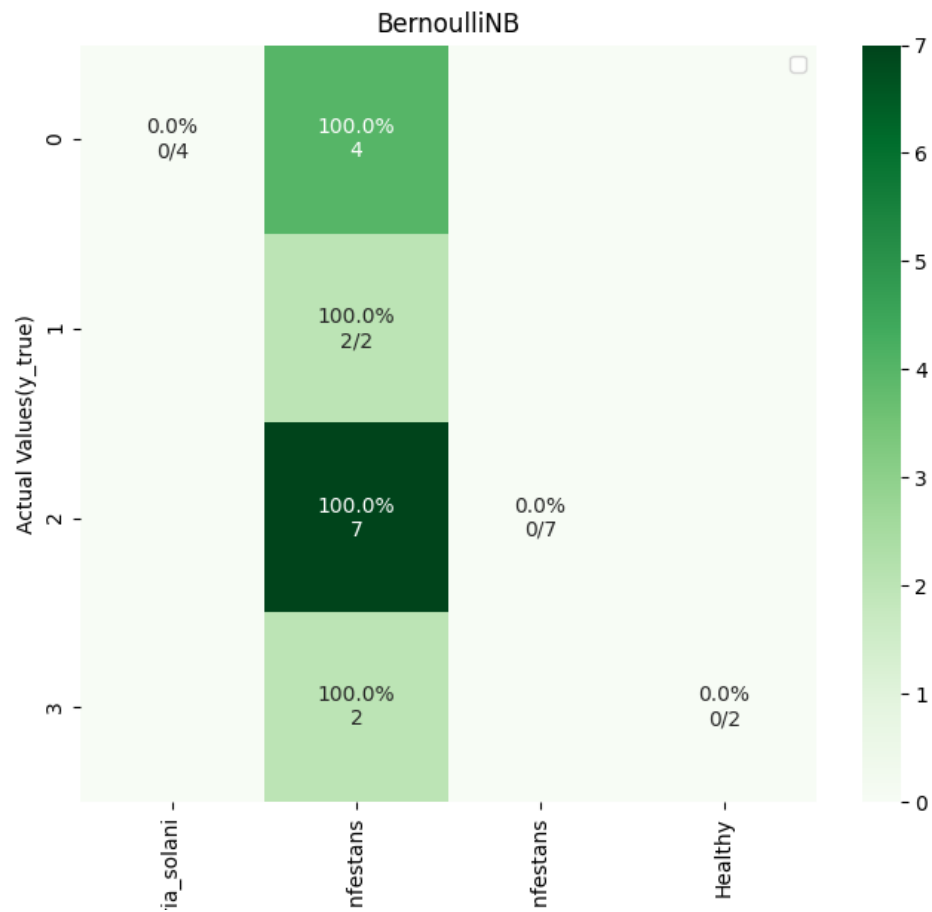


Figure 4.13: Heat Map of Bernoulli Naïve Bayes Classifier

### 3. Decision Tree Classifier :

The classifier arranges an article  $x_i$  by recursively testing for the loads that the highlights naming the inward hubs have in vector  $x_i$ , until a leaf hub is reached. The name of this hub is then allotted to  $x_i$ .

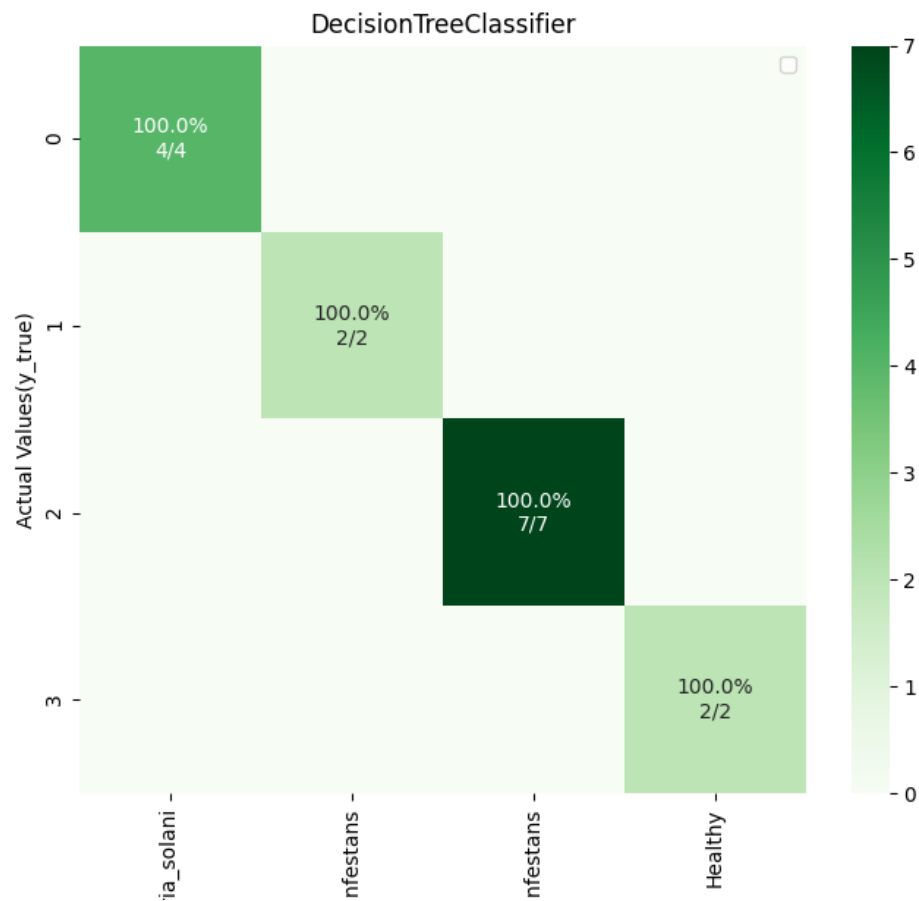


Figure 4.14: Heat Map of Decision Tree Classifier



#### 4. Gaussian Naïve Bayes:

The Gaussian Naive Bayes is one classifier model. It is most commonly used since it is the least difficult and the most well known one.

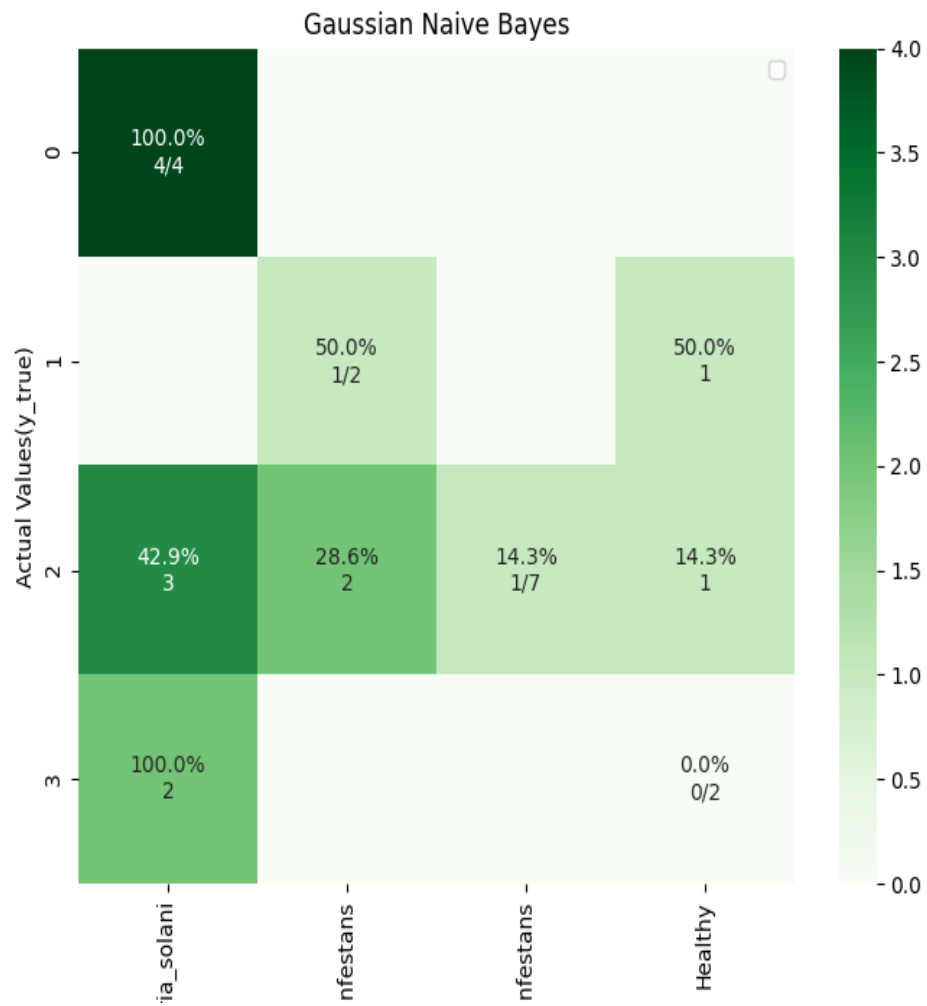


Figure 4.15: Heat Map of Gaussian Naïve Bayes Classifier

## 5. K- Nearest Neighbor Classifier:

k-Nearest neighbors calculation (k-NN) is a non-parametric system used for gathering and backslide. In the two cases, the information contains the k closest getting ready models in the component space. The yield depends upon whether k-NN is used for course of action or backslide.

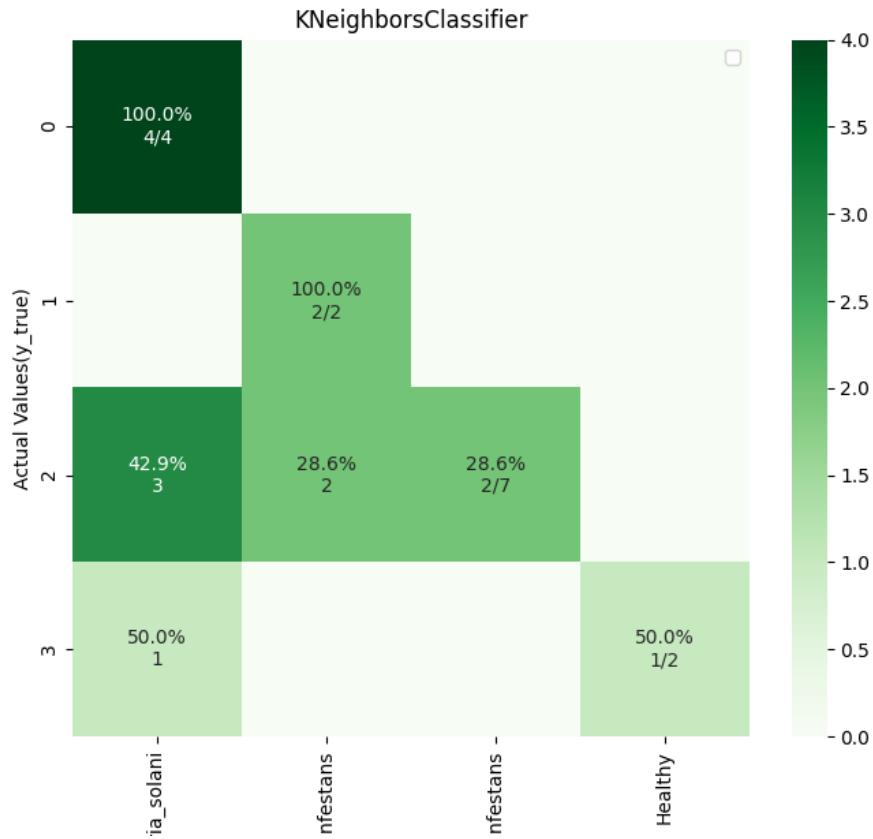


Figure 4.16: Heat Map of K-Nearest Neighbor

## 6. Multinomial Bayes Classifier:

It is classifier that is used to which implies prohibitive opportunity of MNB classifier is used to specific event of a Naive Bayes classifier some features are.

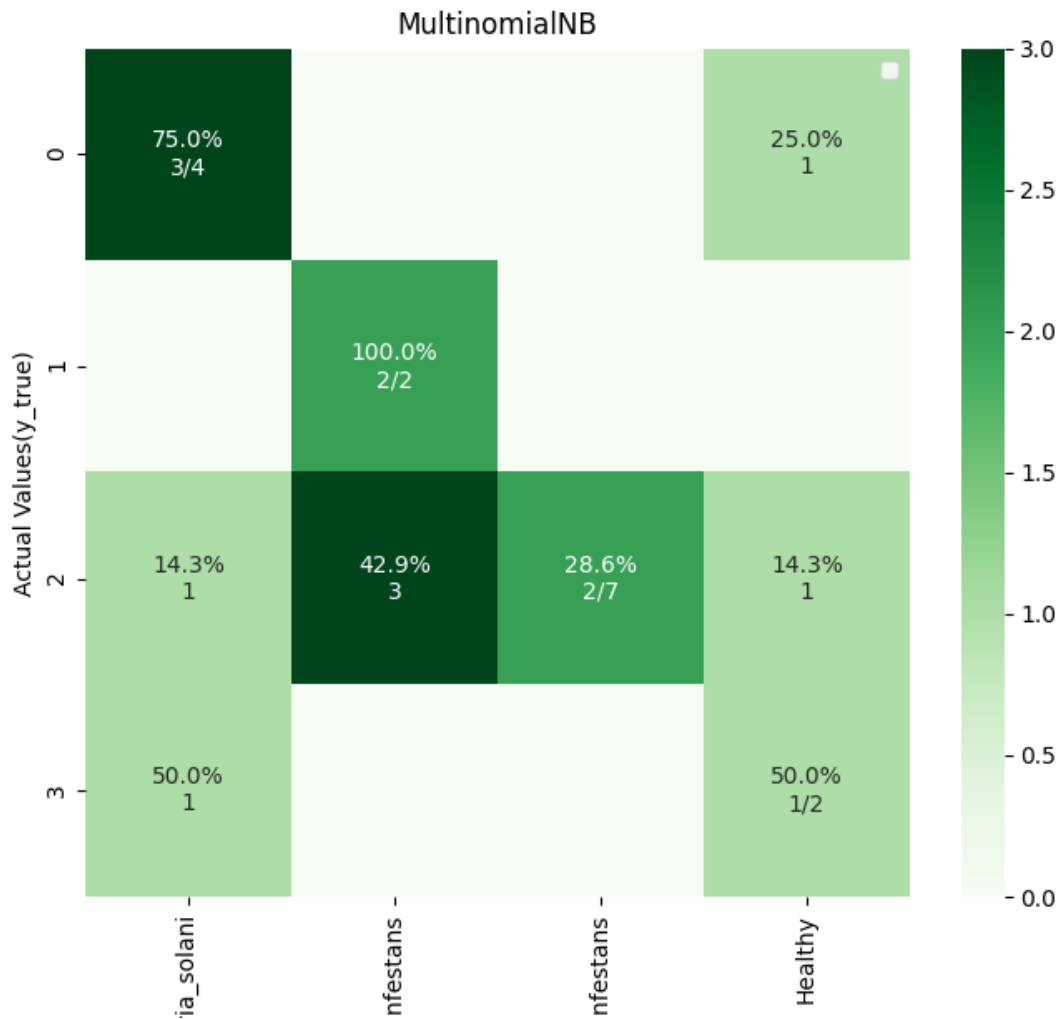


Figure 4.17: Heat Map of Multinomial Naïve Bayes Classifier

## REFERENCES

- [1] Mohanty SP, Hughes DP and Salathé M (2016) Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls.2016.01419
- [2] Sarah J. Pethybridge and Scot C. Nelson(2015), Leaf Doctor: A New Portable Application for Quantifying Plant Disease Severity | *Plant Disease*,ISSN: 0191-2917,e-ISSN: 1943-7692
- [3] Bhimte, Namrata R., and V. R. Thool. "Diseases Detection of Cotton Leaf Spot Using Image Processing and SVM Classifier." 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2018.
- [4] Alexander Johannes, ArtzaiPiconb, Aitor Alvarez Gilab, JoneEchazarrab, Sergio Rodriguez Vaamondeb, Ana Díez, NavajosAmaiaOrtizBarredoc(2017), Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Computers and Electronics in Agriculture* Volume 138, 1 June 2017, Pages 200-209
- [5] Islam, Monzurul, et al. "Detection of potato diseases using image segmentation and multiclass support vector machine." 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE). IEEE, 2017.
- [6] Workflow of Classification, <https://ars.els-cdn.com/content/image/1-s2.0-S1360138518301572-gr2.jpg>
- [7] Plant disease identification and display,[https://www.frontiersin.org/files/Articles/215232/fpls-07-01419-HTML-r2/image\\_m/fpls-07-01419-g001.jpg](https://www.frontiersin.org/files/Articles/215232/fpls-07-01419-HTML-r2/image_m/fpls-07-01419-g001.jpg)
- [8] image segmentation and display, [https://www.frontiersin.org/files/Articles/215232/fpls-07-01419-HTML-r2/image\\_m/fpls-07-01419-g002.jpg](https://www.frontiersin.org/files/Articles/215232/fpls-07-01419-HTML-r2/image_m/fpls-07-01419-g002.jpg)
- [9] Results, *Front plant Sci.*, 22<sup>nd</sup> September 2016, <https://doi.org/10.3389/fpls.2016.01419>
- [10] Visualization of activations in the initial layers of AlexNet architecture, [https://www.frontiersin.org/files/Articles/215232/fpls-07-01419-HTML-r2/image\\_m/fpls-07-01419-g004.jpg](https://www.frontiersin.org/files/Articles/215232/fpls-07-01419-HTML-r2/image_m/fpls-07-01419-g004.jpg)
- [11] Linear Severity estimates from software assessment and actual values., [https://apsjournals.apsnet.org/na101/home/literatum/publisher/aps/journals/content/pdis/2015/pdis.2015.99.issue-10/pdis-03-15-0319-re/20150918/images/medium/pdis-03-15-0319-re\\_f3.gif](https://apsjournals.apsnet.org/na101/home/literatum/publisher/aps/journals/content/pdis/2015/pdis.2015.99.issue-10/pdis-03-15-0319-re/20150918/images/medium/pdis-03-15-0319-re_f3.gif)

# PLAGIARISM REPORT

final Pdd

## ORIGINALITY REPORT

**16%**

SIMILARITY INDEX

**12%**

INTERNET SOURCES

**5%**

PUBLICATIONS

**13%**

STUDENT PAPERS

## PRIMARY SOURCES

<b>1</b>	<b>www.tutorialspoint.com</b> Internet Source	<b>2%</b>
<b>2</b>	<b>Submitted to Jaypee University of Information Technology</b> Student Paper	<b>1%</b>
<b>3</b>	<b>towardsdatascience.com</b> Internet Source	<b>1%</b>
<b>4</b>	<b>www.analyticsvidhya.com</b> Internet Source	<b>1%</b>
<b>5</b>	<b>Submitted to University of Lancaster</b> Student Paper	<b>1%</b>
<b>6</b>	<b>Submitted to Polytechnic of Turin</b> Student Paper	<b>1%</b>

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT**  
**PLAGIARISM VERIFICATION REPORT**

Date: .....18/07/2020.....

Type of Document (Tick):  PhD Thesis  M.Tech Dissertation/  Report  B.Tech Project  Paper

Name: ADITYA MOHAN Department: ECE Enrolment No 161035

Contact No. 9805103197 E-mail. adityamohanjuit@gmail.com

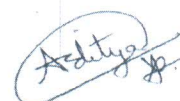
Name of the Supervisor: Dr. NafisUddin Khan

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): PLANT DISEASE DETECTION USING CLUSTERING –BASED SEGMENTATION AND NEURAL NETWORKS

**UNDERTAKING**

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

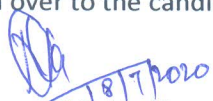
- Total No. of Pages = 60
- Total No. of Preliminary pages = 10
- Total No. of pages accommodate bibliography/references = 1



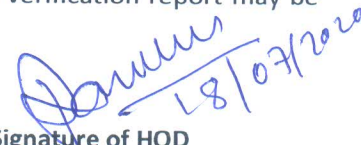
(Signature of Student)

**FOR DEPARTMENT USE**

We have checked the thesis/report as per norms and found **Similarity Index** at .....16.....(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.



(Signature of Guide/Supervisor)



Signature of HOD

**FOR LRC USE**

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Abstract & Chapters Details		
			Word Counts		
Report Generated on	<ul style="list-style-type: none"> <li>• All Preliminary Pages</li> <li>• Bibliography/ Images/Quotes</li> <li>• 14 Words String</li> </ul>		Character Counts		
			Submission ID	Page counts	
				File Size	

Checked by  
Name & Signature

Librarian

**Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at [plagcheck.juit@gmail.com](mailto:plagcheck.juit@gmail.com)**



**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT  
PLAGIARISM VERIFICATION REPORT**

Date: ...18/7/2020.....

Type of Document (Tick):  PhD Thesis  M.Tech Dissertation/ Report  B.Tech Project Report  Paper

Name: Arjun sharma Department: ECE Enrolment No 161040

Contact No. 9805112710 E-mail - sharma.arjun0186@gmail.com

Name of the Supervisor: Dr. Nafis Uddin Khan

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): Plant Disease Detection Using Clustering Based Segmentation And Neural Networks

**UNDERTAKING**

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/ revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages= 61
- Total No. of Preliminary pages =3
- Total No. of pages accommodate bibliography/ references = 3



(Signature of Student)

**FOR DEPARTMENT USE**

We have checked the thesis/ report as per norms and found **Similarity Index** at **16 (%)**. Therefore, we are forwarding the complete thesis/ report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

*(Handwritten signature and date 18/7/2020)*

(Signature of Guide/Supervisor)

*(Handwritten signature and date 18/07/2020)*

(Signature of HOD)

**FOR LRC USE**

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Abstract & Chapters Details	
Report Generated on	<ul style="list-style-type: none"> <li>• All Preliminary Pages</li> <li>• Bibliography/ Images/ Quotes</li> <li>• 14 WordsString</li> </ul>	Submission ID	Word Counts	
			Character Counts	
			Page counts	
			File Size	

Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at [plagcheck.juit@gmail.com](mailto:plagcheck.juit@gmail.com)

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT**  
**PLAGIARISM VERIFICATION REPORT**

Date: .....

Type of Document (Tick):  PhD Thesis  M.Tech Dissertation/ Report  B.Tech Project Report  Paper

Name: Utkarsh Sharma Department: ECE Enrolment No 161097

Contact No. 9459107727 E-mail. sharmautkarsh0902@gmail.com

Name of the Supervisor: Dr. Nafis Uddin Khan

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): PLANT DISEASE DETECTION USING CLUSTERING- BASED SEGMENTATION AND NEURAL NETWORKS

**UNDERTAKING**

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/ revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages = 61



- Total No. of Preliminary pages = 3

- Total No. of pages accommodate bibliography/references = 3

(Signature of Student)

**FOR DEPARTMENT USE**

We have checked the thesis/report as per norms and found **Similarity Index** at .....16.....(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

*(Handwritten signature of Guide/Supervisor)*  
18/11/2020

(Signature of Guide/Supervisor)

*(Handwritten signature of HOD)*  
18/10/2020

Signature of HOD

**FOR LRC USE**

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Abstract & Chapters Details	
	<ul style="list-style-type: none"> <li>• All Preliminary Pages</li> <li>• Bibliography/ Images/Quotes</li> <li>• 14 Words String</li> </ul>		Word Counts	
<b>Report Generated on</b>			Character Counts	
		<b>Submission ID</b>	Page counts	
			File Size	

**Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at [plagcheck.juit@gmail.com](mailto:plagcheck.juit@gmail.com)**