

# Plant Health Monitoring System using Image Processing: A Review

<sup>1</sup>Annie Christi, <sup>2</sup>Brijesh Vala

<sup>1</sup>M.Tech Student, <sup>2</sup>Prof.

<sup>1-2</sup>Dept. of Computer Science Engineering

<sup>1-2</sup>Parul Institute of Engineering & Technology  
Vadodara, India

**Abstract**— Now a day's in India, about 70% of the people are in the agriculture system. Agricultural includes harvesting of the plants, fruits, crops, etc on a large scale. Now, this harvested product should be disease free for healthy production further. Now, hereby diseases in plants cause major production and economic losses. The most standout cause of harm is attack of disease. Rural researchers and scientists assume an imperative job in identifying and discovering solution for plant sicknesses. Farmers experience great difficulties in detection and control on disease with naked eyes. In this paper we survey the need of basic plant disease detection that would facilitate in agriculture. This paper essentially incorporates strategies that utilization picture handling systems to distinguish the plant infections. It likewise incorporates the side effects of the illness in any piece of the plant. This paper displays the different computerized picture handling procedures and its correlation for the identification of illness.

**Keywords**— Image Processing, Plant disease detection, pattern recognition, image processing techniques, Kmean clustering

## I. INTRODUCTION

A plant disease is any abnormal condition that alters the appearance or function of a plant. A plant disease takes place when an organism infects a plant and disrupts its normal growth habits. Symptoms can range from slight discoloration to death. Diseases have many causes including fungi, bacteria, viruses, and nematodes.

### Theoretical Background for Leaf Diseases

#### • Pre-Processing

**1. Histogram [3, 8, and 15]:-** It is an accurate representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable (quantitative variable) and was first introduced by Karl Pearson. It differs from a bar graph, in the sense that a bar graph relates two variables, but a histogram relates only one. To construct a histogram, the first step is to "bin" (or "bucket") the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often (but are not required to be) of equal size.

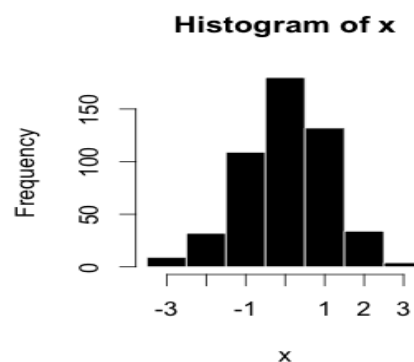


Figure 1: Histogram

In a more general mathematical sense, a histogram is a function  $m_i$  that counts the number of observations that fall into each of the disjoint categories (known as bins), whereas the graph of a histogram is merely one way to represent a histogram. Thus, if we let  $n$  be the total number of observations and  $k$  be the total number of bins, the histogram  $m_i$  the following conditions:

$$n = \sum_{i=1}^k m_i \quad (1.1)$$

**2. Color Model [8]:-** A color model is an abstract mathematical model describing the way colors can be represented as tuples of numbers, typically as three or four values or color components. When this model is associated with a precise description of how the components are to be interpreted (viewing conditions, etc.), the resulting set of colors is called color space. This section describes ways in which human color vision can be modeled.

### • Segmentation

**1. K-means[1,3,14,15]:-** k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

Given a set of observations  $(x_1, x_2, \dots, x_n)$ , where each observation is a  $d$ -dimensional real vector, k-means clustering aims to partition the  $n$  observations into  $k$  ( $\leq n$ ) sets  $S = \{S_1, S_2, \dots, S_k\}$  so as to minimize the within-cluster sum of squares (WCSS) (i.e. variance). Formally, the objective is to find:

$$\arg \min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min \sum_{i=1}^k |S_i| \text{Var } S_i \quad (1.2)$$

Where  $\mu_i$  is the mean of points in  $S_i$ . This is equivalent to minimizing the pairwise squared deviations of points in the same cluster:

$$\arg \min \sum_{i=1}^k \frac{1}{2|S_i|} \sum_{x, y \in S_i} \|x - y\|^2 \quad (1.3)$$

The equivalence can be deduced from identity

$$\sum_{x \in S_i} \|x - \mu_i\|^2 = \sum_{x \neq y \in S_i} (x - \mu_i)(\mu_i - y) \quad (1.4)$$

Because the total variance is constant, this is also equivalent to maximizing the sum of squared deviations between points in different clusters (between-cluster sum of squares, BCSS), which follows easily from the law of total variance.

**2. Otsu's method:-** In computer vision and image processing, Otsu's method, named after Nobuyuki Otsu, is used to automatically perform clustering-based image thresholding, or, the reduction of a graylevel image to a binary image. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal, or equivalently (because the sum of pairwise squared distances is constant), so that their inter-class variance is maximal. Consequently, Otsu's method is roughly a one-dimensional, discrete analog of Fisher's Discriminant Analysis. Otsu's method is also directly related to the Jenks optimization method.

**3. Morphological image processing:-** Binary images may contain numerous imperfections. In particular, the binary regions produced by simple thresholding are distorted by noise and texture. Morphological image processing pursues the goals of removing these imperfections by accounting for the form and structure of the image. These techniques can be extended to greyscale images. Morphological image processing is a collection of non-linear operations related to the shape or morphology of features in an image.

Morphological operations rely only on the relative ordering of pixel values, not on their numerical values, and therefore are especially suited to the processing of binary images. Morphological operations can also be applied to greyscale images such that their light transfer functions are unknown and therefore their absolute pixel values are of no or minor interest. Morphological techniques probe an image with a small shape or template called a structuring element.

When a structuring element is placed in a binary image, each of its pixels is associated with the corresponding pixel of the neighbourhood under the structuring element. The structuring element is said to fit the image if, for each of its pixels set to 1, the corresponding image pixel is also 1. Similarly, a structuring element is said to hit, or intersect, an image if, at least for one of its

pixels set to 1 the corresponding image pixel is also 1. Zero-valued pixels of the structuring element are ignored, i.e. indicate points where the corresponding image value is irrelevant.

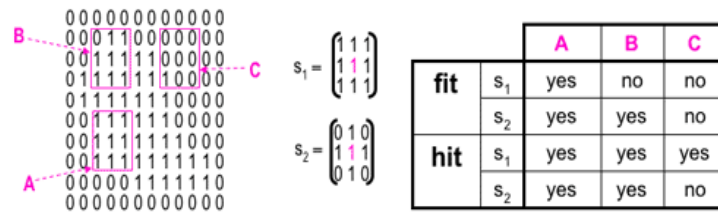


Figure 2: Fitting and Hitting of a Binary Image

## • Feature Extraction

### 1. Shape:-

#### • Area:-

The area was determined by counting the total number of non zero pixels within the image region.

#### • Parametric:-

It was measured by calculating distance between successive Boundary pixels.

#### • Major Access

Returns a scalar that specifies the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region

#### • Minor Access:

Returns a scalar that specifies the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

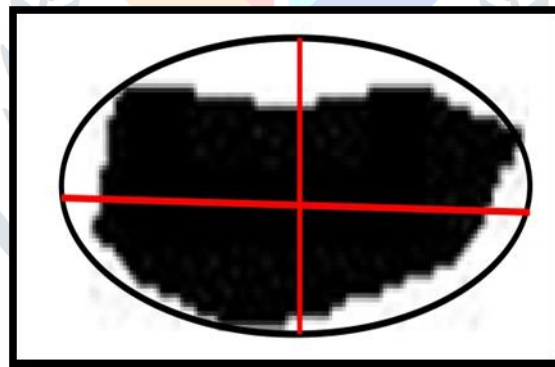


Figure 3: Shapes Features

### 2. Texture:-

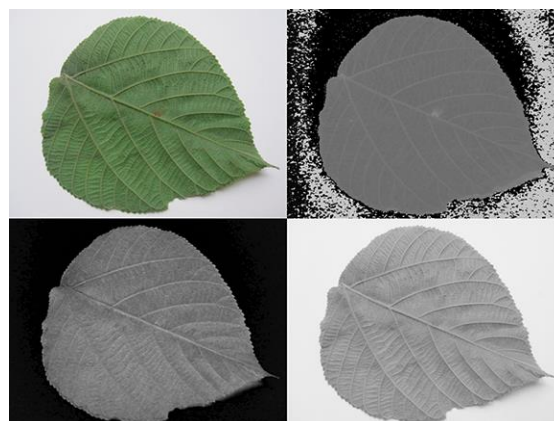


Figure 4: Texture

- **Gabor[1]:-** In image processing, a Gabor filter, named after Dennis Gabor, is a linear filter used for texture analysis, which means that it basically analyzes whether there are any specific frequency content in the image in specific directions in a localized region around the point or region of analysis. Frequency and orientation representations of Gabor filters are claimed by many contemporary vision scientists to be similar to those of the human visual system, though there is no empirical evidence and no functional rationale to support the idea. They have been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.
- **GLCM[3,4,8,12]:-** A co-occurrence matrix or co-occurrence distribution is a matrix that is defined over an image to be the distribution of co-occurring pixel values (grayscale values, or colors) at a given offset.
- **Haralick[1]:-** Haralick is one of the leading figures in computer vision, pattern recognition, and image analysis. Haralick's texture features were calculated using the kharalick() function of the cytometry tool box for Khoros. The basis for these features is the gray-level co-occurrence matrix ( **G** in Equation 1.5). This matrix is square with dimension  $N_g$ , where  $N_g$  is the number of gray levels in the image. Element  $[i,j]$  of the matrix is generated by counting the number of times a pixel with value  $i$  is adjacent to a pixel with value  $j$  and then dividing the entire matrix by the total number of such comparisons made. Each entry is therefore considered to be the probability that a pixel with value  $i$  will be found adjacent to a pixel of value  $j$ .

$$\mathbf{G} = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1, N_g) \\ p(2,1) & p(2,2) & \cdots & p(2, N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g, 1) & p(N_g, 2) & \cdots & p(N_g, N_g) \end{bmatrix}$$

Since adjacency can be defined to occur in each of four directions in a 2D, square pixel image (horizontal, vertical, left and right diagonals - see Figure 1.16), four such matrices can be calculated.

### 3. Color:-

- **Color Moment [1]:-** Color moments are measures that characterize color distribution in an image in the same way that central moments uniquely describe a probability distribution. Color moments are mainly used for color indexing purposes as features in image retrieval applications in order to compare how similar two images are based on color. Usually one image is compared to a database of digital images with pre-computed features in order to find and retrieve a similar Image. Each comparison between images results in a similarity score, and the lower this score is the more identical the two images are supposed to be.

### 4. Vein:-

A vein is a vascular structure (xylem and phloem cells surrounded by the bundle sheath) in a leaf that provides supports for the leaf and transports both water and food. The veins on monocots are almost parallel to the margins of the leaf. The veins of dicots radiate from a central midrib. The veins are the vascular tissue of the leaf and are located in the spongy layer of the mesophyll. The pattern of the veins is called venation. In angiosperms the venation is typically parallel in monocotyledons and forms an interconnecting network in broad-leaved plants. ... A vein is made up of a vascular bundle.



Figure 5: Vein

- **Classification**

**SVM [1, 4, 9, 10, and 14]:-** Support Vector Machines are supervised learning models that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories; an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVMs can efficiently perform a non-linear classification. For this type of SVM, training involves the minimizing of the error function:

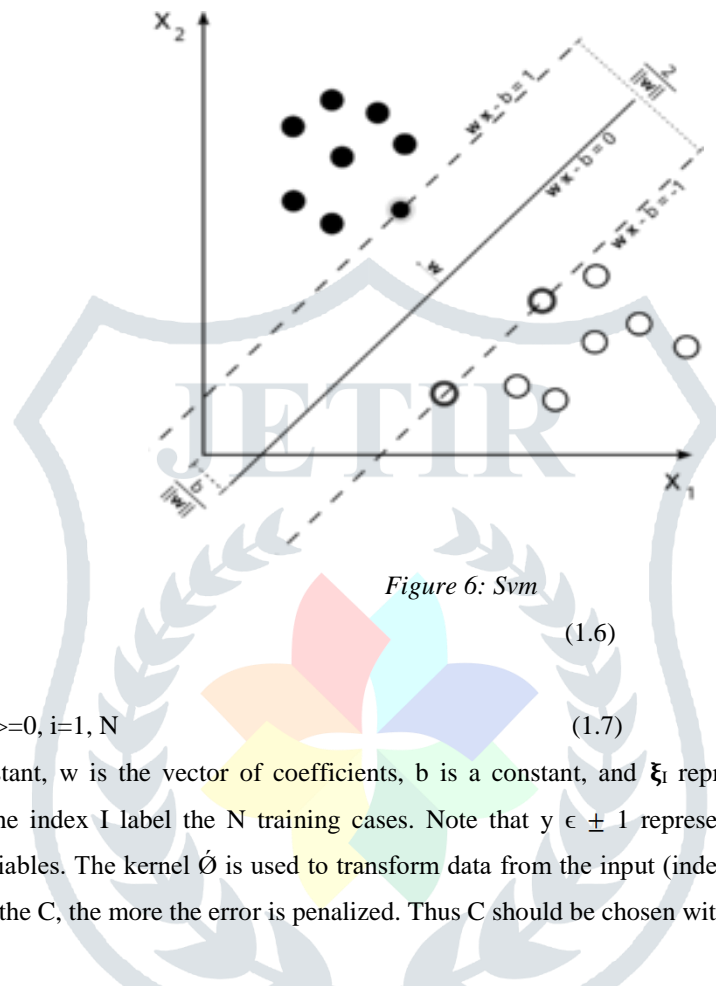


Figure 6: Svm

$$\frac{1}{2}w^T w + C \sum_{i=1}^N \xi_i \quad (1.6)$$

Subject to the constraints:

$$Y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i=1, N \quad (1.7)$$

Where  $C$  is the capacity constant,  $w$  is the vector of coefficients,  $b$  is a constant, and  $\xi_i$  represents parameters for handling nonseparable data (inputs). The index  $i$  label the  $N$  training cases. Note that  $y \in \pm 1$  represents parameters for label and  $x_i$  represents the independent variables. The kernel  $\phi$  is used to transform data from the input (independent) to the feature space. It should be noted that the larger the  $C$ , the more the error is penalized. Thus  $C$  should be chosen with care to avoid over fitting.

- **ANN [3, 4, 10, and 14]:-** Artificial neural networks (ANN) or connectionist systems are computing systems vaguely inspired by the biological neural networks that constitute animal brains. The neural network itself isn't an algorithm, but rather a framework for many different machine learning algorithms to work together and process complex data inputs. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules. For example, in image recognition, they might learn to identify images that contain cats by analyzing example images that have been manually labeled as "cat" or "no cat" and using the results to identify cats in other images. They do this without any prior knowledge about cats, e.g., that they have fur, tails, whiskers and cat-like faces. Instead, they automatically generate identifying characteristics from the learning material that they process.

An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it.

- **RF:-** Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set.



**AIM AND OBJECTIVE OF STUDY****Aim**

- The main aim of Online Leaf Disease Detection is to provide easy and fast way to detect leaf diseases.
- It will capture plant leaf which is having diseases such as virus, bacteria, fungi, nematodes and also segmentation of leaf will be done to detect diseases part.
- The combination of colour, shape and texture feature will be give rise to accuracy and also using machine learning approach classifies them among four types.
- Plant diseases cause a major production and economic losses in the agricultural industry. The disease management is a challenging task. Usually the diseases or its symptoms such as coloured spots or streaks are seen on the leaves of a plant. In plants most of the leaf diseases are caused by fungi, bacteria, and viruses. The diseases caused due to these organisms are characterized by different visual symptoms that could be observed in the leaves or stem of a plant. Usually, these symptoms are detected manually. With the help of image processing, Automatic detection of various diseases can be detected with the help of image processing. Image processing plays a crucial role in the detection of plant diseases since it provides best results and reduces the human efforts.

**Objectives**

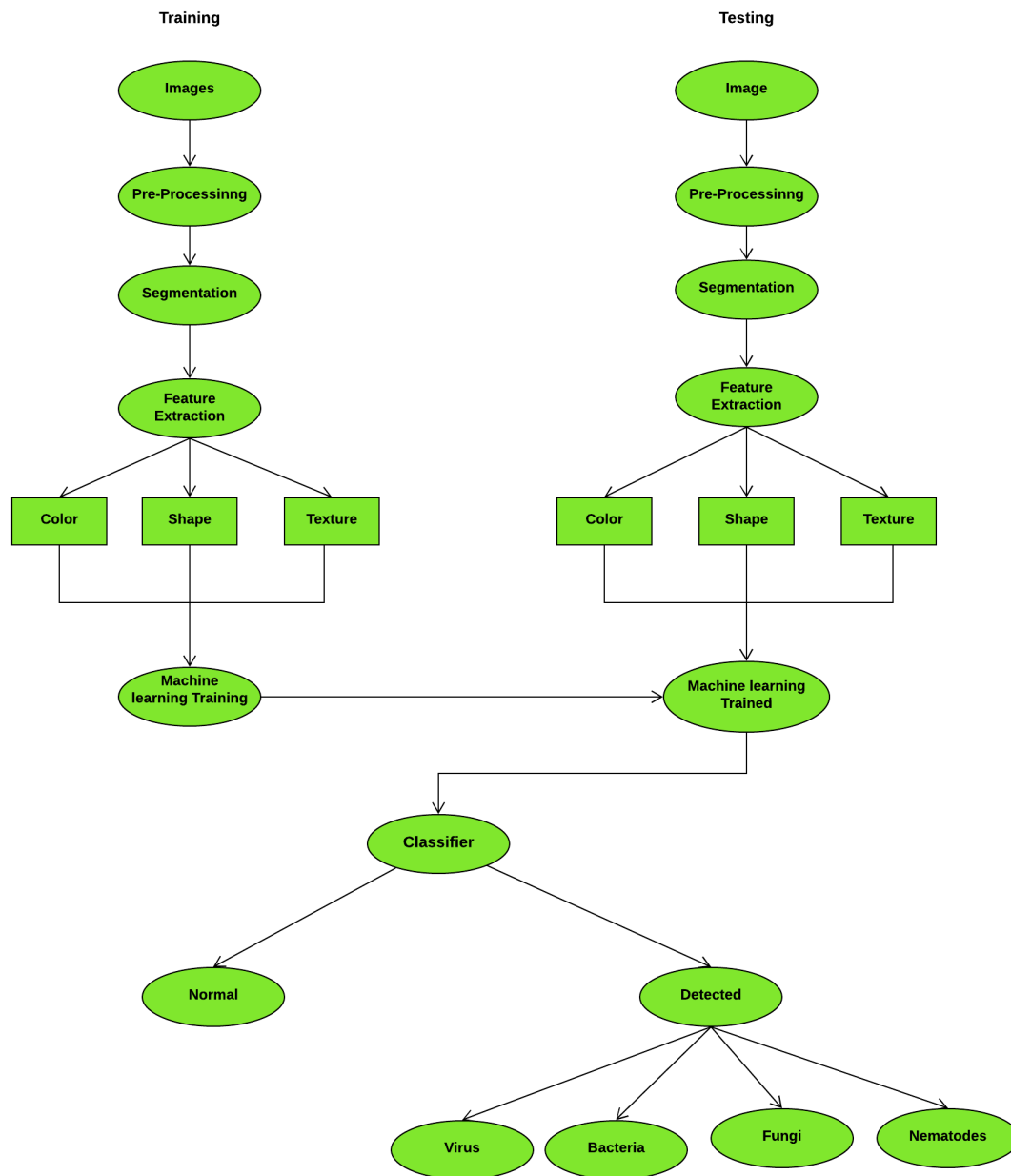
- To define different Common types of disease related to plant.
- To describe types and causes of Plant Disease.
- To improve segmentation using colour and cluster based combine approach.

**Problem Definition**

Identification of leaf diseases it is the important and one of the major problem in early stages. Disease is caused by pathogen which is any agent causing disease. In most of the cases pests or diseases are seen on the leaves or stems of the plant. Therefore identification of plants, leaves, symptoms and finding out the pest or diseases, percentage of the pest or disease incidence, symptoms of the pest or disease attack, plays a key role in successful cultivation of crops. What is wrong with my plant; followed by, what can I do to get rid of the problem? It may be too late to help the specific plant when the question is asked, but proper diagnosis may be extremely important in preventing the problem on other plants or in preventing the problem in the future. Without proper identification of the disease and the disease-causing agent, disease control measures can be a waste of time and money and can lead to further plant losses. Consider various environmental and cultural factors.

## II. METHODOLOGY

## Block Diagram of Proposed System



## Algorithm of Proposed System

## Training

Step 1: Select or upload images and its Label.

Step 2: Apply Pre-Processing using Histogram Equalization and Denoising on whole image datasets.

Step 3: Apply Colour and Cluster Based Combine Segmentation approach.

Step 4: Extract Shape, Colour and Texture Features for all images.

Step 5: Apply machine Learning Approach SVM, RF and make database.

## Testing

Step 1: Select or upload image.

Step2: Apply Pre-Processing using Histogram Equalization and Denoising.

Step 3: Apply Colour and Cluster Based Combine Segmentation approach.

Step 4: Extract Shape, Colour and Texture Features.

Step 5: Apply machine Learning classification Approach SVM, RF using database.

Step 6: Classify Disease type.

### Existing System Vs Proposed System

Parameters	Existing[1]	Proposed
Pre-Processing	No	Yes
Features	Combination of two Features	Hybrid (Shape+Color+Texture+Vein)
Classifier	Binary (Only Two Class)	Multiple Class
Accuracy	Low	High
Space	Big Feature Vector	PCA (Principle for better accuracy)
Time	Low	Moderate

### III. CONCLUSION

In this work, the dataset of various plants have been collected. Features have been extracted for these leaves and are classified using RF classifiers. The texture based features; Shape and Color have been looked upon. The feature set has been formed using all the extracted features. This feature set has been given as the input for the classification part. The Random Forest also shows misclassification results. So, the future work is mainly on normalizing the feature values and reduces the misclassification errors. Further this work can be extended into an application that can be installed in the smart phones and educate people about the existence of the medicinal plants around them and make them aware of its therapeutic values. It also helps the common people, researchers and doctors to easily discriminate between the various medicinal plants without the help of experts.

### REFERENCES

1. SukhvirKaur ,Shreelekha Pandey And ShivaniGoel," Semi-Automatic Leaf Disease Detection Andclassification System For Soybean Culture" ,Iet Image Process., 2018 .
2. HuuQuanCap,KatsumasaSuwa,ErikaFujita,"A Deep Learning Approach For On-Site Plant Leaf Detection",2018 Ieee 14th International Colloquium On Signal Processing & Its Applications (Cspa 2018)
3. AnandR,VeniS,AravinthJ,"An Application Of Image Processing Techniquesfor Detection Of Diseases On Brinjal Leaves Using K-Means Clustering Method",2016 Fifth International Conference On Recent Trends In Information Technology
4. R.Meena Prakash, G.P.Saraswathy, G.Ramalakshmi,K.H.Mangaleswari And T.Kaviya,"Detection Of Leaf Diseases And Classification Using Digital Image Processing",2017 International Conference On Innovations In Information, Embedded And Communication Systems (Iciiecs)
5. JiaShijie, JiaPeiyi, Hu Siping And Liu Haibo,"Automatic Detection Of Tomato Diseases And Pests Based On Leaf Images",Liaoning Provincial Natural Science Foundation (2017 Ieee)
6. Wan MohdFadzil W.M.N, Shah Rizam M.S.B, R.Jailani, NooritawatiM.T,"Orchid Leaf Disease Detection Using Border Segmentation Techniques",2014 Ieee Conference On Systems, Process And Control (Icspc 2014), 12 - 14 December 2014, Kuala Lumpur, Malaysia.
7. Jobin Francis , AntoSahayaDhas D, Anoop B K,"Identification Of Leaf Diseases In Pepper Plants Using Soft Computing Techniques",2013.
8. N.Krithika, Dr.A.GraceSelvarani, "An Individual Grape Leaf Disease Identification Using Leaf Skeletons And Knn Classification",2017 International Conference On Innovations In Information, Embedded And Communication Systems (Iciiecs).
9. Pranjali B. Padol, Prof. Anjali A. Yadav,"Svm Classifier Based Grape Leaf Disease Detection", 2016 Conference On Advances In Signal Processing (Casp) Cummins College Of Engineering For Women, Pune. Jun 9-11, 2016.
10. Pranjali B. Padol, S. D. Sawant, "Fusion Classification Technique Used To Detect Downy And Powdery Mildew Grape Leaf Diseases", 2016 International Conference On Global Trends In Signal Processing, Information Computing And Communication.
11. FarhanaTazmimPinki, Nipa Khatun, S.M. Mohidul Islam, "Content Based Paddy Leaf Disease Recognition And Remedy Prediction Using Support Vector Machine",2017 20th International Conference Of Computer And Information Technology (Iccit), 22-24 December, 2017.



12. R.Meena Prakash, G.P.Saraswathy, G.Ramalakshmi, "Detection Of Leaf Diseases And Classification Using Digital Image Processing", 2017 International Conference On Innovations In Information, Embedded And Communication Systems (Iciiecs).
13. P.R. Rothe, Dr. R. V. Kshirsagar, "Automated Extraction Of Digital Images Features Of Three Kinds Of Cotton Leaf Diseases", 2014 International Conference On Electronics, Communication And Computational Engineering (Icecce).
14. Suresha M And Shreekanth K N, Thirumalesh B V, "Recognition Of Diseases In Paddy Leaves Using Knn Classifier", 2017 2nd International Conference For Convergence In Technology (I2ct).
15. Dr.K.Thangadurai, K.Padmavathi, "Computer Visionimage Enhancement For Plant Leaves Disease Detection", 2014 World Congress On Computing And Communication Technologies (Ieee).

