*International Journal on Applications in Information and Communication Engineering*
*Volume 5 : Issue 1: May 2019, pp 18 – 27 www.aetsjournal.com* ISSN (Online) : 2394 - 6237

# AGRICULTURAL PLANT LEAF DISEASE IDENTIFICATION USING IMAGE PROCESING AND DATA MINING

**R Kalaivani ,** Department of Computer Science and Engineering , Mahendra Engineering College , Namakkal .( Email ID : kalaivani10@gmail.com )
**K R Sathishkumar ,** Assistant Professor , Department of Computer Science and Engineering , Mahendra Engineering College , Namakkal .
( Email ID : sathishkumarr@mahendra.info )

***Abstract***— Plant disease identification is the most important sector in agriculture. Turmeric is one of the important rhizomatous crops grown in India. The turmeric leaf is highly exposed to diseases like rhizome rot, leaf spot, and leaf blotch. The identification of plant diseases requires close monitoring and hence this paper adopts technologies to manage turmeric plant diseases caused by fungi to enable production of high quality crop yields. Various image processing and machine learning techniques are used to identify and classify the diseases in turmeric leaf. The dataset with 800 leaf images of different categories were pre-processed and segmented to promote efficient feature extraction. Machine learning algorithms like support vector machine, decision tree and naïve bayes were applied to train the model. The performance of the model was evaluated using 10 fold cross validation and the results are reported.

## I. INTRODUCTION

Agriculture is one of the emerging topics in data mining.Research in agriculture is aimed towards increase of productivity and food quality at reduced expenditure and with increased profit. In the past fewyears new trends have emerged in the Indian agricultural sector and technological advancements are gradually finding their place.Turmeric is the dried rhizome of Curcuma longa, aherbaceous perennial belonging to the family Zingiberaceae. Curcumin, the most biologically active phytochemical compound is available upto 3% in Turmeric.

Indian turmeric is considered as the best in the world due to its high curcumin content. It is extracted and researched for its well-known range of therapeutic effects.The idea of integrating Information and Communications Technology (ICT) with agriculture sector motivates the development of an automated system for turmeric disease classification. The various diseases that affect turmeric leaf are categorized as leaf spot, leaf blotch and rhizome rot. The cause and symptoms for each of these diseases are described below. Leaf spot of turmeric is the most important disease of turmeric. It has become a major constraint in successful cultivation of turmeric. Symptom appears as brown spots of various sizes on the upper surface of the young leaves.It is caused by the fungi Colletotrichumcapsici. Leaf blotch disease symptom appears as small, oval, rectangular or irregular brown spots on either side of the leaves which soon become dirty yellow or dark brown. The causal organism is the fungiTaphrinamaculans. Foliar symptoms due to rhizome rot appear as light yellowing of the tips of lower leaves which gradually spreads to the leaf blades. The fungiPythiumgraminicolum is the agent causing this disease.

Meenakshi M. Pawar et al., [1], approaches an automatic grading and sorting system for pomegranate. Color texture feature analysis was used for detection of surface defects on pomegranates. Best features were used as an input to Support Vector Machine (SVM) classifier and tests were performed to identify best classification model.K.R.Wang and Shaokun Li [2] created a model of cotton leaf chlorophyll determination based on using the machine vision technology for the colour features of cotton leaf.

The research showed that the BIR values of ROB color system values of chromaticity coordinate and the S values of HIS color system were all significantly correlated with chlorophyll content of cotton leaf. These values could be used to determine the concentration of chlorophyll.Yan Cheng Zhang, et al., [3] tries to identifyand diagnose cotton disease using computer vision.They developed the fuzzy feature selection approach, fuzzy curves (FC) and surfaces (FS) to select features of cotton diseased leaves image. They showed that the effectiveness of features selected by the FC and FS method was much better than that selected by human randomly or other methods.NitinP.Kumbhar et al., [4], explained texture statistics for detecting the plant leaf disease by color transformation structure.RGB is converted into HSV space because HSV was a good color descriptor. Masking and removing of green pixels with pre-computed threshold level was done. Segmentation was performed and they obtained useful segments. These segments were used for texture analysis by color co-occurrence matrix. Finally SGDM texture parameters were compared to texture parameters of normal leaf. PradnyaRavindraNarvekar et al., [5], proposed a system to discuss the effective way used in performing detection of grape diseases through leaf feature inspection. Leaf image was captured and proposed to determine the health status of each plant. The digital images were acquired from the environment using a digital camera.

Then image processing techniques were applied to the acquired images to extract useful features that were necessary for further analysis. After that, several analytical discriminating techniques were used to classify the images according to the specific problem. From the above literature survey a clear outlook about the techniques and methodologies followed in the existing works for detection of plant diseases has been obtained. This assists developing a proposed work of disease classification in turmeric with better performance.

## II. REVIEW OF LITERATURE SURVEY

### 1)LEAF DISEASE DETECTION USING IMAGE PROCESSING AND NEURAL NETWORK

In agriculture research of automatic leaf disease detection is essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect symptoms of disease as soon as they appear on plant leaves. There are the main steps for disease detection of Image Acquisition, Image Pre-processing, Image Segmentation, Feature Extraction and Statistical Analysis. This proposed work is in first image filtering using median filter and convert the RGB image to CIELAB color component, in second step image segmented using the k-medoid technique, in next step masking green-pixels & Remove of masked green pixels, after in next step calculate the Texture features Statistics, in last this features passed in neural network. The Neural Network classification performs well and could successfully detect and classify the tested disease.

**Keywords—** Leaf disease, Image processing, CIELAB color model, SGDM Matrix, ColorCooccurrence Method, k-medoids, Neural Network.

**LIMITATIONS**

✓Complexity of neural network computation is too high.

### 2)FAST AND ACCURATE DETECTION AND CLASIFICATION OF PLANT DISEASES

We propose and experimentally evaluate a software solution for automatic detection and classification of plant leaf diseases. The proposed solution is an improvement to the solution proposed in as it provides faster and more accurate solution. The developed processing scheme consists of four main phases as in. The following two steps are added successively after the segmentation phase. In the first step we identify the mostlygreen colored pixels. Next, these pixels are masked based on specific threshold values that are computed using Otsu's method, then those mostly green pixels are masked. The other additional step is that the pixels with zeros red, green and blue values and the pixels on the boundaries of the infected cluster (object) were completely removed. The experimental results demonstrate that the proposed technique is a robust technique for the detection of plant leaves diseases. The developed algorithms efficiency can successfully detect and classify the examined diseases with a precision between 83% and 94%, and can achieve 20% speedup over the approach proposed in.

**General Terms** Artificial Intelligence, Image Processing.

**Keywords** K-means, SGDM Matrix, Color Co-occurrence Method, HSI, Neural Networks.

**LIMITATIONS**

✓Working on individual pixel increases the complexity of the system and there by hamper the accuracy of the system.

### 3)COLOR TRANSFORM BASED APPROACH FOR DISEASE SPOT DETECTION ON PLANT LEAF

In this paper, an algorithm for disease spot segmentation using image processing techniques in plant leaf is implemented. This is the first and important phase for automatic detection and classification of plant diseases. Disease spots are different in color but not in intensity, in comparison with plant leaf color. So we color transform of RGB image can be used for better segmentation of disease spots. In this paper a comparison of the effect of CIELAB, HSI and YCbCr color space in the process of disease spot detection is done. Median filter is used for image smoothing. Finally threshold can be calculated by applying Otsu method on color component to detect the disease spot. An algorithm which is independent of background noise, plant type and disease spot color was developed and experiments were carried out on different “Monocot” and “Dicot” family plant leaves with both, noise free (white) and noisy background.

**Index Terms—** CIELAB, Color Transform, Plant Leaf Spot Disease, Segmentation and Image Processing.

**LIMITATIONS**

✓ Color transformation does not 100% accuracy as the image color can be manipulated due to wrong angle image capture or brightness and darkness can affect the image color quality which decreases the reliability of system to detect accurate diseases.

**4) LEAF DISEASE DETECTION AND GRADING USING COMPUTER VISION TECHNOLOGY AND FUZZY LOGIC**

In Agriculture, leaf diseases have grown to be a dilemma as it can cause significant diminution in both quality and quantity of agricultural yields. Thus, automated recognition of diseases on leaves plays a crucial role in agriculture sector. This paper imparts a simple and computationally proficient method used for leaf disease identification and grading using digital image processing and machine vision technology. The proposed system is divided into two phases, in first phase the plant is recognized on the basis of the features of leaf, it includes pre-processing of leaf images, and feature extraction followed by Artificial Neural Network based training and classification for recognition of leaf. In second phase the disease present in the leaf is classified, this process includes K-Means based segmentation of defected area, feature extraction of defected portion and the ANN based classification of disease. Then the disease grading is done on the basis of the amount of disease present in the leaf.

**Keywords:** Computer vision, color image segmentation, disease grading, artificial Neural Network (ANN), fuzzy logic.

**LIMITATIONS**

✓In this system they do not have the ability to learn and adapt after solving problem.

✓It requires extensive testing this is the one of the demerits of this system.

**5) DETECTION OF UNHEALTHY REGION OF PLANT LEAVES AND CLASSIFICATION OF PLANT LEAF DISEASES USING TEXTURE FUTURES**

Plant diseases have turned into a dilemma as it can cause significant reduction in both quality and quantity of agricultural products. Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves. The proposed system is a software solution for automatic detection and classification of plant leaf diseases. The developed processing scheme consists of four main steps, first a color transformation structure for the input RGB image is created, and then the green pixels are masked and removed using specific threshold value followed by segmentation process, the texture statistics are computed for the useful segments, finally the extracted features are passed through the classifier. The proposed algorithm’s efficiency can successfully detect and classify the examined diseases with an accuracy of 94%. Experimental results on a database of about 500 plant leaves confirm the robustness of the proposed approach.

**Keywords:** HSI, color co-occurrence matrix, texture, SVM, plant leaf diseases.

**LIMITATIONS**

✓ It requires extensive memory therefore due to this lack of transparency occur

✓ Algorithmic complexity goes high.

**6) CLASSIFICATION OF DISEASED PLANT LEAVES USING NEURAL NETWORK ALGORITHMS**

Agriculture is the mother of all cultures. It played a vital role in the development of human civilization. But plant leaf diseases can damage the crops there may be economic losses in crops. Without knowing about the diseases affected in the plant, the farmers are using excessive pesticides for the plant disease treatment. To overcome this, the detected spot diseases in leaves are classified based on the diseased leaf types using various neural network algorithms. By this approach one can detect the diseased leaf variety and thus can take necessary steps in time to minimize the loss of production. The proposed methodology uses to classify the diseased plant leaves using Feed Forward Neural Network (FFNN), Learning Vector Quantization (LVQ) and Radial Basis Function Networks (RBF) by processing the set of shape and texture features from the affected leaf image. The simulation results show the effectiveness of the proposed scheme. With the help of this work, a machine learning based system can be formed for the improvement of the crop quality in the Indian Economy.

sratio, F_ measure.

**LIMITATIONS**

- ✓ It do not provide explanation and also not used for the little data.

### 7) AGRICULTURAL PLANT LEAF DISEASE DETECTION USING IMAGE PROCESSING

The detection of plant leaf is a very important factor to prevent serious outbreak. Automatic detection of plant disease is essential research topic. Most plant diseases are caused by fungi, bacteria, and viruses. Fungi are identified primarily from their morphology, with emphasis placed on their reproductive structures. Bacteria are considered more primitive than fungi and generally have simpler life cycles. With few exceptions, bacteria exist as single cells and increase in numbers by dividing into two cells during a process called binary fission Viruses are extremely tiny particles consisting of protein and genetic material with no associated protein. The term disease is usually used only for the destruction of live plants. The developed processing scheme consists of four main steps, first a color transformation structure for the input RGB image is created, and this RGB is converted to HSI because RGB is for color generation and his for color descriptor. Then green pixels are masked and removed using specific threshold value, then the image is segmented and the useful segments are extracted, finally the texture statistics is computed. From SGDM matrices. Finally the presence of diseases on the plant leaf is evaluated.

**Keywords:** HSI, Segmentation, Color Co-occurrence Matrix, Texture, Plant Leaf Diseases.

**LIMITATIONS**

- ✓ Time consuming process

## III. PROPOSED SYSTEM

The proposed work focuses on classifying the diseases in the turmeric leaf. Initially, the leaf images are collected for three different diseases namely Leaf spot, Leaf blotch, Rhizome rot and stored in a database for further processing. The database consists of 800 leaf images with 200 diseased leaf images for each disease category and 200 images for normal category. The proposed system includes different phases as follows. In the pre-processing phase, the given image is resized and converted into HIS images. The hue component is alone taken for further processing. In segmentation phase, K-means segmentation technique is used to segment the diseased portion from the original image. The segmentation result is processed and feature vectors are generated including color, texture and shape features. A subset of best features is selected from the feature set for accurate classification using ranking method.

## IV. BLOCK DIAGRAM

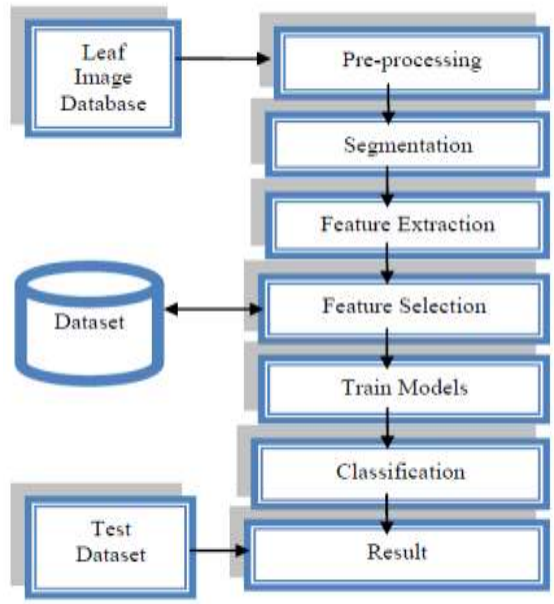



**Figure1 : System Architecture**

The system architecture of the proposed system, in which we aim to solve the diseases identification detection leafs.

### 1) ALGORITHM AND TECHNIQUES K-MEANS CLUSTERING ALGORITHM

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\|x_i - v_j\|\right)^2$$

where,

'‖xi - vj‖' is the Euclidean distance between xi and vj.

'ci' is the number of data points in ith cluster.

'c' is the number of cluster centers.

Algorithmic steps for k-means clustering

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1) Randomly select 'c' cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, 'ci' represents the number of data points in ith cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.
6) If no data point was reassigned then stop, otherwise repeat from step 3).

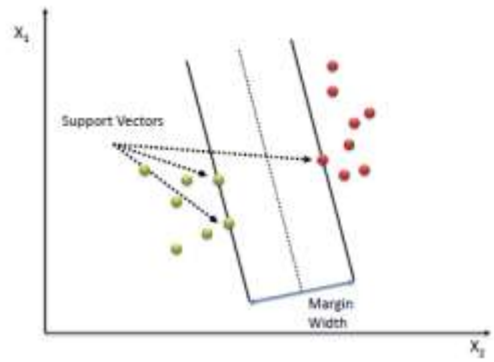
**ADVANTAGES**

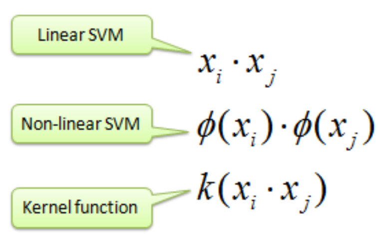
1) Fast, robust and easier to understand.
2) Relatively efficient: O(tknd), where n is # objects, k is # clusters, d is # dimension of each object, and t is # iterations. Normally, k, t, d << n.
3) Gives best result when data set are distinct or well separated from each other.

**2) Support Vector Machine (SVM)**

A Support Vector Machine (SVM) performs classification by finding the hyperplane that maximizes the margin between the two classes. The vectors (cases) that define the hyperplane are the support vectors.

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. A support vector machine constructs a hyper plane or set of hyper planes in a high-dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class called functional margin, since in general the larger the margin the lower the generalization error of the classifier. Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of support vector machines, a data point is viewed as a p-dimensional vector of a list of p numbers, and one wants to know whether one can separate such points with a p − 1-dimensional hyper plane. This is called a linear classifier.




Polynomial

$$k(x_i, x_j) = (x_i.x_j)^d$$

Gaussian Radial Basis function

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

### 3) Decision tree

Decision tree learning is the process of learning decision trees from the labelled training examples. Decision tree classification algorithm generates the o utput as a binary tree like structure called a decision tree, where each non leaf node i.e., internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node or terminal node holds a class label. The topmost node in a tree is the root node. A decision tree model contains rules which are used to predict the target variable. The class label of a new instance is predicted by testing the attribute values of the instance against the decision tree. A path is traversed from the root to a leaf node, which gives the class label of that data. Decision trees can be easily converted into classification rules.

### 4) Naive Bayes

The Naive Bayes Classifier technique is based on Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Naïve Bayes classifiers assume that the effect of a variable value on a given class is independent of the values of other variable. The Naive-Bayes inducers compute conditional probabilities of the classes given the instance and pick the class with the highest posterior. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting.

A Naïve Bayes classifier is a simple probabilistic classifier based on Bayes theorem with strong independence assumptions. General formulation is given by

$$\text{Given classes } \omega_j \text{ and dataset } x$$

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

$$\text{where}$$

$$p(x) = \sum_j p(x \mid \omega_j) P(\omega_j)$$

### 5) Pre-Processing

The Turmeric leaf images are collected manually from Annur, Thondamuthur and Sulur. The images are taken using high resolution camera of 12 mega pixels. The collected images are of different dimensions hence it is essential to convert them to uniform size for efficient pre-processing. The images are resized to 256*256 dimensions using interpolation method.

### 6) RGB to HSI

RGB images are converted into Hue Saturation Intensity (HSI) color` space representation. RGB is ideal for color generation. But HSI model is an ideal tool for color perception. Hue is a color attribute that describes pure color as perceived by an observer. Saturation refers to the relative purity or the amount of white light added to hue and intensity means amplitude of light. After the transformation process, saturation and intensity are dropped since it does not give extra information the hue component is taken for further analysis. The output of HSI conversion is shown in Fig.2.

**Figure 2 : HIS conversion**

### 7) Segmentation

Image segmentation is an important aspect of digital image processing. It may be defined as a process of assigning pixels to homogenous and disjoint regions which form a partition of the image that share certain visual characteristics. The major goal of segmentation is to simplify or change the representation of an image into meaningful image that is more proper and easier to explore.

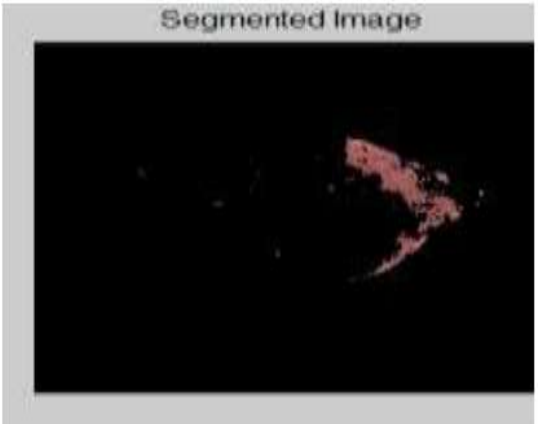



**Figure 3 : Segmentaation of leafs**

Segmentation is essentially a collection of methods that allows spatial partitioning to the close parts of the image as objects. In this proposed work segmentation process is done using kmeans segmentation algorithm.K-Means segmentation algorithm classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural grouping in them. The segmented image is shown in Fig.3.

**8)Feature Extraction**

Drawing out specific features from the preprocessed images is called feature extraction. Feature extraction plays a vital role in data mining. It can be used to improve the classification effectiveness and computational efficiency. Feature extraction is carried out with all the pre-processed leaf images. Three types of features such as color, texture and shape features are extracted in this work in order to extract the prominent features of an image.

**9) Color Features**

Color features play a decisive role in the classification ofleaf image. A color image is a combination of some basic colours. Each individual pixel of a color image is broken down into red, green and blue values. The entire image is represented in three matrices, each corresponding to red, green and blue. The three matrices are arranged in sequential order, to create a three m by n by 3 matrixes. For an image which has a height of 5 pixels and width of 10 pixels the result is 5 by 10 by 3 matrixes for a true colorimages.The RGB color features such as meanR, meanG, meanB are extracted from the leaf images.

**10) Texture features**

The texture features are extracted using Grey Level Co-occurrence Matrix (GLCM). A GLCM is a matrix where number of rows and columns is equal to number grey levels G in an image. It is defined over an image to be the distribution of co-occurring values in the given offset. It is a way of extracting second order statistical features.

**11) Energy**

$$energy(ene) = \sum_i \sum_j g_{ij}{}^2$$

This statistic is also called uniformity or angular second moment. It measures the textural uniformity that is pixel pair repetitions. It detects disorders in textures. Energy reaches a maximum value equal to one. High energy values occur when the gray level distribution has a constant or periodic form. Energy has a normalized range.

**12) Entropy**

$$entropy(ent) = -\sum_i \sum_j g_{ij} \log_2 g_{ij}$$

This statistic measures the disorder or complexity of an image. The entropy is larger when the image is not texturally uniform and many GLCM elements have very small values. Complex textures tend to have high entropy. Entropy is strongly, but inversely correlated to entropy

**13) Contrast**

$$contrast(con) = \sum_i \sum_j (i-j)^2 g_{ij}$$

This statistic measures the spatial frequency of an image and difference moment of GLCM. It is the difference between the highest and the lowest values of a contiguous set of pixels. It measures the amount of local variations present in the image. A low contrast image presents GLCM concentration term around the principal diagonal and features low spatial frequencies.

**14) Variance**

$$Variance\ (var) = \sum_i \sum_j (i - \mu)^2 g_{ij} \text{ where } \mu \text{ is the mean of } g_{ij}$$

This statistic is a measure of heterogeneity and is strongly correlated to first order statistical variable such as standard deviation. Variance increases when the gray level values differ from their mean.

**15) Homogeneity**

$$homogeneity(hom) = \sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij}$$

Homogeneity weights values by the inverse of the Contrast, weight, with weights decreasing exponentially away from the diagonal.

#### 16) Correlation

The correlation feature is a measure of gray tone linear dependencies in the image. GLCM.

Correlation is quite a different calculation from the other texture measures. It is independent of them and can often be used profitably in combination with another texture measure. It also has a more intuitive meaning to the actual calculated values: 0 is uncorrelated, 1 is perfectly correlated.

$$correlation(cor) = \frac{\sum_j \sum_i (ij)g_{ij} - \mu_x\mu_y}{\sigma_x\sigma_y} g_{ij}$$

where $\mu_x, \mu_y, \sigma_x$ and $\sigma_y$ are the means and standard deviations of $g_x$ and g

## V. EXPERIMENTS AND RESULTS

The classification models can be evaluated using various criteria such as accuracy, precision, recall and F-measure.

The formula for calculating accuracy is,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The formula for calculating Precision is,

$$Precision = \frac{TP}{TP + FP}$$

The formula for calculating Recall is,

$$Recall = \frac{TP}{TP + FN}$$

The formula for calculating F-measure is,

**F-measure**=2*(precision*recall)/ (precision+recall)

Where TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

The comparative results indicate that Support Vector Machine based classification model yields a better performance when compared to other models. The performance of the classifiers are assessed using measures such as accuracy, precision, recall, fmeasure and time taken to build the model.

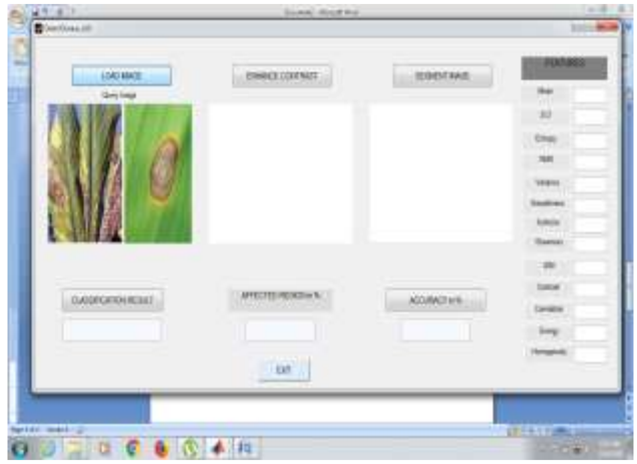


**Figure 4 : Load Query Input Turmeric Leaf Image**

The collected images are of different turmeric leafs hence it is essential to convert them to uniform size for efficient Load pre-processing.

**Figure 5: Contrast Enhanced Image**

Image enhancement techniques have been widely used in many applications of image processing where the subjective quality of images is important for human interpretation. Contrast is an important factor in any subjective evaluation of image quality. Contrast is created by the difference in luminance reflected from two adjacent surfaces.

In other words, contrast is the difference in visual properties that makes an object distinguishable from other objects and the background. In visual perception, contrast is determined by the difference in the colour and brightness of the object with other objects.

Our visual system is more sensitive to contrast than absolute luminance; therefore, we can perceive the world similarly regardless of the considerable changes in illumination conditions. Many algorithms for accomplishing contrast enhancement have been developed and applied to problems in image processing.

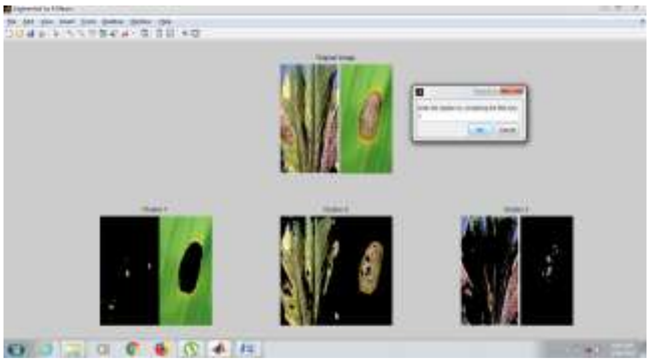


**Figure 6 : Enter The Cluster No Containing ROI**

A Region of Interest (ROI) is an area of an image, which is graphically selected from a window displaying that image, in which you want to focus your image analysis. This area can be used to focus further processing. The following document contains information about ROI functions and the way they are used.

**Figure 7 : Segmentation Process and Features**

Segmentation is the process of dividing a heterogenous market into smaller markets or sub-segments on the basis of some homogenous characteristics. Segmentation yields market segments which are similar in some aspect.

**Figure 8 : Classification Result**

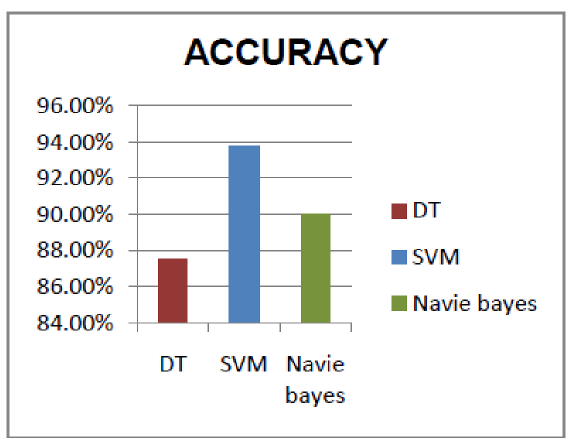
The classification models, the evaluation results are based on the following metrics or measures:Accuracy: is the number of correct predictions over the number of total instances that have been evaluated.

**Figure 9 : Output Accuracy of Disease Identification**

We empirically evaluate four techniques of background removal and three techniques of segmentation. To enable accurate extraction of features, we propose centroid feeding based K-means clustering for segmentation of disease portion from a leaf image. We enhance the output of K-means clustering by removing green pixels in the disease portion. We extract various features under three categories: color, shape, and texture. We use Support Vector Machine (SVM) for multi-class classification. We achieve 93.75% accuracy on training dataset.

**1) PREDICTIVE PERFORMANCE OF CLASSIFIERS**

| Classifier | DT | SVM | Naive Bayes |
|---|---|---|---|
| Precision | 0.8817 | 0.9395 | 0.9354 |
| Recall | 0.875 | 0.9437 | 0.725 |
| F-measure | 0.8783 | 0.9415 | 0.8168 |
| Accuracy | 87.50% | 93.75% | 90% |



# VI. CONCLUSION

The research has modeled disease detection and classification for turmeric leaves. The work involves various image processing and machine learning techniques. The machine learning algorithms such as Support Vector Machine (SVM), Decision Tree (DT) and NB (Naive Bayes) were implemented using MATLAB platform. This analysis was verified by the result of 10-fold cross validation. The study shows that classification of turmeric leaf diseases using Support Vector Machine (SVM) gives better accuracy of 93.75% when compared to other algorithms. In future the research can be extended using different techniques for segmenting the diseased portion of the original leaf.

# References

[1] Meenakshi M. Pawer, Sanman Bhusari, Akshay Gungewar (2012), “Indentification of infected pomegranates using color texture feature analysis”, International journal of computer application(0975-8887).

[2] K.R.Wang, S.K.Li, CT.wang (2006), “Acquired chlorophyll concentration of cotton leaves with technology of machine vision”, ActaAgronvo1.1, pp. 34-40, sin.

[3] Yan Cheng Zhang, et al.,“ Features selection of cotton disease leaves image based on fuzzy feature selection techniques”, IEEE Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, China, 2-4 Nov 2007.

[4] Mr.Nitin, P.Kumbhar, Prof. Sanjay, B. Dhaygude. (2013), “Crop Disease Detection Using CBIR”, International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS).Prof.

[5] Pradnya Ravindra Narvekar, Mahesh Manik Kumbhar, S. N. Patil. (2014), “Grape leaf disease detection & Analysis using SGDM Matix Method”, International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Issue 3.

[6] Yeo Zhang, M.Z. Li (2005), “Nutrition information extraction of the cucumber leaves in the greenhouse based on computer vision technology”,Transactions of the CSAE,vo1.21, pp.102-IOS.

[7] Sanjay, B. Dhaygude, Mr.Nitin, P.Kumbhar (2013), “Agricultural plant Leaf Disease Detection using Image proessing”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering , Vol. 2, Issue 1.

[8] Patil J.K, Raj Kumar(2012), “Feature Extraction of Diseased Leaf image”, Journal of Signal and Image Processing Volume 3, Issue 1.

[9] P.Revathi, M. Hemalatha. (2012), “ Homogenous segmentation based edge detection techniques for proficient Identification of the Cotton leaf Spot Disease”, International Journal of Computer Application Vol.47-No 2.

[10] Mokhled S, Al-Tarawneh (2013), “An Empirical Investigation of Olive Leave Spot Disease Using Auto-Cropping Segmentation and Fuzzy C-Means Classification”, World Applied Science Journal 23 (9): 1207-1211.

[11] Anthony.G, wickramarachchf.N, “An image recognition system for crop disease Identification of paddy fields in Sri Lanka”, Fourth International Conference on Industrial and Information System, ICIIS 2009,Sri Lanka.P 403-407, 2009.

[12] Moshou et al., “Automatic detection of yellow rust in wheat using reflectance measurements and neural networks”, Computer and Electonics in agriculture, v.44, p.173-188, 2004.

[13] Soman K.P, Loganathan R, Ajay V (2009), “Machine Learning with SVM and other Kernel Methods”, PHI, India.

[14] V.Premalatha, Dr. M. G. Sumithra, S. Deepak, P. Rajeswari (2014), “Implementation of Spatial FCM for Leaf Image Segmentation in Pest Detection”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 10, ISSN: 2277 128X.

[15] Manimekalai.K, Vijaya.MS(2014), “Support Vector Machine Based Tool For Plant Species Taxonomic Classification”, Journal of Asian Scientific Research 4(4):159-173.

[16] K. Muthukannan et al., (2015) “Classification of diseased plant leaves using neural network algorithms”, ARPN Journal of Engineering and Applied Sciences vol. 10, no. 4.

[17] Zulkifli Bin Husin et al.,(2012), “Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques”, IEEE International Conference on Intelligent Systems Modelling and Simulation.