Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)
IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

Computer Vision Based Turmeric Leaf Disease Detection and Classification

A Step to Smart Agriculture

Gayatri Kuricheti
Post Graduate Scholar
kgayatri195@gmail.com

Supriya P
Associate Professor
p_supriya@cb.amrita.edu

Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India

Abstract- **Disease identification plays a vital role in agricultural sector. Turmeric being a rhizomatous crop and well known for its therapeutic effects, monitoring such crops is crucial. The turmeric leaves are mainly exposed to diseases like Leaf Spot and Leaf Blotch. The paper develops an algorithm for detecting and preventing the spreading of diseases to the whole crop and results in high quality crop production. The data base of different leaf images was created and processed using k-Means image segmentation and leaf images textural analysis was carried out using GLCM. SVM classifier is used to classify the feature extracted images after ranking their attributes using an information gain algorithm. A GUI has been created to portray the various stages of the image processing algorithm and detect the two leaf diseases.**

Keywords- Information Gain algorithm, GLCM, Texture Analysis, SVM, GUI

I. INTRODUCTION

Turmeric, a product of rhizomatous crop (Curcuma longa) which is native to tropical South Asia is belonging to the ginger family. India is the world's largest turmeric crop producer and consumes 80% of it. In the world, Indian turmeric was described as best medicinal herb because of its implicit qualities and the presence of major bioactive compound which is Curcumin. A city in the state of Tamil Nadu called Erode is the most important trading area for turmeric, for which it is therefore called as "Yellow City". A place in Maharashtra called Sangli is the second place to erode as turmeric trading centre.

In Turmeric cultivation, diseases are the most important problems which reduce the harvest of Turmeric crop. The focused diseases in this work are Leaf Spot and Leaf Blotch. For instant, leaf spot is affected badly in rainy seasons under humid conditions. It causes damage to a greater extent by reducing rhizome size up to 52 percent. The infection of such diseases causes the change in the color and appearance of the turmeric leaves. The existing methods for plant disease detection simply by naked eye observation or laboratory based techniques by experts is time consuming and requires continuous monitoring of plant. Therefore, if the plant monitoring methods can be stored by using some programming language into an automatic module then the process can be error tolerant. Hence image processing plays a vital role in disease detection and analysis. In the proposed work several image processing methods and algorithms are used for the detection of two main turmeric leaf diseases namely Leaf Spot and Leaf Blotch and the results obtained were classified using SVM classifiers.

II. LITERATURE SURVEY

Dr.J.Vijayakumar [1] proposed about the study of turmeric leaf diseases and various methods to identify the diseases in a detailed approach. Various methods used for the identification of diseases are Statistical analysis, Histogram analysis and neural network based disease identification.

Sagar Patil, Anjali Chandavale [2] proposed different methods for plant disease detection. Different types of plant families are monocot and dicot family plants. In [3] it mentioned about different techniques for segmentation and feature extraction of infected plants.

In [4] the author claims about different feature selection techniques for the classification of turmeric plant diseases. The RGB color image is converted into HSI color space and masking is applied to remove the greenness of the leaves. Then the image is processed with K-Means segmentation algorithm. For the feature selection different search algorithms like principal component analysis, information gain and relief-F attribute evaluator were used.

Eapen [5], proposed a method for reducing the noise in the given images and enhancing the edges before processing it with segmentation phase. In the pre-processing stage they concentrated on image resizing, histogram equalization, image cropping and median filtering. In this approach, histogram equalization was used for contrast and textural enhancement of medicinal images.

Madhuri [6] has proposed a method to segment the images by masking the green pixels. The input RGB images which are complicated to segment in three dimensional spaces, therefore for the information to be available in single plane gray level images have been used.

Patil R.V. [7] says that if the number of clusters can be divided in appropriate manner, an algorithm called K-means clustering can give accurate results. They discussed methods for edge detection for calculating the number of clusters. The detection of edges happened using phase congruency. Euclidean distance was used for estimating the clusters. The popular K-means algorithm has been used for the segmentation process. The proposed technique has been implemented in Matlab.

In [8], they discussed about advanced computing to assist the farmers in plant quality production. This method used mobiles to capture diseased cotton leaves. Color feature segmentation has been done to detect disease spots. For the feature extraction, Edge detection was used to detect the diseases. Neural network was implemented for classifying the diseases.

In [9], they detected pest and disease effected leaves by converting RGB to HSV color space and green channel has been extracted using threshold mask. Area, Perimeter features have been extracted. The classifier used is Decision Tree.

In [10], they differentiated between weed and a crop which are carrot and curry leaves. The classifier used is SVM classifier.

III. SYSTEM MODEL OF PROPOSED WORK

The suggested work mainly classifies the diseases in the turmeric leaves. The dataset of 200 turmeric leaves are collected for processing the data. The two diseases considered are leaf spot and leaf blotch as shown in fig. 1 and fig. 2.

Fig. 1. Leaf Spot

Fig. 2. Leaf Blotch

The work includes different phases of image processing techniques, the steps included are

1. Initially for pre-processing state, the input images are to be converted from one color space which is RGB to HSI color space and only the hue component is processed through other phases.

2. Then K-Means Segmentation has been used for extracting the diseased part of the leaf.

3. The output of the segmentation phase is then processed to feature extraction phase where GLCM texture analysis is used.

4. The features are trained and using multi class SVM classifier, the diseases are classified. The system overview of the above mentioned work is as follows

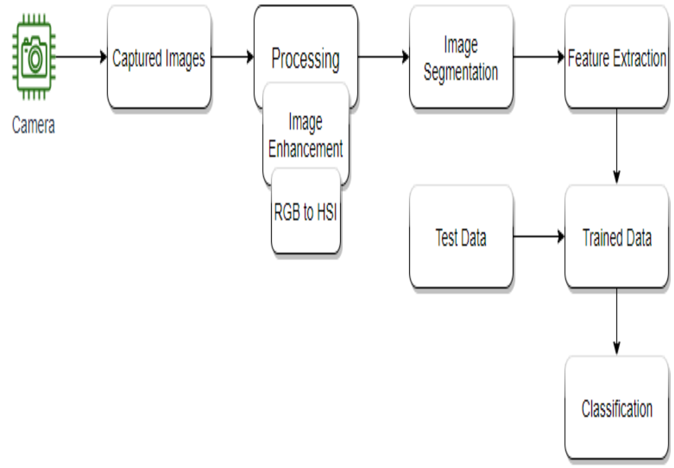

Fig. 3. Block Diagram of proposed Work

IV. METHODOLOGY

A. *Leaf Acquisition and color model Conversion*

Initially, the input images are resized to the same format using interpolation method. The purpose of RGB to HSI conversion is that it makes certain calculations more convenient to provide an intuitive way to identify colors. Research proves that the conversions of various models speed up the image processing with least time delays. Therefore in situations where color description plays a vital role, HSI (Hue Saturation and Intensity) color model is often preferred. Therefore, steps for the conversion of color model and the output for the conversion of leaf image into HSI are as follows

1. First read the RGB image
2. Represent it in the range [0 1]
3. $H = \begin{pmatrix} \theta, if\ B\leq G \\ 360-\theta, if\ B>G \end{pmatrix}$, *where*

$$\theta = \cos^{-1}\left\{\frac{1}{2}[(R-G)+(R-B)][(R-G)2+(R-B)(G-B)\frac{1}{2}\right.$$

4. *S = 1-3R+G+B[min(R,GB,)]*
5. $I = \frac{1}{3}(R+G+B)$

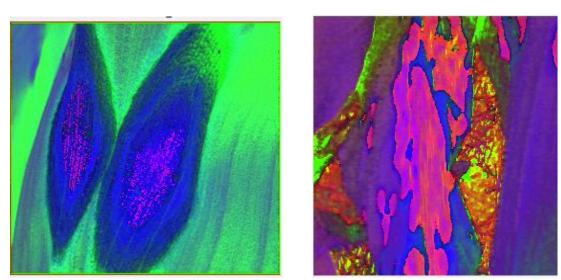

Fig. 4. HSI Conversion for Diseased leaves

B. *Segmentation using k-means Clustering*

According to the survey, clustering methods are considered to be very effective for color extraction. One of such clustering techniques is K-Means [12] [13] [15]. It’s easy to implement and requires low computational efforts. K-Means tries minimizing the sum of distances between cluster centroid and points. The algorithm consists of following steps:

1. Selecting K points as initial clustering centers.
2. Initiate each point of image pixels to nearest cluster depending on its distance from cluster center.
3. It computes mean values for each cluster and make that particular mean as new center.
4. Likewise, repeat 2 and 3 steps till the centroid no longer moves.

The main disadvantage of this algorithm is the need to fix the number of clusters. For the extraction of infected part of the leaf, 3 clusters (representing background, healthy part and the infected part) can be ideal. The disease attacked on the green pigments of leaf changes the color at that portion. The segmentation results are shown where Fig.5 is the input image; Fig.6-8 is the clustered image where one of the clusters will be the required disease portion, after removing all the unwanted part. This Region of Interest (ROI) is then transmitted to the next phase for applying feature extraction algorithms.

Fig. 5. Input Image

Fig. 6. Cluster1

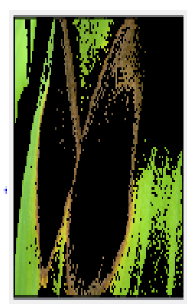

Fig. 7. Cluster2

Fig. 8. Cluster3

A. Feature Extraction and Dataset formation

Texture analysis is a method to define the relationship among the pixels in local area. Gray Level Co-occurrence Matrix is the texture analysis statistical approach which considers the pixels spatial relationship. By finding the pairs of pixel with certain values obtained in specific orientation and distance, GLCM characterizes the textural properties of an image. GLCM is a way to calculate the statistical features of a given image. It can tell specific properties regarding the spatial closeness of gray-levels in an image. This method is best used for classifying texture features. For a given image M*M, the elements of GLCM matrix 'M' are defined as

$$\sum_{x=1}^{K}\sum_{y=1}^{K}\begin{cases}1, if\ I(x,y)=i\ and\ I(x+d_x,y+d_y)=j\\ 0, oterwise\end{cases} \quad (1)$$

As discussed in [11], various image features can be calculated using GLCM matrix. It includes thirteen feature vectors which are further used for disease classification. The features include energy, contrast, correlation, entropy, and mean etc which are described below.

1. Mean

$$\mu=\left(\frac{1}{2k}\left(\sum_{i=1}^{k}\sum_{j=1}^{k}M_{ij}\right)\right) \quad (2)$$

Mean is the basic statistical measure of an image. Mean computes the mean of the image array as rows and as columns and is denoted as μ and it tends to remove the noise of the image.

2. Standard Deviation

It is the statistical index which quantifies mean and also portrays the diversity in the image. It is defined as:

$$\sigma=\sqrt{\left(\frac{1}{k^2-1}\sum_i\sum_j\left(M_{ij}-\frac{1}{2k-1}\sum_{i=1}^{k}\sum_{j=1}^{k}M_{ij}\right)\right)^2} \quad (3)$$

3. Entropy

$$Entropy=\sum_i\sum_j \log_2 M_{ij}\, M_{ij} \quad (4)$$

Entropy shows the amount of information of an image. It measures the randomness of an image texture. When the image is texturally uniform, entropy is large.

4. RMS

RMS Value represents the image distribution of an image and it is given by:

$$RMS\ Value=\left(\frac{1}{2k}\left(\sum_{i=1}^{k}\sum_{j=1}^{k}M_{ij}\right)^2\right)^{1/2} \quad (5)$$

5. Variance

$$Variance=\sum_i\sum_j(i-\mu)^2 M_{ij} \quad (6)$$

Where μ is the mean

Variance is the second order statistical feature which measures the heterogeneity. Variance will increase when there is difference in the grey level values from the mean value.

6. Smoothness

It defines the relative smoothness of intensity of an image and is carried out as replacing every pixel by the weighted average of its neighborhood.

7. Kurtosis

$$K=\sum_{i=0}^{L-1}(z-\mu)^4\, p(z_i) \quad (7)$$

8. Skewness

$$S=\sum_{i=0}^{L-1}(z-\mu)^3\, p(z_i) \quad (8)$$

Where p (z) is the corresponding histogram of the grey levels of an image with i= 0, 1…………L-1

9. Inverse Difference Movement

$$IDM = \sum_i \sum_j \frac{1}{1+(i-j)^2} M_{ij} \quad (9)$$

10. Contrast

$$Contrast = \sum_i \sum_j (i-j)^2 \, M_{ij} \quad (10)$$

11. Correlation

$$Correlation = \frac{\sum_i \sum_j ijM_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (11)$$

Where σ is the standard deviation

Correlation measures the linear dependency of the neighbor pixels of an image with gray levels.

12. Energy

$$Energy = \sum_i \sum_j {M^2}_{ij} \quad (12)$$

Where M is the GLCM matrix; i and j are the rows and columns of the matrix

Energy explains the disorders in textures. It is the sum of the squares of the pixels in the obtained GLCM matrix where M is the matrix. When the pixel repetition is more, energy value is high. Energy can reach a maximum value of one. It is also called angular secondary moment.

13. Homogeneity is defined as

$$H = \sum_i \sum_j \frac{1}{1+(i-j)^2} M_{ij} \quad (13)$$

Such 13 features can be calculated from the GLCM matrix that describes texture of an image as described by Haralick. The feature vector dataset is created for all the disease classes. Among the 13 attributes few features can be overlapped making classification difficult. Therefore, feature selection algorithm is used where it minimizes the number of features that are considered for disease classification.

Feature selection is an algorithm that enables automatic search for best attributes from the trained data. Trained data may contain many features that are irrelevant to the classification task; therefore Information Gain algorithm is used for the ranking of the features. Based on the ranking method, it gives the subset of suitable attributes for the classification. The classifier used for the proposed work is SVM Classifier.

B. Support Vector Machine Classifier

Based on the survey [14], different classifiers such as Back Propagation Neural Network (BPN), Artificial Neural Network (ANN), Fuzzy Logic and Support Vector Machine (SVM) which shows 91% accuracy rate has been chosen for the classification of diseased leaves.

SVM is a supervised learning model that analyzes the data for classification analysis. In the set of trained data, where each example belongs to one or the other group of diseases, such algorithm develops a model that makes new data fit into one or the other category. As with any supervised algorithm, SVM will be trained and then cross validated. Then, the new data will be classified using the trained machine. There are various kernel functions in which linear kernel has been implemented to gain the accurate classification.

V. RESULTS

For result visualization, a MATLAB GUI (Graphical User Interface) has been created that shows the original input image getting converted to HSI color space and the enhancement takes place. Hue and Saturation values will be used and segmented using K-Means Clustering in which K=3 are made appropriate. For the feature extraction phase, GLCM techniques have been used for the extraction of statistical features with necessary formulae. The data set has been created for different leaf diseases and classified using supervised learning algorithm called Support Vector Machine (SVM). The GUI outputs showing diseased leaf and healthy leaf classification are shown in the fig. 9 and fig. 10.

Fig. 9. Leaf Spot Classification

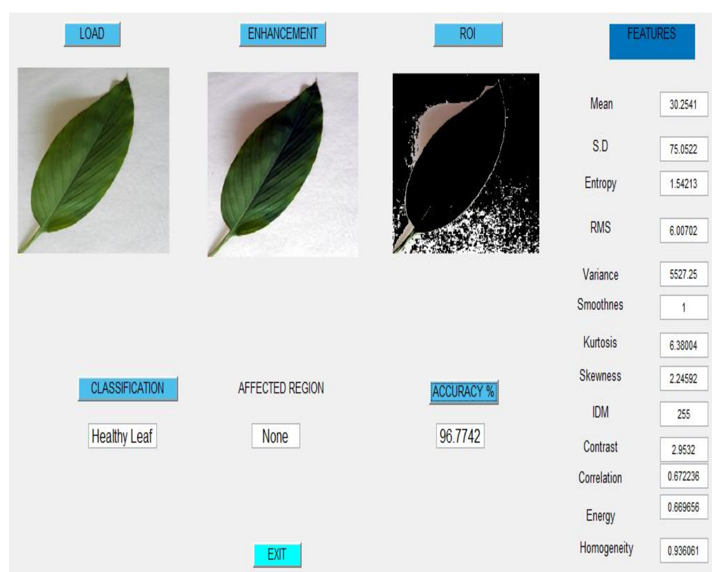

Fig. 10. Healthy Leaf Classification

VI. CONCLUSIONS

The paper was modeled for the detection and classification of turmeric leaves. It involved various processing techniques and classification algorithms for the accurate results. The data set has been created with 200 leaf images using .mat file. SVM was implemented for classification of the turmeric leaf diseases and its accuracy has been tested and found to be high. In future, better automatic algorithms can be implemented using IOT platform with large number of data set.

REFERENCES

[1] Dr.J.Vijayakumar, “Study of Turmeric Plant Diseases and Methods of Disease Identification using Digital Image Processing Techniques,” *International Journal on Future Revolution in Computer Science and Communication Engineering*, vol 3, 2017.

[2] Sagar Patil, and Anjali Chandavale, “A Survey on Methods of Plant Disease Detection,” *International Journal of Science and Research*, vol 4, 2015.

[3] Darshana A, Jharna Majumdar, and Shilpa Ankalaki, “Segmentation Method for Automatic Leaf Disease Detection,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol 3, 2015.

[4] Pream Sudha V, “Feature Selection Techniques for the Classification of Leaf Diseases in Turmeric,” *International Journal of Computer and Technology,* vol 43, 2017.

[5] Eapen M, and Korah R, “Medical image segmentation for Anatomical Knowledge Extraction,” *Journal of Computer Science*, vol 10, 2014.

[6] Jadhav Sanjivani, Lohar Utkarsha, Bhagat Madhuri and Salunke Shubhangi, “Plant Disease Identification System for Android*,” International Journal of Computer Science and Mobile Computing*, vol 5, 2016.

[7] R. Patil and K. Jondhale, "Edge based technique to estimate number of clusters in k-means color image segmentation,” in *Proc. 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, pp. 117-121, 2010.

[8] P. Revathi and M. Hemalatha, “Advance Computing Enrichment Evaluation of Cotton Leaf Spot Disease Detection Using Image Edge detection,” *IEEE*, 2012.

[9] Aparna K and Supriya P, “Precision Agriculture in Maize Fields,” *International Conference on Intelligent Computing and Control Systems*, 2018.

[10] Rani, Anjali K, Supriya P and T V Sarath, “Computer Vision based Segregation of Carrot and Curry leaf plants with weed identification in Carrot Field,” *International Conference on Computing Methodologies and Communication*, pp. 185-188, 2017.

[11] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.

[12] Puri C, and Sukhwinder Singh, “Image segmentation and Classification- A Review,” *International Journal of Innovative Research in Science, Engineering and Technology*, vol 3, (apr) 2014.

[13] F. C. Monteiro and A. Campilho, "Watershed framework to region-based image segmentation,” in *Proc. International Conference on Pattern Recognition*, ICPR 19th, pp. 1-4, 2008.

[14] M. Hameed, M. Sharif, M. Raza, S. W. Haider, and M. Iqbal, "Framework for the comparison of classifiers for medical image segmentation with transform and moment based features," *Research Journal of Recent Sciences*, vol. 2277, p. 2502, 2012.

[15] W. Cui and Y. Zhang, "Graph based multispectral high resolution image segmentation," in *Proc. International Conference on Multimedia Technology (ICMT)*, pp. 1-5, 2010.