

Identification of Anthracnose Disease in Mango Plant Using Clustering and Gray level Co- occurrence Matrix

Mamata Gawand^{1,*}, Soumya Gowda², Sonali Hnade³, Pooja Mahadik⁴, and Prajakta Dere⁵

¹ Ramrao Adik Institute of Technology ,Electronics and Telecommunication department, Nerul, India

² Ramrao Adik Institute of Technology ,Electronics and Telecommunication department, Nerul, India

³ Ramrao Adik Institute of Technology ,Electronics and Telecommunication department, Nerul, India

⁴ Ramrao Adik Institute of Technology ,Electronics and Telecommunication department, Nerul, India

⁵ Ramrao Adik Institute of Technology ,Electronics and Telecommunication department, Nerul, India

Abstract. The aim of this paper is to help farmers to identify disease without observing each and every mango plant by eyes. Farmers can use the pesticides according to the leaf disease identified. For identification of this leaf disease K-means clustering and GLCM (Gray level co-occurrence matrix) technique are used. Affected areas of the leaf are identified more accurately by using K-means clustering. GLCM is used to get the texture features which help for identifying the mango leaf disease. The signs of the diseases on the leaf should be identified at the early for the growth of productivity in mango plant, for this leaf disease identification is important.

1 INTRODUCTION

India is the largest producer of “mangoes” which means farmers require frequently monitoring of the large scale of mango crop field. To reduce the usage of pesticides, accurately, early diagnosis is necessary. It is also important to identify the proper disease. This will reduce the cost of buying pesticides and eliminate the use of pesticides. This paper concentrates on anthracnose disease found in mango plant. Diseases in mango plant are anthracnose, alternaria, and bacterial blight.

1. Anthracnose is most common serious diseases in mango plant. This fungal disease cause the development of dark, sunken spot or lesions. This disease is seed borne and carry over on crop residue in the soil. This fungal is spread in water droplets and worse in warm, humid weather.[1]

2. Alternaria leaf spot symptoms first appear as small, brownish circular spot on surface of the leaves. Later on, high concentration of brown, black spot occurs evenly over the mango leaf. This disease can be controlled by regular field spray program including copper based fungicides.[2]

3. Bacterial blight is also known as apical bud necrosis (mango). This disease affects most of the parts plant including the trunk, branches, shoots, buds, flowers, leaves and fruits. Regularly removing the infected twigs, branches and fruits, and disinfecting working tools will help in preventing the disease.[2]

In this paper for identifying the disease in mango plant MATLAB software is used. For better result the image is acquired from digital camera. For segmentation

K-means clustering is used and GLCM is used for the features extraction.

1.1 K-Means clustering

K-means is a centric based algorithm in which cluster size can be changed. Using k-means algorithm random two centres or pixels are selected from the infected leaf. The centres represent the faulty and faultless regions of the leaf. This is done to identify the infected part of the leaf. Clustering is a technique to form number of groups of objects. K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far as possible from objects in other clusters. [6]

1.2 Gray level co-occurrence matrix (GLCM)

GLCM techniques (gray level co-occurrence matrix) is used for obtaining texture features such as Homogeneity, Energy, Contrast, and Correlation. The statistic features in feature extraction obtained are standard deviation, kurtosis, variance, skewness, root mean square (RMS), entropy, mean, inverse difference matrix (IDM)[1]. The difference matrix element $P(i,j,d,\theta)$ contains the second order mathematical probability values for changes between gray levels ‘i’ and ‘j’ at a exacting displacement distance and at a particular angle(θ).

* Corresponding author: author@e-mail.org

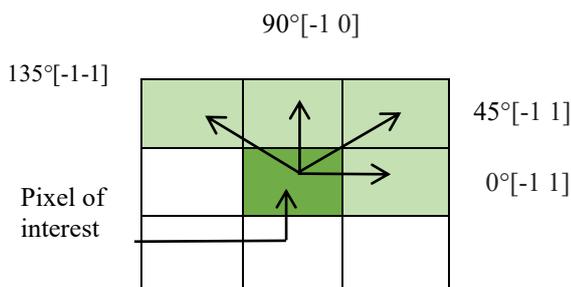


Figure 1. Gray Level Co-occurrence Matrix Angle

Typically GLCM are calculated at four different angles: 0, 45, 90, and 135 degree as shown in figure 1. For each angles different distances can be used $d=1, 2, 3$, etc., and the size of GLCM of 8 bit image: $256*256$ (2^8). 4 features can be calculated from each GLCM. The features are used for texture calculation. Statistical features are required for disease identification.

2 METHODOLOGY

The algorithm design is as shown in the figure 2. The steps taken at each phase are mentioned and they are briefly explained in the sections below. The whole algorithm design can be divided into four phases.

2.1. Image Acquisition:

Camera is used to acquire mango leaf image. Acquired image is RGB that is red, green and blue. The captured image is resize.

2.2 Image Pre-processing

Pre-processing method is used to enhance the input image of leaf for image conversion and filtration of noise.

2.2.1 Contrast enhancing

After capturing the image next step is to apply the blur soften method. Blurring of the image means each pixel of the image gets spread over. When we capture the image some noise is there so blurring the image used to remove some amount of noise. It will help to detect the affected area. Sharpening of the image can be reduced by using blurring of image. Limits to stretch contrast computes the lower and upper limits that can be used for contrast stretching grayscale or RGB image. The limits are returned in lowhigh. By default, the limits specify for bottom 1% and for top 1% of all pixel values.

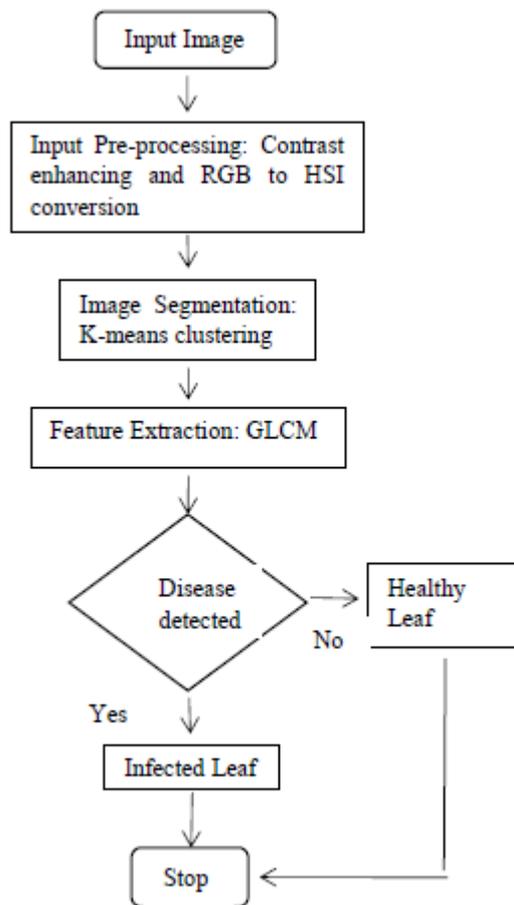


Figure 2. Algorithm of Leaf Disease Identification

2.2.2. Converting the image from RGB to HSI:

Conversion of RGB to HSI is helpful for the colour description. RGB stands the colors in the form of red, green, blue and HSI is the hue saturation intensity. RGB is group of primary colors.

To calculate the intensity average value of RGB is taken.

$$(1/3)*(R+G+B) \tag{1}$$

The colors are the unique features in this method which can be classified for translation of RGB to HSI.

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

$$H = \theta \text{ if } B < G \tag{3}$$

$$H = 360 - \theta \text{ if } B > G \tag{4}$$

$$S = 1 - (3/(R+G+B)) [\min(R, G, B)] \tag{5}$$

$$I = (1/3) (R+G+B) \tag{6}$$

The computation statistics are evaluated using GLCM function and later can be used in objective analysis of Image processing.

The infected portions of the leaves are classified based on the RGB images. The Otsu's threshold is observed and calculated based on the intensities and threshold values. The HSI model of the image is used in segmenting the portion of the image [2].

2.3 Image Segmentation:

Image segmentation is used for easier analysis of image. In image segmentation, K-means clustering is used for dividing the images into clusters. One part of cluster contain image which shows diseased part of the leaf. In this image is dividing into three clusters for better segmentation result.

2.3.1 Image segmentation using K-means clustering

K-means clustering is used for segmentation. The image is converted from RGB colour space to L*a*b* space. The L*a*b* space consists of a luminosity layer 'L*', chromaticity layer 'a*' and 'b*'. All the colour information is in the 'a*' and 'b*' layers.[6]

In this we are using various size of clusters like 3, 6, 8 and 10. Also we are using various co-occurrence angle for the clusters size like 45°, -45°, 30°, 60°, 90° and 180°. The better result of cluster size is getting in the 45°, 60°, 90° and 180° angle. 45° angle of cluster size 3 gives best results. Using this result of clusters we get the better result of GLCM.

2.4 Feature Extraction

2.4.1 Feature extraction using GLCM:

The method of feature selection as well as extraction is very important in this application. The disease can be identified and better result can be observed with respect to features. The detection of diseases are shown with the colour distribution in the image like hardness and roughness in the texture of the image. This shows the infected areas of the leaf.

In this we are using different co-occurrence angles. By changing the angles the cluster output and the GLCM output is changes. Different 13 features like mean, variance, standard deviation (SD), root mean square (RMS), smoothness, contrast, inverse difference movement (IDM) etc. are getting using the GLCM output. The affected area in percentage is shown in the command window. In this paper we are focused on the GLCM for the better result.

3 RESULT

Various leafs from the mango plants is collected. The leaf image is firstly converted into the HSI format. Using the resultant image different size of clusters are obtained. From the cluster image different features like mean, entropy, smoothness, contrast etc., are derived. The features are displayed using the GUI. Using this features leaf disease is detected. After analysis the various clusters of image are observed and whether the leaf is healthy or unhealthy is observed. The affected area is shown in the command window in percentage. In following cases the affected area is 15.0015%. The results of disease identification are shown in four cases.



Figure 3: Input image

The input image is taken from the digital camera is shown in figure 3. This image is mango plant leaf. This leaf has Anthracnose disease.



Figure 4. Contrast enhanced image of the input image

After taking input image the contrast of the image is enhanced as shown in figure 4. It simply highlight the features of the image.

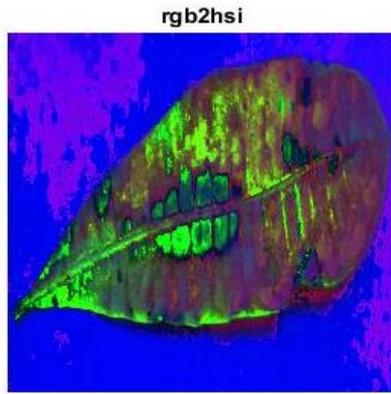


Figure 5. RGB to HSI of the input image

The RGB image is converted into the HSI image as shown in figure 5. HSI image is used in segmenting the portion of the image.

Case 1: The k-means cluster size 3 and co-occurrence angle 45° is taken for simulation.

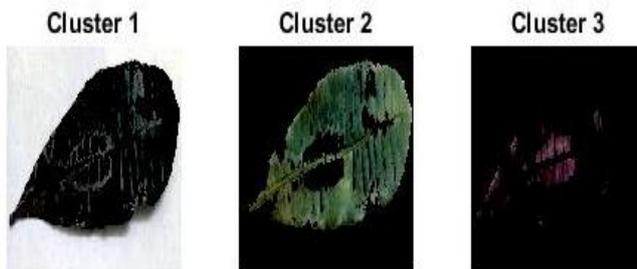


Figure 6. Cluster size 3 and 45° co-occurrence angle

The figure 6 shows the clusters of size 3. From this clusters output we have to select the cluster number which shows the smallest disease. As compared to input image in this the cluster 2 shows the smallest disease.

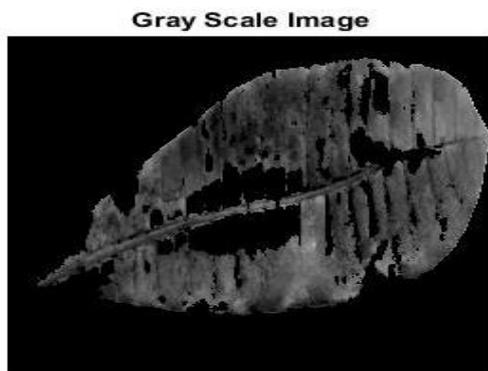


Figure 7. Gray scale image using cluster 2 and 45° co-occurrence angle

From figure 6 the cluster 2 image is transformed into the Gray scale image as shown in the figure 7. In which the disease part is more highlighted using the gray scale image and that image is used for the feature extraction using GLCM.

Case 2: The cluster size 3 and co-occurrence angle 60° using k-means clustering method.



Figure 8. Cluster size 3 and 60° co-occurrence angle

The figure 8 shows the clusters of size 3. As compared to input image cluster 2 shows the smallest disease.

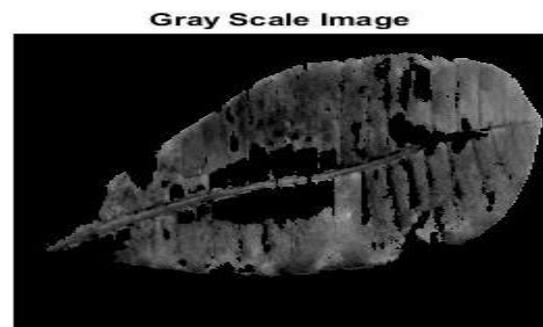


Figure 9. Gray scale image using cluster 2 and 60° co-occurrence angle

From the figure 8 the cluster 2 image is transformed into the Gray scale image as shown in the figure 9. In which the disease part is more highlighted using the gray scale.

Case 3: The cluster size 3 and co-occurrence angle 90° using k-means clustering method.

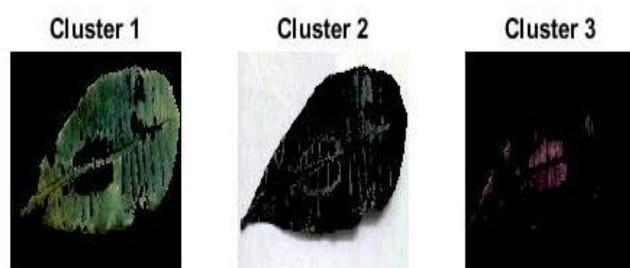


Figure 10. Cluster size 3 and 90° co-occurrence angle

Figure 10 shows the clusters of size 3. As compared to figure 3, the cluster 1 shows the smallest disease.

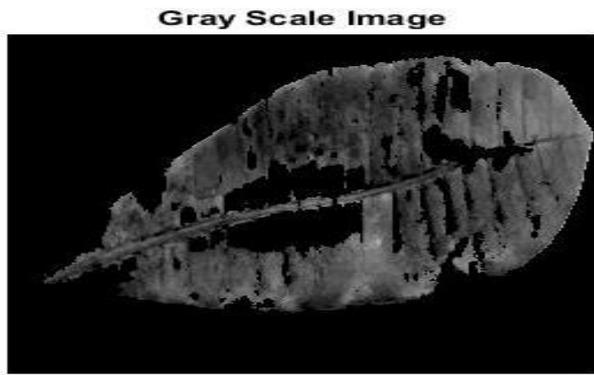


Figure 11. Gray scale image using cluster 1 and 90°co-occurrence angle

From figure 10, the cluster 1 image is transformed into the Gray scale image as shown in the figure 11. In which the disease part is more highlighted using the gray scale.

Case 4: The cluster size 3 of co-occurrence angle 180° using k-means clustering method.



Figure 12. Cluster size 3 and 180 °co-occurrence angle

The figure 12 shows the clusters of size 3. As compared to input image in this the cluster 1 shows the smallest disease.

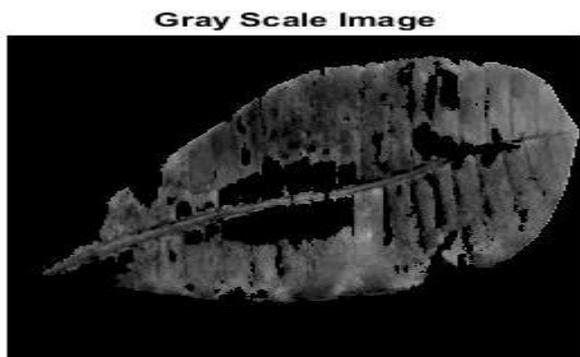


Figure 13. Gray scale image using cluster 1 and 180°co-occurrence angle

From the figure 12 the cluster 1 image is transformed into the Gray scale image as shown in the figure 13. In which the disease part is more highlighted using the gray scale.

Case 1 gives best results.



Figure 14. Segmented image from cluster 2 and extracted features using GLCM

After selecting the cluster 2 of co-occurrence angle 45° the segmented image is shown in GUI window as shown in figure 14. The features are extracted using gray level co-occurrence matrix method are shown in window as shown in figure 14. Using this features we classified the type of disease and accuracy.

In following table 1 the feature extraction formulas are given, using that formulas the values of the 13 features are calculated.

Table 1: Features Extraction Formulas and their values from Cluster size 3 of cluster 2 for co-occurrence matrix

No	Features	Formula	values
1	Contrast	$\sum n \sum m n-m 2 p(n, m, d, \theta)$	0.29185
2	Correlation	$\sum_{n,m} \frac{(n-\mu)(m-\mu)p(n,m)}{n\sigma_n \sigma_m}$	0.86555 6
3	Energy	$\sum n \sum m p(n, m, d, \theta)^2$	0.41681 5
4	Homogeneity	$\sum n \sum m \frac{p(n,m,d,\theta)}{1+ n-m }$	0.92773 5
5	Mean	$\frac{1}{M} \sum Y_i$	24.4796
6	Standard Deviation	$\frac{1}{M} \sum (Y_i - \mu)^2$	37.4463
7	Entropy	$-\sum (p_i \cdot \log_2(p_i))$	3.32329
8	RMS	$\sqrt{(Y_1)^2 + (Y_2)^2 + \dots + (Y_n)^2} / n$	18.72952
9	Variance	$\frac{1}{M} \sum (Y_i - \mu)^2$	1167.84
10	Kurtosis	$\frac{\frac{1}{N} \sum_{j=1}^N (Y_j - Y)^4}{(\frac{1}{N} \sum_{j=1}^N (Y_j - Y)^2)^2}$	3.38699
11	Skewness	$\frac{\frac{1}{M} \sum_j^M = (Y_j - Y)^3}{(\frac{1}{M} \sum_{j=1}^M (Y_j - Y)^2)^{3/2}}$	1.26074
12	IDM	$\sum n \sum m \frac{1}{1 + (n - m)^2} p(n, m)$	255
13	Smoothness	$1 - (1/(1+a))$	1

4 Conclusion

By K means clustering and GLCM technique the disease is identified at its initial stage. Thus pest control tools can be used to work out the problem. This paper is focused on using Co-occurrence method for getting accurate image of clusters and for the feature extraction. In this paper we are using altered size of clusters like 3, 6, 8 and 10. In cluster size 3 we get better results as compared to the other size of clusters. In the cluster size 3 disease is identified. Also we are using different theta angles like 45°, -45°, 30°, 60°, 90°, and 180°. The better results are obtained in 45°, 60°, 90°, and 180°. In this paper we are using co-occurrence angle 45° and cluster size 3. In GLCM output, the part of the disease is highlighted. It only detect the infected area in the leaf.

References

1. G. S.Tumang ,Pests and Diseases Identification in Mango using MATLAB , IEEE, (2019)
2. S.A.Hussain,R.Hasan,S. J. Hussain ,Classification and Detection of Plant Disease using Feature Extraction Methods,IJAER ,pp.4219-4226, **13**,Number 6,(2018).
3. G. Bhogade, R. Mawal, A. Patil, S. R. Desai,Leaf Disease Identification using Image Processing, IJESC, **7**, Issue Number 5,(2017)
4. J. Sethupathy, Veni S., OpenCV Based Disease Identification of Mango Leaves, IJET,Oct-,Volume8,Number5, DOI:10.21817/ijet/2016/v8i5/160805417(2016).
5. Saradhambal.G, Dhivya.R, Latha.S., ,R.Rajesh, Plant Disease Detection And Solution Using Image Classification, International Journal of Pune and Applied Mathematics,**119**,Number 14,pp879-884(2018)
6. S.Raut, A. Fulsunge ,Plant Disease Detection in Image Processing Using MATLAB,IJRSET,**6**,Issue Number 6, DOI:10.156 80/IJRSET.2017.0606034(2017)