HEB

Detection and Classification of Diseases on Guava Fruit Using Image Processing

CASS

Dr.S.Abirami[1], M.Thilagavathi[2]

[1]Assistant Professor: Department of Computer Science and Engineering, Annamalai University, Chidambaram, India

[2]Research Scholar: Department of Computer and Information Science, Annamalai University Chidambaram, India

Address for Correspondence: serviceheb@gmail.com

ABSTRACT

Fruits are a good source of dietary fibre. Fruits contain biologically active substances that protect us from diseases. This paper focuses on the identification and classification of guava fruit diseases. Image processing based approach is used in detection of diseased region of the guava fruit using K-Means clustering segmentation algorithm. The color and the texture features of the fruit images are extracted using HSV color space and Gray-Level Co-occurrence Matrix (GLCM) feature extraction method. Three classes of diseases such as anthracnose, canker and endrot are considered for our study. The classification is done using Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN) classifiers. The experimental result is obtained with an accuracy of more than 90 % with SVM classification.

Keywords—component; formatting; style; styling; insert.

Access this Article Online	
http://heb-nic.in/cass-studies	**Quick Response Code:**
Received on 10/05/2019 Accepted on 29/05/2019@HEB All rights reserved	

I. INTRODUCTION

Fruits are important part of our daily life. They are good in vitamins and minerals that can help in keeping us healthy. Guava is a humble fruit that is rich in vitamin C and several minerals. Guava is beneficiary in lowering the risk of cancer, diabetes friendly, keeps our heart healthy, improves eyesight, good for brain, helps in weight loss, treats cough and cold and acts an immunity booster. Diseases on fruits cause major losses in yield and quality. The classical approach in identification of diseases on fruits is based on the naked eye observation by the experts which is time consuming and expensive. The automatic detection of diseases through machine vision is essential in improving the growth of agriculture in the developing countries.

In [1] the several methodologies utilized for apple fruit disease detection was surveyed. The merits and demerits of region based approaches, edge detection approaches clustering methods were illustrated in a table. The feature extraction methods such as GLCM, Local Binary Pattern, Gabor Filter, Support Vector Machine and k-Nearest Neighbor approaches were surveyed. (Bhavini J.Samajpati et al. 2015). In [2] clustering and color based techniques were analyzed in detection and classification of fruit diseases. The authors concluded that the K-means clustering and SVM provided high accuracy and was widely used in many research works.

The authors of [3] proposed K-means clustering technique for segmentation of diseased part of the fruit image. The images were classified using Support vector machine. The authors of [4] segmented the infected fruit with K-means clustering process. The pixels are clustered, based on their color and spatial features. Then the clustered blocks are merged to a specific number of regions. By doing these steps the computational efficiency was increased and the feature extraction for every pixel can be avoided.

The performance evaluation of K-Means and C-Means clustering algorithms were done by measuring the parameters such as Measure of overlapping, Measure of over segmentation, measure of under segmentation, Dice similarity measure and Error rate. The features are extracted using GLCM and the k-Nearest Neighbor classifier was used to classify the type of diseases on fruit images (Ranjit K N et al. 2016).

Using the proposed work [7] the type and cause of disease on fruits can be easily detected and preventive measures and suggestions can be taken from the system. The authors of [8] presented a solution for the three common diseases such as apple rot, apple scab and apple blotch diseases affecting apple fruit. K-means clustering approach was implemented for segmentation and the disease wise classification was done using Learning Vector Quantization Neural Network yielding more than 95% of classification accuracy.

Plant fruit diseases are crucial causes that reduce the quantity and can degrade the quality of agricultural products. Image processing plays a vital role in detecting the degradation in plants and fruits (Zalak R. Barot et al. 2015). The segmentation of the fruit image was done using K-Means and C-Means approach. The performance evaluation was done between the techniques and higher accuracy of classification was obtained through K-Means approach (Y.H. Sharath Kumar et al. 2016). The interesting region was segmented from the background using K-Means algorithm. First, the partial stretching enhancement was applied to the image to improve the quality of the image. Subtractive clustering method generated the centroid based on the potential value of the data points. The initial centers generated by this approach were used in K-means algorithm. (Nameirakpam Dhanachandra et al. 2015).

II. PROPOSED METHODOLOGY

The steps to be followed in the proposed approach is given in Fig. 1

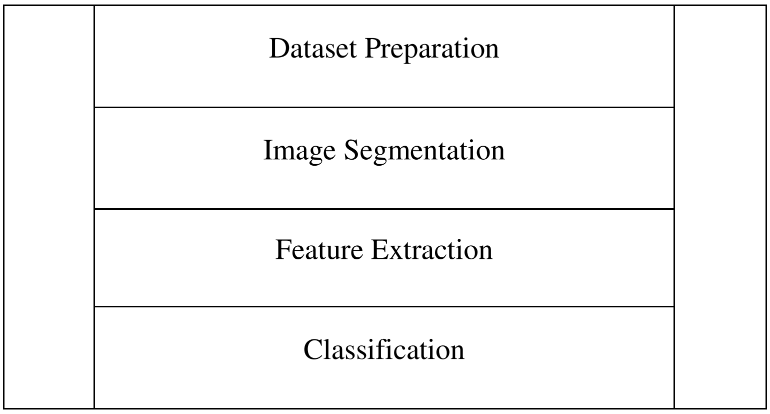

Fig. 1 Basic steps of the proposed approach

A. *Dataset Preparation*

The RGB images of disease affected guava fruits are captured through the camera with high resolution. Totally 60 images are captured with 20 per class and is resized in to 256 X 256. The obtained data is stored as a database for further processing.

B. *K-Means Clustering Segmentation*

Image segmentation is the division of image in to several groups. K-Means clustering algorithm is an unsupervised algorithm that segments the interest area from the background. The image is divided in to discrete regions according to the high similarity of pixel values in each region and high contrast between regions. K-Means clustering is simple and computationally fast when compared to other clustering techniques.

K-Means clustering partitions the data in to a number of groups. It classifies a given set of data in to k number of disjoint clusters. K-means algorithm consists of two separate phases. In the first phase the k centroid is calculated and in the second phase it takes each point to the cluster which has nearest centroid from the respective data point. Euclidean distance method is used to define the distance of the nearest centroid. Once the grouping is done, the new centroid is recalculated for each cluster based on that centroid. Euclidean distance is calculated between each center and each data point and assigns the points in the cluster which have minimum Euclidean distance. Each cluster in the partition is defined by member objects and by its centroid. The centroid for each cluster is the point to which the sum of distances from all the objects in that cluster is minimized. Hence k-means is an iterative algorithm. In the proposed work the input image is segmented in to three clusters as shown in Fig. 2

Fig. 2 Segmented Image of Anthracnose Disease Guava Fruit using K-Means Clustering

The infected region of the segmented image is cropped in to a size of 75 X 75. The region of interest values are extracted from the images as feature vector and are used for classification.

C. Feature Extraction

Feature extraction begins with an initial set of measured data and builds derived features intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps. It is a dimensionality reduction process, where an initial set of raw variables is reduced to more manageable groups.

a) HSV Color Feature Extraction

The color features of the input RGB images are obtained as follows:

Step 1: Read the input RGB image.

Step 2: Convert the RGB image in to HSV (Hue, Saturation and Value) image.

Step 3: Obtain the statistical values such as mean and standard deviation from the H, S and V planes separately.

Step 4: Store the collected features as a dataset.

The RGB to HSV converted image is shown in the Fig. 3.

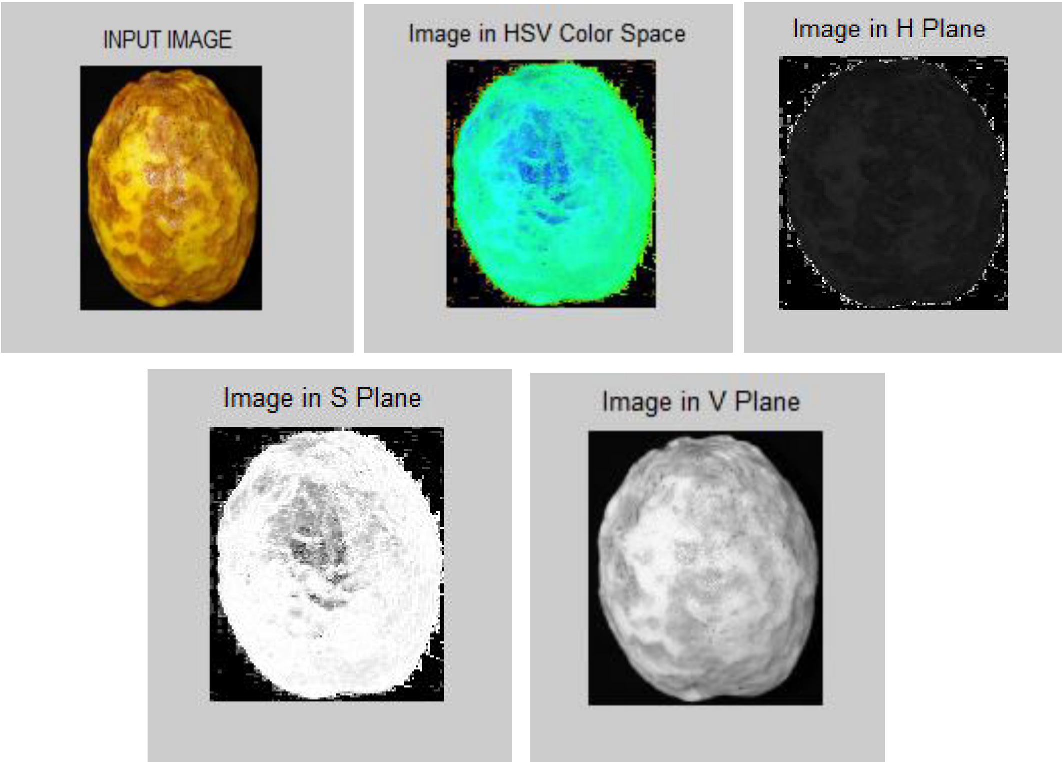

Fig. 3 RGB image converted in to HSV Color Space

b) GLCM Feature Extraction

An image texture is a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Texture gives us the information about the spatial arrangement of selected region of the image. The texture features of the input image are calculated using the Gray-Level Co-Occurrence Matrix (GLCM) feature extraction method. In GLCM, the texture is examined by considering the spatial relationship of pixels. The statistical features such as Contrast, Correlation, Energy and Homogeneity are obtained from this method.

GLCM is a matrix in which the number of rows and columns are equal to the number of gray levels, G of the image. The matrix element $P(i,j|\Delta x,\Delta y)$ is the relative frequency with which two pixels separated by a pixel distance $(\Delta x,\Delta y)$ occur within a given neighborhood, one with intensity 'i' and other with intensity 'j'. The GLCM features are calculated as given below.

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \quad (1)$$

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij}\frac{(i-\mu)(j-\mu)}{\sigma^2} \quad (2)$$

$$Energy = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2 \quad (3)$$

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \quad (4)$$

The color and texture features computed are combined as a feature vector and in fed in to the classifier to classify the type of disease affected on the guava fruit.

D. Classification

Classification is done on the basis of the training set of data containing observations. The classifier maps the input data to a category, thus defining the disease class of the given input image.

a) Support Vector Machine (SVM)

SVM is a supervised learning technique. The main objective of SVM is to define a hyperplane in the N-dimensional feature space that distinctly classifies the data points. The hyperplane having the maximum distance is to be chosen, since the classification can be done with more confidence among the data points. Hyperplanes are the decision boundaries that classify the data points falling on either side of it to different classes.

The classification task in SVM involves with the training and testing data consists of some data instances. Each instance in the training set contains one target value and several attributes. The testing set is given only attributes. The goal of SVM is to produce a model which predicts the target value of data instances in the testing set.

b) k-Nearest Neighbor (k-NN)

k-NN is a non parametric learning algorithm. The data points are separated in to several classes to predict the classification of a new sample point. Data points in the feature space are classified based on the similarity measures. The closeness of the sample data resemble in the training set determines the classification of the given data point.

A positive integer k is specified along with a new sample. K entries in the database which are closest to the new sample are selected. The most common classification of these entries are identified and these values are assigned to the new sample.

III. EXPERIMENTATION

During experimentation we considered 60 images of guava affected with anthracnose, canker and endrot diseases, 20 for each. The affected region of the given input is easily identified through segmentation process. K-Means clustering segmentation algorithm is applied to divide the input image in to three clusters. The illustration of the fruit image in each cluster is shown as in Fig. 4.

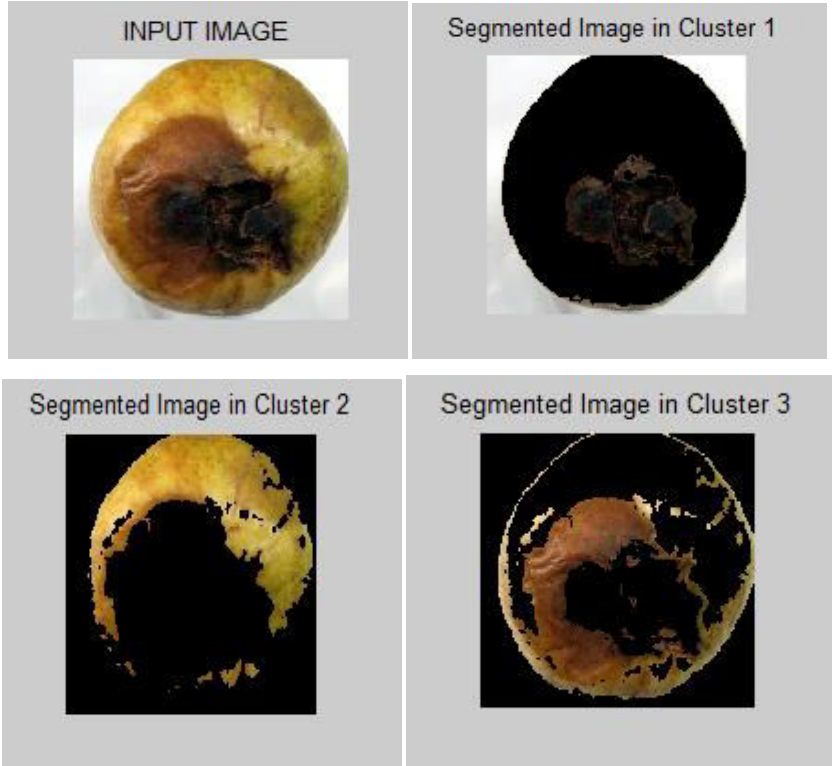

Fig. 4 Result of K-Means clustering segmentation image of endrot disease

The color features are extracted from the segmented image by calculating the statistical values of the H, S, and V planes individually. Table 1 shows the sample data points obtained from the HSV color space.

Table 1 Sample values obtained from HSV planes

Samples	Mean of H	Mean of S	Mean of V	Standard Deviation of H	Standard Deviation of S	Standard Deviation of V
Sample1	0.1021	0.6275	0.4670	0.0885	0.3624	0.3024
Sample2	0.0779	0.4862	0.4355	0.0878	0.4052	0.3869
Sample 3	0.2092	0.5171	0.6176	0.2523	0.3270	0.2370
Sample 4	0.4684	0.2789	0.6317	0.2344	0.2236	0.2201
Sample 5	0.4916	0.2873	0.7082	0.2375	0.1465	0.2512

GLCM texture feature extraction method is implemented on the image to get the parameters as shown in

Table 2

Table 2 Data values obtain from GLCM

Samples	Contrast	Correlation	Energy	Homogeneity
Sample 1	0.0584	0.9910	0.2078	0.9710
Sample 2	0.0593	0.9941	0.2481	0.9716
Sample 3	0.1363	0.9850	0.1005	0.9328
Sample 4	0.0573	0.9903	0.1888	0.9713
Sample 5	0.0448	0.9926	0.2357	0.9776

The classification takes place with 40 training samples and 20 test samples. The performance comparison of the SVM and k-NN classifiers is given in the Table 3.

Table 3 Performance comparison of SVM and k-NN Classifiers

Type of Disease	SVM Accuracy (%)	k-NN Accuracy (%)
Anthracnose	92	85
Canker	90	76
Endrot	94	68

CONCLUSION

In this work, an automated approach is identified for detection and classification of guava fruit disease with the following steps. First segmentation is done to identify the affected area of the fruit image with K-means clustering technique and then the color and texture features are extracted for implementing them with the SVM and k-NN classifiers. The proposed approach results with a greater accuracy of more than 90 % with SVM which is superior to k-NN.

REFERENCES

[1] Bhavini J. Samajpati and Sheshang D. Degadwala, A Survey on Apple Fruit Diseases Detection and Classification, *International Journal of Computer Applications*, Vol. 130, November 2015, pp. 25-32.

[2] Sandesh Raut and Amit Fulsunge, Review on Fruit Disease Using Image Processing Techniques, *International Journal of Innovative and Emerging Research in Engineering*, Vol. 4, 2017, pp. 22-24.

[3] Sulakshana A. Gaikwadi, Kalyani S.Deore, Monali K. Waykar, Priyanka R.Dudhane and Geeta Sorate, Fruit Disease Detection and Classification, *International Research Journal of Engineering and Technology*, Vol. 04, December 2017, pp. 1151-1154.

[4] Shiv Ram Dubey, Pushkar Dixit, Nishant Singh and Jay Prakash Gupta, Infected Fruit Part Detection using K-Means Clustering Segmentation Technique, *International Journal of Artificial Intelligence and Interactive Multimedia*, Vol. 2, 2013, pp 65-72.

[5] Ranjith K N, Chethan H K and Naveena C, Identification and Classification of Fruit Diseases, *International Journal of Engineering Research and Application*, Vol. 6, July 2016, pp. 11-14.

[6] Khaled Arafat, A Novel Isolate of Phyllosticta capitalensis Causes Black Spot Disease on Guava Fruit in Egypt, *Asian Journal of Plant Pathology*, Vol. 12, 2018, pp. 27-37.

[7] Sherlin Varughese, Nayana Shinde, Swapnali Yadav and Jignesh Sisodia, Learning-Based Fruit Disease Detection Using Image Processing, *International Journal of Innovative and Emerging Research in Engineering*, Vol. 3 2016, pp. 96-100.

[8] Abhijeet V.Jamdar and A.P.Patil, Detection and Classification of Apple Fruit Diseases using K-means clustering and Learning Vector Quantization Neural Network, *International Journal of Scientific Development and Research*, Vol. 2, June 2017, pp. 423-429.

[9] Zalak R. Barot, Narendrasingh Limbad, An Approach for Detection and Classification of Fruit Disease: A Survey, *International Journal of Science and Research*, Vol. 4, December 2015, pp. 838-842.

[10] Y.H.Sharath Kumar and G.Suhas, Identification and Classification of Fruit Diseases, *International Conference on Recent Trends in Image Processing & Pattern Recognition*, 2016, pp. 382-390.

[11] Nameirakpam Dhanachandra, Khumanthem Manglem and Yambem Jina Chanu, Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm, *Procedia Computer Science*, 2015, pp. 764-771.

[12] P.Mohanaiah, P.Sathyanarayana and L.Gurukumar, Image Texture Feature Extraction Using GLCM Approach, *International Journal of Scientific and Research Publications*, Vol. 3 May 2013, pp. 1-5.