IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

# Identification of Cucumber Leaf Disease using Image Processing Techniques

**Shrutika.C.Rampure[1*], Vindhya .P. Malagi[2], Ramesh Babu D.R[3]**

[1,2,3]Dept. of CSE, Dayananda Sagar College of Engineering, Bangalore, India

*Corresponding Author: Shrutika235@gmail.com

**Available online at: www.ijcseonline.org**

Accepted: 20/Jun/2018, Published: 30/Jun/2018

***Abstract***— Agriculture is the backbone of Indian economy. Plant disease which mainly affects the leaves is the major constraining factor, which decreases the productivity of cucumber. Farmers are experiencing heavy loss in the yield due to disease attack on leaves. Hence detection and diagnosis of cumber leaf disease at the right time are very essential. Diagnosis of cucumber leaf disease at the early stage helps in preventing heavy loss in the yield. Automatic detection of cucumber disease using image processing techniques helps in monitoring large fields by identifying the diseases as soon as they appear on the leaf. The main purpose of this work is disease identification and classification using image processing techniques. The proposed method mainly comprises of image pre-processing, segmentation using K means clustering to segment the diseased leaf then feature extraction and followed by classification of disease using SRC. The experimental results show that the cumber leaf diseases can be identified more accurately for the proposed work.

***Keywords***—Cucumber leaf disease, K-means Clustering, Sparse representation Classification (SRC)

## I. INTRODUCTION

Agriculture is one of the prime sectors of Indian economy. It is the major source of income where people earn their living directly or indirectly through agriculture. In recent years agricultural issues have become much more critical and severe. Agricultural production is getting reduced due to various factor. One of the key element that contributes major loss is disease attack. The existence of the disease on the plants may result in the consequential loss in quality of agricultural production [1]. When no proper control measures are undertaken then plant disease become one of the major bottlenecks in agricultural production. India has emerged the second largest producer of vegetables. Cucumber is one among them which plays a significant role in our daily meals. Cucumber has originated in India but is now grown in most of the countries. Cucumber (Cucumis sativus L.) belongs to Cucurbitaceae family is an essential greenhouse summer vegetable crop cultivated in India. However, the cucumbers often suffer from pests and diseases that mainly affect leaves which decreases the quality of cucumber production. The symptoms mainly appear on leaf, stem, and fruit. The leaf shows the symptoms in the form of spots or by changing color. Traditional methods such as manual identification using naked eye observation is one of the major approach (adopted in practice) that is used for diagnosing cucumber diseases are slow which requires continuous monitoring for the large field, which may consume more time and require expert advice for identification of disease which may prove to be costly and also the disease recognized by human eye look similar and identical in appearance. So the diagnosis of disease on cucumber leaf disease at the early stage is necessary in order to prevent heavy loss in yield which in turn results in loss of both quality and quantity of agricultural production. So the main aim of this work is to identify and classify the cucumber leaf disease accurately.

The rest of this paper is organized as follows. Section 2 describes the related work. Section 3 describes the methodology. Results and analyses are given in Section 4. Section 5 gives the details about conclusion and future work on improving the cucumber disease recognition algorithm.

## II. RELATED WORK

The symptoms of the plant disease often appear on leaves which are identified by image processing techniques. Author has used local minimum are located the threshold cut-off value is determined from the histogram. This method is useful only when the region of interest is with the wide range of intensity distribution. But the next stage of development could be classifying the diseases region [2]. In order to control disease spread on the leaf, early detection of disease is necessary. The authors describe that for detecting the type of disease various features needs to be extracted. Color features which is one of the most widely used features are obtained by using color moment which can be used for indexing the image based on color and it can be further used for classification various diseases of the leaves using clustering techniques [3]. The disease of cucumber downy mildew, powdery mildew, and anthracnose leaf image are mainly studied. Disease identification was done based on minimum distance method. The proposed method is feasible with the recognition rate of more than 96%. Author has considered only 25 samples of each cucumber disease which could be further increased [4]. Image enhancement is the essential step in digital image processing. Author has mainly focused on how the image is enhanced using adaptive histogram equalization (AHE). The experimental results show that result of AHE has high PSNR and low value of mean square error when compared to traditional histogram equalization [5]. Author has mainly described different local and global contrast enhancement techniques. Histogram Equalization (HE) is not suitable because it changes the brightness and degrades the image. Brightness preserving global techniques can be applied but cannot enhance the local contrast of the image. AHE improves the local contrast of the image but increases the noise. Extended to this CLAHE suppress the noise and improves the brightness to the specific range [6]. The author has mainly demonstrated about the analysing the image and the classification techniques used to identify the olive leaf spot disease using fuzzy c means clustering techniques. Here the RGB was converted to LAB color space to segment the diseased region. Author has done a comparative study on FCM and KMC and the results showed that the average speed for FCM was 6.96 second for 100 images and for K means it was 4.23 second which is comparatively less than FCM [7]. The author has demonstrated the clustering techniques used for analysing information. The proposed system compares the performance of K-means clustering and Fuzzy C-means clustering algorithms. The experimental results show that the maximum number of times k means clustering produces better, faster and clear segmented results when compared to FCM. Hence K- means clustering is the better choice when compared to FCM [8]. Principal Component Analysis (PCA) is applied for feature selection which selects important characteristics from feature component. It then performs feature selection by evaluating the significance of feature component. The experimental results show that PCA is able to considerably reduce the dimensionality of the original image which improves the performance of the classification [9]. The author mainly estimated the application of PCA on a high-resolution image to reduce the dimensionality and the reduced feature images are compared with different variance value. The experimental results showed that the PCA reduces the dimensionality of the digital image which saves the storage while maintaining the principal integrity of image [10]. Here the main focus is to address the additional problem of how to identify the basis vector for representing the data. The major focus is the judicious choice of a dictionary: representation of the test signal should be a sparse linear combination of the training signals. If the identity of the test sample is not known then the problem becomes challenging [11]. A novel approach to automatic species identification using SR is proposed for leaf tooth features. The proposed methodology was feasible for plant species identification since sparse representation classifier helps in identifying the plant species sample. As a future work more complex features of plant leaf tooth need to be studied including shape, color and texture features should be incorporated into SR based plant species identification [12].

In spite of various efforts, the existing system still has various limitations. The extracted features are generally susceptible to orientation, luminescence, and scaling. The step prior to feature extraction is required to describe translation, rotation and scaling factor. Features selected are normally regarded as essential for classification regardless of their actual role. Some of the features may have little importance in recognizing the disease thereby decreasing the performance. Assigning equal importance to all the features results in classification error. All these limitations have motivated to develop an algorithm for segmenting the lesion part of the cucumber leaf.

## III. METHODOLOGY

In this section, the major steps required for cucumber leaf disease identification using image processing techniques have been described as shown below (Fig. 1). There are five modules and nine sub-modules involved in the proposed method.

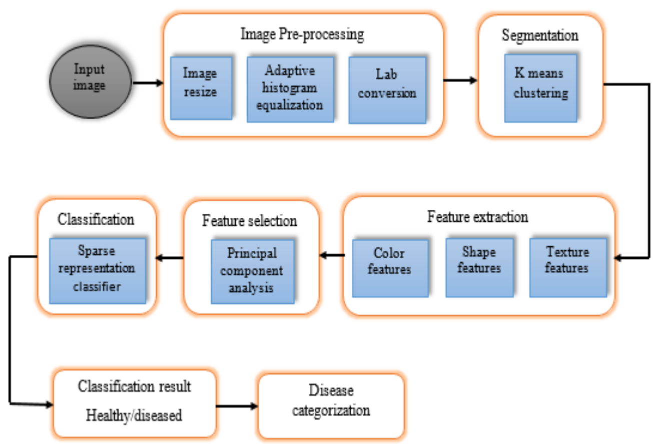



Fig 1. Basic steps to detect the cucumber leaf disease

### A. *Image Preprocessing*

Pre-processing is done to enhance the quality of the image, which includes image resizing, image enhancement, color conversion and histogram equalization. Image resize is done to perform matrix multiplication. Image enhancement is done by using adaptive histogram equalization. Color conversion is done by converting input image from RGB Color Space to L*a*b* Color Space.

Adaptive histogram equalization (AHE):
AHE is one of basic local histogram equalization technique which is used to improve the contrast of the image to obtain a clear view of the image by normalizing intensity value redistribution. It partitions the image into various non-overlapping regions and performs histogram equalization on each individual region. Later the final image is produced by combining the region using bilinear interpolation [13]. Since histogram equalization operates on a global contrast of the image which may also increase the contrast of the background noise. So in order to overcome the disadvantages of the ordinary histogram equalization adaptive histogram equalization is introduced which uses multiple local window size to improve the local contrast of the image in preference to improve the overall contrast of the image.


Figure 2: Image enhancement using AHE

Lab conversion:
It is device independent color model. RGB color space is one the most common color model. Lab conversion was mainly introduced since the lightness and the color components do not depend on each other. It was designed in such a way that the L component closely matches the human vision of perception [14]. In Lab color 'L' stands for the lightness that ranges from black to white, 'a' and 'b' are the color components that range from red to green axis and blue to a yellow axis. The color conversion is done by converting an RGB image to XYZ using the equation:

$$X = 0.4124 * R + 0.3576 * G + 0.1805 * B \tag{1}$$
$$Y = 0.4124 * R + 0.7152 * G + 0.0722 * B \tag{2}$$
$$Z = 0.0193 * R + 0.1192 * G + 0.9505 * B \tag{3}$$

Later XYZ is converted to Lab using the equation:

$$L = 0.2126 * R + 0.7152 * G + 0.0722 * B \tag{4}$$
$$A = 1.4749 * (0.2213 * R - 0.3390 * G + 0.1177 * B) + 128 \tag{5}$$
$$B = 0.6245 * (0.1949 * R + 0.6057 * G - 0.8006 * B) + 128 \tag{6}$$

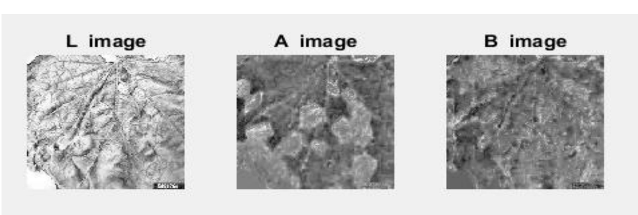



Figure 3: Color conversion from RGB to LAB color space

*B. Segmentation*

Segmentation means identifying the region of interest from the image. Segmentation is broadly classified into two approaches. Discontinuity: Here the image is segmented based on the change in pixel value or abrupt change in intensity value. The algorithm includes edge detection techniques. Similarity: Here the image is segmented based on similarity existing in the region. Algorithms include thresholding, region growing, clustering techniques. Clustering is one of the most common unsupervised machine learning approaches where the pixels are classified based on the properties that such that cluster should have minimum interclass similarity and maximum intra class similarity. Here in our approach K means clustering algorithm is used to perform segmentation

K means clustering:
K means is one of the most popular clustering technique used for segmentation. K means is unsupervised learning which initially requires the 'k' to be specified (here k=4) i.e., the number of clusters to be partitioned. It is an iterative process where the clusters are formed by classifying the pixels based on the set of feature value into k classes or groups. Selecting the CH based on the intensity value having highest count. If there are two CH with same intensity value then take the average of two values and select the average value as the new CH otherwise continue with comparing all the pixel values in the image with corresponding CH and assigning the pixel to a particular cluster based on the minimum distance criterion. Suppose there are two CH's having minimum distance then it is not possible to take both as CH since intra cluster distance should be maximum, so taking an average of two CH's and select average value as new CH. Then after assigning the pixel value to the corresponding cluster recomputed the CH by taking the average of the pixel value and the pixel with the highest value is selected as new CH. Repeat this process until there is no change in the CH selection and the cluster remain unchanged. The segmented results are shown in figure 4 and figure 5.

Following are the steps for K-means clustering algorithms:
Step 1: Initialize the number of the cluster here k=4
Step 2: Selection of cluster head (CH) is based on the pixel having highest intensity count value.
Step 3: Assigning the pixel to closest CH using minimum distance criteria and update the CH by taking the average of all the pixel.
Step 4: Repeat step 2 and 3 until there is no change in the CH

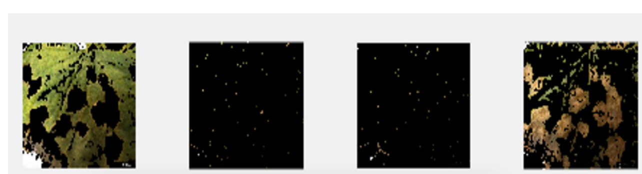


Figure 4: Segmentation using K-means clustering

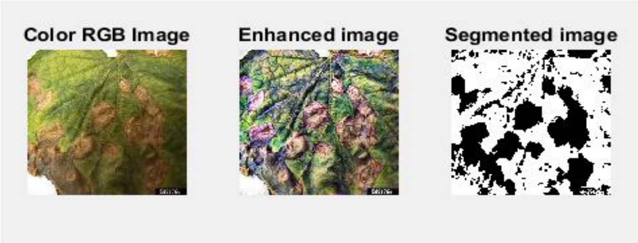



Figure 5: Segmentation using K-means clustering

### C. Feature Extraction

After segmenting the particular region of interest, features are extracted for the diseased region. Feature to be extracted carries maximum information about the image. In order to reduce the computation time during classification transform the input data into a reduced set of representation named as feature vector. Feature extraction plays a vital role which will have major impact for classifying the disease accurately. Color, texture and shape feature are commonly used image features which are extracted for the proposed work.

Color feature extraction:
Color feature are one of the most significant feature which helps forms basis of discrimination when identifying the leaf disease by extracting the significant attributes from color image. When compared to other features color feature are very stable Used color moments for extracting the color features. Color moments is very efficient method to extract the color features for identifying the cucumber leaf disease [15].

Texture features extraction:
Texture is one of the most important property of the leaf image. It generally helps in segmenting the diseased cucumber leaf image into region of interest and also classify the region accordingly. Texture feature gives information about the spatial arrangements of the gray level and intensity of the image. Texture features are usually evaluated for the lesion based segmented cucumber leaf image. Gray level co-occurrence matrix (GLCM) is widely adopted to extract the texture features.

Shape feature extraction:
Shape is the basic visual feature used for image content description. Shape cannot be defined exactly because estimating the similarity between shapes is difficult. Shape features are more stable when compared to color features and better than the texture features. Only by using shape feature extracted for the lesion part high recognition rate cannot be achieved. By combining all the three feature by transforming each color lesion image into gray scale and reducing it by Singular value decomposition (SVD).

### D. Feature selection

Feature selection plays a very important role for reducing the computational workload. Feature selection helps in reducing the dimensions of the feature vector without losing any useful information. Considering all the features for classification without knowing its importance tend to result in classification error. So it is necessary to consider only the prominent features in order to reduce the computational complexity and at the same time maintain reasonable classification performance. Feature selection is done using PCA (Principal Component Analysis).

Principal component analysis:
It is a dimensionality reduction technique which is used to reduce the dimensionality of the feature set where in the variables are correlated with each other by maintaining the variation in feature set up to greater extent. As a layman it can be used for summarizing the information present in the dataset such that the first component in the feature set have highest variance when compared to second, three and so on in the decreasing order. The main goal of PCA is to retain the maximum variance with very few number of principal components.

### E. Classification

Classification means assigning every pixel in the image to different categories or classes. Classification mainly involves two-phase training phase and testing phase. In training phase first, identify the specific features of the image and then isolate those features, based on all these unique descriptions classify the classes into different categories to form a training class. In the testing phase, the feature set categorization is used to classify the image. Here the classification is performed by using sparse representation classifier.

Sparse representation classification:
Sparse representation is pixel-wise classification technique which is widely used for image analysis. Sparse representation is a technique of finding a matrix with very few non zero elements is called sparse representation. The sparse representation classifier uses the learned dictionaries to classify the pixels in the image. First, train the dictionary for each of the classes in a supervised manner. Entire training image data set cannot be used as a dictionary because of large computational cost. So first obtain the feature vector for the entire training set using the extracted feature vector to build a dictionary.

There are two approaches in a sparse representation. First one is sparse coding where learning the sparse coefficients of a test sample from a given basic matrix. The second one is dictionary learning where learning the basic vector from a given training data.

Consider a dictionary of training samples A = [A1, A2, A3,……Ak ] where Ai = [ ai1, ai2,ai3,……ain] for the i-th class and j-th training sample aij , k is the number of classes, ni is the number of training sample for the i-th class and n is the total number of training samples. The test sample y from i-th class lies in the linear span of a training sample is described as

$$y \approx \alpha_{i1},\alpha_{i2}\alpha_{i2},\alpha_{i3}\alpha_{i3},\ldots\ldots\alpha_{in_i}\alpha_{in_i}$$

$$y = A_i\alpha_i \tag{7}$$

Where $\alpha_i$ is the coefficient vector where most of the value are zero since it is sparse. Here y can be represented in terms of training samples as

$$y = Ax \tag{8}$$

Where $x=[0,\ldots.0,\alpha_{i1},\alpha_{i2},\alpha_{i3},\ldots..0,\ldots.0]^T \in R^n$ (9)

x is a coefficient vector where the elements are zero expect for those belonging to the i-th class and the non-zero elements will be associated with the columns of matrix A so and assign the test sample y to the respective i-th class.
Here x can be obtained by solving by using $l_0$-norm. It is expressed as shown below

$$x_0' = argmin||x||_0 \tag{10}$$

Where $||x||_0$ is $l_0$-norm and the problem is solved by finding column vector x, subject to y = Ax and $||x||_0$ is minimized which is equal to the number of non-zero elements in vector x.
Finding the solution to the above sparse representation problem is NP-hard. Using l1-norm which is equivalent to

$$x_1' = argmin||x||_1 \tag{11}$$

Where $||x||_1$ is $l_1$-norm. and the problem is solved by finding column vector x subject to y= Ax.

If there is any error or noise, it may cause non-zero entries in x which corresponds to a column of vector A. Generalized version of the above equation is

$$J(x,\lambda) = \min\{||Ax-y||_2 + \lambda||x||_1\} \tag{12}$$

Where $\lambda$ is a scalar quantity that balances the degree of noise and reconstruction error. Find the test sample y that best matches the training sample by minimizing the sum of squared errors [16].
Later calculate the re-construction residuals on dictionaries using the equation

$$r^m(y) = ||y - D_m x_m||^2,\ m = 1,2,\ldots.q \tag{13}$$

Where xi is the projection coefficient on the i-th class.
Then the test sample y can be classified into the respective class having a minimum residual error.

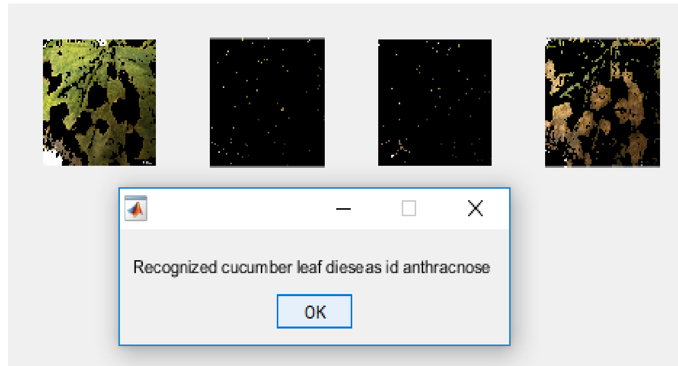



Figure 6: Cucumber disease recognition using sparse representation classification

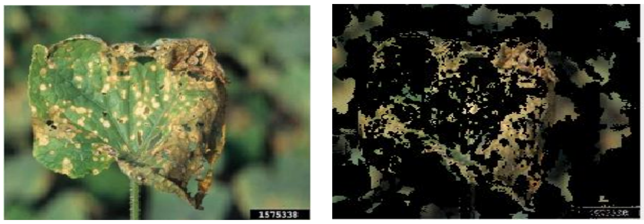
## IV. Results and Discussion

The experimentation is carried out for cucumber leaf images using Matlab and image processing toolbox. The cucumber leaf image is collected from [17] which are stored in JPEG format. The dataset contains 4 diseased cucumber leaves affected by anthracnose, powdery mildew, angular leaf spot, downy mildew and 1 healthy leaf is considered. Cucumber leaf image is pre-processed using AHE and RGB to LAB conversion. Then the LAB converted image is used for segmentation to segment the lesion part using K-means clustering algorithm Figure 7 shows the original cucumber leaf image and lesion segmented image.

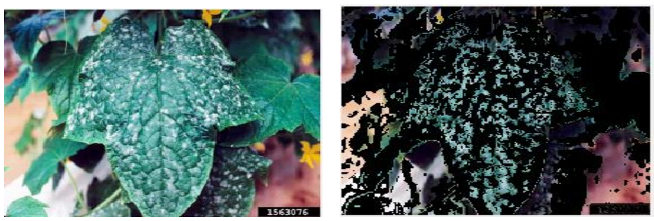
After segmentation color, texture and shape are extracted. In the existing method, the author has combined shape and color features [1]. In proposed method all the three features shape, color and texture feature have been combined the by transforming each color lesion image into grayscale and reducing it by Singular value decomposition (SVD) which is shown in figure 8. Then feature selection is performed for the combined features vector using PCA. The reduced feature vector is given to the sparse representation classifier. The accuracy of SRC obtained is 97.22%.

The cucumber leaf image is reduced to the feature vector, from that some of the images from each class are selected as a training set to construct a dictionary Eq. (7). Remaining vectors are selected as a test set to measure the performance of the proposed work. From the dictionary, calculate the projection coefficient of the test sample Eq. (8) and then calculate the reconstruction residual for each class Eq. (12). Finally, the test sample belonging to the corresponding class is decided by the minimum residual error.

To evaluate the performance of the proposed methodology used various evaluation measures as shown in table 1. The accuracy of our proposed work is better when compared to SVM classifier, the results show that accuracy of SRC is 97.22 while that of SVM is 76.11. Table 2 shows the comparison of SVM and SRC. Figure 9 shows the bar graph of performance measures compared for SVM and SRC.

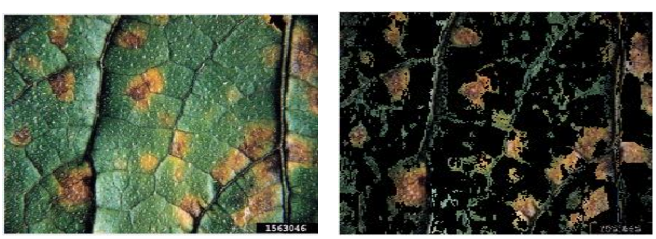


(A) Anthracnose



(B) Powdery mildew

(C) Angular leaf spot



(D) Downy mildew

**(E)** Healthy

Figure 7: Four diseased cucumber leaves and one healthy leaf with corresponding segmented images

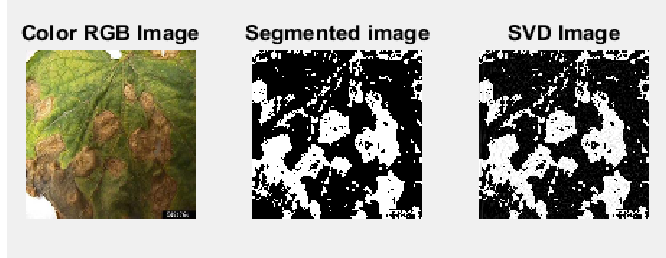



Figure 8: Singular value decomposition of segmented lesion image

Table 1. Evaluation measure of SVM and SRC

| **Evaluation measure** | **SVM** | **SRC** |
|---|---|---|
| Accuracy (%) | 76.11 | 97.22 |
| Precision (%) | 73.07 | 87.36 |
| Sensitivity (%) | 92.65 | 80.00 |
| Specificity (%) | 94.71 | 99.32 |

Table 2. Comparison of SVM and SRC classifier

| **Leaf sample** | **No. of image used for training** | **No. of images used for testing** | **Accuracy (%)** | |
|---|---|---|---|---|
| | | | **SVM** | **SRC** |
| Anthracnose | 54 | 20 | 73.10 | 96.23 |
| Powdery mildew | 03 | 02 | 79.15 | 99.15 |
| Angular leaf spot | 24 | 09 | 75.14 | 98.32 |
| Downy mildew | 17 | 09 | 77.05 | 95.16 |
| Healthy | 30 | 11 | 76.12 | 97.26 |
| Overall accuracy (%) | | | 76.11 | 97.22 |

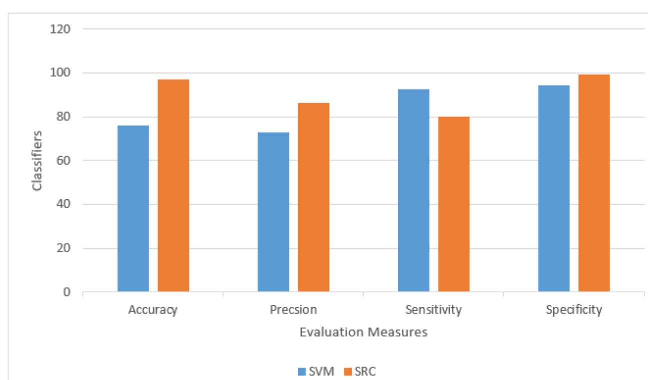



Figure 9: Show the bar graph of performance measures compared for SVM and SRC. Here x-axis represents the

evaluation measure in percentage and the y-axis represents the classifiers.

## V. CONCLUSION AND FUTURE SCOPE

Using image processing techniques it is easier to identify the diseases of the crop. In the proposed method, AHE to enhance the image which not only enhances the low contrast image but also has the capability of reducing noise present in the image with an extended feature as Contrast limited adaptive histogram equalization (CLAHE). Then LAB converted the image for K-means clustering

since LAB image is device independent which is equal to human visual perception. Then all the three features are combined and dimensionality is reduced using PCA. A reduced feature vector is used for sparse representation classifier to identify the disease accurately. There is a scope of improvement in the existing work.

As a future work, cucumber leaf having one or more disease on a single leaf will be considered for classification.

## REFERENCES

[1] X. W. Z. Y. L. Z. Shanwen Zhang, "Leaf image based cucumber disease recognition using sparse representation classification," computers and Electronics in Agriculture 134 (2017) 135–141, p. 7, 2017.

[2] J. S. A. Camargo, "An image-processing based algorithm to automatically identify plant disease visual symptoms," Published by Elsevier Ltd. All rights reserved., p. 9, 2009.

[3] R. K. Jayamala K. Patil, "Color Feature Extraction of Tomato Leaf," International Journal of Engineering Trends and Technology- Volume2Issue2- 2011, p. 3, 2011.

[4] W. X. Dong Pixia, "Recognition of Greenhouse Cucumber Disease Based on Image Processing Technology," Open Journal of Applied Sciences, 2013, 3, 27-31, p. 5, 2013.

[5] a. G. S. J. Pooja, "Adaptive histogram equalization technique for enhancement of coloured image quality," International Journal of Latest Trends in Engineering and Technology, 2017.

[6] Y. K. Shefali Gupta, "Review of Different Local and Global Contrast Enhancement Techniques for a Digital Image," International Journal of Computer Applications , 2014.

[7] M. S. Al-Tarawneh, "An Empirical Investigation of Olive Leave Spot Disease Using Auto-Cropping Segmentation and Fuzzy C-Means Classification," World Applied Sciences Journal 23 (9): 1207-1211, 2013, p. 5, 2013.

[8] D. S. S. D. M. S. Shabari Shedthi B, "Implementation and Comparison of K-Means and Fuzzy C-Means Algorithms for Agricultural Data," in International Conference on Inventive Communication and Computational Technologies, 2017.

[9] Z. G. D. M. Fengxi Song, "Feature selection using principal component analysis," in International Conference on System Science, Engineering Design and Manufacturing Informatization, 2010.

[10] S.C.Ng, "Principal component analysis to reduce dimension on digital image," in 8th International Conference on Advances in Information Technology, IAIT2016, 19-22, 2017.

[11] Y. M. J. M. G. S. S. H. S. Y. John Wright, "Sparse Representation for Computer Vision and Pattern Recognition," in 2010 IEEE, 2010.

[12] X. H. P. L. F. Z. Taisong Jin, "A Novel Method of Automatic Plant Species Identification Using Sparse Representation of Leaf Tooth Features," PLOS ONE, p. 20, 2015.

[13] Y. H. a. H. H. Hyunsup Yoon, "Image Contrast Enhancement based Sub-histogram Equalization Technique without Over-equalization Noise," International Journal of Electrical and Computer Engineering, p. 7, 2009.

[14] G. Z. Libo Liu, "Extraction of the Rice Leaf Disease Image Based on BP Neural Network," in Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on, 2009.

[15] T. L. a. N. L. Peng Guo, "Design of Automatic Recognition of Cucumber Disease Image," Information Technology Journal, p. 9, 2014.

[16] M. C. Saurabh Prasad, "Sparse Representations for Classification of High Dimensional Multi-Sensor Geospatial Data," in ieee, 2013.

[17] "www.ipmimages.org".