



Original papers

Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection



Muhammad Sharif^{a,*}, Muhammad Attique Khan^{a,c}, Zahid Iqbal^a, Muhammad Faisal Azam^a, M. Ikram Ullah Lali^b, Muhammad Younus Javed^c

^a Department of Computer Science, COMSATS Institute of Information Technology Wah Campus, Pakistan

^b Departments of Software Engineering, University of Gujrat, Gujrat, Pakistan

^c Department of Computer Science & Engineering, HITEC University, Museum Road, Taxila, Pakistan

ARTICLE INFO

Keywords:

Citrus fruits
Citrus diseases
Feature selection
Feature extraction
Features fusion

ABSTRACT

In agriculture, plant diseases are primarily responsible for the reduction in production which causes economic losses. In plants, citrus is used as a major source of nutrients like vitamin C throughout the world. However, 'Citrus' diseases badly effect the production and quality of citrus fruits. From last decade, the computer vision and image processing techniques have been widely used for detection and classification of diseases in plants. In this article, we propose a hybrid method for detection and classification of diseases in citrus plants. The proposed method consists of two primary phases; (a) detection of lesion spot on the citrus fruits and leaves; (b) classification of citrus diseases. The citrus lesion spots are extracted by an optimized weighted segmentation method, which is performed on an enhanced input image. Then, color, texture, and geometric features are fused in a codebook. Furthermore, the best features are selected by implementing a hybrid feature selection method, which consists of PCA score, entropy, and skewness-based covariance vector. The selected features are fed to Multi-Class Support Vector Machine (M-SVM) for final citrus disease classification. The proposed technique is tested on Citrus Disease Image Gallery Dataset, Combined dataset (Plant Village and Citrus Images Database of Infested with Scale), and our own collected images database. We used these datasets for detection and classification of citrus diseases namely anthracnose, black spot, canker, scab, greening, and melanose. The proposed technique outperforms the existing methods and achieves 97% classification accuracy on citrus disease image gallery dataset, 89% on combined dataset and 90.4% on our local dataset.

1. Introduction

Fruit plants are a major part of any agro economical society (Yesuf, 2013). Among fruit plants, citrus plants are packed with Vitamin-C, which provide multiple benefits to human health and are also used as a raw material in several agro-industries (Pujari et al., 2013). However, in recent years, citrus production is widely affected by the citrus diseases. These citrus diseases include canker, greasy spot, and black spot. In this regard, several automated solutions have been proposed for the symptom-based detection of citrus plant diseases, and promising results were recorded. Therefore, in present times, researchers are striving to find novel computer-based solutions for the early identification of citrus diseases, from the field of image processing (Kakade and Ahire, 2015; Vishnu and Ranjith, 2015). As the citrus lesion spot detection and classification consists of four major steps including (1) Preprocessing, (2) Segmentation, (3). Feature extraction, and (4) Classification, thus most of the studies carried out in this regard are focusing solution in

these lines.

For the detection of infected regions on plants leaves, Gavhale et al. (Gavhale and Gawande, 2014) compared different classification schemes such as KNN, RBF, PNN, and SVM, commonly used in the field of image processing. Similarly, for disease spot segmentation and classification in citrus fruits and leaves different techniques have been implemented.

For lesion spot segmentation, the K-means clustering (Arivazhagan et al., 2013), Otsu thresholding (Qin et al., 2008), region-based segmentation, region-oriented segmentation (Singh and Misra, 2015) and edge-based segmentation (Malik et al., 2016) techniques have been utilized.

For classification, many feature extraction and supervised machine learning techniques are employed including color feature (Jhuria et al., 2013), texture feature (Bandi et al., 2013), morphological features (Narvekar et al., 2014), SVM classifier (Gavhale et al., 2014), Artificial Neural Networks (ANN) (Qin et al., 2008), Moreover, various other

* Corresponding author.

E-mail addresses: muhammadsharifmalik@yahoo.com, attique@ciitwah.edu.pk (M. Sharif).

techniques including, Linear Discriminant Analysis (LDA), K-nearest Neighbors (KNN), Naive Bayes (NBC), Random Forests (RFT) (Bandi et al., 2013), and Back Propagation Neural Networks (BPNN) (Gavahale et al., 2014), have been also exploited.

The discussed studies reported various shortcomings. For instance, in case of low contrast lesion spots, contrast stretching is necessary, as it affects the lesion segmentation accuracy and degrades the classification accuracy. Moreover, efficient feature extraction and selection approaches are also crucial steps for such applications because each disease has distinct spots and symptoms. In this article, four major problems are considered, which include; (a) problem of local noise due to capturing devices and poor environment; (b) change of orientation; (c) change in shape and color of symptoms and (d) most prominent feature selection.

The above-listed problems are resolved by implementation of a new hybrid technique which is comprised of five major steps: (1) contrast stretching method is implemented to make infection area more visible and to improve the local contrast of lesion spot, (2) lesion spot segmentation, (3) construction of codebook by simple concatenation based fusion of color, texture, and geometric features, (4) selection of most prominent features by feature selection method, which is based on PCA score, entropy and skewness based covariance vector, and (5) Multi-class SVM is utilized for classification. The proposed method is validated for six types of citrus diseases including anthracnose, black spot, canker, scab, greening, and melanose. Our prime contributions in this research are as follow: of this research is enumerated below.

- (a). Preprocessing step to improve the contrast of input image by application of Top-hat filter and Gaussian function, which helps for efficient diseases spot segmentation. The Top-hat filter is used to improve the visual quality of input image, which is further improved by a difference-Gaussian image. The enhanced disease lesion spots are extracted by an efficient weighted segmentation method which uses Chi-square distance and Threshold function. The minimum distance pixels are sorted into the cluster using threshold function and optimized their performance by fusion of HDCT saliency method. For fusion, a limit rule based on pixels of segmented images is used.
- (b). A feature selection method is applied in which PCA score is found out by an extracted codebook vector. Then, an extracted score is fed to entropy and covariance methods separately and each is sorted in ascending order. Finally, the best 100 features are selected from both entropy and covariance vector based on higher feature value.
- (c). For experimental results, two databases are prepared as a combined dataset and local dataset. The combined dataset is prepared by a mixture of Plant Village Dataset and Citrus Images Dataset of Infested with Scale, which is publically available. Furthermore, a local dataset is prepared, which consists of images infected by four types of citrus diseases including anthracnose, black spot, scab, and canker. All images of the local dataset are collected from Sargodha district, Pakistan.

The chronological order of the article is as follows: Section 2 presents the related work. Proposed method is presented in section 3. Experimental results and conclusion are explained in section 4 and 5 respectively.

2. Related work

High importance of citrus fruits in the agriculture economy demands techniques to minimize the damages caused by diseases in citrus plants. Researchers from the domain of image processing and machine learning proposed numerous techniques for the detection and classification of citrus diseases.

Malik et al. (Malik et al., 2016) introduce an automated system for

detection, segmentation, and measurement of citrus fruits. The technique consists of preprocessing, shadow reduction, object separation, K-means clustering, and blob detection phases. In this regard, Gomez et al. (Gómez-Sanchis, 2008) borrowed a technique from the field of computer vision. The technique improves opposing effect created by the shape of spherical objects at image acquisition stage. In (Omid et al., 2010), the authors proposed a technique to measure the mass and volume of four diverse citrus fruits including: tangerines, oranges, limes, and lemons. The technique uses the elementary frustums to evaluate the mass and volume of the fruit. Kumar et al. (Kumar et al., 2015) introduced a technique for sorting and color-based grading of defected citrus fruits. The proposed technique efficiently identifies the type of a citrus fruit from the images of the combined fruits. In the technique, fruits are classified into different groups based on Gray-Level Co-Occurrence Matrix (GLCM) parameters.

Hussam et al. (Ali, 2017) introduced a novel technique for detection and classification of citrus diseases using AE color difference algorithm for the detection of diseased part in the image. The technique uses color, HOG and texture features for classification of the citrus diseases. The results show that the proposed technique has significantly improved performance in terms of accuracy and AUC. Pydipati et al. (Pydipati et al., 2005) presented a machine learning based technique for classification of citrus diseases. The technique exploits color co-occurrence algorithm for the classification. In the technique, four feature models are designed which include 39 texture feature sets. These features are further classified by two methods: neural network based on back propagation and neural network based on radial basis functions. According to the results improved performance in terms of accuracy up to 95% was recorded. Bruno et al. (Wetterich, 2016) introduced a new technique for citrus canker detection. The technique exploits the combination of fluorescence imaging spectroscopy and machine learning techniques such as SVM for canker detection. The introduced technique efficiently detects the citrus diseases like scab and canker and distinguishes between them by an improved accuracy of 97.8%. Deng et al. (Deng, 2016) introduced a novel technique for citrus greening detection based on huanglongbing (HLB) and cost SVM. In the technique, the citrus leaves were collected under the lighting condition, and color, texture, and HOG features were extracted. The authors used PCA for the reduction of extracted features and cost SVM for classification. According to the results, the cost SVM classifier achieved 91.93% accuracy with low cost and minimum computational time, as compared to other classification methods.

Georgina et al. (Stegmayer, 2013) presented a model for automated detection and classification of citrus diseases. The introduced model consists of two stages. In the first stage, a feature selection method is applied which further consists of two primary steps including feature ranking and threshold based best feature selection. In the second stage, the selected features are classified by multiple classification methods such as CART, Naive Bayes and NN using multi perceptron learning and achieved 88% accuracy for all detected diseases. Min et al. (Zhang and Meng, 2011) proposed a new approach for automatic detection of citrus canker from leaf images by the combination of global and local features. The proposed technique uses an improved Ada-boost method to select the most prominent features. These features are, later on, utilized for the disease detection. Finally, color and local texture features are extracted and classified by NN, SVM, and KNN. Gavahale et al. (Gavahale et al., 2014) introduced another techniques. The technique exploits leaves based feature inspection for the early detection of citrus diseases. The technique consists of four phases. In first phase, the input image is captured using a digital camera and then preprocessing like enhancement and color space structure is performed. In second phase, segmentation of the affected region by K-means clustering is performed. In phase three, extraction of GLCM features is carried out. In fourth phase, SVM-based classification is performed to identify the defected leave images.

Most of the widely used segmentation techniques such as K-means

clustering, thresholding, region growing and morphological based techniques have numerous drawbacks. For instance, these methods work poorly when the input image is too much complex, and having low contrast. Also, the GLCM features do not support the high classification rate as compared to color features, because of a different color of disease spots. Moreover, there are many other challenges which degrade the system accuracy. These challenges include poor contrast problem, irregularity in disease part, extraction of efficient features, and selection of most prominent features. A Comparison of proposed algorithm is performed with several existing methods (Zhexin et al., 2007; Abdullah, et al., 2012; Jaganathan and Arumugam, 2014; Yun, 2015; Bulanon, 2013; Kim, 2014; Pourreza, 2016; Qin et al., 2008; Kim, 2009; Zhao, 2009; Qin, 2012; Niphadkar, 2013; Dubey and Jalal, 1405; Sankaran et al., 2010; Mishra, 2012; Li, 2012; Sankaran, 2013; Xiaoling, 2016; Qin et al., 2008) on six identified diseases of a citrus fruit.

3. Proposed methodology

The proposed method is a conjunction of four primary steps consisting of (a) image enhancement, (b) identification of infected region using lesion segmentation, (c) feature extraction and construction of codebook and (d) feature selection and classification. Each process is composed of series of steps as presented in Fig. 1. The detail of each process is given below.

3.1. Preprocessing

Preprocessing is used to enhance the visual quality of input image. It removes several problems like brightness effects, illumination, and issues due to poor contrast. The preprocessing step plays a vital role in the field of image processing because the poor contrast images affect the lesion segmentation accuracy. In this article, a hybrid contrast stretching technique is applied, which is based on Top-hat filter and Gaussian function. Initially, the Top-hat filter is performed on an input image and later on, the lesion contrast is improved by addition of Top-hat filter image and difference-Gaussian image. The enhancement procedure is defined as follows:

Let $I(x,y)$ be the original RGB image having dimension $256 \times 256 \times 3$. The top-hat filter (Thapar, S. and S. Garg, Study and Implementation of Various Morphology Based Image Contrast Enhancement Techniques. International Journal of Computing Business

Research, 2012) is performed on the original image which enhances the infected regions in the query image as follows:

$$T_{\text{hat}}(x, y) = I(x, y) - I(x, y) \circ S \quad (1)$$

Where, $T_{\text{hat}}(x, y)$ in Eq. (1) denotes the top-hat filtered image and S is structuring element, whose initial value is 9. The value of S is heuristically initialized after several iterations. As the contrast of diseased part depends on the value of S , so the selection of S improves the visual quality of diseased part. Higher value for S increases the visual contrast of an image. After that, a Gaussian function is implemented on the original image with following formula:

$$G(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2)$$

In Eq. (2), I denotes the original image, μ is the mean of input image, and σ denotes the scaling parameter, which is calculated as:

$$\sigma = \sqrt{E(I) - [E(I)]^2} \quad E(I) = \sum_{i=1}^N \frac{I_i}{N} \quad (3)$$

The original image is subtracted from the Gaussian function with following formula:

$$T_n(u, v) = G(x, y) - I(x, y) \quad (4)$$

The maximum and minimum values are calculated from $T_n(x, y)$ and their results are added in the top-hat filtered image to handle the background region with following formulas in (5) and (6):

$$T_{\text{new}}(x, y) = T_{\text{hat}}(x, y) + \alpha \quad (5)$$

$$\alpha = \frac{\max(T_n(x, y)) + \min(T_n(x, y))}{2} \quad (6)$$

Finally, the enhanced image $T_{\text{new}}(x, y)$ is added with Gaussian image to obtain a new enhanced image. The new image is significantly better than original image, as shown in Fig. 2. This image can now handle the brightness, and poor contrast related problems. Further, the infected regions are more visible as compared to the original image.

$$I_F(x, y) = T_{\text{new}}(x, y) + G(x, y) \quad (7)$$

3.2. Lesion segmentation

Image segmentation is an important step in pattern recognition and image processing based applications. In plants, the lesion segmentation

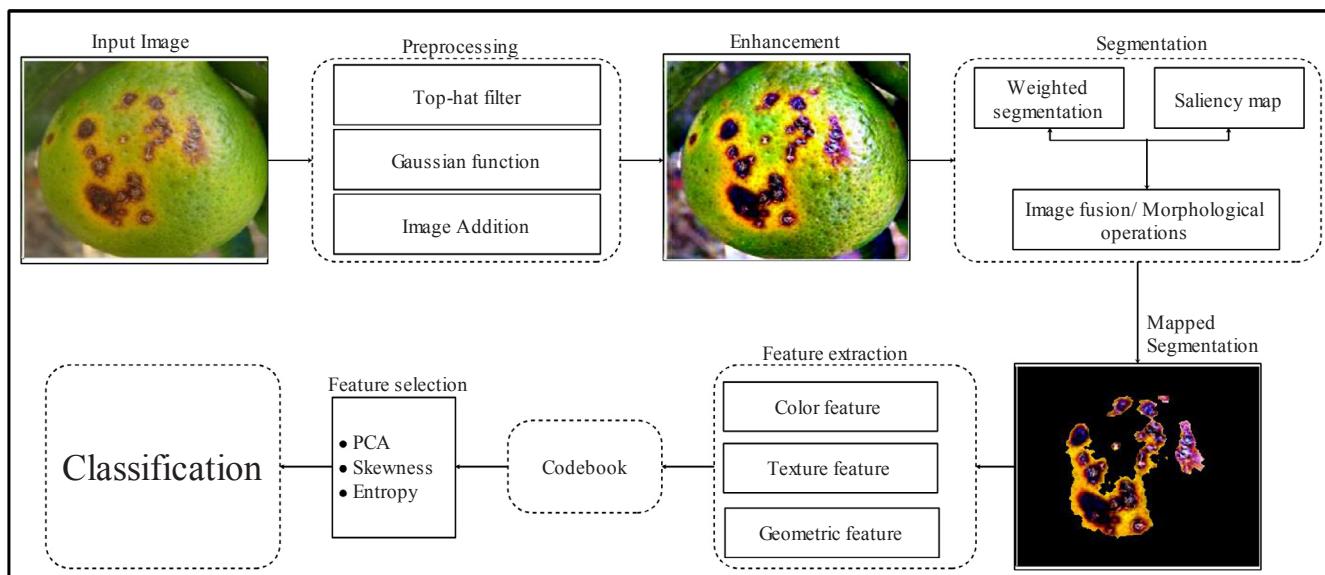


Fig. 1. System architecture for identification and classification of infected plants diseases.

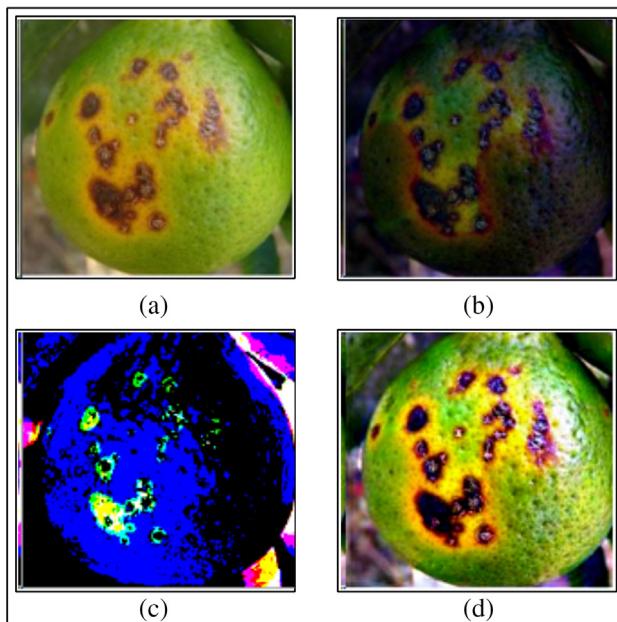


Fig. 2. Image enhancement results: (a) original image; (b) top-hat filtering image; (c) Gaussian image with original image; (d) final enhanced image.

step is used for the detection of lesion spot in the infected images automatically. To handle the automatic detection of lesion spots in the images, a weighted segmentation technique based on the Chi-square distance and threshold function is implemented. Moreover, the weighted segmentation method is optimized with existing High Dimensional Color Transform (HDCT) based saliency (Kim, 2016) technique to obtain a new segmented image, which has better results as compared to saliency segmentation method.

3.2.1. Weighted Lesion Segmented

Let $I_F(x, y)$ be the enhanced RGB image. Initially, the Chi-Square distance between rows and columns is calculated as follows:

$$I_F(\chi^2) = \sum_{i=1}^r \sum_{j=1}^c \frac{(x_{ij} - \mu)^2}{\sigma^2} \quad (8)$$

Where x_{ij} is the pixels at line i and column j , and r and c are the number of lines and columns of the image; I_F be the enhanced; μ and σ^2 denote the mean of enhanced image and the variance of enhanced image respectively. The purpose of measuring the Chi-square distance is to find the minimum distance between rows and columns pixels as compared to the background. A threshold function is set based on the skewness value of the enhanced image as follows:

$$F(w) = \begin{cases} i & \text{if } \chi^2 > Sk \\ 0 & \text{if } \chi^2 < Sk \end{cases} \quad (9)$$

Where in Eq. (9), χ^2 represents the Chi-square distance between pixels of enhanced image and Sk denotes the skewness value of enhanced image. Sk is calculated as follows:

$$Sk = \frac{(\mu - \text{mid})}{\sigma} \quad (10)$$

From the threshold function, it is clear that the output must be 1 (in case of infected region) when Chi-square distance is greater than the skewness value. Otherwise, it represents the background. This method produced better results as compared to Otsu segmentation, Expectation-Maximization (EM) segmentation, and active contour segmentation as shown below in Fig. 5. The comparison of weighted segmentation with saliency-based techniques (Kim, 2016) is shown in the Fig. 3, which shows that the weighted segmentation method performed well as

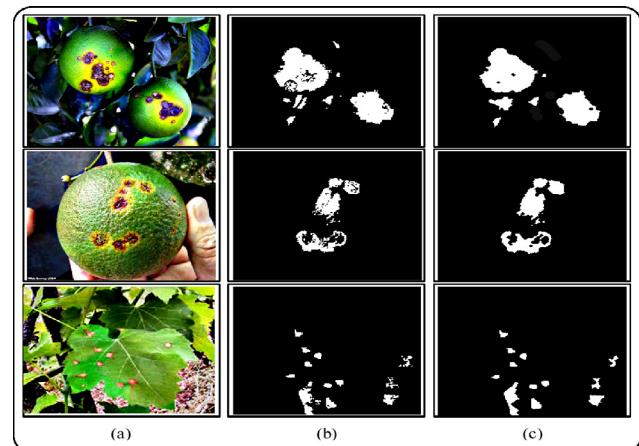


Fig. 3. Comparison of weighted segmentation with saliency-based technique. (a) Original image; (b) saliency-based method; (c) proposed method results.

compared to saliency method.

After that, the weighted segmentation method is optimized by existing HDCT based saliency method to improve their lesion spot segmentation accuracy.

3.2.2. Optimized weighted segmentation

In this section, the weighted segmentation results are optimized with HDCT based saliency segmentation (Kim, 2016). The major aim of optimization is to further improve the segmentation accuracy. Therefore, the fused segmented and saliency image are based on a new limit rule in which common pixels are included.

Let Z be a sample images database having n number of pixel values. Let z_1 denotes the number of pixels value of weighted segmented image $I_w(x, y)$, z_2 denotes the pixel values of saliency mapped image $I_{map}(x, y)$, and z_3 denotes the common points between both segmented images.

The output of the fused frame must be same, when

$$\sum_{i=1}^Z \Phi(I_w(x, y), I_{map}(x, y)) = 1 \quad (11)$$

The Eq. (11) shows that the output is foreground when both pixels are same. We implement an activation function, which is defined as:

$$\text{Activation} = \begin{cases} 1 & \text{if } \lim_{\lambda \rightarrow 0} \frac{|\lambda + \frac{1}{2}| + 1}{2} \geq 1 \\ 0 & \text{if } \lim_{\lambda \rightarrow 0} \frac{|\lambda + \frac{1}{2}| + 1}{2} < 1 \end{cases} \quad (12)$$

Where λ denotes the pixel values in segmented images and Φ is a coefficient between weighted segmentation and saliency mapped image. Eq. (12) shows that if $|\lambda + \frac{1}{2}| + 1$ is greater than 1, then the output pixel is spot lesion otherwise it shows the background.

Finally, morphological operations are performed such as closing, area removal and filling operation to make segmented image more efficient as shown in Fig. 4(d) and (e). The proposed approach presents very good results for all the selected types of citrus diseases and automatically detects the diseased part in the image with improved accuracy (Figs. 4–6).

3.2.3. Analysis

In this section, the proposed segmentation method is analyzed on total 90 citrus disease images. The selected images are collected from Citrus Disease Image Gallery Dataset (Dewdney and Timmer, 2009; Citrus Diseases Image Gallery, 2017) that includes both infected fruits and leaf images. For measuring the segmentation accuracy, the ground truth images are provided by an expert in this area. The segmented

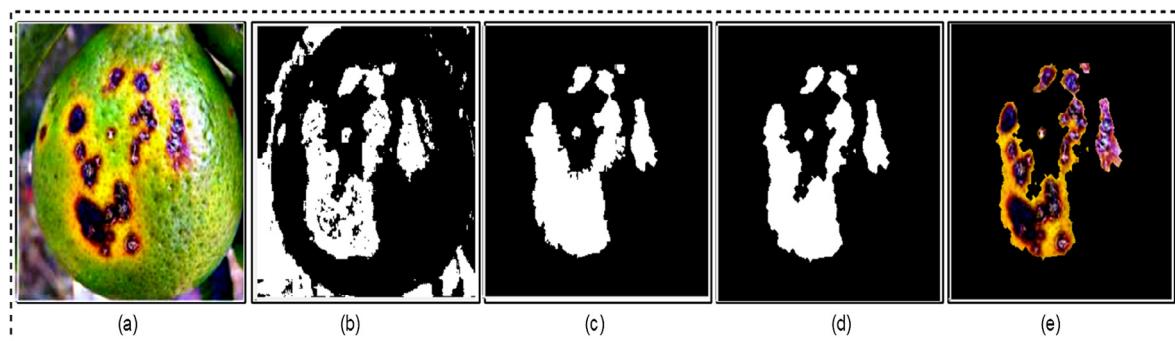


Fig. 4. Segmentation results: (a) Enhanced image; (b) Saliency image; (c) Weighted segmentation; (d) Fused image; (e) Mapped image.

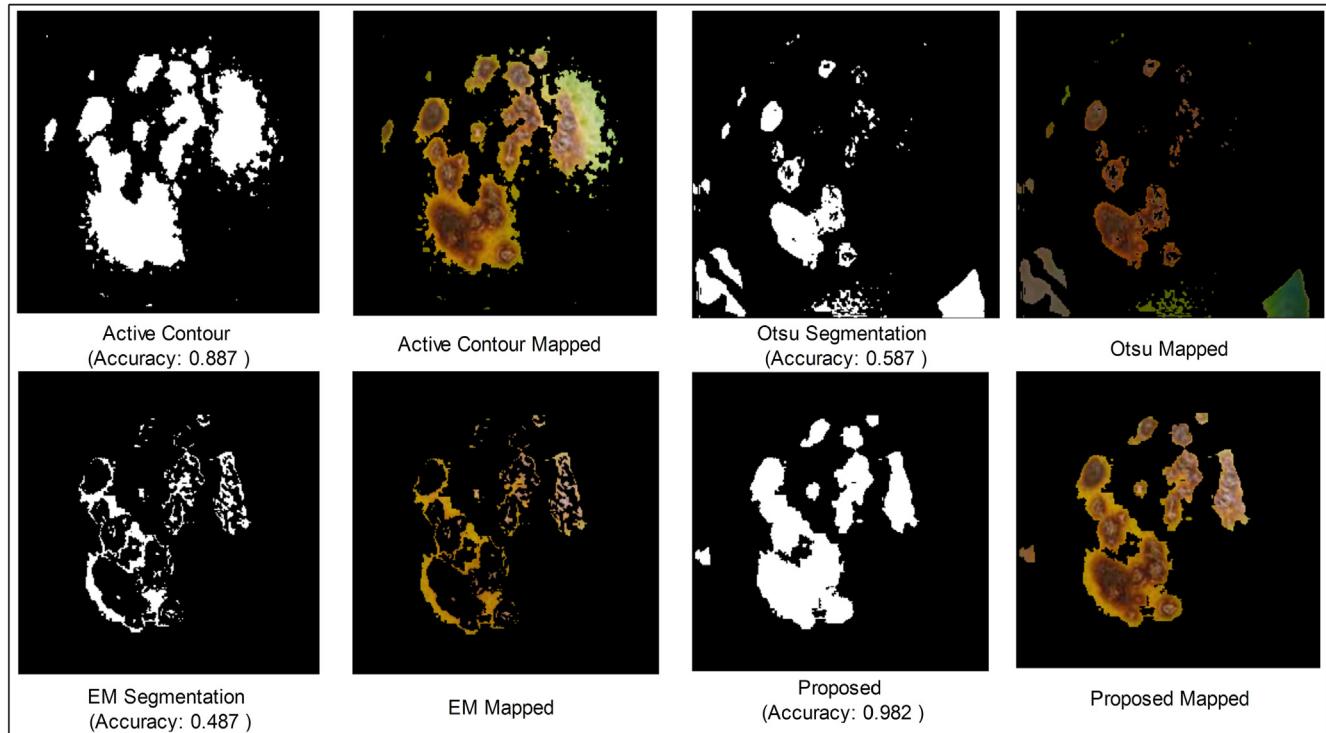


Fig. 5. Comparison of proposed segmentation method results with EM, Otsu, and active contour methods on complex images.

results are compared with their ground truth images. In lesion segmentation step, a new weighted segmentation has been proposed. Later on, Saliency-based segmentation is performed and results are fused with our proposed hybrid technique. This strategy produced better results as compared to separately applied segmentation technique for all types of selected diseases. The proposed method performs well for several difficult images as compared to existing methods such as saliency method as their effects are shown in Fig. 4. Moreover, some other segmentation methods are tested on a sample diseased image such as Otsu, EM, and active contour and their effects are shown in Fig. 5.

Figs. 4 and 6 shows some sample segmentation results of proposed algorithm. The detailed segmentation results with ground truth images are presented in Fig. 7 with their accuracy. From Figs. 3–6, it is clear that the proposed segmentation method accurately identifies the infection region in the image. In Fig. 7, the segmentation method produces maximum accuracy 99.68% and average accuracy is 93%. The details of some images used in this section are shown in Fig. 10(a). A similarity measure activation function is defined for the testing purpose of the segmentation algorithm.

This function compares the proposed segmentation results with ground truth images as follows.

Let f is fused image pixel values and g is fused image pixel values and i is ground truth image, and then the similar pixels are recalculated as: i is ground truth image, and then the similar pixels are calculated as:

$$A(F) = \begin{cases} 1 & \text{if } (F_{(i,j)} == G_{(i,j)}) \\ 0 & \text{if } (F_{(i,j)} \neq G_{(i,j)}) \end{cases} \quad (13)$$

Hence the final accuracy is computed as:

$$\text{Final}(val) = \frac{\text{Truepositives} + \text{Truenegatives}}{\text{Allpixelsofbothimages}} \quad (14)$$

Where in (14), the true positive denotes the similar pixels in both segmented and ground truth images and true negative denotes those pixels which are not similar to each other. Hence, accuracy is calculated by comparing both segmented and ground truth images and by dividing their addition on total number of pixels. The segmentation accuracy results of proposed method are presented in Table 29 and comparison results with Otsu segmentation, HDCT Saliency method, and EM segmentation are presented in Table 28. The average segmentation accuracy of Otsu segmentation, EM segmentation, HDCT saliency method and proposed segmentation method is 83.954%, 66.199%, 90.169%, and 92.433%, respectively. The comparison of proposed segmentation

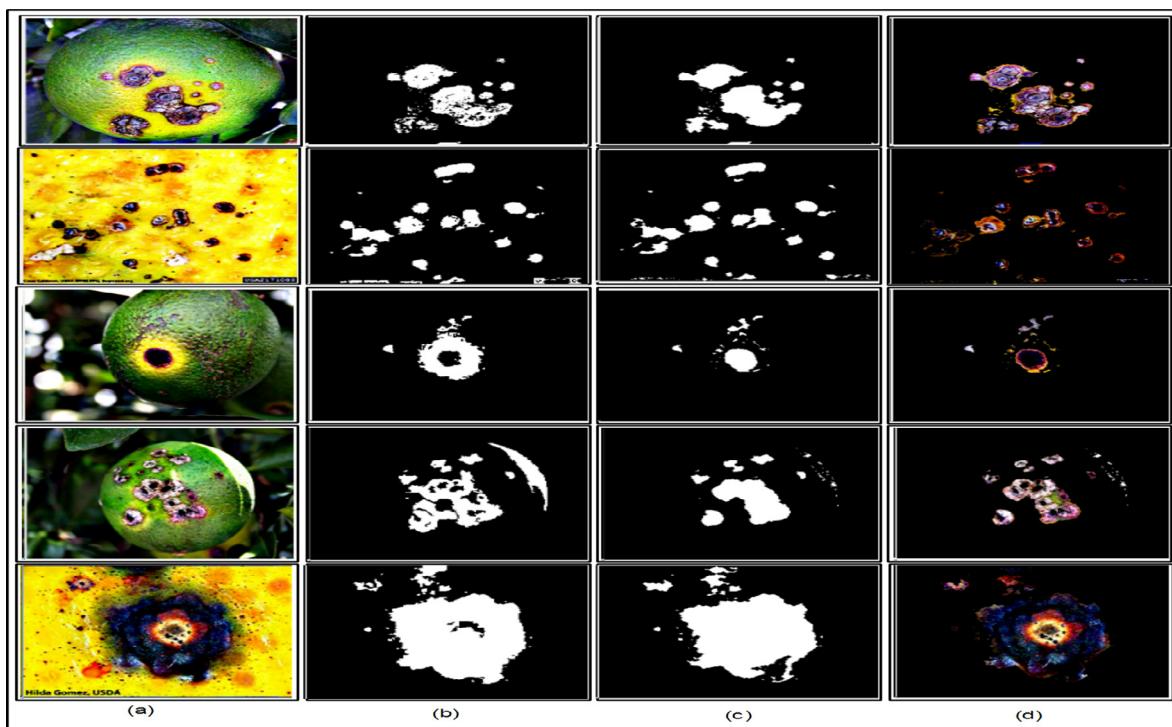


Fig. 6. Proposed segmentation results. (a) Enhanced image; (b) Weighted segmentation; (c) Fusion image; (d) Identification of infection region.

method is done on 20 images, which are collected from Citrus Disease Image Gallery database. The proposed average segmentation accuracy from Table 29 shows that the proposed method performs significantly well as compared to the existing methods.

3.3. Feature extraction

Feature extraction plays a vital role in the field of computer vision and pattern recognition for representatin of an image. The feature extraction have many applications such as video surveillance, agriculture,

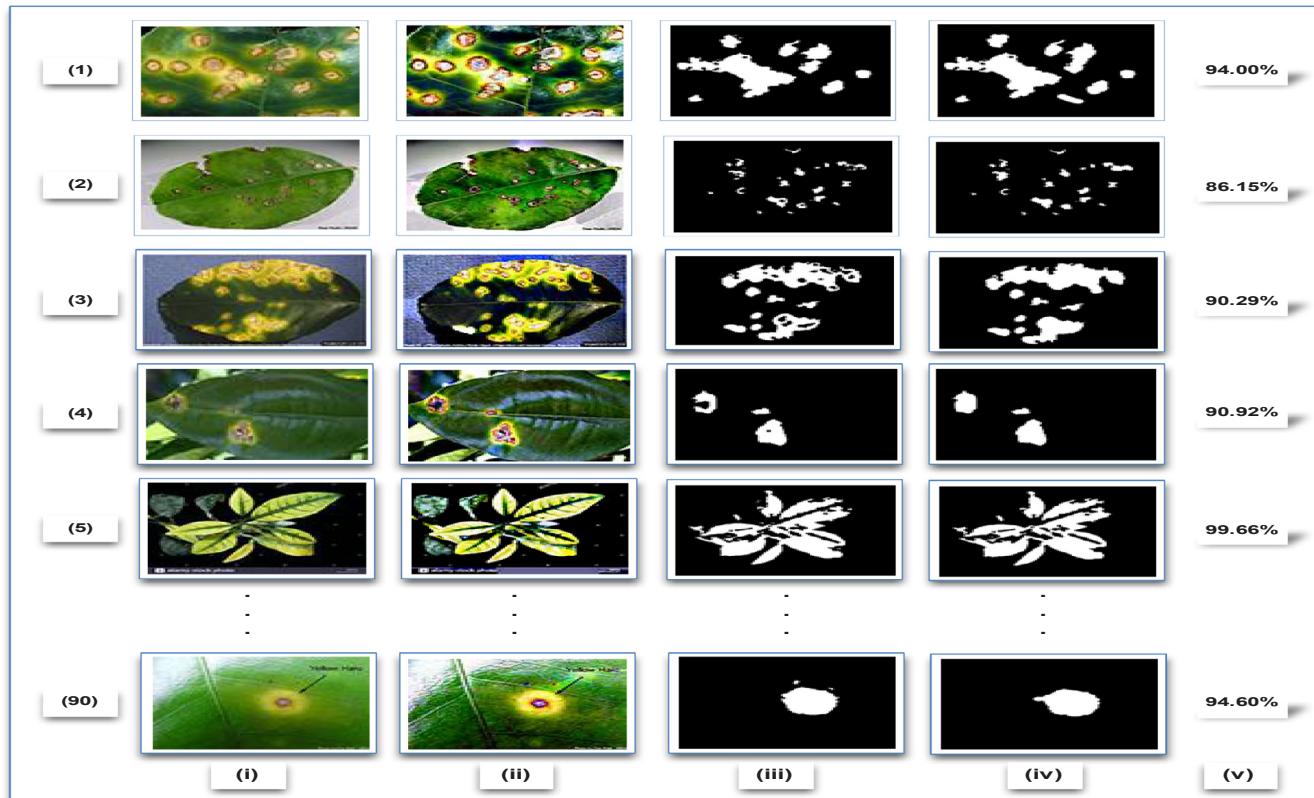


Fig. 7. Segmentation results with ground truth image: (i) Input image (ii) Contrast image (iii) Final image (iv) Ground truth image (v) Accuracy.

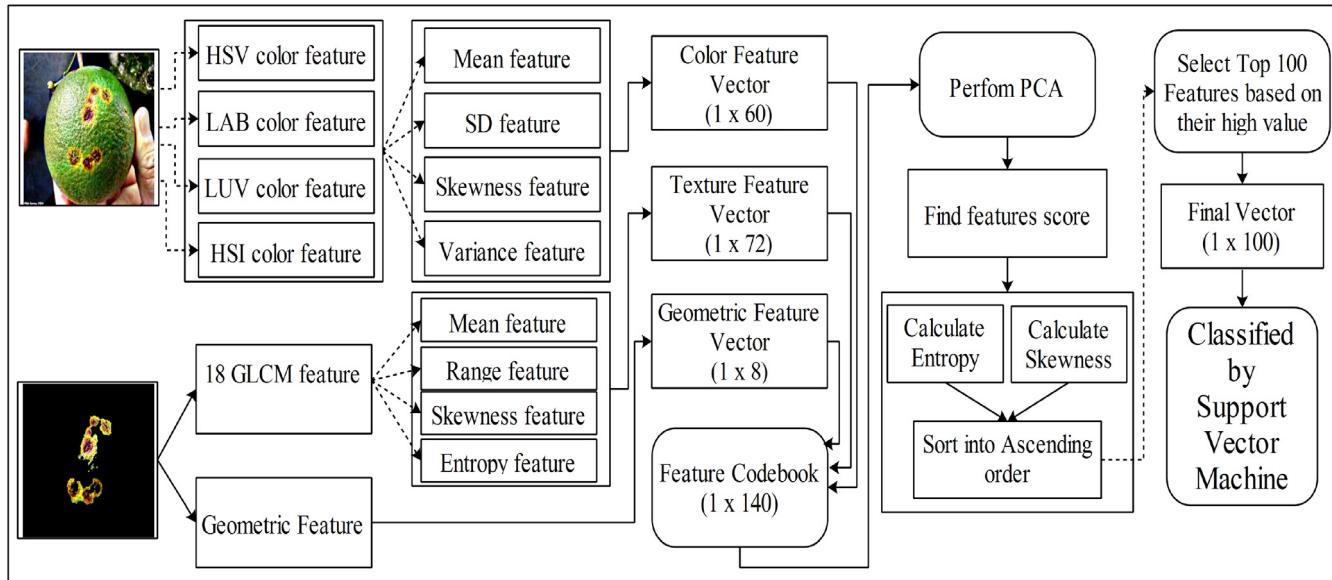


Fig. 8. Flow diagram for Codebook construction and then feature selection.

medical and robotics. In this section, the major issue is efficient feature extraction and selection for accurate classification of infected regions. To solve this problem, a codebook is constructed by extraction of three types of features including texture (Jolly and Raman, 2016), color (Naik and Sivappagari, 2016), and geometric (Ahmed et al., 2016). The extracted features are stored into a codebook and best features for classification are selected based on their maximum score achieved after application of our proposed feature selection approach. The flow diagram of feature extraction is shown in Fig. 8.

3.3.1. Color features

Firstly, the color features are extracted from enhanced images, which are obtained from Section 3.1. The color features are more important for the classification of infected region in the image because each lesion spot has different color. These features are obtained by the use of different types of color spaces and parameters. In this work, five types of color spaces are used for color features: Enhanced RGB, HSV, HSI, LAB, and LUV. Each color space produces distinct information as shown in Fig. 7.

RGB is an important color space and mostly used in many computers vision applications such as agriculture, medicine, and few more. It is easy to apply and need to require for different color spaces which are much apart for processing. The RGB to HSV and RGB to HSI models hold non-linear conversions of RGB color space. These color spaces are well perceptive and machine dependent and mostly used in the area of color utilization. Moreover, in RGB to LAB and RGB to LUV models are also non-linear color spaces and its conversions are reversible.

The enhanced images are utilized for extraction of color features from color spaces. The main idea behind these features is to obtain the maximum lesion spot information from each selected color space. Hence, the fusion of these features generates a feature vector which consists of much information as compared to an individual selection of color space. Each color space is divided into their respective three channels (i.e. RGB have red, green and blue channels) and calculate the mean, standard deviation, entropy, and skewness for each channel. The clear description of extracted color features from selected color spaces is shown in Fig. 8, which explains that the enhanced RGB image and four-color spaces are utilized for features extraction. All the features are finally fused in one vector by simple concatenation and obtain a feature vector of size 1×60 . For example, the calculation of mean, standard deviation, entropy, and skewness for each channel is extracted as follows:

Let $I_F(x, y)$ is an enhanced image having three dimensions. Let R denote the red channel, which is extracted from RGB.

$$\bar{X} = E(X) = \sum_{i=1}^n \frac{(X_i Y_j)}{n} \quad (15)$$

$$\sigma = \sqrt{E(X^2) - [E(X)]^2} \quad (16)$$

Where,

$$E(X^2) = \frac{1}{n} [(1)^2 + (2)^2 + \dots + (n)^2] \quad (17)$$

$$= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} \quad (18)$$

$$E(X) = \frac{(n+1)(2n+1)}{6} \quad (19)$$

$$\text{Entropy} = - \sum_{i,j} I(X_i, Y_j) \log \frac{I(Y_j)}{I(X_i, Y_j)} \quad (20)$$

$$\text{Skewness} = \frac{E(x - \bar{X})^3}{\sigma^3} \quad (21)$$

\bar{X} in Eq. (15) denotes the mean feature, n denotes the total number of pixels in the Red channel, X_i and y_j denotes the rows and column pixel of red channel respectively and σ denotes the standard deviation.

3.3.2. Geometric features

The geometric feature also plays an important role in many applications such as robotics, vision system, and agriculture in finding their geometric registration. In this article, 8 geometric features are extracted as given in Table 1. These features are extracted from the segmented spot lesions to find the global information and local information such as

Table 1

Extracted geometric features for the construction of codebook.

Feature Name	Description
'Area	$\sum_{i=1}^n \sum_{j=1}^m A[i, j]$
Filled Area	Infected regions in the image
Aspect Ratio	$\frac{\text{Width}}{\text{Height}}$
Orientation	$\tan^{-1} \left(\frac{y}{x} \right)$
Perimeter	$2l + 2w$
Extent	<i>PleaseCheck</i>
Solidity	<i>PleaseCheck</i>
Major axis, and Minor axis	$x_1 + x_2, \sqrt{(x_1 + x_2)^2 - d}$

Table 2

Extracted GLCM features for construction of codebook.

Feature Name	Equation	Feature Name	Equation
Auto Correlation	<i>PleaseCheck</i>	Homogeneity 1	<i>PleaseCheck</i>
Contrast	<i>PleaseCheck</i>	Homogeneity 2	<i>PleaseCheck</i>
Correlation 1	<i>PleaseCheck</i>	Maximum	<i>PleaseCheck</i>
		Probability	
Correlation 2	<i>PleaseCheck</i>	Sum of Squares (Variance)	<i>PleaseCheck</i>
Cluster Prominence	<i>PleaseCheck</i>	Sum Average	<i>PleaseCheck</i>
Cluster Shade	<i>PleaseCheck</i>	Sum Entropy	$-\sum_{x=2}^{2\phi^{\Gamma}-2} I_{x+y}(x) \log(I_{x+l}(x))$
Dissimilarity	<i>PleaseCheck</i>	Sum Variance	$\sum_{x=2}^{2\phi^{\Gamma}-2} (x - \bar{\phi}^H) I_{x+y}(x)$
Energy	<i>PleaseCheck</i>	Difference Variance	$\sigma^2(P_{x-y})$
Entropy	<i>PleaseCheck</i>	Difference Entropy	$\sum_{x=0}^{\phi^{\Gamma}-1} I_{x-l}(x) \log(I_{x-l}(x))$

lesion shape, size, and position. The geometric features include: area, aspect ratio, filled area, major and minor axis, extent, perimeter and solidity. These features play a major role in improving the accuracy of disease classification, when these are fused with color and texture features. The fusion process is shown in Fig. 8. Hence, the final dimension of extracted geometric vector is 1×8 .

3.3.3. Texture features

The texture features are extracted from the segmented image. In texture feature extraction, 18 GLCM features (Albrightsen, 2008) are extracted as presented in Table 2.

Originally, the Harlick features are 14 in number (Haralick and Shanmugam, 1973) but we add 4 new features. Four new features such as cluster prominence and shade, homogeneity, and energy are added in to achieve best classification accuracy. The cluster shade and cluster prominence features can be modified by a sum histogram problem and homogeneity feature can be calculated from difference histogram problems. The energy texture feature calculates the orders in an image and gives the sum of the square element. The energy value must be high when a window is appropriately organized. Then calculate 4 sub features (Mean, range, skewness, and entropy) from each extracted GLCM feature. The main idea behind is - to measure the level of irregularity, midpoint, smallest interval between diseased pixels, and shape of disease spots.

The purpose of texture feature extraction is to observe the combination of image intensities at the distinct positions, measure the level of irregularity, midpoint, smallest interval between diseased pixels, and shape of disease spots, which are relative to each other in the image. The number of pixels in the image are classified into first order, second order, and third order. The range is calculated by the difference between the maximum and minimum value of extracted feature as follows:

$$\text{Range} = \max(\eta_F) - \min(\eta_F) \quad (22)$$

Where in (22), η_F denote the GLCM feature. Other parameters like mean, skewness and entropy are calculated by Eqs. (15), (20), and (21). Hence, against one GLCM feature, four features are obtained. Thus, the dimension of extracted GLCM vector is 1×72 and the detailed description of extracted GLCM features is presented in Table 3.

3.3.4. Codebook construction

The codebook is constructed after extraction of three types of features including color, geometric and texture. These features are constructed into a codebook by simple concatenation based fusion method. The feature fusion means to fuse multiple features or description of

Table 3

Description of training and testing samples using selected datasets.

Citrus Type	Total Images	Training Images	Testing Images
Citrus Disease Image Gallery Dataset (Citrus Diseases Image Gallery, December 20, 2017)			
Anthracnose	100	50	50
Black Spot	80	40	40
Canker	120	60	60
Scab	100	50	50
Greening	100	50	50
Melanose	70	35	35
Healthy	100	50	50
Combined Dataset (Guth et al., 2017) (Cubero, 2016)			
Black Spot	40	20	20
Canker	40	20	20
Anthracnose	45	23	22
Greening	5507	2500	2500
Our Dataset			
Black spot	120	60	60
Anthracnose	130	65	65
Canker	100	50	50
Scab	90	45	45
Healthy	140	70	70

multiple features in one vector, which contains more information as compared to individual feature vector. In literature, several types of feature fusion techniques have been implemented such as discriminant correlation analysis (Haghigat et al., 2016) and CCA (Sun, 2005). However, we do not have large number of features therefore; we used simple concatenation based feature fusion. The utilized method maximizes the correlation across extracted set of features and omit between class variations at the same time. Hence, the size of feature codebook is 1×140 , which is later optimized by a new feature selection method. In feature selection method, the best features are selected for final classification.

3.4. Feature selection

In computer vision and machine learning applications, the high dimensional features increase the system execution time and space requirement for processing. Therefore, the feature selection techniques are applied to find a subset of most relevant features from irrelevant data, where irrelevant denotes the redundant features. In this article, a hybrid feature selection technique is implemented based on PCA score, Entropy, and Covariance. The major aims of this method are (a) to improve the classification accuracy by solving the problem of over fitting; (b) to improve the prediction performance for a supervised method such as SVM etc.; and (c) to remove the redundancy between features. In the proposed approach, initially, the fused codebook features are fed to PCA as input to find out their score value. Then entropy and covariance are performed on score vector separately and sorted into ascending order. Thereafter, the top 100 features are selected from each and fused in parallel based on the maximum value. The detailed description of entropy and covariance is given below.

3.4.1. Covariance & entropy

The covariance finds out the mutual changeability between extracted features and also measure the point in which features are linearly linked to each other. The PCA is performed on fused codebook vector having dimension 1×140 and calculate the score of each feature vector. After that, covariance is calculated between each score vector and sort into ascending order. The covariance is found as follows:

Let S_i and S_j denote the score value of two features i, j where $(i, j \in \Delta)$. The Δ denotes the feature codebook having dimension 1×140 . Then calculate covariance and entropy features by using $\Delta^{1 \times N}$, where each

time value of features i, j is changed. Then calculate covariance between features i, j and this process is iterated upto N.

$$\text{Cov}(S_1, S_2) = E[(S_1 - E(S_1))(S_2 - E(S_2))] \quad (22)$$

$$= E(S_1S_2 - S_1E(S_2) - S_2E(S_1) + E(S_1)E(S_2))$$

$$= E(S_1S_2) - E(S_1)E(S_2) - E(S_1)E(S_2) + E(S_1)E(S_2) \quad (23)$$

$$\text{Cov}(S_1, S_2) = E(S_1S_2) - E(S_1)E(S_2) \quad (24)$$

Where $E(S_1) = \sum_i S_{1i}g(x)$ and $E(S_2) = \sum_j S_{2j}h(y)$. The $g(x)$ and $h(y)$ denotes the feature vectors and denotes the final covariance vector. This vector is sorted in an ascending order as follows:

The entropy vector is calculated with PCA score value as follows:

$$ENT(S_1, S_2) = - \sum_{i,j} \frac{\text{Score}(S_i, S_j) \log \frac{\text{Score}(S_i, S_j)}{\text{Score}(S_i, S_j)}}{\text{Score}(S_i, S_j)} \quad (26)$$

$$FV_{ENT}(S_1, S_2) = \sum_{j=1}^{140} \{ENT(S_1, S_2), 'Ascending'\} \quad (27)$$

Where, denotes the entropy ascending vector and denotes the entropy ascending vector and denotes the covariance ascending vector. Finally, select the top 100 from both vectors and fed into one vector based on comparison to each other. The comparison means that the highest value feature is selected from both vectors for final classification. The final selected vector is fed to M-SVM for classification of selected diseases. For this purpose, one against all (Liu and Zheng, 2005) method is used. The cost function of this method is given below:

$$\text{minimize: } k(\tilde{w}_i, \zeta_j^i) = \frac{1}{2} \|\tilde{w}_i\|^2 + D \sum_{j=1}^N \zeta_j^i \quad (29)$$

subject to:

Where if $y_j = i$ else . For classification, a sample φ_x is classified as in class i^* for which f_i^* generates largest value

4. Experimental results

The performance of the proposed algorithm is evaluated on Citrus Diseases Image Gallery Dataset (Dewdney and Timmer, 2009; Citrus Diseases Image Gallery, 2017), Plant Village Dataset (Guth et al., 2017), Oranges infested with scale Dataset (Cubero, 2016) and own collected citrus dataset. The system performance is tested in three experiments on selected datasets. In first experiment, 670 images are collected from Citrus Diseases Image Gallery Dataset as sample images are shown in Fig. 9(a) and their results are evaluated using 10-fold cross validation. In second experiment, the combined dataset is generated from Plant Village Dataset and Oranges Infested with Scale Dataset having total images 5632 of four diseases (i.e. citrus black spot, canker, anthracnose, and greening). The sample images are shown in Fig. 9(c).

In third experiment, the recognition results are obtained on our own collected database of four types of diseases (black spot, anthracnose, scab, and canker) and healthy images. The sample images of our own collected dataset are shown in Fig. 9(b). 10-fold cross validation is performed on every experimental test. Four classification methods (Weighted K-Nearest Neighbor (W-KNN), Ensemble Boosted Tree (EBT), Decision Tree and Linear Discriminant Analysis (LDA)) are used for comparing the proposed classification results. Seven measures are implemented for examining the performance of proposed algorithm as True Positive Rate (TPR), False Positive Rate (FPR), False Negative Rate (FNR), Positive Prediction Value (PPV), False Discovery Rate (FDR),

Area under Curve (AUC) and accuracy. All work is implemented on MATLAB 2016a, using a personal computer, which has a 64bit operating system and 8 GB RAM.

4.1. Characteristics of selected citrus diseases

The detailed description about technical characteristics of selected citrus diseases is described below.

4.1.1. Anthracnose

Anthracnose is an initial colonizer of damaged and elderly tissue. The plants grow on stagnant copse in the cover, and it grows short distance by rain spray, dew, and overwatering. The leaf symptoms of citrus are a more or less broadside, smooth region, and a rich brown in color with a prominent purple perimeter. Moreover, the fruit symptoms are usually damaged by other factors such as sunburn, chemical high temperature, pest damage, and expanding storage time. The color of anthracnose lesion spot is brown having 1.5 mm or greater diameter (Timmer and Graham et al., 2000; Citrus Diseases Image Gallery, 2017). The sample image of anthracnose is shown in Fig. 10.

4.1.2. Black spot

The citrus black spot also abbreviated as CBS is available at the time when the plant is sensitive and the atmosphere is favorable for disease. The symptoms of the citrus leaf and fruits are small, round, and under dangerous spots with gray squares. The color of black spot lesion is a dark brown having spot diameter of 0.12 to 0.4 in. as shown in Fig. 10 (Timmer and Graham et al., 2000; Citrus Diseases Image Gallery, 2017).

4.1.3. Canker

Citrus canker spot is coated by wind-driven rain. On the citrus leaf, the range of lesion is 2 to 10 mm in size and put circles around the leaf. On the other hand, the range of fruit lesion will be 1 to 10 mm in size and these lesions differ in size from each other. Mostly, the lesions are enclosed by water soaked and yellow halo. The color of this lesion is dark brown and black as shown in Fig. 10 (Timmer and Graham et al., 2000; Citrus Diseases Image Gallery, 2017).

4.1.4. Scab

In citrus leaf and fruits, the scab acne is a composite of fungal and organism tissue. These acnes are hardly raised and pink to light brown in color. The color of scab lesion is yellow-brown and dirty grey as shown in Fig. 10 (Timmer and Graham et al., 2000; Citrus Diseases Image Gallery, 2017).

4.1.5. Melanose

Melanose is a saprophyte in which the severity is defined by the number of inoculum on dead wood in the tree cover. The symptoms of the leaf appear as small brown spot which becomes filled with a reddish-brown gum. On the other hand, the fruit symptoms depend on the age of the fruit at infection time. The symptoms appear outside on the citrus leaf and fruit (Timmer and Graham et al., 2000; Citrus Diseases Image Gallery, 2017). The melanose sample image is given in Fig. 10.

4.2. Citrus disease image gallery dataset

This dataset contains 1000 images of several citrus diseases that include anthracnose, canker, scab, melanose, leprosis and few more. The gallery images consist of both fruits and leaf having image dimension of 100×150 with resolution 96 dpi as shown in Fig. 9 (a) and Fig. 10. In this article, six citrus disease images are selected including anthracnose, scab, black spot, melanose, greening, and canker. The selected diseases are initially evaluated by healthy images and finally classified by all selected diseases. The detail evaluation results are presented next.

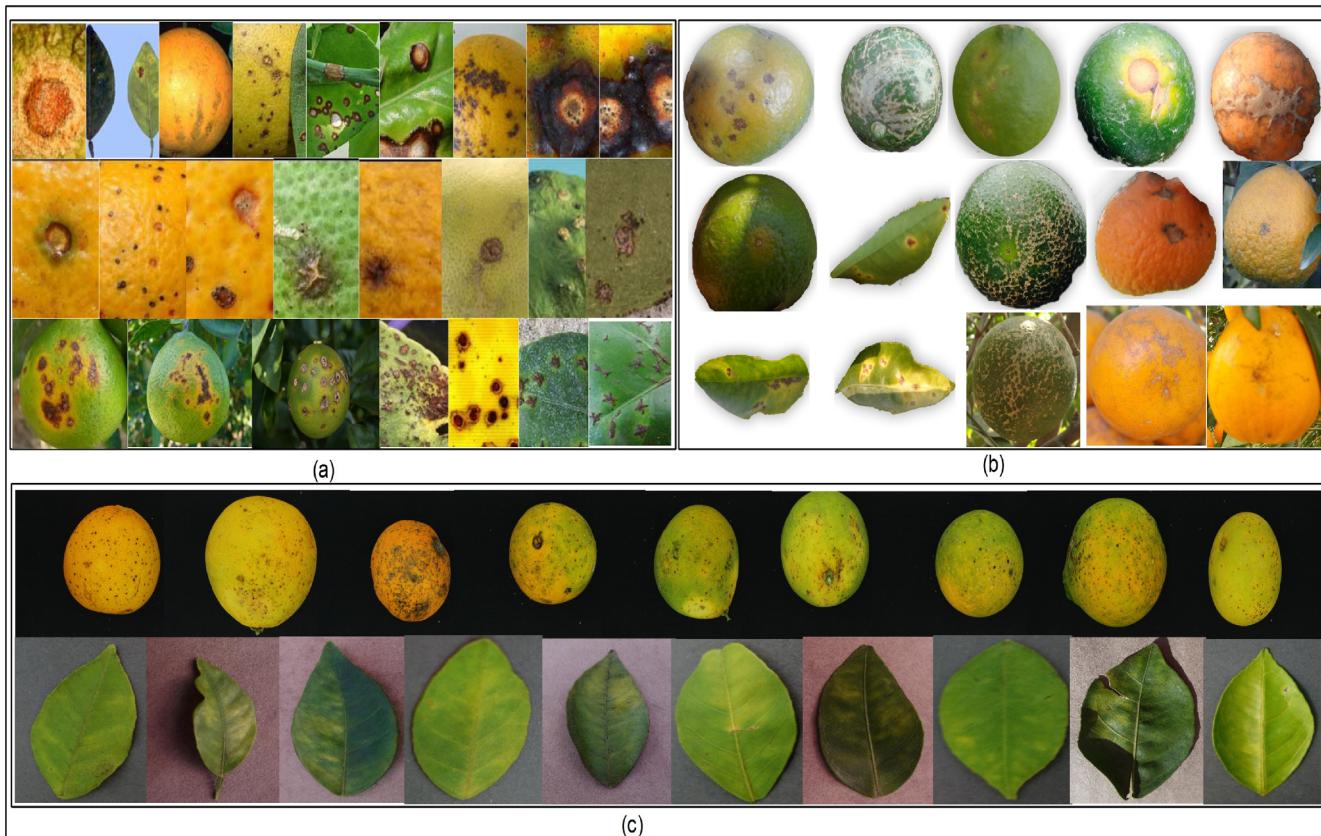


Fig. 9. Sample images of selected diseases. (a) Citrus Diseases Image Gallery Dataset ([Citrus Diseases Image Gallery, December 20, 2017](#)); (b) Our owned dataset, and (c) combined dataset ([Guth et al., 2017](#)) ([Cubero, 2016](#)).

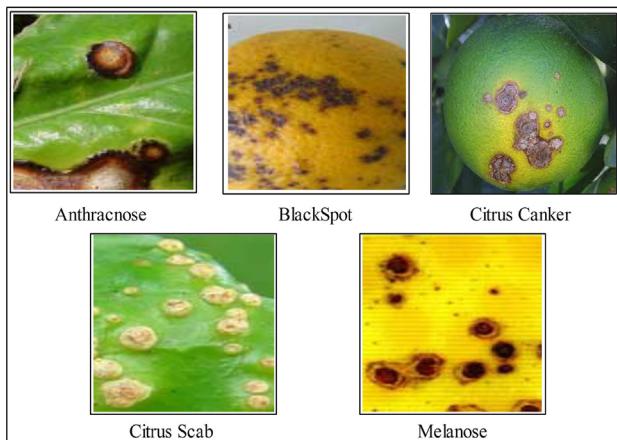


Fig. 10. Sample images for Citrus Diseases Image Gallery Dataset ([“Citrus Diseases Image Gallery, December 20, 2017”](#)).

4.2.1. Anthracnose disease

Anthracnose is a fungal disease caused by the ‘*Colletotrichum*’. Anthracnose creates the dark fungal spores, premature leaf drops, staining and twigs to dieback symptoms on the citrus leaf and fruits. For experimental results, 100 unhealthy anthracnose sample images are selected for training and testing. The selection of training and testing samples are described in [Table 3](#).

[Table 4](#) shows the comparison of the proposed algorithm with four states of the art classifiers. Proposed method achieved maximum accuracy of 96.9%. Minimum FPR is 0.030, which is better than other state-of-the-art methods. The performance of proposed method is presented with confusion matrix in [Table 5](#).

Table 4
Performance measures on anthracnose.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.969	0.030	97.0	3.1	2.9	0.9447	96.9%
W-KNN	0.959	0.040	96.1	4.1	3.8	0.9569	95.9%
EBT	0.938	0.061	93.8	6.2	6.1	0.9863	93.8%
DT	0.958	0.041	95.8	4.1	4.1	0.9574	95.9%
LDA	0.959	0.041	95.9	4.1	4.0	0.9449	95.9%

Table 5
Confusion matrix of anthracnose disease.

Leaf Types Image	No. of images	Unhealthy (%)	Healthy (%)
Unhealthy	50	100%	0%
Healthy	50	6.1%	93.9%

Furthermore, the proposed algorithm performance is compared with existing techniques of anthracnose (as shown in [Table 6](#)). This shows that proposed method achieved good results as compared to existing methods.

4.2.2. Black spot disease

Black spot is a fungal lesion created by *guignardia citricarpa*. This disease infects a citrus plant around subtropics weather and causes the reduction in both fruits quality and quantity. For experimental results, total 80 black spot images are selected. The 50:50 strategy is adopted for training and testing as described in [Table 3](#).

[Table 7](#) depicts the comparison of proposed algorithm with four state-of-the-art classifiers having maximum accuracy of 98.7% and lowest FPR of 0.013, which shows that proposed method performs

Table 6
Comparisons of existing techniques of anthracnose.

Ref.	Year	Accuracy
Zhexin et al. (2007)	2007	96.6%
Abdullah, et al. (2012)	2012	67%
Jaganathan and Arumugam (2014)	2014	81%
Yun (2015)	2015	91.08%
Proposed	–	96.9%

Table 7
Performance measures on black spot.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.987	0.013	98.8	1.3	1.2	0.9993	98.7%
W-KNN	0.936	0.063	96.3	6.4	6.4	0.9361	93.6%
EBT	0.979	0.025	97.5	2.6	2.5	0.9976	97.4%
DT	0.975	0.025	97.4	2.6	2.5	0.9657	97.5%
LDA	0.921	0.078	93.5	7.7	6.5	0.9960	92.3%

better as compared to other classification methods. The performance of proposed method is confirmed by their confusion matrix that is shown in Table 8.

The comparison with existing techniques is listed in Table 9, which shows that proposed method perform significantly better than compared methods.

4.2.3. Canker disease

Canker disease is created by a microorganism called pathogen and it effects the leaves. Canker disease can transmit easily from one leaf to other. Its infection damages fruits, leaves, and stems of citrus plants including grapefruit, lime, and oranges. For experimental results, 120 Canker images are selected based on 60:60 proportions for training and testing.

Table 10 shows the comparison of proposed algorithm with four classification methods. Maximum accuracy of 99.1% is obtained using the proposed method with lowest FP rate of 0.009. From Table 10, it is clear that proposed method performs well as compare to other classification methods. The performance is confirmed by the confusion matrix listed in Table 11. Furthermore, Table 12 shows a comparison with existing techniques. It is clear that the proposed method shows significantly good results.

4.2.4. Citrus scab disease

Scab disease causes a serious issue to all citrus families. This disease attack leaves of citrus plants such as sour orange, carizzo citrange, rough lemon and rangpur lime. For leaf, scab pimple is a combination of host and fungal diseases. This pimple is slightly raised and converts to lightly brown in color. For experimental results, 100 citrus scab images are selected. For training and testing, we make a strategy of proportion 50:50. It means that 50 images are being employed for training process and 50 images are being employed for testing process as described in Table 3.

The results of proposed algorithm are compared with other classification methods as shown in Table 13. Maximum accuracy of 97.6% is achieved, which is confirmed by their confusion matrix (shown in Table 14). The proposed algorithm results on citrus scab are compared with existing methods as presented in Table 15. From Table 15, it is

Table 8
Confusion matrix of black spot disease.

Leaf Type Images	No. of images	Unhealthy (%)	Healthy (%)
Unhealthy	40	97.4%	2.6%
Healthy	40	0%	100%

Table 9
Comparisons of existing techniques of black spot.

Ref.	Year	Accuracy
Bulanon (2013)	2013	92%
Kim (2014)	2014	97%
Pourreza (2016)	2016	94%
Proposed	–	98.7%

Table 10
Performance measures on canker.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.991	0.009	99.2	0.9	0.8	0.9942	99.1%
KNN	0.982	0.018	98.4	1.7	1.6	1	98.3%
EBT	0.965	0.034	96.5	3.5	3.4	0.9981	96.5%
DT	0.982	0.018	98.4	1.7	1.6	1	98.3%
LDA	0.975	0.025	97.4	2.6	2.6	1	97.4%

Table 11
Confusion matrix of canker disease.

Leaf Types Image	No. of images	Unhealthy (%)	Healthy (%)
Unhealthy	60	100%	0%
Healthy	60	1.8%	98.2%

Table 12
Comparisons of existing techniques of canker.

Ref.	Year	Accuracy
Qin et al. (2008)	2008	92.7%
Kim (2009)	2009	96%
Zhao (2009)	2009	93.3%
Qin (2012)	2012	95.3%
Proposed	–	99.1%

Table 13
Performance measures on citrus scab.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.976	0.023	97.6	2.4	2.4	0.9715	97.6%
W-KNN	0.950	0.049	95.5	4.7	4.5	0.9506	95.3%
EBT	0.929	0.070	92.9	7.1	7.1	0.9813	92.9%
DT	0.942	0.058	94.0	5.9	5.9	0.9315	94.1%
LDA	0.963	0.036	96.5	3.5	3.4	0.9520	96.5%

Table 14
Confusion matrix of citrus scab disease.

Leaf Types Image	No. of images	Unhealthy (%)	Healthy (%)
Unhealthy	50	97.4%	2.6%
Healthy	50	2.2%	97.8%

Table 15
Comparisons of existing techniques of citrus scab.

Ref.	Year	Accuracy
Zhao (2009)	2009	93.3%
Jhuria et al. (2013)	2013	90%
Niphadkar (2013)	2013	95.7%
Dubey and Jalal (1405)	2014	93%
Proposed	–	97.6%

Table 16
Performance measures on greening.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.968	0.031	97.1	3.1	2.8	0.9414	96.8%
W-KNN	0.938	0.062	93.8	6.2	6.2	0.9380	93.8%
EBT	0.948	0.052	94.9	5.2	5.1	0.9489	94.8%
DT	0.958	0.042	96.0	4.1	4.0	0.9310	95.8%
LDA	0.948	0.051	94.9	5.2	5.1	0.9481	94.8%

clear that the proposed method performs significantly better as compared to existing methods.

4.2.5. Greening disease

The greening disease is known as a yellow dragon disease and it is created by bacteria pathogen. The greening disease is difficult to control in the affected plants. The affected plants yield green fruits and are unsuitable for sale. Once a plant becomes affected, there is no cure and the effected plant dies. The proper care of plants including weed control, foliar nutrition, irrigation, soil applied fertilizer and efficient psyllid can maintain the plants productive.

For experimental results, 100 greening images are selected based on 50:50 proportions for training and testing. Table 16 depicts the comparison results of proposed method with four classification techniques. From Table 16, it is clear that proposed method achieved good results having accuracy of 96.8% and FP rate of 0.031, which is confirmed by the confusion matrix in Table 17. In Table 18, proposed method is compared with existing greening disease methods. From Table 18 it is clear that the proposed method perform significantly better as compared to existing methods.

4.2.6. Melanose disease

Melanose disease is created by Diaporthe citri. Its severity can be found by the inoculum on dead wood in the trees. A small brown spot on leaves and fruits appears in the symptom of this disease. These spots are saturated with reddish-brown gum and raised on the leaves surface.

For experimental results, 70 melanose images are selected based on the proportion of 35:35. It means that 35 images are being utilized for training process and remaining 35 images are being utilized for the testing process. Table 19 depicts the comparison results of proposed method with four other classification methods. Maximum accuracy of 97.1% is achieved, which is confirmed by the confusion matrix, shown in Table 20. The comparison of proposed method with the existing method is done in Table 21, which shows that proposed method performed well as compared to existing methods.

4.2.7. Final classification

In this section, the classification results are obtained on all selected diseases as described in Table 22 having maximum classification rate of 95.8% on proposed method with AUC 0.98, FNR 4.2 and FPR is 0.02. The lowest accuracy is achieved on LDA 93.5% with FPR 0.03 and FNR is 6.5. The proposed classification results are confirmed by their confusion matrix in Table 23. The final classification results show that the proposed method performs significantly well as compared to other reported.

Table 17
Confusion matrix of greening disease.

Leaf Types Image	No. of images	Unhealthy (%)	Healthy (%)
Unhealthy	50	98.0%	2.0%
Healthy	50	6.4%	95.6%

Table 18
Comparisons of existing techniques of greening.

Ref.	Year	Accuracy
Sankaran et al. (2010)	2010	95%
Mishra (2012)	2012	90%
Li (2012)	2012	90%
Sankaran (2013)	2013	87%
Xiaoling (2016)	2016	92%
Proposed	–	96.8%

Table 19
Performance measures on melanose.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.972	0.027	97.1	2.9	2.8	0.9713	97.1%
W-KNN	0.942	0.055	94.6	5.8	5.4	0.9444	94.2%
EBT	0.957	0.042	95.6	4.3	4.4	0.9797	95.7%
DT	0.958	0.041	95.8	4.3	4.1	0.930	95.7%
LDA	0.916	0.083	92.3	8.7	7.7	0.9292	91.3%

Table 20
Confusion matrix of melanose disease.

Leaf Types Image	No. of images	Unhealthy (%)	Healthy (%)
Unhealthy	35	100%	0%
Healthy leaf	35	5.8%	94.2%

Table 21
Comparisons of existing techniques of melanose.

Ref.	Year	Accuracy
Qin et al. (2008)	2008	96%
Qin et al. (2008)	2008	92.7%
Zhao (2009)	2009	93.3%
Niphadkar (2013)	2013	95.7%
Pujari et al. (2013)	2013	91%
Proposed	–	97.1%

Table 22
Classification results of all six types of diseases.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.956	0.02	95.3	4.2	4.68	0.98	95.8%
W-KNN	0.935	0.03	93.3	6.2	6.7	0.97	93.8%
EBT	0.943	0.02	94.3	5.5	5.65	0.98	94.5%
DT	0.941	0.02	93.9	5.5	6.08	0.98	94.5%
LDA	0.934	0.03	93.2	6.5	6.75	0.97	93.5%

Table 23
Confusion matrix of all six types of diseases. A (Anthracnose), B (Black spot), C (Canker), S (Scab), G (Greening), and M (Melanose).

Disease	A	B	C	S	G	M
A	100%					
B		92.5%		2.0%		1.0%
C			100%			
S				100%		
G				2.7%	2.0%	95.3%
M				6.0%	6.0%	2.0%
						86.0%

4.3. Combined dataset

In this section, we combined the four diseases of two citrus image datasets such as Plant Village Dataset and Images Database of Infested with Scale. The combined dataset consists of total 5632 including four

Table 24

Classification results on combined dataset.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.897	0.032	90.25	11.0	9.75	0.972	89.0%
W-KNN	0.877	0.042	88.25	12.8	11.75	0.960	87.0%
EBT	0.885	0.040	88.75	11.50	11.25	0.970	88.5%
DT	0.865	0.045	87.0	13.50	13.0	0.975	86.5%
LDA	0.852	0.050	86.0	14.7	14.0	0.932	85.3%

Table 25

Confusion matrix of combined dataset.

Disease Class	Disease Class			
	Anthracnose	Black Spot	Canker	Greening
Anthracnose	72%	28%		
Black Spot	13%	87%		
Canker			100%	
Greening				100%

citrus diseases such as anthracnose, blackspot, canker, and greening. The sample images are shown in Fig. 9 (c). For evaluation results, 50:50 strategy is adopted and all results are obtained using 10-fold cross validation. The maximum classification results are obtained on SVM having classification rate of 89.0% with FPR = 0.032. The execution time of SVM on proposed method is 3.467 images per second. The classification results are described in Table 24, which is confirmed by their confusion matrix given in Table 25. The classification results confirmed that the proposed method performed significantly well on combined dataset.

4.4. Our own dataset

The new dataset consists of total 580 images of four types of citrus diseases. The citrus diseases are greening, black spot, canker, scab, and healthy. Each collected image has dimension 520x530 with 300 dpi. The images are collected in outdoor sequence; therefore, it is a challenging database because of background complexity. The sample images are shown in Fig. 9(b). Moreover, the annotated images with their respective labels are shown in Fig. 11. For experimental results, 50: 50 strategy is used using 10 fold cross-validation. The maximum classification accuracy of 90.4% is achieved on proposed method as presented in Table 26, which is confirmed by their confusion matrix given in Table 27. The per-image execution time of proposed method is 3.379 s and average execution time of all selected classification methods is 12.267 s per image.

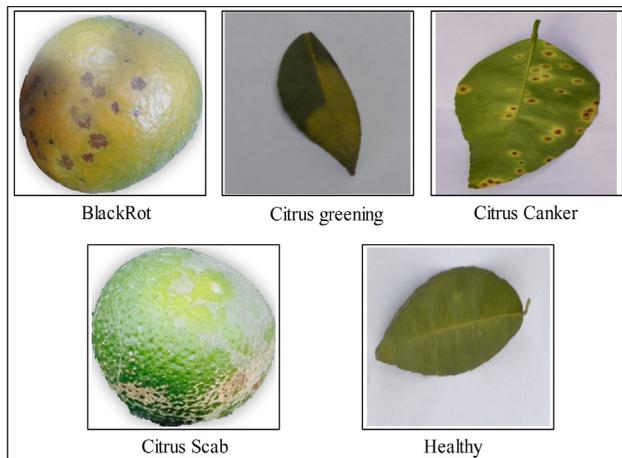


Fig. 11. Annotated samples of our own dataset.

Table 26

Classification results on Own recorded database.

Classifiers	TPR	FPR	PPV	FNR	FDR	AUC	Accuracy
Proposed	0.904	0.024	89.8	9.6	10.2	0.976	90.4%
W-KNN	0.848	0.038	85.2	15.2	14.8	0.964	84.8%
EBT	0.864	0.030	86.2	13.6	13.8	0.972	86.4%
DT	0.856	0.036	85.8	14.4	14.2	0.966	85.6%
LDA	0.832	0.042	83.6	16.8	16.4	0.894	83.2%

Table 27

Confusion matrix of Own recorded database.

Disease Class	Disease Class				
	greening	Black Spot	Canker	Scab	Healthy
greening	88%	8%			8%
Black	4%	84%			12%
Canker			100%		
Scab	8%	11%		81%	
Healthy					100%

5. Discussion

In general, the proposed technique comprised of four major steps: preprocessing, lesion segmentation, features extraction, and classification (as shown in Fig. 1). Initially, the contrast of input image is improved as shown in Fig. 2(d), which is utilized for lesion segmentation using proposed optimized weighted segmentation method as shown in Figs. 3, 4, and 6. In Fig. 5, compared the performance of proposed segmentation algorithm with EM, Otsu, and active contour and best accuracy is obtained 97.8%. Fig. 6 shows the comparison of proposed segmentation results with their ground truth image. Moreover, the accuracy of proposed segmentation method is compared with some existing techniques such as EM, Otsu, active contour, and saliency as given in Tables 28 and 29. From Table 29, segmentation method is tested on two different datasets including Citrus Diseases Image Gallery dataset and Plant Village dataset and obtained segmentation accuracy 98.5086 and 96.56%, respectively. Moreover, the output of proposed method along with the original image, which is taken from Plant Village dataset is shown in Fig. 12.

Table 28

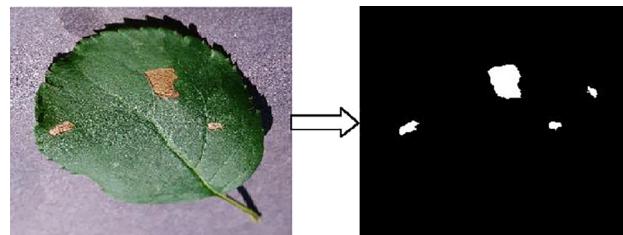
Segmentation comparison with existing methods on 20 images.

Comparison of segmentation accuracy				
Image #	Otsu Method	EM Segmentation	Saliency Method	Proposed
1	90.234	65.231	92.245	98.0132
2	78.345	67.567	93.005	97.4836
3	79.567	78.654	93.221	97.1459
4	83.230	69.999	89.999	96.6217
5	81.220	65.456	94.456	93.2438
6	89.321	78.345	88.888	92.4239
7	74.239	58.235	86.456	88.7387
8	88.321	56.221	80.000	96.4678
9	84.777	69.001	93.221	94.0273
10	89.001	70.245	93.678	89.8438
11	91.213	58.925	90.421	93.9654
12	76.124	57.145	89.990	92.0888
13	72.156	68.888	92.556	95.7983
14	81.780	66.003	91.335	95.4292
15	81.254	65.156	90.346	87.6195
16	90.901	62.345	86.667	91.6642
17	83.478	79.999	86.337	82.852
18	87.267	80.213	86.212	81.7988
19	86.543	56.113	90.215	87.4753
20	90.112	50.256	93.999	96.0184
AVG	83.954	66.199	90.162	92.435

Table 29

Segmentation accuracy results.

Image No	Accuracy (%)	Image No	Accuracy (%)	Image No	Accuracy (%)	Image No	Accuracy (%)
(a)							
1	98.0132	20	96.0184	39	97.8224	58	97.1106
2	97.4836	21	93.2269	40	91.3963	59	94.6013
3	97.1459	22	94.2562	41	94.3351	60	96.1559
4	96.6217	23	98.4293	42	98.5086	61	86.6012
5	93.2438	24	81.4153	43	96.8544	62	87.4497
6	92.4239	25	92.0286	44	91.3327	63	88.6792
7	88.7387	26	91.5013	45	88.3498	64	85.9226
8	96.4678	27	83.151	46	90.3054	65	95.792
9	94.0273	28	96.4661	47	86.5575	66	89.4099
10	89.8438	29	95.1316	48	97.4818	67	91.4281
11	93.9654	30	88.6619	49	91.5132	68	91.239
12	92.0888	31	80.876	50	91.6597	69	93.9629
13	95.7983	32	95.6731	51	86.1566	70	95.0958
14	95.4292	33	98.3855	52	95.961	71	92.1419
15	87.6195	34	92.7112	53	91.2761	72	94.7185
16	91.6642	35	94.6493	54	94.0075	73	95.6711
17	82.852	36	95.2686	55	90.9285	74	93.3446
18	81.7988	37	90.5878	56	93.9371	75	88.7121
19	87.4753	38	90.1722	57	90.2964	76	98.2456
(b) Plant Village Dataset							
1	90.23	6	90.99	11	86.56	16	88.12
2	91.11	7	90.01	12	90.45	17	87.76
3	87.78	8	92.45	13	94.45	18	92.24
4	95.44	9	89.99	14	93.35	19	94.44
5	96.56	10	88.45	15	95.45	20	94.48

**Fig. 12.** Proposed segmentation effects on Plants Village Dataset.

Thereafter, three types of features such as color, texture and geometric are fused for construction of a codebook vector and then feature selection method is performed for selecting the most prominent features as shown in Fig. 8.

Finally, the selected features are fed to SVM and obtain classification results are 96.9%, 98.7%, 99.1%, 97.6%, 95.8%, and 97.1% for anthracnose, black spot, canker, citrus scab, greening and melanose leaves, respectively. The proposed performance measures results are shown in Table 4, 7, 10, 13, 16, and 19, which is confirmed by their confusion matrix in Table 5, 8, 11, 14, 17, and 20. The comparisons of the proposed algorithm with existing techniques are presented in table 6, 9, 12, 15, 18, 21 and it shows the authenticity of the proposed method. Moreover, the proposed method is tested on two other datasets such as combined dataset and our local dataset. Obtained recognition accuracy of 89.0% and 90.4% is achieved as presented in Table 24 and 26, which is confirmed by their confusion matrix given in Table 25 and 27. All results indicate that the proposed method performs significantly well as compared to existing methods and also shows improved performance on M-SVM.

6. Conclusion

In this article, a hybrid approach is proposed for automatic detection and classification of six types of citrus diseases based on an optimized weighted segmentation and feature selection methods. The pre-processing step enhances the contrast of lesion spot and extracts these spots with a new optimized weighted segmentation method. After that,

a codebook of three types of features (color, texture, and geometric) is constructed. A feature selection method is applied to feature codebook and best features are selected. The selected features are classified by M-SVM. Obtained results show that the preprocessing method improves the contrast of lesion part, which in turn, improves the segmentation accuracy. We obtained average segmentation accuracy of 92.435% on all datasets used. Moreover, the fusion of color, texture, and geometric features are performed as compared to the individual features. For each citrus disease, we performed a classification process and compared the proposed results with four classification methods (i.e. EBT, KNN, DT, and LDA) obtaining the best accuracy of 95.8%. Also, a comprehensive comparison shows that the proposed method performs significantly well as compared to existing methods. In future, we would like to construct a deep model and apply the model on the selected citrus datasets, as the deep learning performed significantly well in the field of computer vision. However, we need a big citrus dataset for it.

References

- Abdullah, N.E., et al. 2012. A characterization of watermelon leaf diseases using Fuzzy Logic. in Business, Engineering and Industrial Applications (ISBEIA), 2012 IEEE Symposium on 2012. IEEE.
- Ahmed, N., Khan, U.G., Asif, S., 2016. An automatic leaf based plant identification system. *Sci. Int.* 28 (1), 427–430.
- Albrechtsen, F., 2008. Statistical texture measures computed from gray level cooccurrence matrices. Image processing laboratory, department of informatics, university of oslo, 5.
- Ali, H., et al., 2017. Symptom based automated detection of citrus diseases using color histogram and textural descriptors. *Comput. Electron. Agric.* 138, 92–104.
- Arivazhagan, S., et al., 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agric. Eng. Int.: CIGR J.* 15 (1), 211–217.
- Bandi, S.R., Varadharajan, A., Chinnasamy, A., 2013. Performance evaluation of various statistical classifiers in detecting the diseased citrus leaves. *Int. J. Eng. Sci. Technol.* 5 (2), 298–307.
- Bulanon, D.M., et al., 2013. Citrus black spot detection using hyperspectral image analysis. *Agric. Eng. Int.: CIGR J.* 15 (3), 171–180.
- Citrus Diseases Image Gallery, December 20, 2017. [Online] Available <http://idtools.org/id/citrus/diseases/gallery.php>. [Accessed: Dec. 20, 2017].
- Cubero, S., et al., 2016. Automated systems based on machine vision for inspecting citrus fruits from the field to postharvest—a review. *Food Bioprocess Technol.* 9 (10), 1623–1639.
- Deng, X., et al., 2016. Citrus greening detection using visible spectrum imaging and C-SVC. *Comput. Electron. Agric.* 130, 177–183.

- Dewdney, M., Timmer, L., 2009. Citrus production and diseases in Ghana. Citrus Industry. Available online: <http://www.crec.ifas.ufl.edu/academics/faculty/dewdney/PDF/Ghanatripstory.pdf>. Retrieved February, 2009. 19: p. 2014.
- Dubey, S.R., A.S. Jalal, Adapted approach for fruit disease identification using images. arXiv preprint arXiv:1405.4930, 2014.
- Gavahale, K.R., Gawande, U., 2014. An overview of the research on plant leaves disease detection using image processing techniques. *IOSR J. Comput. Eng. (IOSR-JCE)*. 16(1): p. 10–16.
- Gavahale, K.R., Gawande, U., Hajari, K.O., 2014. Unhealthy region of citrus leaf detection using image processing techniques. In *Convergence of Technology (I2CT), 2014 International Conference for IEEE*.
- Gómez-Sanchis, J., et al., 2008. Automatic correction of the effects of the light source on spherical objects. An application to the analysis of hyperspectral images of citrus fruits. *J. Food Eng.* 85 (2), 191–200.
- Guth, F.A., Ward, S., McDonnell, K.P., 2017. Autonomous disease detection in crops using deep learning. *Biosyst. Food Eng. Res. Rev.* 22, 209.
- Haghigat, M., Abdel-Mottaleb, M., Alhalabi, W., 2016. Discriminant correlation analysis for feature level fusion with application to multimodal biometrics. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on IEEE*.
- Haralick, R.M., Shanmugam, K., 1973. Textural features for image classification. *IEEE Trans. Syst., Man, Cybernetics* 6, 610–621.
- Jaganathan, V., Arumugam, S., 2014. Powdery mildew disease identification in karpoori variety of betel vine plants using histogram based techniques. *Adv. Image Video Process.* 2 (5), 63–75.
- Jhuria, M., Kumar, A., Borse, R., 2013. Image processing for smart farming: Detection of disease and fruit grading. In *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on IEEE*.
- Jolly, P., Raman, S., 2016. Analyzing Surface Defects in Apples Using Gabor Features. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on IEEE*.
- Kakade, N.R., D.D. Ahire, A review of grape plant disease detection. *Int. Res J. Eng. Technol. (IRJET)*, 2015. 20(5).
- Kim, D.G., et al., 2009. Classification of grapefruit peel diseases using color texture feature analysis. *Int. J. Agric. Biol. Eng.* 2 (3), 41–50.
- Kim, D., et al., 2014. Citrus black spot detection using hyperspectral imaging. *Int. J. Agric. Biol. Eng.* 7 (6), 20.
- Kim, J., et al., 2016. Salient region detection via high-dimensional color transform and local spatial support. *IEEE Trans. Image Process.* 25 (1), 9–23.
- Kumar, C., Chauhan, S., Alla, R.N., 2015. Classifications of citrus fruit using image processing-GLCM parameters. In *Communications and Signal Processing (ICCP), 2015 International Conference on IEEE*.
- Li, X., et al., 2012. Spectral difference analysis and airborne imaging classification for citrus greening infected trees. *Comput. Electron. Agric.* 83, 32–46.
- Liu, Y., Zheng, Y.F., 2005. One-against-all multi-class SVM classification using reliability measures. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on 2005*. IEEE.
- Malik, Z., et al., 2016. Detection and Counting of On-Tree Citrus Fruit for Crop Yield Estimation. *IJACSA International Journal of Advanced Computer Science and Application*. 7(5).
- Mishra, A., et al., 2012. Identification of citrus greening(HLB) using a VIS-NIR spectroscopy technique. *Trans. ASABE* 55 (2), 711–720.
- Naik, M.R., Sivappagari, C.M.R., 2016. Plant leaf and disease detection by using HSV features and SVM classifier. *Int. J. Eng. Sci.* 3794.
- Narvekar, P.R., Kumbhar, M.M., Patil, S., 2014. Grape leaf diseases detection & analysis using SGDM matrix method. *Int. J. Innovative Res. Comput. Commun. Eng.* 2 (3), 3365–3372.
- Niphadkar, N.P., et al., 2013. Estimation of citrus canker lesion size using hyperspectral reflectance imaging. *Int. J. Agric. Biol. Eng.* 6 (3), 41–51.
- Omidi, M., Khojastehnazar, M., Tabatabaeefar, A., 2010. Estimating volume and mass of citrus fruits by image processing technique. *J. Food Eng.* 100 (2), 315–321.
- Pourreza, A., et al., 2016. Spectral characteristics of citrus black spot disease. *HortTechnol.* 26 (3), 254–260.
- Pujari, J.D., Yakkundimath, R., Byadgi, A.S., 2013. Automatic fungal disease detection based on wavelet feature extraction and PCA analysis in commercial crops. *Int. J. Image, Graph. Signal Process.* 6 (1), 24.
- Pydipati, R., Burks, T., Lee, W., 2005. Statistical and neural network classifiers for citrus disease detection using machine vision. *Trans. ASAE* 48 (5), 2007–2014.
- Qin, J., et al., 2012. Development of a two-band spectral imaging system for real-time citrus canker detection. *J. Food Eng.* 108 (1), 87–93.
- Qin, J., et al. Classification of citrus peel diseases using color texture feature analysis. In: *Food Processing Automation Conference Proceedings*, 28–29 June 2008, Providence, Rhode Island. 2008. American Society of Agricultural and Biological Engineers.
- Qin, J., et al. 2008. Detecting citrus canker by hyperspectral reflectance imaging and PCA-based image classification method. In *SPIE Defense and Security Symposium*. 2008. Int. Soc. Optics Photonics.
- Qin, J., et al. 2008. Classification of citrus peel diseases using color texture feature analysis. In *American Society of Agricultural and Biological Engineers-Food Processing Automation Conference 2008*.
- Sankaran, S., et al., 2013. Huanglongbing (citrus greening) detection using visible, near infrared and thermal imaging techniques. *Sensors* 13 (2), 2117–2130.
- Sankaran, S., Ehsani, R., Etxeberria, E., 2010. Mid-infrared spectroscopy for detection of Huanglongbing (greening) in citrus leaves. *Talanta* 83 (2), 574–581.
- Singh, V., Misra, A., 2015. Detection of unhealthy region of plant leaves using Image Processing and Genetic Algorithm. In *Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in*. 2015. IEEE.
- Stegmayer, G., et al., 2013. Automatic recognition of quarantine citrus diseases. *Expert Syst. Appl.* 40 (9), 3512–3517.
- Sun, Q.-S., et al., 2005. A new method of feature fusion and its application in image recognition. *Pattern Recogn.* 38 (12), 2437–2448.
- Thapar, S., Garg, S., 2012. Study and Implementation of Various Morphology Based Image Contrast Enhancement Techniques. *Int. J. Comput. Business Res.*
- Timmer, Lavern Wayne, J.H. S.M., Graham, Compendium of citrus diseases. No. 634.3 C737 2000. American Phytopathological Society, St. Paul, Minn.(EUA), 1988.
- Vishnu, S., Ranjith, A., 2015. Ram, Plant disease detection using leaf pattern: A review. *Int. J. Innovative Sci., Eng. Technol.* 2 (6), 774–780.
- Wetterich, C.B., et al., 2016. Detection of citrus canker and Huanglongbing using fluorescence imaging spectroscopy and support vector machine technique. *Appl. Optics* 55 (2), 400–407.
- Xiaoling, D., et al., 2016. Detection of citrus Huanglongbing based on image feature extraction and two-stage BPNN modeling. *Int. J. Agric. Biol. Eng.* 9 (6), 20.
- Yesuf, M., 2013. Pseudocercospora leaf and fruit spot disease of citrus: Achievements and challenges in the citrus industry: A review. *Agric. Sci.* 4 (07), 324.
- Yun, S., et al., 2015. PNN based crop disease recognition with leaf image features and meteorological data. *Int. J. Agric. Biol. Eng.* 8 (4), 60.
- Zhang, M., Meng, Q., 2011. Automatic citrus canker detection from leaf images captured in field. *Pattern Recogn. Lett.* 32 (15), 2036–2046.
- Zhao, X., et al., 2009. Digital microscopic imaging for citrus peel disease classification using color texture features. *Appl. Eng. Agric.* 25 (5), 769–776.
- Zhexin, Cen, Baoju, Li, Yanxia, Shi, Haiyang, Huang, Jun, Liu, Ning Fang, Liao, Jie, Feng, 2007. Identification of cucumber anthracnose and brown spot based on color statistics of color images. *J. Horticulture* 34 (6), 1425–1430.