Original Research Paper

Detection of Plant Leaf Diseases Using K–mean++ Intermeans Thresholding Algorithm

[1]Kittipol Wisaeng (Member IEEE) and [2]Worawat Sa–Ngiamvibool

[1]*Technology and Business Information System Unit,
Mahasarakham Business School, Mahasarakarm University, Mahasarakham 44150, Thailand*
[2]*Department of Electrical and Computer, Faculty of Engineering, Mahasarakham University, Mahasarakham 44150, Thailand*

Corresponding Author:
Kittipol Wisaeng
Technology and Business
Information System Unit,
Mahasarakham Business
School, Mahasarakarm
University, Mahasarakham
44150, Thailand
Email: Kittipol.W@acc.msu.ac.th

Abstract: In the field of agricultural information, the plant leaf disease detection is highly important for both farmer life and environment. To improve the accuracy of plant leaf disease detection and reduce the image processing time, the improved K–mean++ clustering and intermeans thresholding method are proposed in this study. The proposed algorithms are used for training and testing diseases in plant leaf images in two different databases. Of the proposed methods, the intermeans algorithm will be selected based on different thresholding values. The optimal value of thresholding-i.e., the intermeans algorithm-will help increase the accuracy and speed of classifying diseases in plant leaf images. This method will be also used with unseen images of plant leaf. The experimental result of the detection of plant leaf diseases achieves an average detection accuracy of 98.10%. When compared with the results based on standard K–mean clustering, the current method gives better results around 23.20%. The proposed algorithm is more effective than the standard algorithms for detecting plant leaf diseases, as well as the reduction in cots in the computational power of computers.

Keywords: Plant Leaf Disease, K–mean++ Clustering, Intermeans Thresholding

Introduction

Plant leaf detection is an important method for future recognition of plant leaf diseases. In Thailand, grapes are considered to be important food and one of the most widely produced crops. However, grape's leaf diseases are still problematical for its growth. If there is an effective method available for early screening, the possibility of plant leaf diseases will decrease. Normally, the procedures of plant leaf disease segmentation and detection require human involvement. There have been a number of methods, both traditional methods and soft computing methods, proposed for detecting plant leaf diseases. However, these methods still need some image processing steps which result in time consumption. This is because the image processing method requires properties such as the following: Size, color, shape, and texture.

Generally, the image processing method used for detecting plant leaf diseases is divided into two classes, namely; (1) Traditional method and (2) Soft computing method. In the first class, some of the traditional methods widely used for segmenting plant leaf diseases are thresholding method, edge–based segmentation, clustering algorithm and region–based segmentation. In the second class, which is soft computing or special theory based segmentation, there are such methods as fuzzy C–mean clustering algorithm, neural network, physically based segmentation, region growing method, K–mean algorithm and genetic algorithm, used for plant leaf disease detection. Each of these two classes has both advantages and disadvantages which can be beneficial for the current study.

In this study, an automatic detection of plant leaf diseases is proposed. The proposed algorithm uses the combination of K–mean++ clustering and intermeans thresholding method for the classifying and detecting diseases present on the plant leaf images. The proposed method comprises three stages as follows:

(1) The efficiency of the K–mean++ (KM++) clustering algorithm is used to group attributes believed to the pixels of diseases and non-diseases. This stage is beneficial for feature extraction process. When used for segmenting plant diseases, KM++ clustering

method is potential and powerful for the coarse segmentation stage

(2) The result from the coarse segmentation stage will be used to fine segmentation for obtaining the plant leaf disease candidates pixels by using an intermeans thresholding method

(3) The result from intermeans thresholding method will be transferred to binary images in order to measure the performance of the proposed methods at the pixel-based evaluation. As well, the result from intermeans thresholding method will be compared with the performance of the work in the literature

Related Work

Plant leaf disease detection is a complicated task to perform. The previous studies have used both traditional methods and soft computing algorithms to detect the infected plant leaves. These methods and algorithms are summarized in Table 1. Table 1 show both various kinds of plants and different algorithms which have been used for detecting diseases. For instances, SVM classifier is by (Islam *et al.*, 2017; Patil *et al.*, 2017) for classifying potato leaf diseases. The same classifier is also proposed by (Hossain *et al.*, 2018) to segment tea leaf diseases while (Masazhar and Kamal, 2017) adopts SVM classifier for identifying palm oil leaf diseases. Interestingly, the classifier of SVM is also widely applied by different scholars such as (Krithika and Veni, 2017) for cucumber leaf disease classification, (Padol and Yadav, 2016) for grape leaf disease detection, (Neumann *et al.*, 2014) for classifying beet leaf diseases and (Zhou *et al.*, 2014) for Cercospora leaf disease identification. Even though SVM classifier has been presented by a number of researchers, the method still has a certain disadvantage. The disadvantage is that it is time consuming in the training process. Moreover, when applied to classify different plant diseases, the performance of the segmentation stage shows no robustness. Another method, Neural Network (NN) algorithm, coupled with shuffle frog leap algorithm, is used by (Guo *et al.*, 2018) for classifying plant diseases. Other researchers using NN for their classifications of plant diseases include (Kurale and Vaidya, 2018; Wang *et al.*, 2012; Prajwala *et al.*, 2018). In addition, (Anand *et al.*, 2016) uses standard K–means clustering method for identifying Brinjal leave diseases.

After having tried out the reviewed methods, it is found that the current algorithms of plant leaf disease detection are manual, expensive; require highly trained personal to facilitate the process of searching disease from a large number of plant leaf images. Clustering algorithms have also been proposed as a possible solution to the segmentation and detection of plant leaf problems. However, the main problem with clustering algorithms is determining the number of clusters to use. Machine learning algorithms have been proposed as suitable to segment of plant leaf disease, but the disadvantage of machine learning algorithms is to take a time to training process and require many predetermined features. Then, to classify the detected regions into disease and non-disease, the standard K–mean clustering algorithm was investigated. Afterward, to improve the accuracy of plant leaf disease detection and reduce the image processing time, the improved K–mean++ clustering and intermeans thresholding method are proposed in this study.

Materials and Methods

This section focuses on plant leaf disease detection by using a combination of KM++ clustering and intermeans thresholding method. In the coarse segmentation stage, KM++ clustering method with the optimal weight is applied to classify diseases on plant leaf images. The KM++ clustering algorithm's efficiency is applied for clustering features assumed to be the diseases and non-diseases pixels. Afterward, by using an intermeans thresholding method, to obtain the plant leaf disease candidates pixels, the result from the coarse segmentation stage is utilized in order for fining segmentation. Finally, the binary images will be obtained by using the optimal thresholding.

Table 1: Review plant leaf disease detection by using traditional methods and soft computing algorithms

Author	Algorithms	Area
Islam *et al.* (2017)	Image segmentation and multiclass SVM	Potato diseases
Hossain *et al.* (2018)	SVM	Tea leaf diseases
Masazhar and Kamal (2017)	Multiclass support vector machine classifier	Palm oil leaf disease
Krithika and Veni (2017)	SVM	Cucumber leaf
Neumann *et al.* (2014)	SVM	Beet leaf diseases
Zhou *et al.* (2014)	SVM	Cercospora leaf
Guo *et al.* (2018)	Pulse coupled neural network with shuffle frog leap algorithm	Plant diseases
Kurale and Vaidya (2018)	NN classifier	Leaf diseases
Prajwala *et al.* (2018)	Convolutional neural networks	Tomato leaf diseases
Anand *et al.* (2016)	KCM	Brinjal leaf diseases
Trongtorkid and Pramokchon (2018)	Rule–based model	Mango diseases
Korkut *et al.* (2018)	Machine learning methods	Plant leaf diseases

Table 2: Dataset for plant leaf disease image.

| Class images	Total number of images	Number of images for training			Number of images for testing		
		10%	20%	30%	90%	80%	70%
Grape leaf	1,000	100	200	300	900	800	700
Apple leaf	276	27	55	82	249	221	194
Maize leaf blight	986	98	186	295	888	789	691
Pepper bell bacterial spot	998	99	198	299	899	799	699
Peach bacterial spot	2,298	229	458	689	2,069	1,839	1,609
Potato leaf	1,001	100	200	300	901	801	701
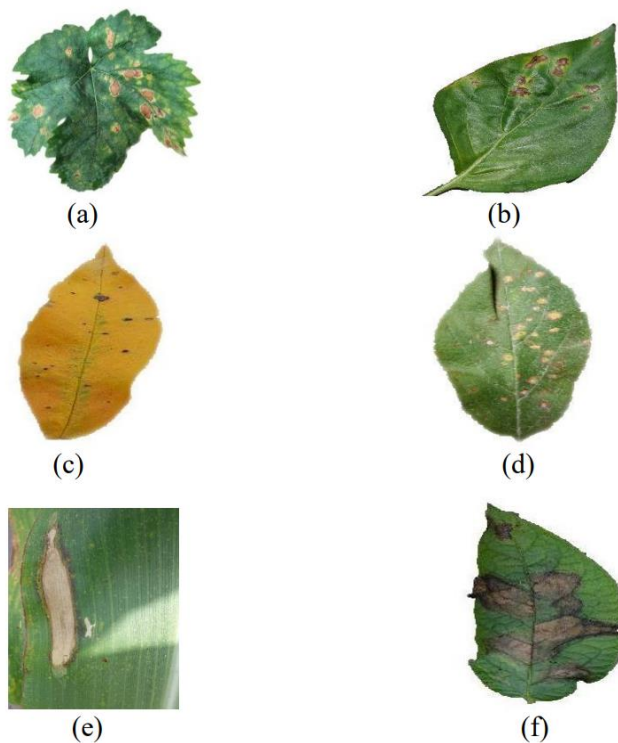
Fig. 1: Six common plant leaf diseases, (a) grape leaf images, (b) pepper bell Bacterial spot, (c) peach Bacterial spot, (d) apple rust, (e) maize leaf blight, (f) potato leaf blight

A. Dataset

At all stages of plant leaf diseases study, an appropriate dataset is required. The dataset is collected from two different databases. A total of 1, 0 0 0 grape leaf images are collected from local databases and the rest are from Plant Village databases, including apple frogeye spot, peach bacterial spot, pepper bell Bacterial spot, maize northern leaf blight and potato leaf blight. The plant leaf dataset contains six kinds of plant diseases, which are shown in Fig. 1. To evaluate the performance of the proposed methods, 6,559 plant leaf images are examined. At the training stage, these images are divided into three different proportions of 10, 20, 30%, respectively. The remaining images are reserved for testing stage (90, 80 and 70%). Table 2 presents the dataset of plant leaf disease images.

B. Plant Leaf Diseases

Plant leaf, like human and animals, suffer from diseases. Hence, the diseases have certain affects to the normal growth of a plant. An accurate diagnosis of plant leaf diseases is vital for the current study. Therefore, plant diagnosis needs considering the following three criteria.

First, by naked eyes, look for diseases such as unwanted spots, blight, rust, and dead areas on the part of the plant leaf. Next, recognize the characteristics of plant diseases. Finally, differentiate disorders and diseases in plants. Whereas disorders are caused by some environmental problems, diseases are the properties affecting the other parts of the plant like leaf and fruit. In this study, as shown in Fig. 2, spots and blight disease detection is mainly focused on.
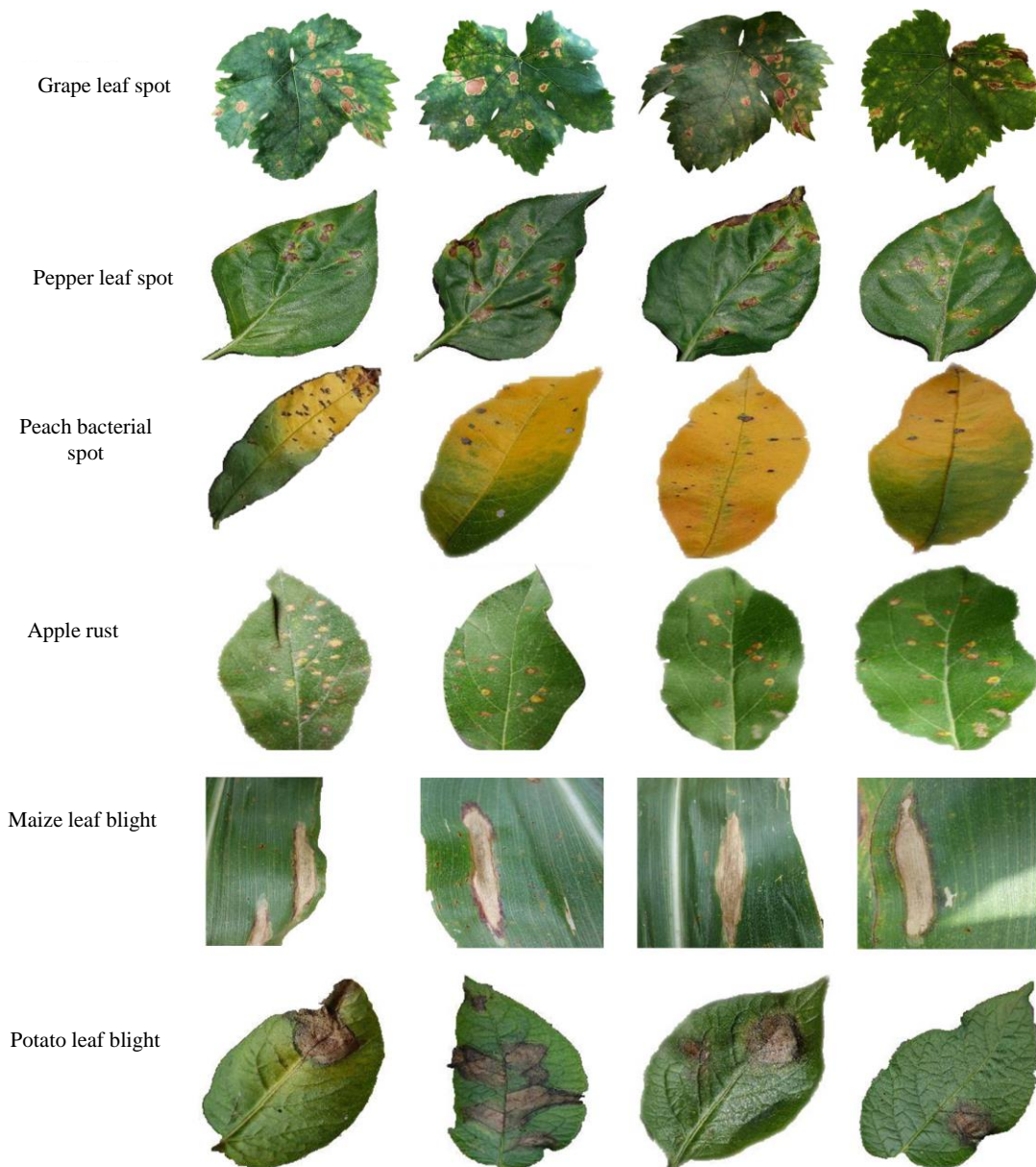
Fig. 2: Example of plant spots and blight disease

Proposed Methods

The combination of computational intelligence methods is used in the proposed method in order to focus on detecting plant leaf diseases. We choose the standard K–mean clustering algorithm to segment of plant disease task. Afterward, a fast version of the KM++ clustering algorithm was investigated.

The standard K–mean clustering, the simplest and originally from signal processing, is a popular unsupervised machine learning algorithm. The standard K–mean clustering's efficiency is being a powerful tool for initializing the weight for training process that can correctly classify different diseases on plant leaf image with high segmentation speed and accuracy. In the interest of confirming the accuracy of the plant leaf disease detection in the dataset, an intermeans thresholding method is used to fine segmentation and compare with agricultural experts examining plant leaf images.

A. Coarse Segmentation Using Standard K–Mean Clustering Algorithm

In this subsection, the standard K–means clustering method is used to classify disease and non–disease plant leaf images. The Standard K–means clustering method is

advantageous as it can be devised which does not require a training dataset (Dhanachandra and Chanu, 2015). In the current study, various values of k are used for coarse segmentation of plant diseases. After that, the values of k from the lowest to the highest are tested to find the most suitable parameter values. Finally, the optimal parameters are estimated in order to identify plant leaf diseases. The pixels are grouped around centroids which use the minimizing objective and are defined as Equation (1):

$$V = \sum_{i=1}^{k} \sum_{x_i \in S_i} \left(x_j - \mu_i \right)^2 \tag{1}$$

where, k refers to the cluster S_i, $i = 1,2,…,k$ and μ_i denotes the average point of all points $x_j \in S_i$. An iterative version of the clustering algorithm is developed in this study. The clustering algorithm is tested by using a 2–dimension image. The standard K–mean clustering algorithm process is as follows:

1. Make initial guesses and let the initial matrix V to the identity matrix
2. Segment the image of plant leaf by using the linear discriminant criterion
3. Repeat the aforementioned steps until no change of the image's cluster labels is found
4. Compute the clusters of centroid intensities by Equation (2):

$$c^i := \arg \min_j \left\| x^{(i)} - \mu_j \right\|^2 \tag{2}$$

5. The new centroid for each of the clusters is computed by using Equation (3):

$$\mu_j := \frac{\sum_{i=1}^{m} 1\left\{ c_{(i)} = j \right\} x^{(i)}}{\sum_{i=1}^{m} 1\left\{ c_{(i)} = j \right\}} \tag{3}$$

where, i refers to all iterates in the intensities, j refers to all iterates in the centroids and μ_i is the centroid intensities. The results segmented by using the standard K–means clustering method to the plant leaf image are shown in Fig. 3. In this process, the values of $k = 2$, run 246 iterations, which is used as the initial values for starting the segmentation. Fig. 3b shows the plant leaf disease image segmented by using $k = 2$ and gives satisfactory results.

B. Proposed Method: Coarse Segmentation Using K–Mean++ Clustering Method

The disadvantage of the standard K–mean clustering algorithm for plant leaf disease detection is that it cannot identify true areas of diseases. Therefore, the effectiveness of KM++ clustering algorithm (Arthur and Vassilvitskii, 2006), believed to be one of the most effective tools and guaranteed to give suitable results of image segmentation, is used to replace the standard K–mean clustering method.

In the testing process, let $D(x)$ be the shortest distance from a data pixel x to the closest center. Then, the KM++ clustering algorithm is defined below:

1. Take one center c_1, chosen uniformly at random from X
2. Take a new center c_{i1}, choosing $x \in X$ with probability $D(x)^2 / \sum_{x \in X} D(x)^2$
3. Repeat Step 2 until the value of k centers is retrieved
4. Repeat Steps 2-4 with the standard K–means clustering algorithm

In Step 2, "D^2 weighting" is the weighting used. In the final segmentation error of standard K–means clustering method, this seeding algorithm gives out considerable improvements. Despite the fact that the initial parameter selections take extra time, the KM++ converges very fast for segmentation stage and thus lowers the computation time. Six images of plant leaf image are tested by using different parameters of KM++ clustering method. The values of these parameters are $k = 2$, $k = 4$ and $k = 8$ respectively. See segmentation results in Fig. 4. Consider Fig. 4, the value $k = 4$ represent the optimal number of leaf disease segmentation present in plant image. As a result, the value of $k = 4$ is used for segmentation of plant leaf diseases in all databases.
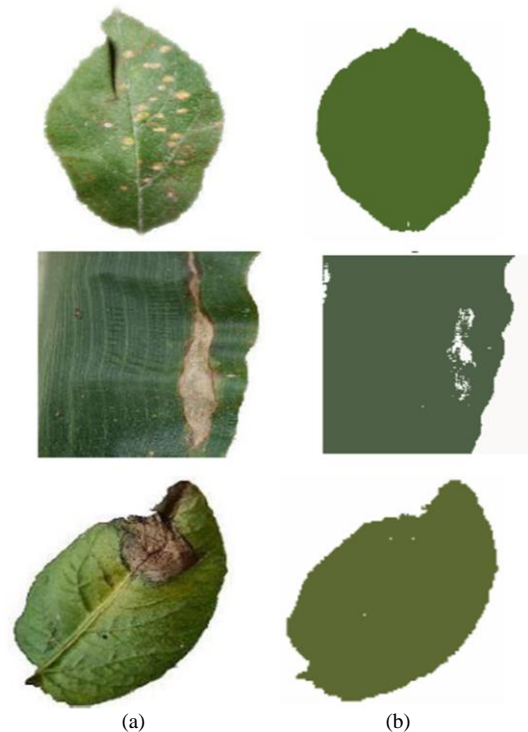

Fig. 3: Segmentations of the plant leaf image into 2 classes. (a) original image, (b) result displayed as an image of k = 2
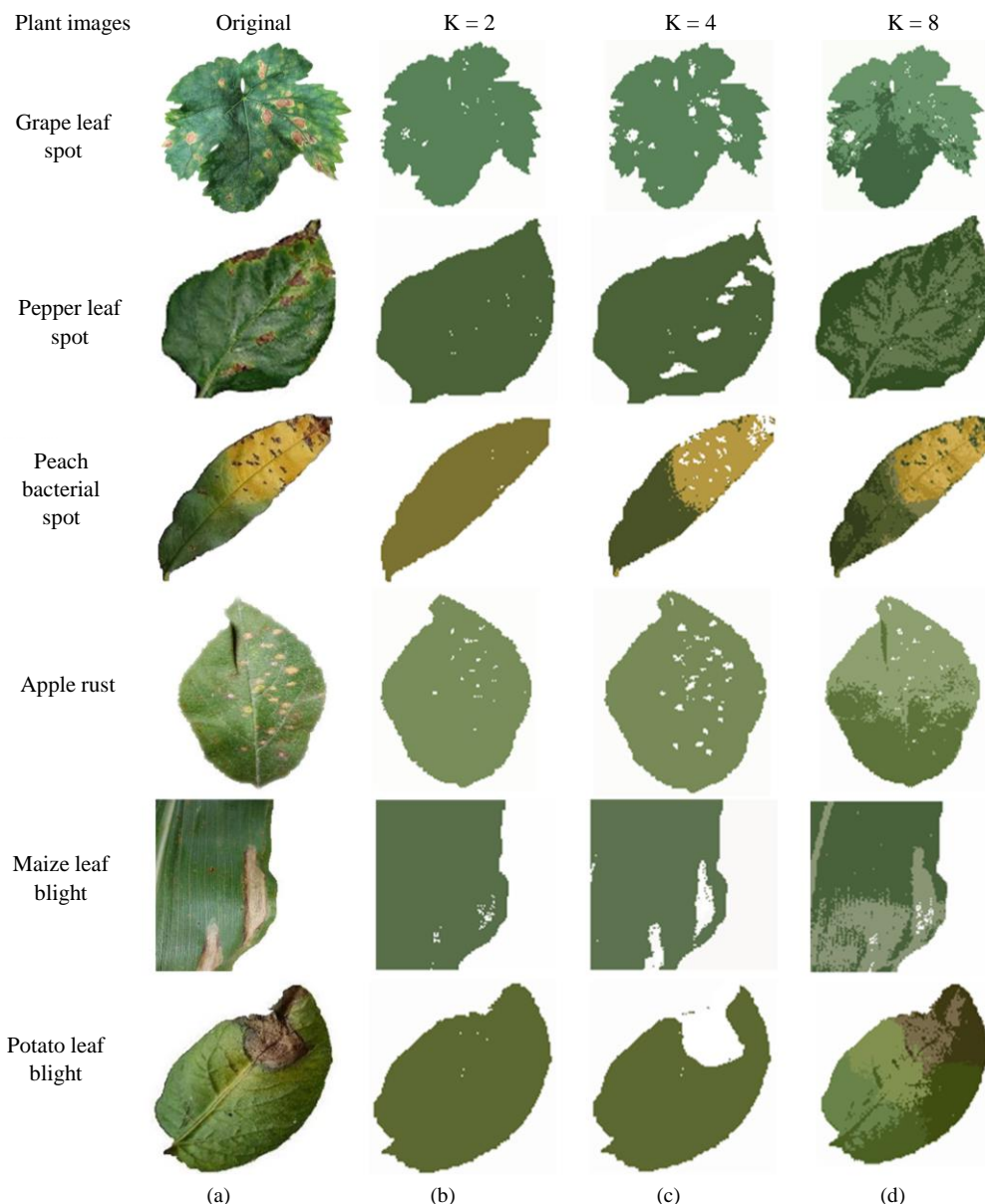
Fig. 4: Results for plant leaf disease segmentation using KM++ clustering method (a) Original image (b) result displayed as an image of k = 2, (c) result displayed as an image of k = 4 and (d) result displayed as an image of k = 8

C. Fine Segmentation Using Thresholding Method

By using thresholding method, fine segmentation is frequently the critical step in the analysis of plant leaf diseases. Nevertheless, the results of fine segmentation by using automatic thresholding methods show low accuracy.

Therefore, to improve the results, manual intervention is considered an appropriate technique. Thresholding methods generally used in previous studies for binary image transformation consist of two categories-global and local thresholding methods. Global thresholding method generally depends on maximizing variances between the classes and minimizing the error within the classes (Otsu, 1975; Zhang and Hu, 2008; De Albuquerque *et al*., 2004; Kapur *et al*., 1985). Local thresholding method uses spatial features of a neighborhood in an image (Niblack, 1985; Sauvola and Pietikäinen, 2000; Chang *et al*., 2000; Ray and Saha, 2007; Chuang *et al*., 2011). Image thresholding methods may also provide promising results for image segmentation with different image datasets. However, image thresholding methods generally take more time when applied for high–resolution images. This paper studies one of the image thresholding methods called intermeans thresholding method, which uses a segmentation of plant leaf image.

Building the Training Set of 120th Intensity Histogram Thresholding (120iht)

In an intensity histogram thresholding, there exists different values of RGB color components useful for segmenting plant leaf diseases. In this stage, an Intensity Histogram Thresholding (IHT) value is used for separating disease and non-disease on plant leaf images. To transform RGB color images to grey images, Equation (4) is presented:

$$\sum_{k=0}^{t} h_k \geq \frac{n^2}{2} > \sum_{k=0}^{t-1} h_k, \qquad (4)$$

where, t represent a grey value and the threshold value, respectively. In thresholding method, any pixel values in the plant leaf images have the range of values between 0 to 255. In this stage, the threshold values of 98, 110, 112 and 120 are chosen for testing on grape leaf images. The results of different threshold values are given in Fig. 5. The threshold has segmented the

image into two predominant types (white pixels represent plant leaf disease and dark pixels represent background) successfully. With a threshold of 120, Fig. 5d looks most outstanding as most of the connected pixels have been correctly classified.

Building the Training Set of 138th Intensity Histogram Thresholding (138iht)

The previous section has shown that the threshold value of 120 is able to utmost classify disease and non–disease. Nonetheless, the identified pixels of disease are not exactly the true diseases. This is because the white pixel areas representing the disease areas have been expanded causing false positive values to have increased. In this stage, a range of values of t = 142, 149, 157 and 215, respectively, is selected for testing. The result of the experiment on grape leaf images is shown in Fig. 6. Between the threshold values of 149 and 215, the threshold value of 157 gives the most suitable result for separating disease and non-disease on the grape leaf images.

Fig. 5: Example of three classification by thresholding for the grape leaf image, (a) segment with t = 98, (b) segment with t = 110, (c) segment with t = 112 and (d) segment with t = 120

Fig. 6: Example of three classification by thresholding for the grape leaf image, (a) segment with t = 142, (b) segment with t = 149, (c) segment with t = 157 and (d) segment with t = 215

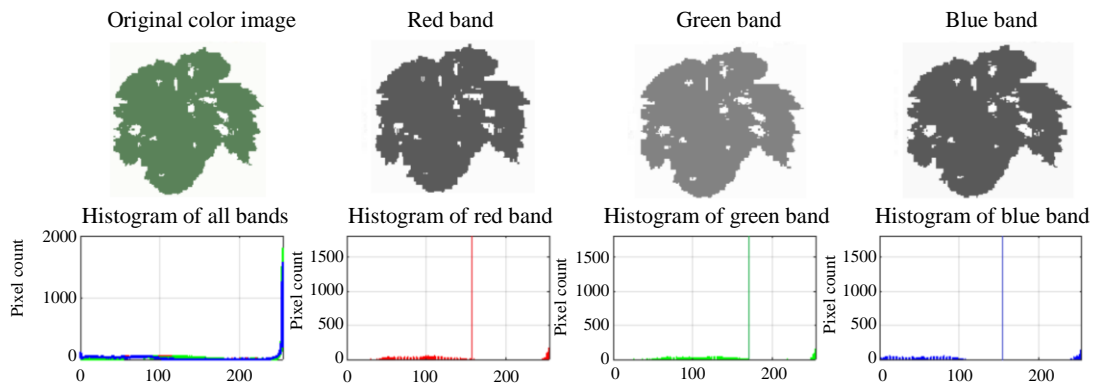**Fig. 7:** RGB histogram of grape leaf image

1243

Histogram–Intermeans Thresholding Algorithm

The histogram values of 120IHT and 138IHT in the previous sections are considered the most suitable values of the two experiments. However, considering the results in Fig. 5 and 6, false positive rate occurs when the threshold value of 120IHT is used. In a similar vein, the threshold value of 138IHT causes the pixels of diseases to partially be removed. It is therefore assumed that the most suitable value of threshold may be between 120IHT and 138IHT. In this study, such value is thus defined as intermeans thresholding algorithm (Ridler and Calvard, 1978). This algorithm is repositioned to lie exactly halfway between the two means. The algorithm can be calculated in the following manner:

1. Let the initial value of t for segmentation of plant leaf images based on the mean pixel value, which is calculated by Equation (5):

$$\sum_{k=0}^{t} h_k \geq \frac{n^2}{2} > \sum_{k=0}^{t-1} h_k, \tag{5}$$

where, h is the histogram of pixel values (h_0, h_1, …, h_N), h_k represents the number of pixels in plant leaf images with greyscale value k and n^2 denotes the number of pixels in an image by $n \times n$

2. Compute the average pixel value in each cluster by considering the threshold values less than or equal to t using Equation (6):

$$\mu_1 = \frac{\sum_{k=0}^{t} h_k}{\sum_{k=0}^{t-1} h_k} \tag{6}$$

Similarly, the threshold values greater than t can be calculated by using Equation (7):

$$\mu_2 = \frac{\sum_{k=t+1}^{N} k h_k}{\sum_{k=t+1}^{N} h_k} \tag{7}$$

where, N represents the maximum pixel value (usually 255)

3. Re–compute the value of t as half–way between the two means by using Equation (8):

$$t = \left[\frac{\mu_1 + \mu_2}{2} \right], \tag{8}$$

4. Repeat Steps (2) and (3) until the value of t stays unchanged. In this stage, Equation (5) – (8) are used

to find the intermeans RGB histogram of plant leaf images

In Fig. 7, assuming that the values of RGB histogram are 167.72, 187.37 and 164.42, respectively, the total pixel value is 519.51 divided by 3 which is 173.17 and is called the intermeans thresholding algorithm. Fig. 8b shows the result of plant leaf diseases by using the intermeans thresholding algorithm. Finally, to detect plant leaf diseases is to identify the region boundaries. In this stage, apply Canny edge operator by using the value of 5 to label edge pixels presented in black lines shown in Fig. 8c.

Experiments and Results

In the proposed method, the algorithm is developed on MATLAB 2018b working on a system with 1.8 GHz Core i5 CPU processor having 8 GB RAM. To evaluate the performance of the algorithms, different two datasets of plant leaf images are used. The first dataset consists of 1,000 grape leaf images (300 for training and 700 for testing) selected from local databases; the second dataset consists of 5,559 images selected from crowdAI.org categorizing among the different 5 sets of diseases. The results of the binary images are used for evaluating the performances of the proposed method. The correct measurement of the method for classifying the detected areas of disease and non–disease on the plant leaf images is evaluated using three quantitative evaluation parameters that are based on the statistical performance and segmented image using the algorithm. Once the binary images are assigned as the reference images, it is possible to calculate the correctness of all plant leaf images by comparing with ground–truth images.

The three quantitative evaluation parameters are specificity, sensitivity, and accuracy by using Equation (9), (10) and (11), respectively:

$$Specificity = \frac{TP}{TN + FN} * 100 \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} * 100 \tag{10}$$

where True Positive (TP) represents the pixel numbers of disease exactly detected, False Positive (FP) denotes the pixel numbers of disease incorrectly detected, True Negative (TN) shows the pixel numbers of disease exactly misdetected and False Negative (FN) represents the pixel numbers of disease incorrectly misdetected. Thus "Specificity" was defined as percentage of non–disease pixels correctly detected, and "Sensitivity" was defined as percentage of disease pixels correctly detected as disease pixels. The performance of results

of binary images is measured by using TP, FP, TN, and FN. The specificity and sensitivity values are between 0 and 100 when the result is equivalent to 100 means perfect segmentation. For the accuracy, the weighted average of all values is calculated by Equation (11). Also, "Accuracy" was defined as overall pixel–based success rate of the method. Noticeably, therefore, the values of accuracy occur at the interval (0, 100) and the larger values of accuracy correspond to the higher clustering quality:

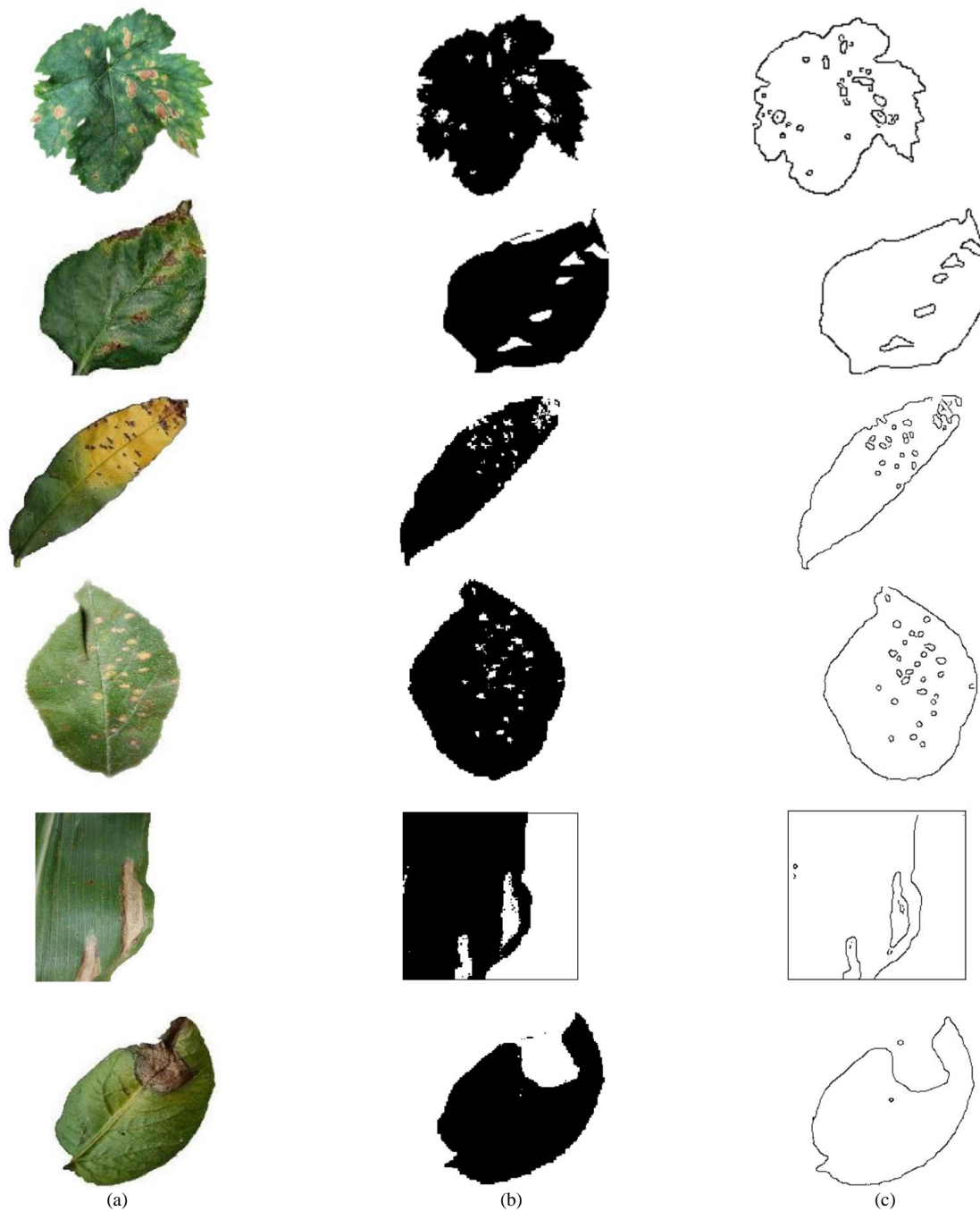$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} * 100 \tag{11}$$

Fig. 8: Detections of the plant leaf image, (a) original images, (b) segmented image using KM++ clustering method and intermeans thresholding, (d) edge boundaries output from Canny operator

Detection of Plant Disease Results

The occurrence of diseases on plant leaf's can be caused by a number of different features including areas, shapes, colors and sizes. Therefore, the accurate detection of plant leaf diseases depends on these features. To group, search and coarse segment plant leaf diseases, the KM++ clustering method is used. Then, to improve the results obtained from the KM++ clustering method, the intermeans thresholding method has been applied (Fig. 8). The performance evaluation of the proposed method is done based on three quantitative evaluation parameters known as SE, SP and AC, respectively. For validating the detection of plant leaf diseases, the performance of standard K–mean clustering and KM++ intermeans thresholding method is considered. The average of AC value of 98.10% for the KM++ intermeans thresholding method shows higher accuracy for detecting grape leaf diseases when applied on different images compared to that of 74.90% for standard K–means clustering method. The results show that if the KM++ clustering method is combined with intermeans thresholding method then the accuracy results for detection of diseases increases. As in the case of plant leaf disease detection form crowdAI.org, the comparative results based on SN, SP and AC for standard K–mean clustering and KM++ intermeans method are given in Table 4.

According to Table 3, the KM++ clustering method gives the best results by using intermeans thresholding method on the testing datasets. The proposed algorithms achieve the average SE, SP and AC of 98.46, 97.12 and 98.10% respectively showing higher performance than standard K–mean clustering technique in local database. The results show that the proposed methods outperform standard K–mean clustering algorithm around 23.20% in terms of accuracy measurement. These results are also the best results in overall experimentation from local databases. The performances of the proposed method are, respectively, 98.66, 98.20 and 98.58% by using

accuracy of measure for training the sizes of 10, 20 and 30%, respectively. Therefore, the experiments' results show that 30% training set size gives the best results. The average values of different correct measures are shown in Table 4.

Classification of Accuracy Measures

The Considering only the classification of accurate measures in the proposed method, Table 4 provides a sample confusion matrix of the best experimental results by using 30% for training dataset. The classification accuracy of plant leaf diseases is calculated by using Equation (9), (10) and (11), respectively.

For example, for sensitivity rate, assume that sensitivity measures are 3,917/4198*100 = 93.30%, specificity measures are 24,080/24,178*100 = 98.66% and accuracy measures are (3,917+24,080)/28,376 *100 = 98.66%, respectively. Based on this information, if the application selects KM++ clustering algorithm for plant leaf diseases, it is also acceptable in terms of intermeans thresholding method. Therefore, the results in Table 5 may be a proper performance indicator. The intermeans thresholding method helps improve the performance of detection in the binary stage.

Time Analysis

The run–time performance of the proposed method is compared by using a grape leaf image with 750×560 pixels from local database and images of apple leaf, maize leaf, peach leaf, pepper leaf, potato leaf with 256×256 pixels each form plant village disease classification challenge organized by CrowdAi.org. In Table 5, the timings of plant leaf disease detection for each image in second are compared. Considering Table 6, the combination of standard K–mean clustering and 138th intensity thresholding takes more time than the other algorithms. While the proposed method is built for coarse segmentation by using KM++ clustering and the intermeans thresholding method can be binarized in 11.04 sec, this timing is considered suitable for detection system of plant leaf diseases.

Table 3: Comparative results for detection of plant leaf disease upon SN, SP and AC for standard K–mean clustering and KM++ intermeans thresholding algorithms

Disease name	Standard K–mean clustering			KM++ intermeans thresholding		
	SN (%)	SP (%)	AC (%)	SN (%)	SP (%)	AC (%)
Grape leaf spot	74.80	75.10	74.90	98.40	97.12	98.10
Apple rust	70.01	69.80	69.90	98.12	98.06	98.09
Maize leaf blight	72.08	74.10	72.80	97.40	97.78	97.52
Peach leaf spot	75.10	75.06	75.08	98.44	98.50	98.47
Pepper leaf spot	73.90	74.12	74.09	98.10	97.92	98.05
Potato leaf blight	74.12	74.10	74.08	98.12	97.80	97.82
Average	**73.33**	**73.71**	**73.47**	**98.09**	**97.73**	**98.00**

Table 4: Correctness measure results of the experiments for different database by using KM++ intermeans thresholding algorithms

Disease name	Training size 30%			Training size 20%			Training size 10%		
	SN (%)	SP (%)	AC (%)	SN (%)	SP (%)	AC (%)	SN (%)	SP (%)	AC (%)
Grape leaf spot	98.30	98.47	98.66	97.11	98.07	98.20	98.40	98.90	98.58
Apple rust	98.15	98.60	96.10	98.06	98.12	98.10	97.81	97.92	97.88
Maize leaf blight	98.08	98.90	98.38	97.82	97.98	97.90	96.92	97.12	97.10
Peach leaf spot	98.92	98.30	98.47	98.20	97.91	98.09	98.10	97.52	97.74
Pepper leaf spot	98.50	98.12	98.32	98.10	98.12	98.14	97.70	97.80	97.78
Potato leaf blight	98.10	98.08	98.47	97.80	97.78	97.68	98.11	98.10	98.06
Average	**98.46**	**98.41**	**98.06**	**97.84**	**97.98**	**98.01**	**97.84**	**97.89**	**97.85**

Table 5: Example of confusion matrix analysis in the proposed method

Labels	Diseases	Non–Diseases	Total	Detection
Diseases	3,917 (TP)	281 (FP)	4,198 (TP+FN)	93.30% Sensitivity
Non–diseases	98 (FP)	24,080 (TN)	24,178 (FP+TN)	99.47% Specificity
Totals	4,015	24,361	28,376	98.66% Accuracy

Table 6: Timing of coarse segmentation, fine segmentation, and classification algorithm

Category	Algorithm	Time per image (seconds)
Coarse segmentation	Standard K–mean clustering	24.126
	KM++ clustering	9.239
Fine segmentation	120th intensity thresholding	1.936
	138th intensity thresholding	1.972
	Intermeans thresholding	1.807
Classification	K–mean (120th intensity thresholding)	26.062
	K–mean (138th intensity thresholding)	26.098
	K–mean (Intermeans thresholding)	25.933
	KM++ (120th intensity thresholding)	11.175
	KM++ (138th intensity thresholding)	11.211
	KM++ (Intermeans thresholding)	11.046

Conclusion and Future Work

This paper has proposed a new method for segmentation and detection by using KM++ clustering algorithm to combine different kinds of thresholding methods. The algorithm framework is constructed by using some image features such as color, size, shape, and intensity of the plants images that are labeled by the experts. The classes of diseases correspond to a binarization method, which is proper for the plant images. Then, a different thresholding parameter for a given test plants leaf image is selected by using the same algorithm and the intermeans thresholding is applied to the plants leaf image to generate binary image.

Three different thresholding parameters to combine KM++ clustering is used in this study to evaluate the performance of the plants leaf diseases detection. Knowing the performance of individual thresholding parameters is helpful to understand how much improvement can be made with the intermeans thresholding method. Therefore, several results can be concluded at the end of this work as follows:

1) 120th and 138th intensity thresholding methods may not be adequate for detecting plant leaf diseases.
2) Using the intermeans thresholding method provides the best performance for KM++ clustering method with SN = 98.46%, SP = 97.12% and AC = 98.10% on the average for local databases.
3) Intermeans thresholding method can be considered the best in terms of being robust and feasible as it generated binary images for plant leaf images unlike any other thresholding techniques.
4) The success of KM++ clustering algorithm depends on the success of the intermeans thresholding methods which are selected for the detection of plant leaf diseases. The proposed methods outperform the standard K–mean clustering techniques only around 23.20% with 70% of testing data from local database in our experiments.

For most cases, the proposed methods outperformed other algorithms. It is however difficult to verify the completeness based on the KM++ intermeans thresholding algorithm.

When the proposed methods are used in an automatic detection of plant leaf diseases, incorrect classifications may lead improper decision making. Another issue is that, with regard to the choice of the suitable thresholding algorithm in the training process, the thresholding method generated good results for a specific image. This significant similarity between the proposed methods and other thresholding methods is the difficult training for some images. Therefore, it is believed that this is the main reason why the optimal algorithms for detecting of plant leaf diseases have not been reached. The experimental results show that it is possible to improve detection accuracy by increasing the diversity of pooling operations, the reasonable addition of a KM++ intermeans thresholding of the model parameters.

In future work, the KM++ intermeans thresholding algorithm is to be applied on other features in order to build a more advanced algorithm to construct the algorithm and improve the detection accuracy. Meanwhile, more types of plants leaf diseases will be identified, with a combination of a new algorithm for the training and testing of the algorithm.

Acknowledgement

Authors Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Anand, R., Veni, S., & Aravinth, J. (2016, April). An application of image processing techniques for detection of diseases on brinjal leaves using k-means clustering method. In 2016 international conference on recent trends in information technology (ICRTIT) (pp. 1-6). IEEE.

Arthur, D., & Vassilvitskii, S. (2006). k-means++: The advantages of careful seeding. Stanford.

Chang, S. G., Yu, B., & Vetterli, M. (2000). Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Transactions on image Processing, 9(9), 1522-1531.

Chuang, M. C., Hwang, J. N., Williams, K., & Towler, R. (2011, September). Automatic fish segmentation via double local thresholding for trawl-based underwater camera systems. In 2011 18th IEEE International Conference on Image Processing (pp. 3145-3148). IEEE.

CrowdAI. Accessed: Jan 20, 2019. https://www.crowdai.org/challenges/plantvillage–disease–classication–challenge

De Albuquerque, M. P., Esquef, I. A., & Mello, A. G. (2004). Image thresholding using Tsallis entropy. Pattern Recognition Letters, 25(9), 1059-1065.

Dhanachandra, N., & Chanu, Y. J. (2015). Image segmentation method using k-means clustering algorithm for color image. Advanced Research in Electrical and Electronic Engineering, 2(11), 68-72.

Guo, X., Zhang, M., & Dai, Y. (2018, November). Image of plant disease segmentation model based on pulse coupled neural network with Shuffle Frog Leap Algorithm. In 2018 14th International Conference on Computational Intelligence and Security (CIS) (pp. 169-173). IEEE.

Hossain, M. S., Mou, R. M., Hasan, M. M., Chakraborty, S., & Razzak, M. A. (2018, March). Recognition and detection of tea leaf's diseases using support vector machine. In 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA) (pp. 150-154). IEEE.

Islam, M., Dinh, A., Wahid, K., & Bhowmik, P. (2017, April). Detection of potato diseases using image segmentation and multiclass support vector machine. In 2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE) (pp. 1-4). IEEE.

Kapur, J. N., Sahoo, P. K., & Wong, A. K. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. Computer vision, graphics, and image processing, 29(3), 273-285.

Korkut, U. B., Göktürk, Ö. B., & Yildiz, O. (2018, May). Detection of plant diseases by machine learning. In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Krithika, P., & Veni, S. (2017, March). Leaf disease detection on cucumber leaves using multiclass Support Vector Machine. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 1276-1281). IEEE.

Kurale, N. G., & Vaidya, M. V. (2018, July). Classification of Leaf Disease Using Texture Feature and Neural Network Classifier. In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1-6). IEEE.

Masazhar, A. N. I., & Kamal, M. M. (2017, November). Digital image processing technique for palm oil leaf disease detection using multiclass SVM classifier. In 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) (pp. 1-6). IEEE.

Neumann, M., Hallau, L., Klatt, B., Kersting, K., & Bauckhage, C. (2014, August). Erosion band features for cell phone image-based plant disease classification. In 2014 22nd International Conference on Pattern Recognition (pp. 3315-3320). IEEE.

Niblack, W. (1985). An Introduction to Digital Image Processing, 215 Strandberg Publishing Company. Copenhagen, Denmark.

Otsu, N. (1975). A threshold selection method from gray-level histograms, Automatica, 11, 23–27.

Padol, P. B., & Yadav, A. A. (2016, June). SVM classifier based grape leaf disease detection. In 2016 Conference on advances in signal processing (CASP) (pp. 175-179). IEEE.

Patil, P., Yaligar, N., & Meena, S. M. (2017, December). Comparision of Performance of Classifiers-SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images. In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-5). IEEE.

Prajwala, T. M., Pranathi, A., SaiAshritha, K., Chittaragi, N. B., & Koolagudi, S. G. (2018, August). Tomato leaf disease detection using convolutional neural networks. In 2018 Eleventh International Conference on Contemporary Computing (IC3) (pp. 1-5). IEEE.

Ray, N., & Saha, B. N. (2007, September). Edge sensitive variational image thresholding. In 2007 IEEE International Conference on Image Processing (Vol. 6, pp. VI-37). IEEE.

Ridler, T. W., & Calvard, S. (1978). Picture thresholding using an iterative selection method. IEEE trans syst Man Cybern, 8(8), 630-632.

Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. Pattern recognition, 33(2), 225-236.

Trongtorkid, C., & Pramokchon, P. (2018, February). Expert system for diagnosis mango diseases using leaf symptoms analysis. In 2018 International Conference on Digital Arts, Media and Technology (ICDAMT) (pp. 59-64). IEEE.

Wang, H., Li, G., Ma, Z., & Li, X. (2012a, May). Application of neural networks to image recognition of plant diseases. In 2012 International Conference on Systems and Informatics (ICSAI2012) (pp. 2159-2164). IEEE.

Zhang, J., & Hu, J. (2008, December). Image segmentation based on 2D Otsu method with histogram analysis. In 2008 International Conference on Computer Science and Software Engineering (Vol. 6, pp. 105-108). IEEE.

Zhou, R., Kaneko, S. I., Tanaka, F., Kayamori, M., & Shimizu, M. (2014). Disease detection of Cercospora Leaf Spot in sugar beet by robust template matching. Computers and electronics in agriculture, 108, 58-70.