# Detection and Classification of Plant Diseases Using Image Processing and Multiclass Support Vector Machine

**Article** · April 2020

DOI: 10.14445/22312803/IJCTT-V68I4P102



Murtaza Khan
Umm Al-Qura University

# Detection and Classification of Plant Diseases Using Image Processing and Multiclass Support Vector Machine

Murtaza Ali Khan

*Department of Computer Science & Umm Al-Qura University*
*Makkah Al-Mukarramah, Saudi Arabia*

***Abstract*** — *Identification of plant disease is very important to prevent the loss and keep the harvest healthy. Determination of plant disease via visual monitoring is difficult and time consuming. In this paper, we described a method of detection and classification of plant disease using image processing and machine learning techniques. We used standard images of leaves of several types of plants to test our method. Initially, our method segments the input image to isolate disease parts of the leaf. Then we obtain various features from the diseased affected segmented image. Finally, we classify leaves into healthy and disease types based on its features using Multiclass Support Vector Machine (SVM) classifier. Experimental results indicate that our method yields very high accuracy rate for detection and classification of plant diseases.*

**Keywords —** *Detection, Classification, Plant Diseases, Image Processing, and Multiclass Support Vector Machine*

## I. INTRODUCTION

Diseases of plants are major cause of plant damage and consequently agriculture and economic loses. Timely identification of plant disease is a critical factor to make harvest healthy and fruitful. The most common approach for identification of diseases of plants is visual observation by experts. But this approach can be time consuming or difficult due to lack of experts at the sites of cultivation. Image processing methods can be effective for continuous monitoring and detection of plant diseases. In this paper, we describe a machine learning technique for detection and classification of diseases of plants from images of leaves. A specific class of plant disease creates certain patterns on the leaves of the plant. We used machine learning techniques to train the system about the classes of plant diseases. Then during testing, image analysis of patterns is used to identify the class of disease from trained set of data. We used a three phases framework to implement the complete system. First, image segmentation is performed to identify the diseased regions. Then, features are extracted from segmented regions using standard feature extraction techniques. Finally, image features are used for classification and identification of different diseases using Multiclass Support Vector Machine (SVM) classifier.

The rest of the paper is organized as follows: Overview of related work is presented in Section 2. The proposed approach is illustrated in Section 3. Selected simulation results are presented in Section 4. Concluding remarks are in Section 5.

## II. RELATED WORK

Several researchers proposed methods to identify plant diseases using different types of images such as RGB imaging, La*b*, multispectral, etc. A general survey about recognition and classification of plant leaves' diseases using image processing techniques is presented by [5]. A recent method that measure the severity of phytophthora root rot disease in avocado trees is described by [12]. In this method, a smart-phone camera is used to capture the RGB images of avocado trees with varying degrees of canopy decline. Then multi-spectral imagery from high spatial resolution satellites is used to indicate disease severity or canopy decline in avocado trees into high, medium, and low ranges. An algorithm for detection of powdery mildew disease from a cherry leaf images is proposed by [7]. The method removes background from the image and then extract diseased portion using morphological operators and intensity-based threshold. A color sensing and image processing-based method to detect the severity of soybean plant foliar disease is described by [13]. This method extract YCbCr Channels from input RGB image and then uses opening and closing bi-level morphological operation to smooth area of infected region. A mobile client–server architecture for leaf disease detection using Gabor wavelet transform (GWT) and gray level co-occurrence matrix (GLCM) is described by [11]. In this method, a mobile client captures and process the leaf image, segments diseased patches, and transmits to the server. The server uses GWT-GLCM for feature extraction and k-Nearest neighbor for classification. The result is sent back to the user's screen via an SMS.

Image segmentation partitions an image into multiple segments/images with the goal of meaningful analysis and extraction of region of interest (ROI). Several authors proposed general segmentation algorithms [8], [1], [2], [9], [10] et al.

that can be used to segment images of leaves. We used k-means algorithm of [1] to segment the images of leaves. This algorithm improves the classic k-means algorithm of Lloyd [8] by a randomized seeding and have time complexity of Θ (log k). A multiphase segmentation method is proposed by [2]. This method uses iterative binary segmentation. At each iteration, the region of pixels with darker mean value of intensity is separated from other regions using intensity function, eigenvector of Hessian matrix, and Curvelet.

We used support vector machines (SVM) for classification and identification of plant diseases via image properties. In machine learning, SVM are supervised learning models with associated learning algorithms to analyze and classify data [3], [4], [14] et al. The original SVM algorithm was invented by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963, while the current version of the algorithm is published in 1995 by Corinna Cortes and Vapnik [3]. The earlier version of SVM have the restriction that training data can be separated without errors while the new version of the algorithm extends it to non-separable training data [3]. A semi-supervised support vector machine classifier ($S^3$VM) based on active learning and context information is presented by [6]. First a semi-supervised learning method uses active learning to select unlabeled samples as the semi-label samples. Then the context information is exploited to further expand the selected samples and relabel them, along with the labeled samples train $S^3$VM classifier.

## III. PROPOSED METHODOLOGY

This section describes the methodology of the proposed system. Figure 1 shows the block diagram of the system. In the block diagram, there are two major parts, one is training and the other is testing. In the training part, we train the SVM classifier using set of sample images of leaves. Each leaf belongs to a known class of disease and we assign a label (an integer value) to each disease class. We segment each leaf and extract the region of interest (ROI), i.e., the part of the leaf that has disease patterns. This ROI is essentially a complete image but unwanted information is removed from it. Then we compute the features of each ROI. Features are computed using image analysis methods. Finally, we create a database of these training sample images. The database contains disease class, its label, and feature vector for each sample image. In the testing part, we take a leaf of unknown disease class, segment it to find region of interest (ROI), compute the feature vector of ROI. The we pass this information to classifier. The SVM classifier, compares the feature vector of ROI of unknown test image with the pre-built database. Based on best match, it predicts the class of disease of the test image. Either we are using the system in the training or testing stage, the system has following three main steps:

A. Segmentation
B. Feature extraction
C. Classification

Now we describe the details of each of these steps in the subsequent sub-sections.

### *A. Segmentation*

In general, the disease affected regions of the leaf differ in color and texture than the healthy regions and background. Our is aim is to separate the disease affected and healthy regions. Background is not region of interest and it is removed too by masking out the background. The segmentation algorithm segments the input RGB image into three segments. Each segment is also in RGB color-space and same in size as the original image. But each segment contains different regions of the original image. One of the segment contains region of interest (ROI), i.e., the region that contains disease part of the original image.

In general, the K-means clustering algorithm is used for classification of object based on a set of features into K number of classes. The classification of object is done by minimizing the sum of the squares of the distance between the object and the corresponding cluster. Following is the general algorithm for K –means clustering:

1. Pick center of K cluster, either randomly or based on some heuristic.
2. Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center.
3. Again, compute the cluster centers by averaging all the pixels in the cluster. Repeat steps 2 and 3 until convergence is attained.

We used K-means clustering algorithm for segmentation of image into three segments as follows:

Algorithm for segment the input leaft image using K-means clustering is following.

1. Convert input image from RGB Color Space to L*a*b* color-space.
2. Extract a*b*, i.e., chroma components from the image in L*a*b* color-space.
3. Apply K-means clustering to chroma components to get three clusters (segments).
4. Map the three clusters (segments) back in to three segmented images in RGB colors-space.

Figure 2 shows the segmentation of Bacterial Blight leaf. In the figure, we showed original image, first, second and third segments. It can be notice from the figure that the segmentation process isolated the disease affected region in the third segment.

### *B. Feature extraction*

An image is characterized and classified by its features. After the process of segmentation, we have three segments in the RGB color space. One of the segment contains the disease affected part and it is our region of interest (ROI). Next among all the segments, the user must choose the disease affected segment by visual inspection of three segments. This part of our system is not automatic and user intervention is necessary. Once the user chooses the disease affected segment, the system must find its features. We extracted thirteen features from every leaf image. Nine features (mean, standard deviation, entropy, root mean square (RMS), variance, smoothness, kurtosis, skewness, and inverse difference) are determined from the disease affected segment in RGB color space. The remaining four features (contrast, correlation, energy, and homogeneity) are determined after converting disease affected segment in to grayscale intensity image. The conversion process eliminates the hue and saturation information while retains the luminance channel.

Table 1 shows the mathematical expressions of the thirteen features. In Table 1 features 1 – 9 are computed using the segment image in RGB color space, while features 10 – 13 are determined using grayscale image.

### C. Classification

Classification of selected disease segment image is preformed using Multiclass Support Vector Machine (SVM). SVM is machine learning technique used for classification. For a given a set of training samples, each sample is labeled as belonging to one of the two classes. An SVM training algorithm builds a model that assigns new samples to one class or the other. Multiclass SVM aims to assign labels to instances by using support vector machines, where the labels are drawn from a finite set of several elements. The common approach to build Multiclass SVM is to reduce the single Multiclass problem into multiple binary classification problems. In our case, Multiclass SVM is reduces to following optimization problem for a given training set of instance label pairs $(x_l, y_l)$, $l=1\ldots i$, where $x_l \in Rn$ and $x_l \in \{1, -1\}^i$.

$$\min_{v,e,\psi} \frac{1}{2} v^T v + C \sum_{l=1}^{i} \psi_l$$

where $y_l = (v^T \phi(x_l) + e) \geq 1 - \psi_l$, $\psi \geq 0$.

We constructed a Multiclass SVM model, first we train it for samples leaves with predetermined features of diseases affected leaves and healthy leaves. Each types of diseases affected leaves belongs to one separate class and there is a class of healthy leaves. Using Multiclass SVM model, we build the data for a set of training sample images and save it in a file. During the testing phase, the system must find the disease class of a leaf image. First, we segment the test image and choose the region of interest (ROI), then we compute its feature vector. Finally, we pass this information to classifier. The Multiclass SVM classifier, compares the feature vector of ROI of unknown test image with the pre-built database. Based on best match, it predicts the disease class of test image.

## IV. EXPERIMENTS AND RESULTS

The proposed algorithm is applied to data set of 148 images. We divided the images into two sets: (1) the training set consists of 73 images and (2) the testing set consists of 75 images. The data set five types of leaves, one type is of healthy leaves and four types of leaves with diseases namely Alternaria Alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot. Testing accuracy of the classification is 92.8571%.

Table 2 shows single set of results for each type of disease and healthy leaves. In the table, each row has three columns, the first column shows the original sample images, the second column shows RGB segments, and the third column shows grayscale segments. First row of Table 2 shows healthy leaves while rows two to five show disease affected leaves.

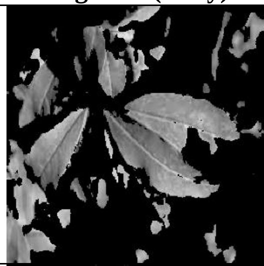
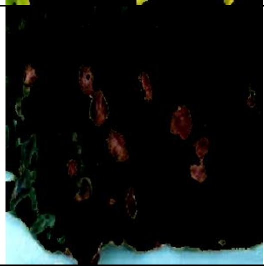
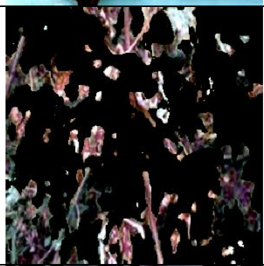
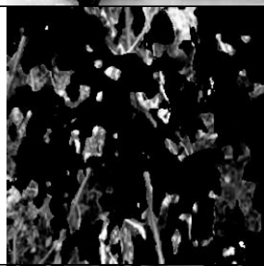
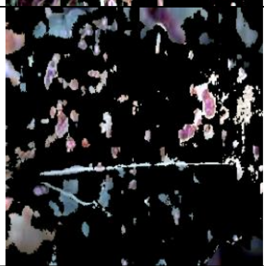
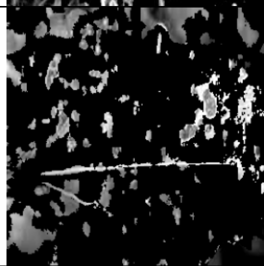
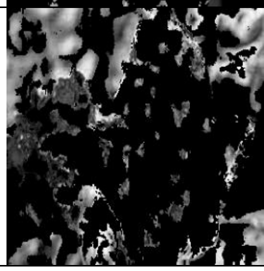
## V. CONCLUSION

In this paper, we described a framework for detection and classification of plant disease. We used Multiclass Support Vector Machine (SVM) as a classifier during training and testing phases. Each type of disease and healthy leaves are assigned a unique label. We used image segmentation to identify the diseased affected regions of a leaf. Then we extracted the standard features from diseased affected segmented image. Finally, we used features to classify leaves into healthy and disease types using Multiclass Support Vector Machine (SVM). Experimental results show that our proposed framework yields very high accuracy rate (92.8571%.) and can be used in real-world for detection and classification of plant disease.

**TABLE I**
**Mathematical Expressions of Features**

| | Feature | Expression |
|---|---|---|
| 1 | Mean =M | $\sum_{i=0}^{N-1} g(i)P(g(i))$ |
| 2 | Standard Deviation =S | $\sqrt{\sum_{i=0}^{N-1} (g(i)-M)^2 P(g(i))}$ |
| 3 | Entropy | $\sum_{i=0}^{N-1} P(g(i))log_2(P(g(i)))$ |
| 4 | RMS | $\sqrt{\frac{1}{NxN}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(g(i,j)-I)^2}$ |
| 5 | Variance | $\sum_{i=0}^{N-1}(i-\mu)^2 p(i)$ |
| 6 | Smoothness | $\sum_i\sum_j \frac{1}{1+(i-j)^2} g_{i,j}$ |
| 7 | Kurtosis | $\frac{1}{S^k}\sum_{i=0}^{N-1}(g(i)-M)^3 P(g(i))$ |
| 8 | Skewness | $\frac{1}{S^3}\sum_{i=0}^{N-1}(g(i)-M)^3 P(g(i))$ |
| 9 | Inverse Difference | $\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{p_{i,j}}{1+(i-j)^2}$ |
| 10 | Contrast | $\sum_i\sum_j (i-j)^2 g_{i,j}$ |
| 11 | Correlation | $\frac{\sum_i\sum_j (ij)g_{i,j}-\mu_x\mu_y}{\sigma_x\sigma_y}$ |
| 12 | Energy | $\sum_i\sum_j g2_{i,j}$ |
| 13 | Homogeneity | $\sum_i\sum_j \frac{1}{1+(i-j)^2} g_{i,j}$ |

**TABLE III**
**Sample images of leaves with disease affected RGB and grayscale segments.**

| | **Original Image** | **Disease affected Segment (RGB)** | **Disease affected Segment (Gray)** |
|---|---|---|---|
| Healthy Leaf |  |  |  |
| Alternaria Alternata |  |  |  |
| Anthracnose |  |  |  |
| Bacterial Blight |  |  |  |
| Cercospora Leaf Spot |  |  |  |

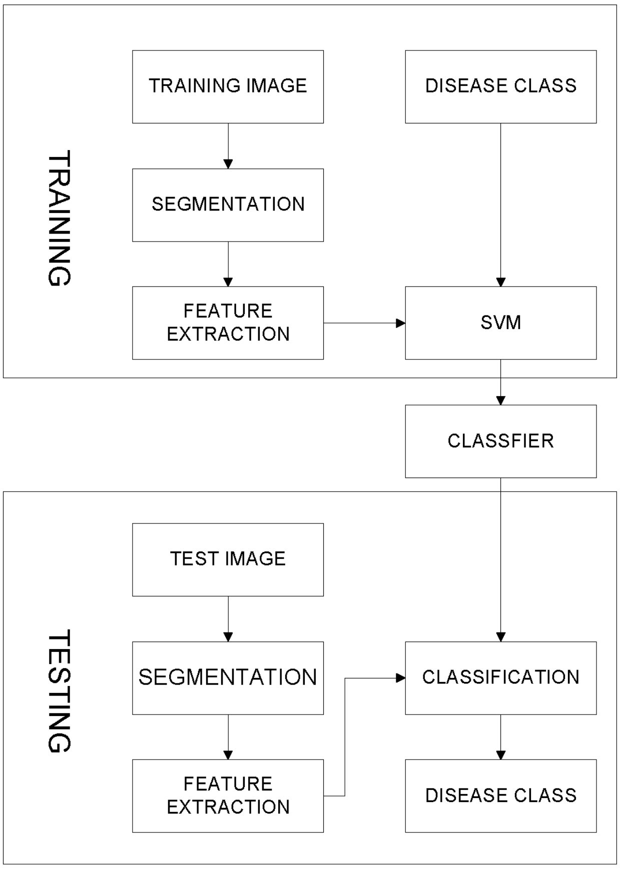



Fig 1: Block diagram of the system

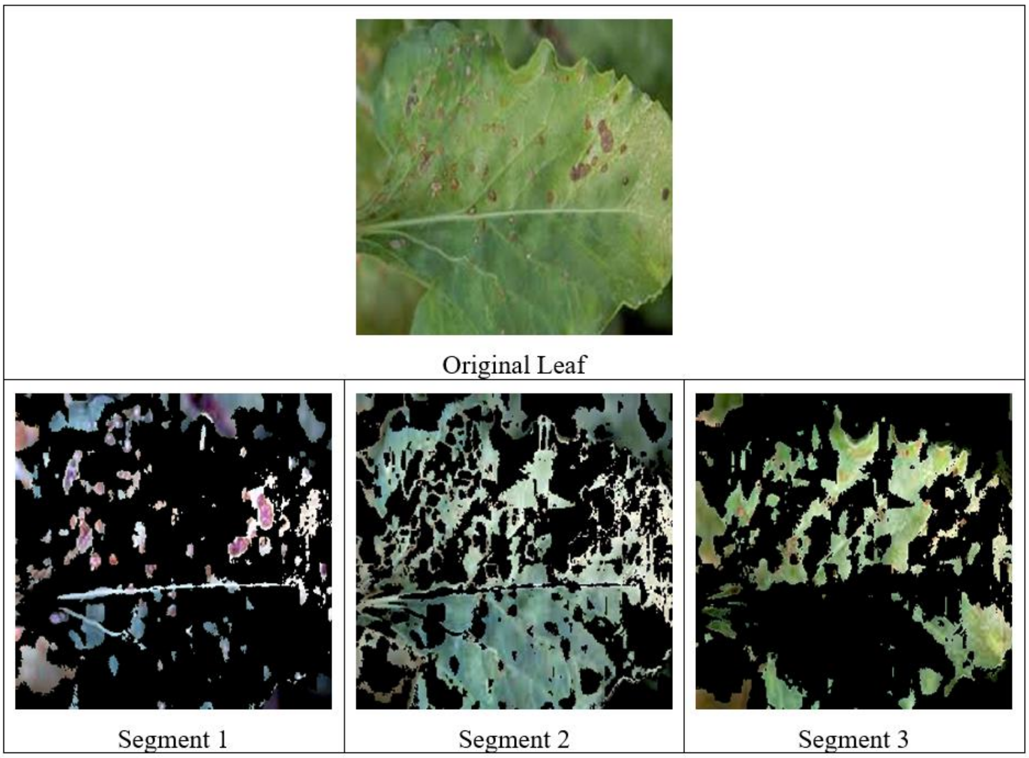



Fig 2: Segmentation of Bacterial Blight leaf: original image, segment 1, segment 2, and segment 3. Disease affected region is isolated in the segment 3.

## REFERENCES

[1] Arthur, D., Vassilvitskii, S. K-means++: "*The advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*", SODA '07, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA 2007; 1027-1035. URL http://dl.acm.org/citation.cfm?id=1283383.1283494.

[2] Cigaroudy, L.S., Aghazadeh, N. "*A multiphase segmentation method based on binary segmentation method for gaussian noisy image*". Signal, Image and Video Processing 2017; 11(5): 825-831. URL https://doi.org/10.1007/s11760-016-1028-9.

[3] Cortes, C., Vapnik, V. "*Support-vector networks*". Machine Learning 1995; 20(3): 273-297. URL https://doi.org/10.1007/BF00994018.

[4] Cristianini, N., Shawe-Taylor, J. "*An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*". Cambridge University Press, 2000. URL https://doi.org/10.1017/CBO9780511801389.

[5] Dhingra, G., Kumar, V., Joshi, H.D. "*Study of digital image processing techniques for leaf disease detection and classification*". Multimedia Tools and Applications, 2017. URL https://doi.org/10.1007/s11042-017-5445-8.

[6] Gao, F., Lv, W., Zhang, Y., Sun, J., Wang, J., Yang, E. "*A novel semi-supervised support vector machine classifier based on active learning and context information*". Multidimensional Systems and Signal Processing 2016; 27(4): 969-988. URL https://doi.org/10.1007/s11045-016-0396-1

[7] Gupta, V., Sengar, N., Dutta, M.K., Travieso, C.M., Alonso, J.B. "*Automated segmentation of powdery mildew disease from cherry leaves using image processing*". In: 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI) 2017; 1-4. URL https://doi.org/10.1109/IWOBI.2017.8006454

[8] Lloyd, S. "*Least squares quantization in PCM*". IEEE Transactions on Information Theory 1982; 28(2): 129-137 URL https://doi.org/10.1109/TIT.1982.1056489.

[9] Mignotte, M. "*Segmentation by fusion of histogram-based k-means clusters in different color spaces*". IEEE Transactions on Image Processing 2008; 17(5): 780-787 URL https://doi.org/10.1109/TIP.2008.920761.

[10] Nabeel, F., Asghar, S.N., Bashir, S. "*Optimizing image segmentation by selective fusion of histogram based k-means clustering*". In: 2015 12th International Bhurban Conference on Applied Sciences and Technology (IBCAST) 2015; 181-185. URL https://doi.org/10.1109/IBCAST.2015.7058501.

[11] Prasad, S., Peddoju, S.K., Ghosh, D. "*Multi-resolution mobile vision system for plant leaf disease diagnosis*". Signal, Image and Video Processing 2016; 10(2): 379-388. URL https://doi.org/10.1007/s11760-015-0751-y

[12] Salgadoe, A.S.A., Robson, A.J., Lamb, D.W., Dann, E.K., Searle, C. "*Quantifying the severity of phytophthora root rot disease in avocado trees using image analysis*". Remote Sensing 2018; 10(2). URL http://www.mdpi.com/2072-4292/10/2/226.

[13] Shrivastava, S., Singh, S.K., Hooda, D.S. "*Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation*". Multimedia Tools and Applications 2015; 74(24), 11467-11484. URL https://doi.org/10.1007/s11042-014-2239-0.

[14] Stoean, C., Stoean, R. "*Support Vector Machines and Evolutionary Algorithms for Classification*". Springer International Publishing 2014. URL https://doi.org/10.1007/978-3-319-06941-8