International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V68I4P102

Volume 68 Issue 4, 5-11, April 2020

Original Article

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.

Received Date: 24 February 2020
Revised Date: 11 April 2020
Accepted Date: 13 April 2020

Abstract - *Identification of plant disease is very important to prevent 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 their features using the Multiclass Support Vector Machine (SVM) classifier. Experimental results indicate that our method yields a very high accuracy rate for the detection and classification of plant diseases.*

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

I. INTRODUCTION

Diseases of plants are a major cause of plant damage and consequently agriculture and economic losses. Timely identification of plant disease is a critical factor to make harvest healthy and fruitful. The most common approach for the identification of diseases of plants is a visual observation by experts. But this approach can be time-consuming or difficult due to the 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 the 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 a 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 the classification and identification of different diseases using a Multiclass Support Vector Machine (SVM) classifier.

The rest of the paper is organized as follows: An 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 the recognition and classification of plant leaves' diseases using image processing techniques is presented by [5]. A recent method that measures the severity of Phytophthora root rot disease in avocado trees is described by [12]. In this method, a smartphone camera is used to capture the RGB images of avocado trees with varying degrees of canopy decline. Then multispectral 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 the detection of powdery mildew disease from cherry leaf images is proposed by [7]. The method removes the background from the image and then extract the diseased portion using morphological operators and intensity-based threshold. A colour sensing and image processing-based method to detect the severity of soybean plant foliar disease is described by [13]. This method extracts YCbCr Channels from the input RGB image and then uses opening and closing bi-level morphological operation to smooth the area of the infected region. A mobile client-server architecture for leaf disease detection using Gabor wavelet transform (GWT) and grey 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 it to the server. The server uses GWT-GLCM for feature extraction and k-Nearest neighbour 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 a 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 the 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 has a 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 a darker mean value of intensity is separated from other regions using intensity function, an eigenvector of Hessian matrix, and Curvelet.

We used support vector machines (SVM) for the 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 was published in 1995 by Corinna Cortes and Vapnik [3]. The earlier version of SVM has 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^3VM) 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 semi-label samples. Then the context information is exploited to further expand the selected samples and relabel them, along with the labelled samples train S^3VM 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 a 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 the disease class, its label, and feature vector for each sample image. In the testing part, we take a leaf of an unknown disease class, segment it to find a region of interest (ROI), compute the feature vector of ROI. We pass this information to the classifier. The SVM classifier compares the feature vector of ROI of an unknown test image with the pre-built database. Based on the best match, it predicts the class of disease of the test image. Whether we are using the system in the training or testing stage, the system has the 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 colour and texture from the healthy regions and background. Our is aim is to separate the disease affected and healthy regions. The background is not a 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 colour space and the same in size as the original image. But each segment contains different regions of the original image. One of the segments contains a region of interest (ROI), i.e., the region that contains the diseased part of the original image.

In general, the K-means clustering algorithm is used for the classification of an object based on a set of features into K number of classes. The classification of an 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:

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

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

The algorithm for the segment of the left input image using K-means clustering is the following.

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

Fig. 2 shows the segmentation of the Bacterial Blight leaf. In the figure, we showed the original image, first, second and third segments. It can be noticed 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 colour space. One of the segments 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 colour space. The remaining four features (contrast, correlation, energy, and homogeneity) are determined after converting disease affected segment into a grayscale intensity image. The conversion process eliminates the hue and saturation information while retaining the luminance channel.

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

C. Classification

Classification of selected disease segment image is performed using Multiclass Support Vector Machine (SVM). SVM is a machine learning technique used for classification. For a given set of training samples, each sample is labelled 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 building Multiclass SVM is to reduce the single Multiclass problem into multiple binary classification problems. In our case, Multiclass SVM is reduced to the 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 sample leaves with predetermined features of diseases affected leaves and healthy leaves. Each type of diseases affected leaves belongs to one separate class, and there is a class of healthy leaves. Using the 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 the classifier. The Multiclass SVM classifier compares the feature vector of ROI of an unknown test image with the pre-built database. Based on the best match, it predicts the disease class of the test image.

IV. EXPERIMENTS AND RESULTS

The proposed algorithm is applied to a 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 Alternate, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot. The testing accuracy of the classification is 92.8571%.

Table 2 shows a 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. The first row of Table 2 shows healthy leaves, while rows two to five show disease affected leaves.

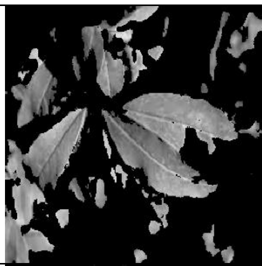
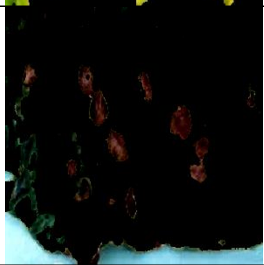
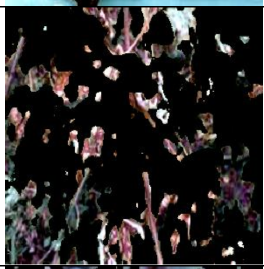
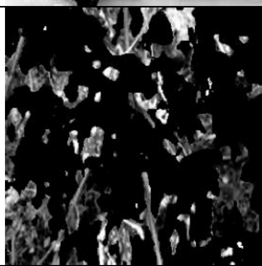
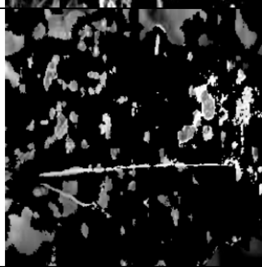
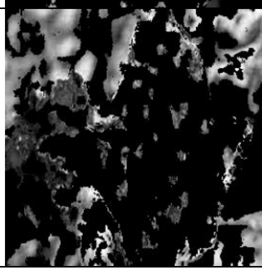
V. CONCLUSION

In this paper, we described a framework for the detection and classification of plant disease. We used Multiclass Support Vector Machine (SVM) as a classifier during the 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 the diseased affected segmented images. 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 a very high accuracy rate (92.8571%.) and can be used in the real world for the detection and classification of plant disease.

Table 1. 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 2. 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			

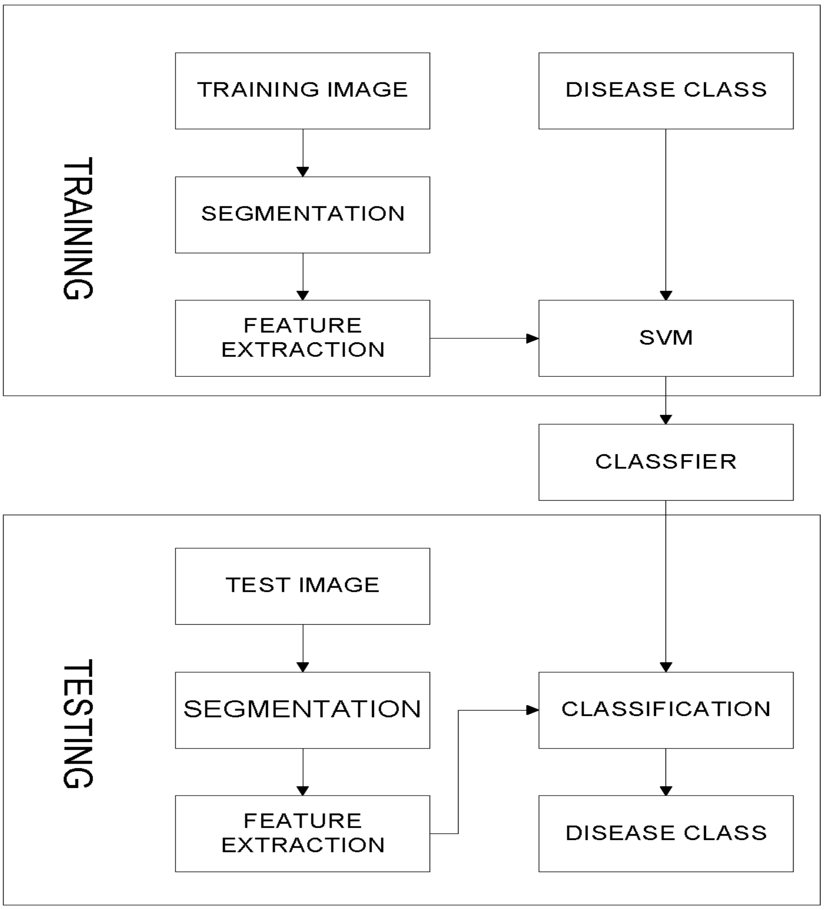


Fig. 1 Block diagram of the system

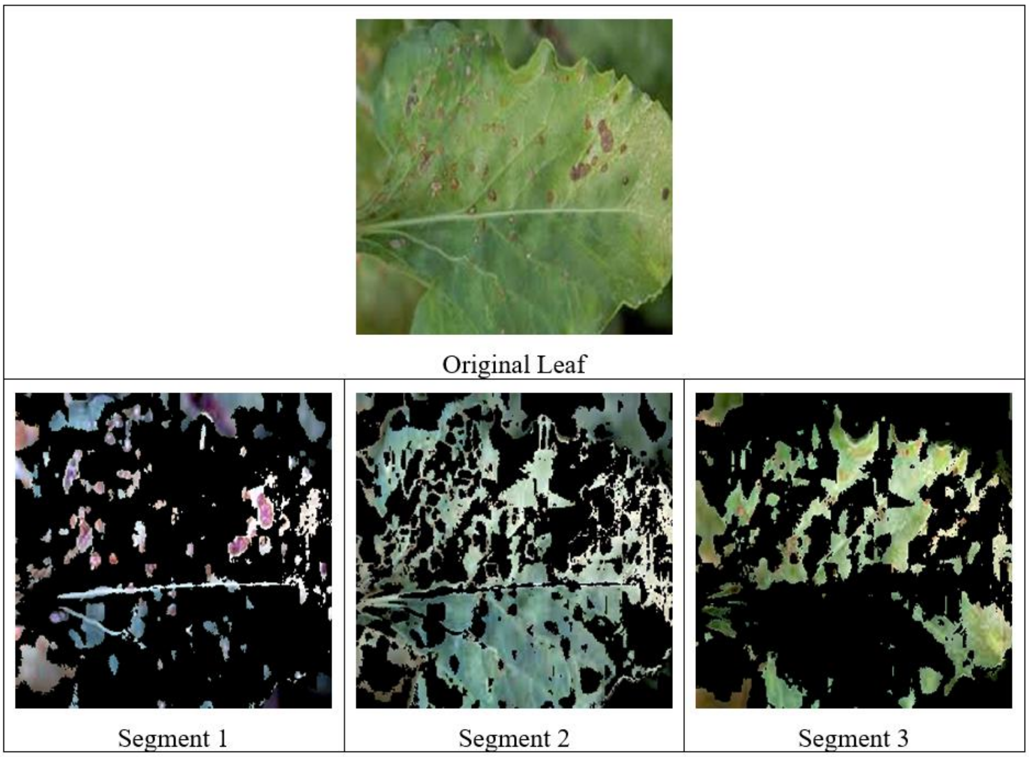


Fig. 2 Segmentation of Bacterial Blight leaf: original image, segment 1, segment 2, and segment 3. Disease affected region is isolated in segment 3.

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 , 11(5) (2017) 825-831. URL Https://Doi.Org/10.1007/S11760-016-1028-9.

[3] Cortes, C., Vapnik, V. Support-Vector Networks. Machine Learning , 20(3) (1995) 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, 27(4) (2016) 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, 28(2) (1982) 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 Colour Spaces. IEEE Transactions On Image Processing, 17(5) (2008) 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 , 10(2) (2016) 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 , 10(2) (2018). 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 , 74(24) (2015) 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