An Intelligent Technique for Grape Fanleaf Virus Detection

Mojtaba Mohammadpoor1*, Mohades Gerami Nooghabi2, Zahra Ahmedi1

1 Electrical & Computer Engineering Department, University of Gonabad, Gonabad (Iran)  
2 Department of Agricultural and Natural Resources, University of Gonabad, Gonabad (Iran)

Received 8 April 2019 | Accepted 3 January 2020 | Published 3 February 2020

ABSTRACT

Grapevine Fanleaf Virus (GFLV) is one of the most important viral diseases of grapes, which can damage up to 85% of the crop, if not treated at the right time. The aim of this study is to identify infected leaves with GFLV using artificial intelligent methods using an accessible database. To do this, some pictures are taken from infected and healthy leaves of grapes and labeled by technical specialists using conventional laboratory methods. In order to provide an intelligent method for distinguishing infected leaves from healthy ones, the area of unhealthy parts of each leaf is highlighted using Fuzzy C-mean Algorithm (FCM), and then the percentages of the first two segments area are fed to a Support Vector Machines (SVM). To increase the diagnostic reliability of the system, K-fold cross validation method with k = 3 and k = 5 is applied. After applying the proposed method over all images using K-fold validation technique, average confusion matrix is extracted to show the True Positive, True Negative, False Positive and False Negative percentages of classification. The results show that specificity, as the ability of the algorithm to really detect healthy images, is 100%, and sensitivity, as the ability of the algorithm to correctly detect infected images is around 97.3%. The average accuracy of the system is around 98.6%. The results imply the ability of the proposed method compared to previous methods.

I. Introduction

RECENT advances in agricultural technology have led to increase demand for development of non-destructive diagnostic methods. Spectrometry and imaging techniques are approaches used to detect disease and stress in trees and plants. By using image processing and machine learning methods, we can accurately detect and diagnose the disease at very low cost, and increase production.

Vine diseases are diverse, but the Grape Fanleaf Virus (GFLV) is the most harmful grape loss in the world, with losses of up to 85% of the crop. The disease has been reported in most temperate regions of the world. There are several common laboratory methods for identifying viruses, for instance Kaur et al. have conducted a study of various computer vision applications that classify images of plant leaves to detect diseases [1]. Lots of researches on the use of neural networks in detection of plant diseases has been performed. For example, Beeshish et al. in [2] have classified hereditary diseases using a post-propagation neural network. Mahmoudi et al. have used visual machine techniques to evaluate the color and appearance properties of the leaves and used these properties to identify two walnut diseases. They have reached to 95% accuracy in classifying the plant diseases [3]. Due to the small differences between the infected and healthy images, the convolutional neural network (CNN) has used to identify plant diseases in recent years [4]-[5]. Due to the complexity of these networks and their needs for many images for training, common neural networks are still interesting. For example, Shah, Nikhil and Sarika Jain in [6] have used artificial neural networks to detect cotton leaf diseases. In [7] Hosseini et al. have provided a system to detect fungal infection of white fish powder and anthracnose of cucumber leaves with image processing techniques and artificial neural networks. Their method was consisted of three steps: segmentation, separation of damaged parts of the leaf and classification.

In [8] Omrani et al. have proposed an Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict four varieties of apple plants by processing their leaf images. For this purpose, after collecting image dataset of leaf samples, they extracted morphological, color and texture features. Their results showed that ANFIS could classify the leaves successfully. The precision of their method for experimental classification was between 83% to 95%. Al-Hiary et al. have classified the leaf symptoms of the diseases by using K-means clustering and neural networks [9]. Menukaewjinda et al. in [10] has used the backward propagation neural networks (BPNN) for a competent grape leaf color, texture features. Their results showed that ANFIS could classify the leaves successfully. The precision of their method for experimental classification was between 83% to 95%. Al-Hiary et al. have classified the leaf symptoms of the diseases by using K-means clustering and neural networks [9]. Menukaewjinda et al. in [10] has used the backward propagation neural networks (BPNN) for a competent grape leaf color, but not for a specific grape disease. Subsequently, several studies were conducted to develop this method and algorithms, but all were based on specific symptoms of plant diseases and not a defined disease, such as viruses [11]. Dubey, et al. in [12] have proposed K-Means clustering segmentation technique to detect infected fruit part. Belkhodja and Hamdadou have also proposed a computer aided detection system for detecting breast masses in [13]. Pujari, Yakkundimath and Byadgi have used SVM and ANN for classification of plant disease in [14].

In this paper, we first discuss the virus, its identification and its experimental diagnostic methods in sections II. By introducing useful...
techniques, including segmentation, classification and cross-validation in section III, the proposed method will be described based on them in section IV. The results will be displayed and discussed in section V, and the document will be concluded in section VI.

II. Disease Description

The Grapevine fanleaf virus (GFLV) is one of the most important grape diseases that causes the leaves to be severely distorted, asymmetrical, hollowed out and wrinkled, and exhibit sharp dentures. Other symptoms of this virus are yellowish color and delayed veins [15]-[17]. The virus is transmitted by an ectoparasite nematode - (Xiphinema Index) [18]. The virus is widespread around the world but its origin is Iran [19]-[20].

There are some Virus detection methods. Some of the common methods used in most of laboratories are discussed in this section.

A. Electron Microscope

Electron microscope method was used to observe virions and detect them in the vector [20]. Due to the low concentration and heterogeneous distribution of the virus in plant tissues, this method has not been used for identification [15].

B. Use of Indicator Plants

One of the methods used to detect viral diseases of plants are indicator plants [21]. The best indicator for detecting GFLV is Vitis rupestris St. George, which is a mosaic marker, and can determine oily spots, deformities and felt leaves [22]. However, this method takes a long time and does not cause symptoms at low virus concentrations.

C. Serological Methods

The concentration of GFLV in the warm season is significantly reduced in plant tissues, hence viral concentrations are below the detection threshold of serological tests [20].

D. Enzyme-linked Immunosorbent Assay

Enzyme-linked immunosorbent assay (ELISA) is one of the most commonly used methods for detecting GFLV in plant tissues [23]. In this technique, the virus in the plant can be determined based on the antigenic tendency of an antigen to attach to a specific antibody [24]. Due to higher virus concentrations in young tissues, the use of young leaves for this method is better than in older ones [20]. Despite the effectiveness of this method, it is not practically applicable in many cases for the diagnosis of plant viruses due to its time consuming procedure [25].

E. Molecular Methods

Reverse Transcriptase- Polymerase Chain Reaction (RT-PCR) is the most sensitive method for detecting GFLV. Unlike other methods of identification, which depend on the age of the tissue, variety, season and virus concentration, PCR have the ability to accurately diagnose the disease independently.

Phenolic and polysaccharide compounds in grape leaves limit the release of pure RNA virus, therefore, the ability of PCR will be reduced. Despite significantly improvements in genomic extraction methods, it cannot be said that their errors reached to zero [20]. Hybridization of a nucleic acid using a probe is another sensitive method for detecting GFLV [22]. To achieve this, special probes against the genome of the virus have been developed [23]. Also this method has some disadvantages. Considering the fact that the intensity of reaction of the probes is low due to the concentration of the virus, tracking would be difficult in the warm seasons, when the concentration of the virus in the tissues decreases.

III. Materials and Methods

A digital image is actually a two-dimensional signal or, in other words, a data matrix that is created by measuring the reflected light from an object. Each of the image components is called a pixel. In gray level images, the minimum pixel value is zero, which represents a completely black dot. The maximum of each pixel is usually 255, which represents a completely white pixel. In color images using the RGB color standard, each pixel has three values ranging from 0 to 255, indicating the red, green, and blue colors, respectively.

A. Clustering

Clustering is an unsupervised learning method, which plays an important role in data mining, machine learning, and pattern recognition. This is an information process that divides data samples into several categories, called clusters, based on similarities between them. Various criteria can be considered as a measure of similarity, for example, one can use the distance criterion for clustering and consider objects closer to each other as a cluster. This type of clustering is called distance clustering. Clustering methods can be divided into hierarchical and separated delimiters. Hierarchical algorithms use the similarity criterion, and at each stage they divide the data into two categories and ultimately create a tree structure of this unit called dendrogram. Separation algorithms directly group data into several clusters. These algorithms are divided into hard (or exclusive) and soft (or fuzzy) algorithms. In strict clustering, the input sample belongs only to a cluster, while in soft clustering, its membership for each cluster is determined by a number between zero and one [26].

B. Fuzzy Clustering and FCM Algorithm

The FCM Clustering Algorithm is the basic algorithm of the segmentation methods. This algorithm is always of interest to researchers because of its advantages, such as its simple structure, ease of implementation, fast convergence and its need for a small storage space. Its simplicity is due to the fact that each cluster is represented by a center of gravity or an average value. One of the drawbacks of this algorithm is that the weight of the functions is constant through the entire clustering process. To overcome this drawback, various strategies have been proposed for adapting the weight of functions during the clustering process. In [27] Zhi et al. have developed a clustering algorithm based on C-means with automatic weighting functions during clustering.

C. Support Vector Machines (SVM)

The most common neural network techniques often focus on improving the structure of the neural network, in order to minimize estimation error and the number of neural network errors, but in a specific form of them, known as Support Vector Machines (SVM), it focuses solely on reducing the operational risk associated with inadequate performance. The SVM network structure has much in common with Multilayer Perceptron (MLP) neural networks, and in practice the main difference is in the learning style. Since this method is used in this paper, it is discussed in more detail below [29]. The support vector machines in their simplest form, linear SVM, consist of a cloud page that separates positive and negative sample sets with maximum distances (Fig. 1).

In general, this problem can be considered in an n-dimensional space in which the data are divided into two categories. In this case, instead of separating lines, a hyperplane separator will be used. In general, in the hyperplane type separator we will have:

\[ \sum w_i x_i + b = 0 \]  \hspace{1cm} (1)

It can be expressed in the following way:

\[ w \cdot x + b \]  \hspace{1cm} (2)
Fig. 1. SVM as a cloud page for linear separation of samples in data space [29].

Where $w$ is the vector weight of perpendicular vector to the hyperplane and $b$ is the initial value. In this display, $u = 0$ refers to the separator line itself, and the closest points are located on the plates $u = \pm 1$. In fact, with the assumption of the separation of two positive and negative data classes, the boundary vectors will be placed on the following hyperplanes:

$$w \cdot x + b = \pm l$$  \hspace{1cm} (3)

The area between these two hyperplanes is called the margin. Fig. 2 shows the two-dimensional state with the assumption that the initial value is negative.

Fig. 2. Two-dimensional distance [25].

As shown in Fig. 2, space is divided into two categories of samples with the following characteristics:

$$w \cdot x + b \geq l \text{ for } y_i = 1$$ \hspace{1cm} (4)

$$w \cdot x + b \leq -l \text{ for } y_i = -1$$ \hspace{1cm} (5)

The above formula can be combined as follows:

$$y_i (w \cdot x + b) - l \geq 0 \ \forall i$$ \hspace{1cm} (6)

In this case, the distance to the source vertically (the closest distance) for the points on the hyperplane $w \cdot x + b = l$ is:

$$|l - b|/|w|$$ \hspace{1cm} (7)

And in the same way, the distance from the source vertically to the points on the cloud plate $w \cdot x + b = -l$ is equal to:

$$|l - b|/|w|$$ \hspace{1cm} (8)

On the other hand, the source distance to the separator hyperplane is equal to:

$$|b|/|w|$$ \hspace{1cm} (9)

So the smallest distance between this hyperplane and any of the pages is as follows:

$$d_+ = d_- = l / |w|$$ \hspace{1cm} (10)

Therefore, the margin, as the distance between the two superimposed pages is as follows:

$$d_+ + d_- = 2 / |w|$$ \hspace{1cm} (11)

In this way, the maximization of the margin can be expressed in the form of the following optimization equation:

$$\min |w|^2$$ \hspace{1cm} (12)

subject to: $y_i (w \cdot x + b) - l \geq 0 \ \forall i$

D. K-fold Validation Method

There are various methods for validating the algorithms applied on databases to ensure that the results are valid in all circumstances and are not depend on selection of training and testing parts. K-fold is one of the common methods for validating classification algorithms. In this method, training and testing data are divided into K subsets. Of the subsets, one is left for validation, and K-1 subsets are used for training. This procedure is repeated K times so that all data is used exactly once for validation purposes. Finally, the final result is considered as the average of all round results [30].

IV. Proposed Method

In order to detect GLFV disease using image processing, a method is proposed and implemented in accordance with the flowchart shown in Fig. 3. Accordingly, after collecting practical images, some preliminary processing steps are performed on them, for example their background is removed, and their image intensity is improved. Images are divided into two categories: healthy and infected leaves, based on visual inspection by experts, and confirmed by molecular testing by RT-PCR. Then, the FCM algorithm is performed to segment the images. As the separation of healthy parts from unhealthy on a green page is better, this color page is considered as the input to the next step of the proposed algorithm.

![Flowchart of the proposed method](image-url)
A. Data Collection and Validation

The leaves used in this research belong to the Kashmar region of Khorasan-Razavi province, Iran, and were captured by a Sony DSC-N2 camera during Spring and May 2013. Leaves are divided into two groups (healthy and infected) based on the results of molecular testing with RT-PCR at Ferdousi University of Mashhad laboratory. Totally 92 images are collected of healthy and infected leaves. Examples of these images are shown in Fig. 4.

![Fig. 4. Right leaf: healthy, Left leaf: completely infected with the virus.](image)

V. RESULTS AND FINDINGS

The result of applying different parts of the proposed algorithm is presented in this section.

A. Background Removal

Due to the small changing of criteria such as light, background color and camera height, etc. during taking pictures of leaves, the background of the image may cause some difficulties, hence all leaves backgrounds are deleted. Fig. 5 shows a sample image before and after of background removing procedure.

![Fig. 5. (a) a sample of original image of a partially infected leaf, (b) the same image after removing its background.](image)

B. Contrast Improving of Images

Improving the intensity of an image is usually desirable. Its use in this research is to improve the characteristics of yellow and cream mosaics, as clear signs of a viral disease in infected leaves. A sample image after improving the intensity is shown in Fig. 6. Comparing Fig. 6 and Fig. 5, one can clearly see the difference between the two images. In Fig. 6 yellow and cream spots in the image are brighter than other parts.

![Fig. 6. Sample infected leaf image after contrast improving.](image)

C. Applying Segmentation Algorithms

In this section, the FCM algorithm is applied to the green plane of the image of previous stage. The number of sections is considered as three. Fig. 7 shows the results of the algorithm.

![Fig. 7. Results of the algorithm.](image)
As shown in Fig. 7 and in comparison with Fig. 6, the contaminated parts of image appeared in varying degrees in the first and second parts, so that the completely contaminated parts are appear in the first segment, and the semi-infected parts in the second segment. Viewing the actions of the algorithm on some other healthy or infected images indicates the ability of FCM algorithm to distinguish infected parts from healthy parts. Therefore, it seems that the percentage of infected and healthy parts in different segments can be considered as a decision factor for classifying the leaves into healthy or infected categories.

**Fig. 8** shows the percentage of the two segments for all images. As implied, all infected leaves (numbers 1-74) have some values in both segments. The average values are 31.5% and 60.5% for segments I, II, respectively. On the other hand, healthy leaves have some values in segment I, averagely 30.5%, while they don’t have any value in segment II.

### Classification Results

In this section, the support vector machines algorithm (SVM) is performed over the previous stage dataset. To increase the reliability of the proposed algorithm, k-fold cross validation method with k = 3 is used. Based on this technique, the data in both classes are divided into 3 equal parts, and each time the two parts are used to train the network, and the remaining part is used for testing the model. This operation is repeated three times, so that all parts are considered for evaluation once. As a result, the mean value of the round results is calculated. The final results are presented in Table I.

| Real Classes | Infected | Healthy |
|--------------|----------|---------|
| Infected     | 97.28    | 2.72    |
| Healthy      | 0        | 100     |

In order to see the effect of K value, the system is launched by selecting K=5. The detection result is shown in Table II.

| Real Classes | Infected | Healthy |
|--------------|----------|---------|
| Infected     | 97.33    | 2.67    |
| Healthy      | 0        | 100     |

The values shown in Tables I and II could be defined as below:

- **True Positive (TP):** Infected leaves correctly identified as infected.
- **False Positive (FP):** Healthy leaves incorrectly identified as infected.
- **True Negative (TN):** Healthy leaves correctly identified as healthy.
- **False Negative (FN):** Infected leaves incorrectly identified as healthy.

As shown in Table II, the TN indicator of the proposed method is 100%, meaning that all healthy leaves are predicted correctly, and the system does not have any wrong prediction among healthy leaves. The TP indicator for the proposed method is 97.33%, and FP is 2.67%, implying that among the infected leaves, only 2.67% of them are incorrectly predicted.

In addition to the above parameters, two other parameters can be defined as below:

- **Sensitivity:** Sensitivity, or true positive rate, refers to the system’s ability to correctly predict infected leaves which do have the condition.
  \[
  SENS = \frac{TP}{TP+FN}
  \]  

- **Specificity:** Specificity, or true negative rate relates to the system’s ability to correctly reject healthy leaves which do not have the condition.
  \[
  SPEC = \frac{TN}{TN+FP}
  \]

From table II, the sensitivity and specificity of the proposed method can be obtained as 1 and 0.974, respectively. Also, overall average accuracy is achieved as 98.66%, which shows better results comparing to existing works such as [30].

### Conclusions and Future Work

Considering the importance of using machine vision methods to detect GLFV disease, this paper presents a method based on a combination of the FCM segmentation algorithm and the support vector machines (SVM) algorithm. Firstly, real images were collected in the Khorasan, Iran region. After applying some pre-processings, each image was divided into three segments using the FCM algorithm, and the percentage of the contaminated part to the healthy part for two first segments is considered as input to the SVM algorithm. The results of applying the SVM algorithm showed that the proposed algorithm is able to separate infected leaves from healthy leaves with an overall...
accuracy of 98.6%. This paper showed that the proposed method for detecting diseases has good potential and is capable of detecting infected plants only by processing their leaves, which overcomes the existing limitations. Since other existing methods for detecting this virus are time-consuming and cost-effective, the use of this method can detect GFLV disease in time, and reduce the cost of diagnosis, which will decrease economic losses.

Considering the ease of preparation of images from any objects such as plants, trees, leaves etc., the proposed model has no limitation in distinguishing their different types, after learning suitably. Therefore, complete datasets in each way is desirable. Developing android or IOS based versions of the model could be very useful for farmers.

References

[1] Kaur, Sukhvir, Shreelakha Pandey, and Shivani Goel. “Plants disease identification and classification through leaf images: A survey.” Archives of Computational Methods in Engineering 26.2 (2019): 507-530.

[2] Al Bashish, Dheeb, Malik Braik, and Suleiman Bani-Ahmad. “Detection and classification of leaf diseases using K-means-based segmentation and.” Information Technology-Journal 10.2 (2011): 267-275.

[3] Mahmoudi, M. Khazaei, J., Vahdati, K., Taleb, M. “Identification of walnut diseases using visual machine techniques”, first National Congress of Agricultural Science and Technology, Zanjan university, Iran, 2011, in Persian.

[4] Zhang, Shanwen, et al. “Cucumber leaf disease identification with global pooling dilated convolutional neural network.” Computers and Electronics in Agriculture 162 (2019): 422-430.

[5] Sun, Xiaoxiao, et al. “Image Recognition of Tea Leaf Diseases Based on Convolutional Neural Network.” arXiv preprint arXiv:1901.02694 (2019).

[6] Shah, Nikhil, and Sarika Jain. “Detection of Disease in Cotton Leaf using Artificial Neural Network.” 2019 Amity International Conference on Artificial Intelligence (AICAI). IEEE, 2019.

[7] HOSSEINI, H., ZAMANI D. MOHAMMAD, and A. ARBAB. “A recognition system to detect Powdery Mildew and Anthracnose fungal disease of cucumber leaf using image processing and artificial neural networks technique.” (2018): 15-28.

[8] OMRANI, ELHAM, et al. “Identification of apple leaf varieties using image processing and adaptive neuro-fuzzy inference system.” (2015): 67-75.

[9] Al-Hiary, Heba, et al. “Fast and accurate detection and classification of plant diseases.” International Journal of Computer Applications 17.1 (2011): 31-38.

[10] Meunkaewjinda, A., et al. “Grape leaf disease detection from color imagery using hybrid intelligent system.” 2008 5th international conference on electrical engineering/electronics, computer, telecommunications and information technology. Vol. 1. IEEE, 2008.

[11] Wang, Haiguang, Guanlin Li, Zhanhong Ma, and Xiaolong Li. “Image recognition of plant diseases based on principal component analysis and neural networks.” In 2012 8th International Conference on Natural Computation, pp. 246-251. IEEE, 2012.

[12] Dubey, Shiv Ram, et al. “Infected fruit part detection using K-means clustering segmentation technique.” International Journal of Interactive Multimedia and Artificial Intelligence 2.2 (2013): 65-72.

[13] Belkhodja, Leila, and Djamilah Hamdadou. “MCAD: Computer Aided System for Breast Masses Detection based on Immune Recognition.” International Journal of Interactive Multimedia and Artificial Intelligence 5.5 (2019): 97-108.

[14] Pujari, Devashish, Rajesh Yakkinumthath, and Abdulmunaf S. Byadgi. “SVN and ANN based classification of plant diseases using feature reduction technique.” International Journal of Interactive Multimedia and Artificial Intelligence 3.7 (2016): 6-14.

[15] Sutic, Dragoljub D., Richard E. Ford, and Malisa T. Tosic. Handbook of plant virus diseases. CRC Press, 1999. 535p.

[16] Vuitenz, A. 1970. Fanleaf of grapevine. In: N. W. Fraizer. Virus disease of small fruits and grapevine. University of California, Berkeley, plant pathology. 217-228.

[17] Martelli G.P. 1993. Grapevine degeneration fanleaf. In: Martelli G.P.