2021 International Symposium on Electrical and Electronics Engineering (ISEE) | 978-1-6654-1487-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISEE51682.2021.9418680

Citrus Leaf Disease Detection and Classification Using Hierarchical Support Vector Machine

Hanh Dang-Ngoc*†, Trang N. M. Cao*, Chau Dang-Nguyen*†
*Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam
†Vietnam National University Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam
Email: {*hanhdn, trang.cao1814391, chaudn*}@hcmut.edu.vn

***Abstract*—In this paper, an effective framework is proposed to classify four types of citrus leaf diseases out of healthy leaves through leaf features inspection. Four considered citrus leaf diseases are canker, sooty mold, greening and leaf-miner. Our framework consists of three main stages of pre-processing, feature extraction and classification. Since image analysis of citrus leaf disease is based on texture, color and shape features through leaf region, the region of main leaf is first segmented from the complex background in the pre-processing stage. Then leaf features are extracted in different color spaces and prominent ones are chosen based on feature distribution analysis. The selected features are fed to the SVM classification stage to detect and classify the diseases. In our study, we propose a hierarchical SVM classification model based on leaf features analysis to improve the accuracy rate. The numerical results prove the effectiveness of our proposed hierarchical SVM model over the multi-class SVM model. Our proposed hierarchical SVM model gives the infected leaf detection rate of 92.5%, and a high accuracy rate of 91.76%.**

***Index Terms*—citrus leaf diseases, leaf segmentation, classification, hierarchical SVM, multi-class SVM.**

I. Introduction

The citrus plants such as lemons, oranges, limes and grapefruits are not only the important constituents of Vietnam's agricultural economy but also the commonly grown fruits all over the world. However, diseases of citrus plants are inevitable, which cause a hazardous threat to productivity and degrade the quality. Therefore, effective and fast leaf disease detection plays an important role in the early treatment to reduce the effect of infection and to prevent the spread to other plants. Early detection of disease symptoms can be done by plant physiologists through visual observation of each suspicious tree; or carrying molecular biology in laboratory [1], [2]. The lack of experts in this area is limiting the wide identification of the disease, other methods of biochemistry, molecular biology are costly and time consuming.

In the prime of the internet of things (IoT), smart agricultural techniques have recently been attracted to be deeply studied. Image processing, which has been applied in many aspects of life, can be used to automatically measure affected areas of disease and determining the difference in the color, shape and texture distribution of the affected area [2]–[6]. The symptoms of disease can be found in various parts of leaves, steams or fruit but detection on leaves can be applied for early warning and proper treatment [1]. Detection of unhealthy or infected leaves has been studied in [6], whereas [2], [3] have studied classification framework to classify a number of citrus leaf diseases. The main differences between their frameworks are the features extraction and features selection steps. The inspected color, texture, and geometric features of citrus leaf diseases consists of principal component analysis (PCA) score, entropy, and skewness based covariance vector [3]; the color co-occurrence method (CCM) textural features [2]; gray-level co-occurrence matrix (GLCM) based contrast, correlation, energy and homogeneity features [6], etc. Sharif and others [3] propose a complex hybrid method for detection and classification of diseases in citrus plants, including both leaves and fruit. Their method has been proven to be effective over a dataset of well-captured images which capture only leaves/fruits or focus only on infected regions. Pydipati and others [2] presented a research used to identify diseased and normal citrus leaves under laboratory conditions. Those above methods might not be feasible in case of practical application of in-field automatically monitoring trees day-by-day without taking off samples of leaves; or under different environments of complex background while taking images.

In this paper, we propose a framework for improving the accuracy of citrus leaf diseases detection system including main leaf region segmentation from complex background of practical captured images; feature analysis based on hierarchical support vector machine classification. The first stage of automatic segmentation of the main leaf region is applied to ensure the appropriate inputs for the classifier in practical application of taking images of suspected leaves in different environments. Then, feature engineering of extracting and selecting features from different color spaces is processed to obtain a small subset of prominent features. Finally, citrus leaf diseases are detected and classified using multi-class SVM and our proposed hierarchical SVM in classification stages.

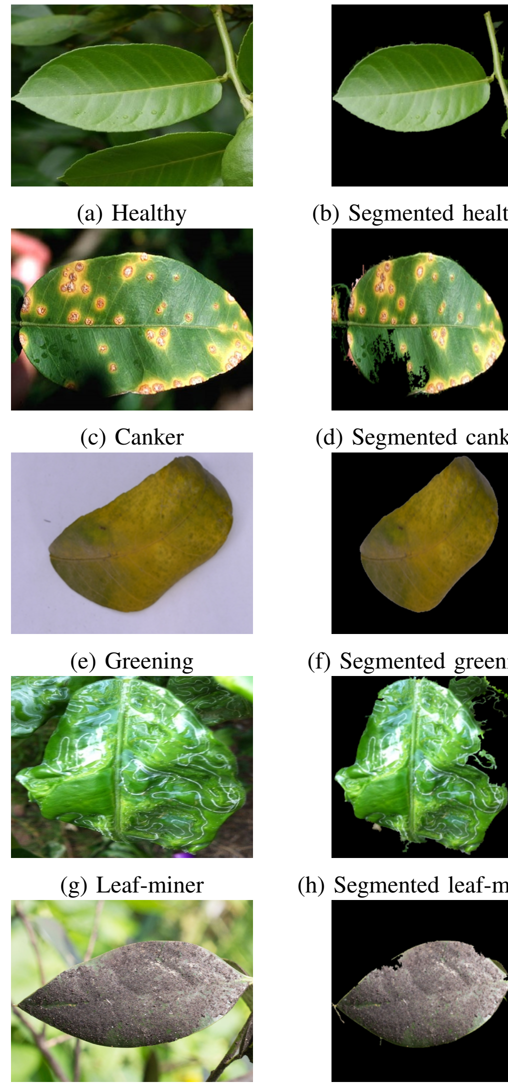
The remainder of this paper is organized as follows. Section II describes our proposed framework of main leaf segmentation; features extraction; multi-class and hierarchical classification using SVM. The experimental results and discussion are then carried out in Section III. Finally, Section IV is the conclusion that remarks the contributions of our work.

II. Methodology

Three main stages in our proposed framework are (i) main leaf region segmentation; (ii) features extraction and selection using features distribution analysis; (iii) multi-class SVM and hierarchical SVM based citrus leaf disease detection and classification. Due to noises during the capturing process such as complex background or overlap leaves, the segmentation stage plays an important role in improving initial images for detection systems by making the interest region in the image more visible compared to the original image. The second stage of features extraction and selection is the most challenging step in computer vision based applications since it significantly affects the accuracy of the final classification stages. In our study, we propose a hierarchical SVM classification model based on leaf features analysis to improve the classification rate.

A. Citrus Leaf Diseases

Four types of citrus leaf diseases are considered to be detected out of healthy leaves are citrus canker, citrus greening, citrus leaf-miner, and sooty mold. The infection results can be defoliation, die-back, premature leaves, fruit drop and at last the citrus plants will be unproductive or likely to achieve low quality for the market. The symptoms of these diseases can be seen as in Fig. 1. Citrus canker, a bacterial disease of citrus plant leaves, is known as a leaf-spotting disease [7], as shown in Fig. 1(c), in which often lesions will be encircled by a water-soaked margin and surrounded yellow nodes. Greening disease is caused by a bacteria pathogen as shown in Fig. 1(e). Citrus leaf-miner is the mining insect that commonly attacks citrus leaves, creates shallow tunnels, or mines, in young leaves of citrus trees as shown in Fig. 1(g). Sooty-mold as shown in Fig. 1(i) is the result of massive, superficial fungal growth most commonly occurring on leaves, fruit, twigs, and small branches.



(a) Healthy (b) Segmented healthy
(c) Canker (d) Segmented canker
(e) Greening (f) Segmented greening
(g) Leaf-miner (h) Segmented leaf-miner
(i) Sooty mold (j) Segmented sooty-mold

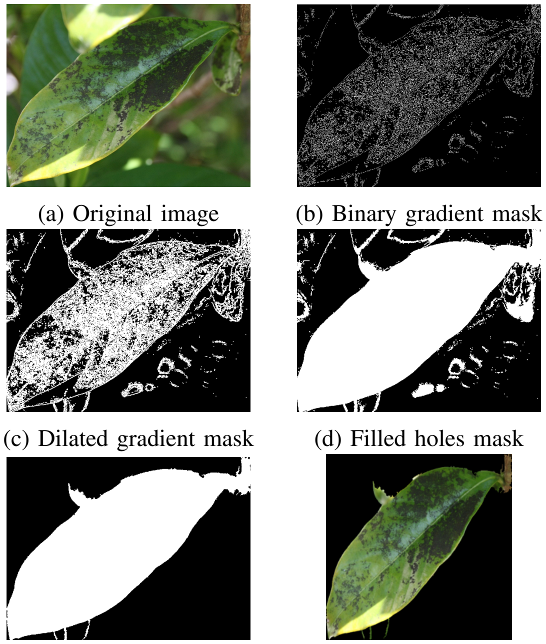
Fig. 1: Examples of successful segmented leaves.

B. Main Leaf Region Segmentation

Due to the lack of real infected citrus plants, images of healthy and infected leaf were collected from many sources. Our dataset of citrus leaf images consists of (i) standard dataset, which capture leaf samples in same environment condition of the laboratory[1]; and (ii) citrus leaf images acquired from many sources online because of the unavailable of those citrus leaf disease images in the standard dataset. Images of healthy and four considered citrus leaf diseases are collected from many sources, which might have complex background and overlap leaves. Therefore, the region of main leaf needed to be segmented from the background and other leaves, which is an important step in our leaf image processing based application. This first stage includes (i) main leaf segmentation from complex background using Sobel operator and a tuned threshold values; and (ii) image resize to 256x256 pixels for standardizing the dataset for next stages.

Since the object being segmented differs greatly in contrast from the background image, changes in contrast can be detected by operators that calculate the gradient of an image. In our framework, the Sobel operator is used to calculate the threshold value, later tune the threshold value and choose the most appropriate value for all classes of the training database, then use edge detection to obtain a binary mask as in Fig. 2(b). The binary gradient mask shows lines of high contrast in the image, but these lines do not quite delineate the outline of the object of interest. The binary gradient mask is then dilated using linear structuring elements such as the vertical structuring element followed by the horizontal structuring element, results in Fig. 2(c). Then Fig. 2(d) shows the result

[1]Citrus leaves dataset: Rauf, Hafiz Tayyab; Saleem, Basharat ALi ; Lali, M. Ikram Ullah ; khan, attique; Sharif, Muhammad; Bukhari, Syed Ahmad Chan (2019), "A Citrus Fruits and Leaves Dataset for Detection and Classification of Citrus Diseases through Machine Learning", Mendeley Data, V2, doi: 10.17632/3f83gxmv57.2



(a) Original image (b) Binary gradient mask

(c) Dilated gradient mask (d) Filled holes mask

(e) Final mask (f) Segmented image

Fig. 2: Segmentation steps.

after filling interior gaps inside the dilated gradient mask. Any small uninterested areas in filled holes mask are removed to obtain a final mask as shown in Fig. 2(e). Finally, all segmented images are automatically resized to 256x256 pixels in order to standardize the database Fig. 2(f). Some successful segmented images of healthy and infected leaves are shown in Fig. 1.

C. Features Extraction

Since each disease has its own features which can be best described in individual color space, considered features will be extracted from the segmented RGB images, and other color spaces which are HSI , HSV, L*a*b and YCbCr. In particular, there are four texture features of contrast, energy, correlation and homogeneity; four color features of mean, standard deviation, skewness and kurtosis; and three shape features of smoothness, variance and entropy are extracted from each color space.

1) Texture Features: Gray-level co-occurrence matrix (GLCM) is the statistical method of investigating texture which considers the spatial relationship of pixels. The GLCM functions characterize the texture of images by computing the spatial relationship among the pixels in the images [6]. Leaf image is first converted into gray-scale image then features are extracted as an array of offsets which describe pixel relationships of varying direction and distance. Four GLCM based features are contrast, energy, homogeneity and correlation. Contrast is a measurement value of the intensity contrast between a pixel and its neighbor over the whole image, while correlation is a measure of how correlated a pixel is to its neighbor over the whole image. Energy is the sum of squared elements in the normalized GLCM, and homogeneity measures the closeness of the feature distributions in the GLCM to the GLCM diagonal.

2) Color Features: Since each color space produces distinct information, the color features of mean, standard deviation, skewness and kurtosis values, are obtained by using different color spaces and parameters. Mean calculates the average of all pixel values and replaces the center pixel value in the destination image with that result. Standard deviation shows how much the data points are close or spread out over a large range of the average mean value. Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable, while kurtosis is a measure of the shape of the probability distribution of a real-valued random variable.

3) Shape Features: The shape is the main feature of image depiction. The shape features are divided into two types: first is regions based feature and the second is boundaries based feature [4]. In our model, the shape features are related to the leaf geometry and are calculated using morphological operations. Smoothness is the measure of relative intensities in the segmented region of the image and can be calculated by the equation in [8]. The measure of smoothness will return values of 0 for a region of constant intensities, and values of 1 when the region exhibits maximum disproportion in intensities. Variance is a measure of how far a set of numbers is spread out. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the expected value. In particular, the variance is one of the moments of a distribution. In that context, it forms part of a systematic approach to distinguishing between probability distributions. Entropy is a measure of randomness in the values of image pixels.

4) Features Selection: Although the color, texture and shape features play a major role in classification of plant diseases, using all of them might decrease the system classification performance in terms of classification rate (due to irrelevant features) and time consuming (due to high number of features). Based on the analysis of distributions of each features changes between healthy leaf samples each disease, prominent features with good discrimination or less overlap are selected to feed into SVM classifier.

D. Multi-class SVM

SVM is utilized for a non-linear classifier. In our framework, two models of multi-class SVM and hierarchical SVM are applied in the final classification stage. The multi-class SVM classification model used distribution analysis of each features changes between four diseases and healthy leaves. Based on visual and statistical analysis using the distribution of each feature to show how it changes between different diseases, the good features were selected as prominent features. Some features have distributed as large gaps from one class to another four classes are also chosen as prominent features since they can help to classify that class from others. Examples of distribution of two selected features are shown in Fig. 3. The plots have the red lines at median value, the boxes have

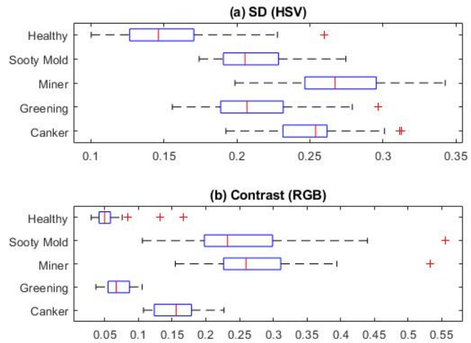


Fig. 3: Distribution of selected features for multi-class SVM.

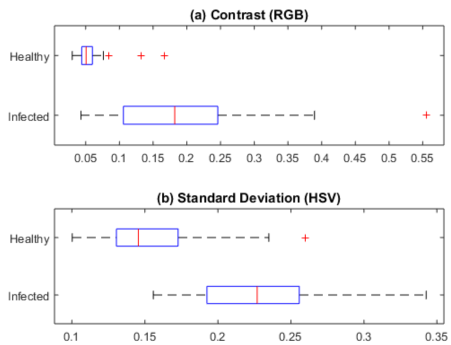


Fig. 4: Distribution of selected features for basic classifier.

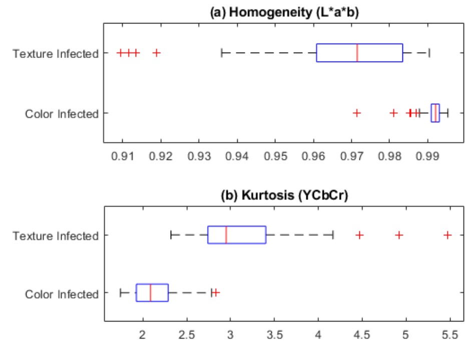


Fig. 5: Distribution of selected features for middle classifier.

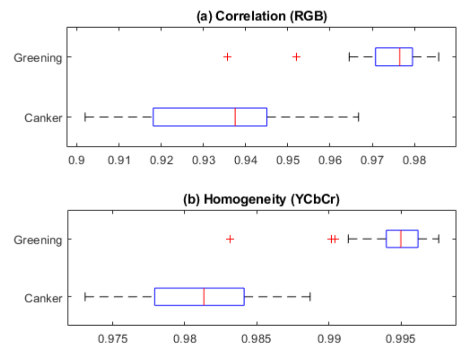


Fig. 6: Distribution of selected features for final classifier 1.

lines at lower and upper quartile, while the whiskers show the extent of the rest of the data, and outliers for data beyond the ends of the whiskers.

E. Proposed Hierarchical SVM Classification System

As seen in Fig. 3, the distributions show the considerable overlap between 5 classes, which means that it is not an easy task to construct a classifier to detect and determine leaf diseases. Therefore, we propose a hierarchical SVM classification system based on analyzing the distribution of each feature change between four diseases and healthy leaves. In the first classifier, infected leaves are detected out of healthy ones by dividing investigated images into two basic classes of healthy and infected. As examples, Fig. 4 shows how contrast from RGB color space and standard deviation from HSV color space give good distribution to classify healthy leaves and infected leaves. All infected labeled images are continued to go through the middle classifier, where they are divided into two middle combined-classes of color infected and texture infected. In our system, color infected class is defined for a combined class of canker and greening, while texture infected class is defined for a combined class of leaf-miner and sooty mold. We combine each two diseases into combined-classes based on the fact that some features are distributed into group of two diseases with good discrimination and non-overlap. As examples, Fig. 5 shows how kurtosis from YCbCr color space and homogeneity from L*a*b color space give good distribution for this important step.

Color infected labeled images continue to go through final classifier 1, where they are divided into two final classes of canker and greening. Fig. 6 shows how correlation from RGB color space and homogeneity from HSV color space give good distribution for the final classifier 1. Whereas, texture infected labeled images continue to go through the final classifier 2, where they are divided into two final classes of leaf-miner and sooty mold. As examples, Fig. 7 shows how smoothness from HSI color space and correlation from RGB color space give good distribution for the final classifier 2.

III. Results and Discussion

Our proposed framework for citrus leaf disease classification is modeled and evaluated through a dataset of total 340 samples of citrus leaf images as in Table I, which consists of healthy ones and four types of infected leaves (i.e. citrus canker, greening, leaf-miner, and sooty mold). The diseases can be at different phases of severity and taken under various environments of background and light. The overall performance of our system was calculated in terms of classification rate as shown in Table II and Table III with detailed discussion as follows.

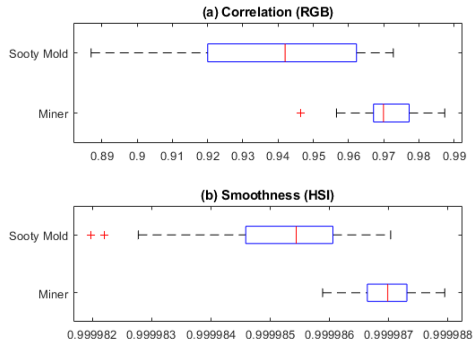


Fig. 7: Distribution of selected features for final classifier 2.

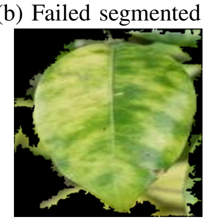
TABLE I: Dataset for training and testing phases.

	Healthy	Canker	Greening	Miner	Sooty mold
Training	*50*	*30*	*30*	*30*	*30*
Testing	*50*	*30*	*30*	*30*	*30*

(a) Original image 1

(b) Failed segmented 1

(c) Original image 2



(d) Failed segmented 2

Fig. 8: Example of failed segmented images.

A. *Main Leaf Region Segmentation*

Fig. 1 shows the step-by-step results of the main leaf region segmentation stage, while Fig. 2 shows some successfully segmented images of healthy leaf and four different diseases. Our segmentation stage is based on a proper chosen tuned threshold value of 0.32, which can not be easily satisfied for all images with different complex backgrounds in our collected dataset. Fig. 8 shows examples of failed segmented images with the defined threshold value. In Fig. 8(b), not only non-leaf region but even a part of the main leaf is removed because of unwanted brightness in image captured condition, while complex background of overlap leaves with same color and low intensity cause failed segmentation with more region of other leaves in captured image, as in Fig. 8(d). However, those failed segmented images show the acceptable segmenting result of our method since the segmented region still has an infected region or a little more region of the same leaf color. Our proposed segmentation method is proven that it is applicable for images which are automatically captured under outdoor conditions in practical IoT application (i.e., step of segmentation multiple leaves image into number of single leaf images should be considered). Therefore, all training and testing images of our dataset are main leaf region segmented automatically using our segmentation model before going through next stages.

B. *Feature Selection and Classification*

To evaluate the effect of our framework quantitatively, the accuracy rate is chosen as the performance metric. We apply our framework through a testing dataset; then count all classified results to calculate the accuracy rate. All testing images go through the main leaf region segmentation stage to have segmented images. Total 55 considered features are extracted from each segmented image. Here, we present selected features for two multi-class SVM classification models and discuss the numerical results.

1) Multi-class SVM Classification: There are 19 features are selected in total to feed the multi-class SVM classification system for classifying 5 classes which are healthy leaves, canker, greening, miner and sooty mold. Those chosen features are correlation and variance from HSI color space; contrast, correlation, standard deviation and variance from HSV color space; contrast, correlation, mean, standard deviation and variance from L*a*b color space; contrast, correlation, homogeneity and variance from YCbCr color space; contrast, homogeneity, standard deviation and variance from RGB color space.

TABLE II: Confusion matrix of multi-class SVM system.

	Healthy	Canker	Greening	Miner	Sooty mold
Healthy	*39*	2	8	1	0
Canker	0	*30*	0	0	0
Greening	0	5	*25*	0	0
Miner	0	0	0	*30*	0
Sooty Mold	1	0	0	3	*26*
Accuracy rate	*88.24%*				

Table II has shown testing results of 5 classes using multi-class SVM classifier. With 19 chosen features, we obtain a total accuracy rate of 88.24% for 5 classes.

2) Hierarchical SVM Classification: Our proposed hierarchical SVM classification model shows its effectiveness with the total accuracy rate of 91.76%, with detailed results are shown in Table III.

There are 13 selected features for basic classifier, including homogeneity and correlation from HSI color space; contrast, variance, standard deviation and homogeneity from HSV color space; homogeneity, correlation and contrast from YCbCr color space; homogeneity and contrast from RGB color space; variance and correlation from L*a*b color space. Distribution of these features shows good discrimination or less overlap between healthy and infected classes. These features have proven effectiveness with high accuracy rate of 94.12% with the infected leaf detection rate of 92.5% in particular (111 successful detected samples over total 120 inflected samples as in Table III).

TABLE III: Confusion matrix of hierarchical SVM system.

Basic Classifier		
	Healthy	*Infected*
Healthy	**49**	1
Infected	9	**111**
Basic accuracy rate	160/170 = 94.12%	
Middle Classifier		
	Color Infected	*Texture Infected*
Color Infected	**54**	0
Texture Infected	0	**57**
Middle accuracy rate	111/111 = 100%	
Final Classifier 1 (Color Infected)		
	Canker	*Greening*
Canker	**30**	0
Greening	1	**23**
Final 1 accuracy rate	53/54 = 98.15%	
Final Classifier 2 (Texture Infected)		
	Miner	*Sooty Mold*
Miner	**27**	2
Sooty Mold	1	**27**
Final 2 accuracy rate	54/57 = 94.45%	
Total accuracy rate	***91.76%***	

All of infected labeled images are then divided into two combined-classes of color infected class and texture infected class using 24 features of variance, kurtosis, standard deviation, correlation and entropy from HSI color space; contrast, and correlation from HSV color space; smoothness, energy, kurtosis and mean from YCbCr color space; homogeneity, kurtosis, contrast, variance, standard deviation, skewness, smoothness, energy, entropy and mean from RGB color space; contrast, correlation and homogeneity from L*a*b color space. Those features show good discrimination or less overlap between combined classes of leaf-miner and sooty mold and combined class of canker and greening, which have successful middle accuracy rate of 100% as in Table III. This absolute result comes from specific traits of these 2 groups of diseases such as yellow scabrous spot and yellow veins of canker and greening, respectively, or silvery wobbly lines and black, powdery fungus as shown in miner and sooty mold symptoms, respectively.

Each combined class labeled images continues to go through final classifier 1 and final classifier 2. Here we select 6 features to classify canker and greening in the final classifier 1, including contrast, correlation and homogeneity from YCbCr color space; homogeneity, correlation and contrast from RGB color space. In the final classifier 2, 6 features used to classify leaf-miner and sooty mold are variance from YCbCr color space; variance, correlation and standard deviation from RGB color space; smoothness and mean from HSI color space. Distribution of these 12 features show good discrimination or less overlap between each pair of canker and greening; miner and sooty mold, which can be used to define exact diseases from combined class. The accuracy rate of the final classifier 1 and the final classifier 2 are 98.15% and 94.45%, respectively. All numerical results prove the effectiveness of our hierarchical SVM classification model with higher accuracy of 91.76% compared to multi-class SVM classification model.

By looking into failed classified images, we found a healthy young leaf with yellow-green color which is wrong detected as yellow symptom of greening disease Fig. 9(a); or severe leaf-miner infected leaf have darker region which is wrong detected as dark smooth region in sooty-mold leaf Fig. 9(b).

(a) Wrong detected Healthy (b) Wrong detected Miner

Fig. 9: Example of failed classified images.

IV. Conclusion

In this paper, a hierarchical SVM based classification system has been proposed and numerically evaluated using our collected dataset of citrus leaf images. The high accuracy rate proved the efficiency of our framework. The feature extraction and selection play an important role in the system accuracy rate. Our study can be used as an initial study for deploying an automatic system of infected tree detection and diseases classification, which can give fast warnings to have early treatment.

Acknowledgment

This research is funded by Ho Chi Minh City University of Technology – VNUHCM, under grant number T-DDT-2019-29. We would like to thank Ho Chi Minh City University of Technology (HCMUT), Vietnam National University Ho Chi Minh City (VNU-HCM) for the support of time and facilities for this study.

References

[1] A. Batool, Y. Iftikhar, S. Mughal, M. Khan, M. Jaskani, R. M. Abbas, and I. Khan, "Citrus Greening Disease–A major cause of citrus decline in the world–A Review," *Horticultural Sci. (HORTSCI)*, vol. 34, pp. 159–166, Jan. 2007.

[2] R. Pydipati, T. Burks, and W. Lee, "Identification of citrus disease using color texture features and discriminant analysis," *Comput. and Electron. in Agriculture*, vol. 52, no. 1, pp. 49–59, Jun. 2006.

[3] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Comput. and Electron. in Agriculture*, vol. 150, pp. 220 – 234, 2018.

[4] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," *Comput. and Electron. in Agriculture*, vol. 153, pp. 12–32, Oct. 2018.

[5] N. Agrawal, J. Singhai, and D. K. Agarwal, "Grape leaf disease detection and classification using multi-class support vector machine," in *2017 Int. Conf. on Recent Innovations in Signal Process. and Embedded Syst. (RISE)*, 2017, pp. 238–244.

[6] R. M. Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," in *2017 Int. Conf. on Innovations in Inform., Embedded and Commun. Syst. (ICIIECS)*, 2017, pp. 1–4.

[7] A. Das, "Citrus canker-A review," *J. of Applied Horticulture*, vol. 5, pp. 52–60, Jun. 2003.

[8] G. Saleem, M. Akhtar, N. Ahmed, and W. Qureshi, "Automated analysis of visual leaf shape features for plant classification," *Comput. and Electron. in Agriculture*, vol. 157, pp. 270 – 280, 2019.