

A Hybrid Combination of Multiple SVM Classifiers for Automatic Recognition of the Damages and Symptoms on Plant Leaves

Ismail El Massi¹(✉), Youssef Es-saady¹, Mostafa El Yassa¹, Driss Mammass¹, and Abdeslam Benazoun²

¹ IRF-SIC Laboratory, Ibn Zohr University, B.P. 80 000 Agadir, Morocco
ismail.elmassi@gmail.com, melyass@gmail.com,
{y.essaady,mammass}@uiz.ac.ma

² Department of Plant Protection, Hassan II Institute of Agronomy and Veterinary, Complex of Agadir, Temara, Morocco
abbenazoun@gmail.com

Abstract. A machine vision system is reported in this study for automatic recognition of the damages and symptoms on plant leaves from images. The system is based on a hybrid combination of three SVM classifiers including an individual classifier, which is used in parallel with a serial combination of two classifiers. The individual classifier adopts two types of features (texture and shape) to discriminate between the damages and symptoms. In serial architecture, the first classifier adopts the color features to classify the images; it considers the damages and/or symptoms that have a similar or nearest color belonging to the same class. Then, the second classifier is used to differentiate between the classes with similar color depending on the shape and texture features. A combination function is provided for comparing the decision of the individual classifier and of the serial architecture in order to achieve the final decision that represents the class of the form to be recognized. The tests of this study are carried out on six classes including three types of pest insects damages and three forms of fungal diseases symptoms. The results, with an overall recognition rate of 93.9%, show the advantages of the proposed method compared to the other existing methods.

Keywords: Hybrid combination of classifiers · Automatic recognition · SVM · Damages · Symptoms · Plant leaves · Image

1 Introduction

The agriculture in Morocco has several issues. The damages of pest insects and symptoms of fungal diseases are among the major problems. They cause the diminution of the quality and quantity of the agricultural products. Therefore, significant economic losses can be result to farmers. In this context, the pattern recognition and machine vision methods can be used to reduce the losses.

These methods provide support decision systems that are used as means of diagnosis and recognition of a phytosanitary problem from images of the infected plant [11]. In recent years, several approaches have been proposed in the literature concerning the recognition of the damages and symptoms [7]. Some of the principal approaches are published in [1–6, 11]. However, there is no satisfactory solution available with the existing methods, because the complexity of the studied system in which some damages and/or symptoms can have similar characteristics (Color, texture, shape). This similarity, in characteristics, between classes makes the classification more difficult.

In the present study, a method is proposed to reduce the problems that experienced by the previous approaches. The suggested method is based on a hybrid combination of classifiers technique [8] that consists on combining an individual classifier in parallel with a serial architecture of two classifiers. In that case, a combination function is proposed to compare the decision of the individual classifier and of the serial architecture in order to obtain the final decision which represents the class of the form (damages or symptoms) in the input image. The main objective of the adopted technique is to improve our previous method [3] that based only on a serial combination of two classifiers, and to reduce its classification errors, which are occurred especially by the first classifier that uses only the color in the classification of the images. The errors of our previous method are reduced, in fact, by the individual classifier, used in parallel with the sequential architecture, which does not adopt color features, and on the other hand it uses only texture and shape features. The improvement of our previous method includes also the features extraction step, in which in the proposed method we adopt more attributes in texture, shape and color features. Moreover, we use SVM method [12] for classification instead neural networks, because of its simplicity, and it also gives significant results.

This study focuses on six classes including, on the one hand, the damages of three pest insects (Leaf miners, Thrips and *Tuta absoluta*), and on the other hand, the symptoms of three fungal diseases (Early blight, Late blight and Powdery mildew), which are among the major challenges of the vegetables crops in Souss Massa region (located in the south of Morocco).

2 Related Works

In this section, we present some existing works in automatic recognition of the damages and symptoms on plant leaves, which are used in this study to contribute to the realization of our proposed method.

In [1], Camargo et al. implemented a machine vision system for automatic recognition of three classes including one class represented by the damages of a type of pest insect (Green stink), and two classes represented by two forms of fungal diseases (*Bacteria angular* and *Ascochyta blight*). Their approach is based on a set of features including color, texture, shape, lacunarity, fractal dimension and Fourier descriptors, which are needed to achieve the classification by SVM method. They are tested their method on 117 images, in which the recognition rate was 93%.

In [2], Wang et al. proposed an approach for automatic recognition of two fungal diseases symptoms (Downy mildew and Powdery mildew), based on neural networks in classification and on four sorts of features including fractal dimension, texture, color and shape. The recognition rate was 97% on a database of 83 images (50 of Downy mildew and 35 of Powdery mildew).

In [3], we proposed an approach for automatic recognition of four classes including the damages of two pest insects (Leaf miners and *Tuta absoluta*) and symptoms of two fungal diseases (Downy mildew and internal Powdery mildew). This method is based on a serial combination of two neural networks classifiers. The tests were carried out on a database of 200 images including 50 images in each class of the four adopted classes.

In [4], Al Bashish et al. introduced a system for the recognition of five diseases including Early scorch, Cottony mold, Ashen mold, Late scorch and Tiny whiteness, which mainly attack the cotton. Their system is based on the Haralick texture features and neural networks for classification. They tested the system on a database of 192 images of six classes (5 of diseases symptoms and one class represents the normal leaves). The global recognition rate was 93%.

3 Proposed System Design

The proposed system is based on hybrid combination of three SVM classifiers: A serial combination of two classifiers, which is used in parallel with an individual classifier. The system was preliminary designed as follow (Fig. 1).

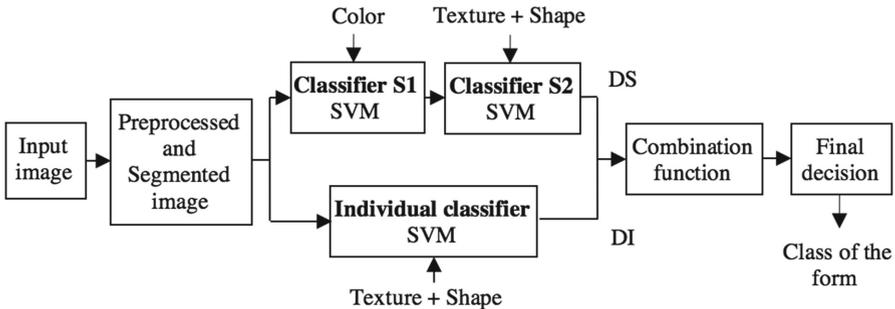


Fig. 1. Design of proposed system.

3.1 Image acquisition

As a first step in our work, the collection of the images in order to build the database. These images are needed to train and test the system. They are captured using a digital camera in several farms with the help of an expert in the agricultural field. Other images are downloaded from the Internet in order to increase the database size and to have diverse environments. Figure 2 shows some images of the database.

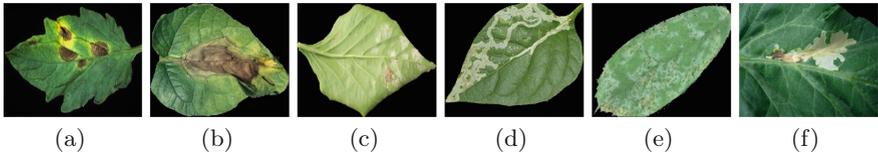


Fig. 2. Some images of the database. The symptoms: (a) Early blight, (b) Late blight, (c) Powdery mildew. The damages: (d) Leaf miners, (e) Thrips, (f) *Tuta absoluta*.

3.2 Preprocessing and Segmentation

The input image should to be preprocessed in order to improve its quality and to facilitate the segmentation and analysis steps. The adopted preprocessing methods in this work include resizing and filtering. At first, the image is resized with a standard size. Then, a filter median is applied as filtering method to reduce the noise in the image. The noise is generally due to acquisition process.

After preprocessing, the image is segmented in order to extract the infected area from leaf area. K-means clustering method [9] is used in this context for segmentation of the input image. This method is the most known and the most used in the previous works, since it gives good results in segmentation of the colored images (see [3,4,11] for more detail about segmentation with k-means clustering method). In this study, the k-means clustering algorithm segments the input image into k clusters ($k=5$ in our case) in which one of them containing the majority of the infected area. Figure 3 shows the segmentation results of two images of two infected leaves with Leaf miners damages and Late blight symptoms using the k-means clustering technique.

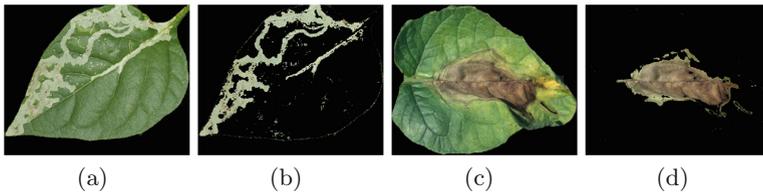


Fig. 3. Segmentation result of two images of two leaves infected by damages and symptoms. (a) Leaf infected by Leaf miners damages and (b) cluster containing the damages area. (c) Leaf infected by Late blight symptoms and (d) cluster containing symptoms area.

3.3 Features Extraction and Selection

The features extraction is the next necessary step to carry out. It consists of representing the segmented image on a vector of fixed features that should be distinct and relevant for the classifier performance. The selected features in this study include color, texture and shape features.

Color moments [10] method is used in this work for color features extraction. This method is adopted because of its ease of use and it provides important results. Color moment method is defined by three moments: Mean, Standard deviation and Skewness. These moments are extracted in this study for each R/G/B component of RGB color space and for each H/S/V component of HSV color space. In total, 18 color features are calculated.

For texture, we adopted Grey Level Co-occurrence Matrix (GLCM) [11] method. This method is the most adopted for texture features in the majority of previous works, since it gives extra informations for dictimation between the damages and symptoms. GLCM is a statistical analysis tool of an image in gray levels, which measures the distribution of gray levels in the image based on the spatial relations of pixels. Haralick introduces 14 attributes of texture based on GLCM. Five attributes were only used in this study including Contrast, Energy, Entropy, Homogeneity and Correlation. These five attributes are the most used in previous works, because they are relevant and distinct. In this work, these five attributes are calculated for each component R/G/B and for each component H/S/V, in which 30 texture features are extracted in total.

Twelve shape features are adpted in this study including Area, Perimeter, Circularity, Complexity, Solidity, Extent, Major axis length, Minor axis length, Eccentricity, Centroid and Diameter. We associate in this work the most used shape attributes in previous works, and then the results are significant.

3.4 Classification

The adopted classifiers in this study are based on SVM method [12]. They are used in a hybrid combination, in which a serial architecture of two classifiers is used in parallel with an individual classifier. Then, a combination function is provided to compare the decision of the serial architecture (DS) and of the individual classifier (DI) in order to achieve the final decision (FD) that represents the class of the damages or symptoms in the input image. Figure 4 demonstrates the architecture of the serial combination of two classifiers: The first classifier S1 uses the color to classify the images; it considers that the damages and/or symptoms, with similar or nearest color, belonging to the same class. For example, in our case, the damages of Leaf miners, Tuta absoluta and Thrips, and the symptoms of Powdery mildew are placed in the same class that is named class of

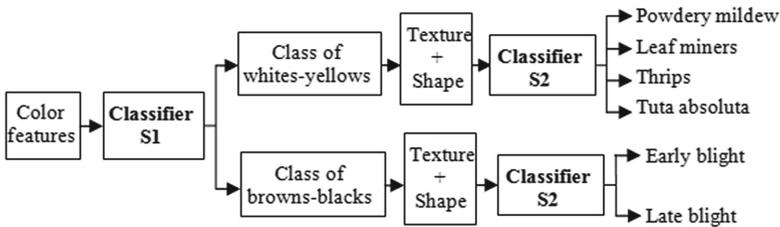


Fig. 4. Architecture of the serial combination

whites-yellows. Also, the symptoms of Early blight and of Late blight are placed in the same class which is named class of browns-blacks. Then, the second classifier S2 is used to differentiate between the classes with similar color depending on the texture and shape features. It has two cases: the first one consists of discriminating between the classes within the whites-yellows class; then, in the other case, the second classifier differentiates between the classes that belong to the browns-blacks class.

The individual classifier adopts texture and shape features, and based on the one-against-one strategy [13] of SVM for the classification of the six chosen classes. The color features have not been used in the individual classifier. This is because of the similarity between some classes in their color. Moreover, the objective of this classifier consists of reducing the errors occurred by the serial combination, especially by its first classifier that adopts only the color in classification of the images. Therefore, the individual classifier is used here to correct the incorrectly classified images in serial combination (Fig. 5).

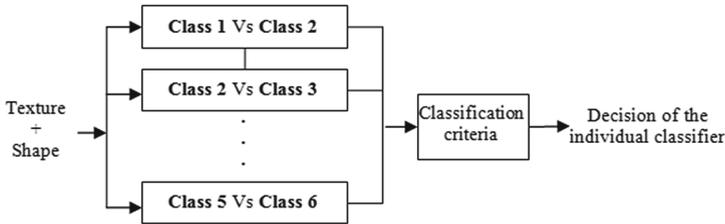


Fig. 5. Architecture of the individual classifier based on the one-against-one strategy of SVM.

The choice of the combination function plays an important role to obtain good results, and therefore having a more efficient system. Figure 6 shows the algorithm of the proposed combination function. When the decisions DS (Decision of serial architecture) and DI (Decision of individual classifier) are equal, then the final decision FD is the same. In the other case, when they are different, FD is obtained by applying a binary classification, based on SVM with the texture and color features, for discriminating between the two decisions DS and DI. This proposed algorithm which gives the good results in this study in the case where one of the two decisions DS and DI is incorrect and the other is correct.

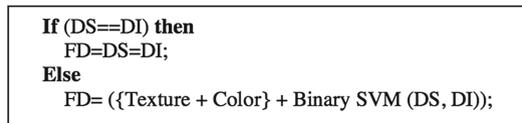


Fig. 6. Algorithm of the combination function

It is used to correct the wrongly classified instance of the two decisions. The limitations of this combination function are when the two decisions are incorrect then the final decision is also incorrect.

4 Experimental Results

In order to test and evaluate our method, a comparison is carried out with the three existing methods [1–3] (see Sect. 2). For that, experiments are carried out on 284 images of the six chosen classes (48 Early blight, 41 Late blight, 46 Powdery mildew, 58 Leaf miners, 38 Thrips, 53 Tuta absoluta).

As a first step, the three previous methods are implemented and tested on our database of images. Then, the obtained results are compared and analyzed for designing our proposed approach that gives the highest recognition rate.

In this experiment, the dataset is divided into two subsets: a set of 202 images (70%) used for training, and a set of 82 images (30%) for the test (Table 1). For dividing the dataset, we used the hold-out cross validation technique [14] that automatically generates the indices of the training and test set from the outputs vector of the six classes.

Table 1. Dividing the dataset into training and test set

Classes	Dataset	Training	Test
Early blight	48	34	14
Late blight	41	29	12
Powdery mildew	46	33	13
Leaf miner	58	41	17
Thrips	38	27	11
Tuta absoluta	53	38	15
Total	284	202(70%)	82(30%)

Table 2 demonstrates the results of the individual classifier and the serial architecture compared to the global result of the hybrid combination. The results are represented by the recognition rate, which is expressed in percentage with the ratio between the number of correctly classified images and the total images used for test. In this table, we see that the two decisions of individual classifier and serial architecture have the same recognition rate which is 87,80%, since the two decisions have the same number of correctly classified images (72 images from 82). The global recognition rate of the hybrid combination using the proposed function combination is 93,90%.

Table 3 illustrates the results of the individual classifier and serial architecture depending on the number of classified images used for test. From this table, we notice that there are some coincident errors but the majority is always correct and there exists one of the two decisions that produces the correct answer.

Table 2. Results of the individual classifier, serial architecture and hybrid combination.

Individual classifier	Serial architecture	Hybrid combination
87,80 %	87,80 %	93,90 %

Table 3. Decision result of the individual classifier and serial architecture for 82 testing images.

Decision of the individual classifier and serial architecture	Number
Two are correct	67
One is correct	10
Two are incorrect	5
Total	82

Table 4. Results of proposed method compared to the three existing methods [1-3]

Approach	Classification	Rate of recognition
Camargo et al. [1]	SVM	82,93 %
Wang et al. [2]	Neural network	79,27 %
El Massi et al. [3]	Neural network	85,36 %
Proposed method	SVM	93,90 %

Table 5. Comparison between the proposed method and the three existing methods depending on the correctly (CC) and incorrectly (IC) classified images

		El Massi [3]		Wang [2]		Camargo [1]		Σ	%
		CC	IC	CC	IC	CC	IC		
Proposed method	CC	68	9	63	14	67	10	77	93,90%
	IC	2	3	2	3	1	4	5	6,10%
	Σ	70	12	65	17	68	14		
	%	85,37%	14,63%	79,27%	20,73%	82,93%	17,07%		

Table 4 shows the comparison between the proposed method and the three existing methods depending on the adopted classification method and the obtained recognition rate.

In Table 5, we demonstrate the comparison between the proposed method and the three existing methods depending on the correctly and incorrectly classified images. The examination of this table shows that the difference between our method and the three methods is statistically significant.

Figure 7 demonstrates the results per class. The recognition rate is labeled for each class above the black curve that represents the proposed method. This figure shows that the three classes Powdery mildew, Leaf miners and Tuta absoluta

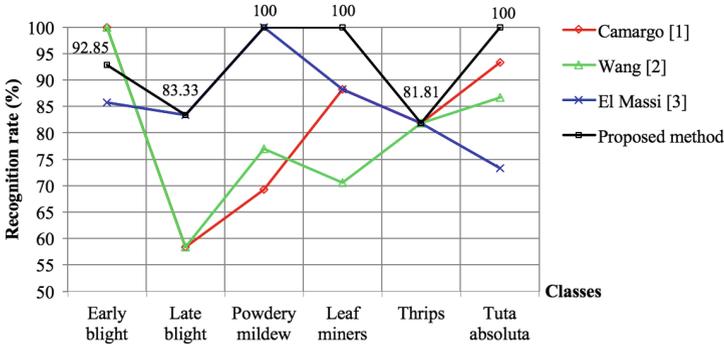


Fig. 7. Results per class of the proposed method compared to the three previous methods.

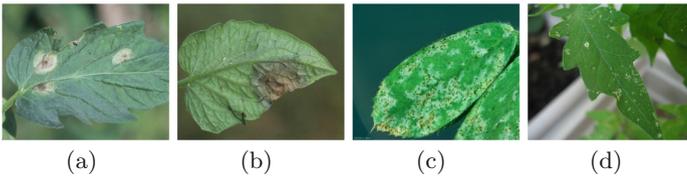


Fig. 8. Difference, in characteristics, between forms of the same class. (a) and (b): Two forms of Late blight symptoms. (c) and (d): Two forms of Thrips damages.

are totally classified (100%) by the proposed method. In the case of Early blight class, the rate is 92,85%, which is less than that of Camargo [1] and Wang [2].

The comparison, carried out in this study, indicate that the results obtained by our method, based on the hybrid combination of classifiers, are significant and encouraging. The examination of Fig. 7 demonstrates that the two classes Late blight and Thrips, with the recognition rate of 83,33% and 81,81% respectively, cause the diminution of the global accuracy of the proposed method. This diminution is due to, on the one hand, the number of images, used in these two classes, that are less than those used in the others, and on the other hand, to their complexity, in which it is difficult to discriminate between forms of the same class. Also, the complexity of these two classes is shown in similarity of the recognition rate, which is the same in the proposed method and our previous method [3]. Figure 8 shows two forms of Late blight symptoms and two forms of Thrips damages in two different stages of development. It appears, through the images, the difference, in the characteristics, between forms of the same class.

5 Conclusion and Future Work

Automatic recognition of the damages and symptoms on plant leaves is the aim of this study. The proposed system is based on a hybrid combination of classifiers technique, with the goal of reducing the classification problems that experienced by the previous methods. These problems are generally due to the complexity of the system of damages and symptoms. The results of this study show that the proposed method would be of importance to use as tool of diagnosis and phytosanitary problem recognition from images of the infected plant.

In future work, we plan to improve the proposed method with the use of other features in order to reduce the classification errors, and therefore obtaining more efficient system. We are also looking for increasing the database of images with the use of other damages and symptoms that attack the entire plant.

References

1. Camargo, A., Smith, J.S.: An image pattern classification for the identification of disease causing agents in plants. *Comput. Electron. Agric.* **66**(2), 121–125 (2009)
2. Wang, H., Li, G., Ma, Z., Li, X.: Image recognition of plant diseases based on back propagation networks. In: *IEEE International Congress Image and Signal Processing (CISP)*, pp. 894–900. IEEE (2012)
3. El Massi, I., Es-Saady, Y., El Yassa, M., Mammass, D., Benazoun, A.: Serial combination of two classifiers for automatic recognition of the damages and symptoms on plant leaves. In: *3th IEEE World Conference on Complex System (WCCS)*. IEEE, Marrakech (2015)
4. Al Bashish, D., Braik, M., Bani-Ahmad, S.: Detection and classification of leaf diseases using k-means clustering based segmentation and neural networks based classification. *Inf. Technol. J.* **10**(2), 267–275 (2011)
5. Tian, Y., Zhao, C., Lu, S., Guo, X.: SVM-based multiple classifier system for recognition of wheat leaf diseases. In: *IEEE Conference on Dependable Computing (CDC)*, pp. 189–193. IEEE (2010)
6. El Massi, I., Es-saady, Y., El Yassa, M., Mammass, D., Benazoun, A.: Automatic recognition of the damages and symptoms on plants leaves using parallel combination of two classifiers. In: *13th IEEE International Conference Computer Graphics, Imaging and Visualization (CGIV)*. IEEE, Beni Mellal (2016)
7. Vipinadas, M.J., Thamizharasi, A.: A survey on plant disease identification. *Int. J. Comput. Sci. Trends Technol.* **3**(6), 129–135 (2015)
8. Woniak, M., Graa, M., Corchado, E.: A survey of multiple classifier systems as hybrid systems. *Inf. Fusion* **16**, 3–17 (2014)
9. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no.14, pp. 281–297 (1967)
10. Patil, J.K., Kumar, R.: Color feature extraction of tomato leaf diseases. *Int. J. Eng. Trends Technol.* **2**(2), 72–74 (2011)
11. Sannakki, S.S., Rajpurohit, V.S., Nargund, V.B., Kulkarni, P.: Diagnosis and classification of grape leaf diseases using neural network. In: *4th IEEE International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1–5. IEEE (2013)

12. Bekkari, A., Idbraim, S., Elhassouny, A., Mammass, D., El yassa, M., Ducrot, D.: SVM and haralick features for classification of high resolution satellite images from urban areas. In: Elmoataz, A., Mammass, D., Lezoray, O., Nouboud, F., Aboutajdine, D. (eds.) ICISP 2012. LNCS, vol. 7340, pp. 17–26. Springer, Heidelberg (2012)
13. Hsu, C.W., Lin, C.J.: A comparison of methods for multi-class support vector machines. *IEEE Trans. Neural Netw.* **13**(2), 415–425 (2002)
14. Sylvain, A.: A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79 (2010)