Automatic Recognition of Plant Leaves Using Parallel Combination of Classifiers

Lamis Hamrouni, Ramla Bensaci, Mohammed Lamine Kherfi, Belal Khaldi, Oussama Aiadi

Automatic recognition of plant leaves using parallel combination of classifiers

Lamis Hamrouni[1,2] ✉@, Ramla Bensaci[1,2], Mohammed Lamine Kherfi[1,3] ,Belal Khaldi[1,2], Oussama Aiadi[1,2]

[1] Dept. Computer Science and Information Technologies, University of Kasdi Merbah Ouargla, 30 000, Algeria

[2] LAGE Laboratory, University of Kasdi Merbah Ouargla, 30 000 Algeria.

[3] LAMIA Laboratory, Université du Québec à Trois-Rivières, 3351, boul. des Forges, C.P. 500, Trois-Rivières, Canada.

lamis0215@gmail.com, ramla.bensaci@gmail.com, khaldi.belal@univ-ouargla.dz, aiadi.oussama@univ-ouargla.dz

Abstract. Because they are exploited in many fields such as medicine, agriculture, chemistry and others, plants are of fundamental importance to life on earth. Before it can be used, a plant need to firstly be identified and categorized. However, a manual identification task requires time, and it is not an easy task to do. This is because some plants look visually similar to the human eye, whereas some others may be unknown to it. Therefore, there has been an increasing interest in developing a system that automatically fulfils such tasks fast and accurate. In this paper, we propose an automatic plant classification system based on a parallel combination technique of multiple classifiers. We have considered using three widely known classifiers namely Naïve Bayes (NB), K-Nearest Neighbour (KNN) and Support Vector Machine (SVM). Our system has been evaluated using the well-known Flavia dataset. It has shown a better performance than those obtained using only one classifier.

Keywords: Morphological features, parallel classifiers, leaf classification, plant leaves, image recognition.

1   Introduction

Plants play an important role in our life, without them there will be no existence of the earth's ecology. They are widely exploited in our life such as in food, breath, health, and even in industry fields such as medicine, economic agriculture and so on. There are millions of plants species, some of them are subject to the danger of extinction [1]. Therefore, there is an urgent need for identifying plant species.

Traditionally, botanists classify plants using molecular biology and cellular features of leaves. Nevertheless, this task is very tedious, requires time and need the presence of expertise which is not available in all times. Additionally, an expert on one species or family may be unfamiliar with another. Subsequently, fast and accurate automatic plant identification system is highly needed.

Plants are usually classified using their leaves, stems, fruits or flowers. Leaves seem to be the most suitable parts that can be used to identify a plant. This is due to their availability in all seasons. In addition, leaves flatness makes it easy to be represented by the computer in some 2D.

In recent years, plant identification techniques have become a hot topic of research [2]. Authors in [3] have exploited the visual features of leaves in combination with Random Forest (RF) and Linear Discriminant Analysis (LDA) to classify and identify 30 plant species. In [4] a similar technique, but using Artificial Neural Networks (ANN) this time, has been opted for to identify 12 plant species. Pedro et al. [5] have focused on developing a system that can automatically identify medicinal plants such as herbs, shrubs and trees. As a classifier, they use ANN then Support Vector Machine (SVM).

Reader should notice that the former works use only one classifier in their systems. Thus, some other works have tried to improve them by combining and using more than one classifiers in the system. In [6] as instance, authors suggest using a serial combination of two SVMs. Their main idea was to devote one classifier for color features, and the other one for both shape and texture features. Their evaluation has been carried out on a dataset that contains six diseases classes. The system reports an 87.7% accuracy. In [7], authors have used a parallel combination of two classifiers namely, ANN and SVM. The first one was devoted to texture, color and shape features, whereas, then the second one uses shape and texture features. The evaluation has been carried out on a dataset that is composed of six diseases classes and they reported 91.46% accuracy.

In this paper, we introduce a system that parallel combines three classifier namely SVM, NB and KNN. As features, we consider extracting shape features (i.e., morphological features) from the leaves. The evaluation of our system has been accomplished using the well-known Flavia dataset. More details about the proposed system will be given in the next section.

The rest of the paper is organized as follow: section 2 present the architecture of our system and then discuss the used morphological features. In Section 3, we conduct experimentations on the proposed system and report results. Finally, we draw some conclusions.

2   Proposed system

In order to achieve a better performance, our system consists in a parallel combination of three classifiers namely SVM, NB and KNN. These classifiers are trained using a set of morphological features that we extract from the leaves. In Fig. 1, we illustrate a general scheme that resumes the different main stages of our system.
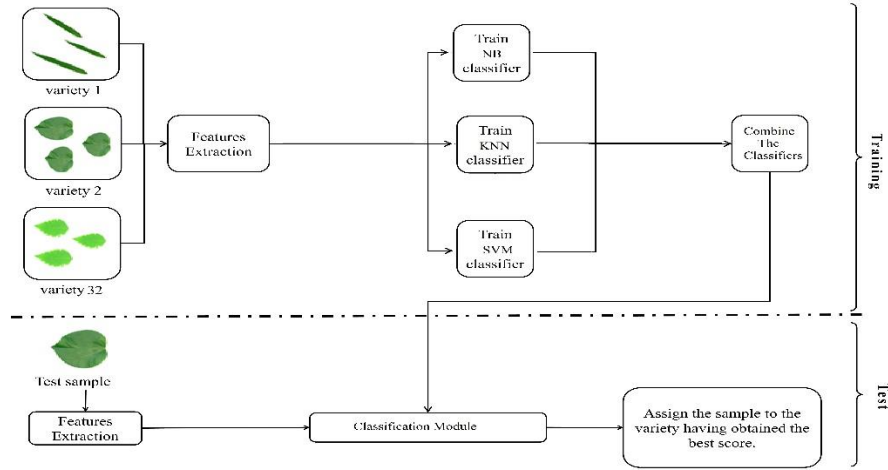
Fig. 1. Architecture of the proposed system.

2.1   preprocessing

The preprocessing stage is, generally, responsible for applying a set of treatment (e.g., noise reduction, rotation, transformation, etc.) on the image before employing it for features extraction. In our work, we firstly converted the original color images to gray-level then to binary images. Thereafter, a smoothing filter is applied to these binary images to reduce the noise. The steps involved in pre-processing are illustrated in Fig. 2.
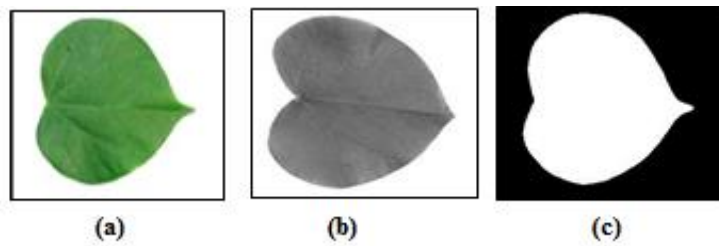

Fig. 2. Preprocessing stage (a).Image in RGB color space,(b).Gray-level image, (c).Binary image.

2.2   Features extraction

This stage aims to transform the objects into a vector of numeric values (i.e., feature vector). There are many types of features that can be extracted from an image, such as shape [8], texture [9], and color [10] features.

In this stage and after having the original image transformed into binary, we extract a set of shape features that describe the morphology of a leaf. Morphological features are obtained by extracting the basic geometrical properties [11] of the leaf such as: diameter, area, perimeter, major and minor. Fig. 3 shows an example of some geometrical features extracted from a leaf image.
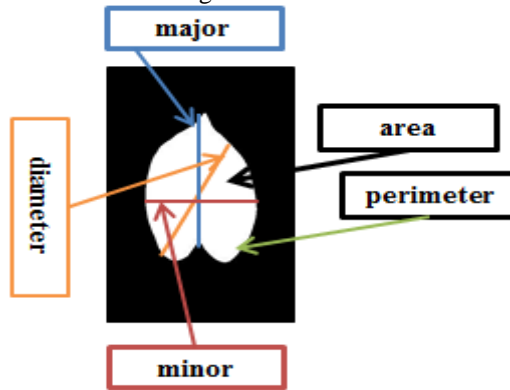

Fig. 3. Basic geometrical features

In our method, we extract the following five morphological features:

1. Diameter: is the longest distance between two points of the leaf contour D.
2. Area: is the number of pixels that constitute the area of the leaf.
3. Major axis length: is the distance between two terminal points orthogonal to minor axis length LP.
4. Minor axis length: is the longest distance orthogonal to major axis length WD.
5. Perimeter: the number of pixels at the margin of the leaf P.

In addition and based on these features, we extract another set of digital morphological features that were introduced by the authers of [11] [12], which are:.

6. Aspect Ratio: is defined as the ratio of major axis length LP to minor axis length WP. It is also called Eccentricity or Slimness. It is given by Aspect Ratio=LP/WP.
7. Perimeter Ratio of Physiological length & width: this features is the ratio of perimeter leaf and the sum of major and minor axis length, given by PRPW=P/(LP+WP).
8. Perimeter Ratio of Diameter: it is the ratio of perimeter to the diameter, given by PRD=P/D.
9. Rectangularity: The similarity between the leaf and a rectangle, given by R=(Lp*Wp)/A.
10. Narrow Factor: the ratio of the diameter D and length Lp (i.e., NF = D/Lp).
11. Circularity: The ratio involving the area A of the leaf and the square of its perimeter P, given by C = 4πA/P2.
12. Solidity: The ratio between *A* the area of the leaf and *Ach* the area of a convex hull, given by S = A/Ach.

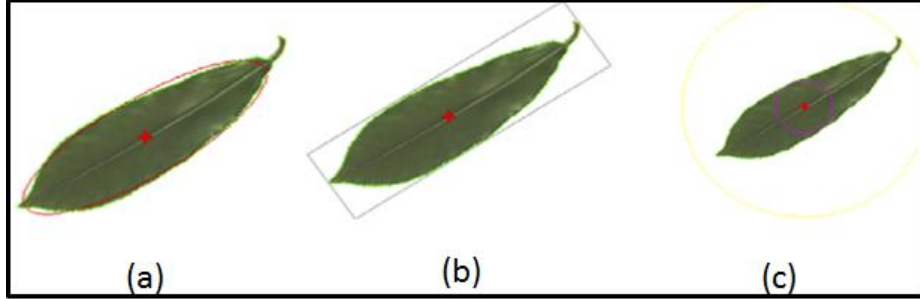Fig. 4 shows some of morphological features.

Fig. 4. Morphological features. (a). Form ellipse, (c). Rectangularity (d). Circularity

2.3 Classification

Classification or categorization is, generally, the process in which images are recognized, differentiated, and understood. The classifier needs to, firstly, be subjected to a set of labeled data (i.e., training set). Then, test samples will be provided to the classifier in order to recognize them.

In our work, a portion of the feature vectors, which are extracted in the previous stage, will be used to train our system that consists in three types of classifiers: KNN [11,12], NB[13] and SVM[14,15]. These three classifiers are combined in parallel. The parallel approach allows the different classifiers to operate independently of each other. The results of each classifier are then merged in order to obtain a higher recognition rate.

K-Nearest Neighbor (KNN)

KNN[11] is a simple yet very effective classification method. For a given test sample s, KNN mainly consists in determining the k closest training example to this s. It then states the class C that has the max membership degree to s as a class of s. the similarity could be extracted using the next formula (1).

$$d(x_r, x_s) = \left[\sum_{i=1}^{P} c_i\left(x_{ri}, x_s i\right)\right]^{\frac{1}{2}} \qquad (1)$$

Support Vector Machine (SVM).

SVM is a supervised classifier that has a great effectiveness especially with high dimensional data [14]. Formally, SVM constructs a hyperplane (alt. hyperplanes) that has the highest distance to the nearest training-data point of any class [15].

Naive Bayes (NB)

Naive Bayes classifiers [13] are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features [13].

$$P(^c/_x) = \frac{P(^x/_c)P(c)}{P(x)} \qquad (2)$$

where:

— $P(^c/_x)$: is the posterior probability of class c (i.e., target) given predictor x (i.e., attributes).
— P(c): is the prior probability of class.
— $P(^x/_c)$: is the likelihood, which is the probability of predictor given a class.
— P(x) : is the prior probability of predictor.

2.4   Combining classifiers

The main aim of combining classifiers is to improve the accuracy. Several works have been proposed in this context [6]. Classifier combination schemes could be roughly categorized into three main approaches, namely: sequential, parallel and hybrid combination.

2.4.1 Sequential combination

The sequential combination consists in placing one classifier after the other. It simpler words, the outcome of one classifier will be the input of another. Such a cascade structure helps to improve the decision taken from the previous step by including reliable samplers or excluding unreliable ones [6].

2.4.2 Parallel combination

In parallel combination, the different classifiers operate independently of each other. Then, the obtained results are fused together, by some method, to produce a final decision [7].

2.4.3 Hybrid combination

The hybrid combination scheme takes the advantages of the two previous schemes (i.e., sequential and parallel combination) in order to reach a more reliable decision. It illustrates the two aspects of the combination which are: in on the one hand reducing all possible classes and in the other hand finding the consensus between all classifiers, in order to reach a final decision [19].

In our system, we combine the classifiers in parallel way and we opted for voting-based methods [16]. In such module, each classifier provides a certain number of votes (i.e., potential classes for a given image). The final decision is then made based on these votes using the following formula:

$$E(x) = \begin{cases} c_i \ si \ \sum_i e(i) = \max_{c_i \in \{1,\ldots,M\}} \sum_j e(j) \geq \alpha K \\ else \ reject \end{cases}$$ (3)

where K is the number of the combined classifiers, α is a threshold that represents the needed number of votes for the same class to be relevant.
Voting-based method could be categorized into two main categories, the simple majority vote and the weighted vote.

In simple majority vote [17], each classifier votes for one class to be relevant to an input image. The final decision will, then, be made regarding the number of votes for each class. The relevant class is the one with the heights number of votes. Although this method is simple and effective, it suffers from the problem of rejection if all classes have the same votes number [18].

In the other hand, the weighted vote method associates each classifier with some coefficient (i.e., weight) that indicates the importance of the corresponding classifier in the combination. Selecting weights for different classifiers is a critical process and highly affects the quality of the results.

To take advantage of both the former voting methods, we suggest combining them in one module. In our used module, the system tries, firstly, majority vote method. If a conflict accrues using majority vote, then, the system opts for weighted vote method. The following algorithm resumes these two steps.

```
If (DSVM==DKNN==DNb) Then
      FD ← DSVM+DKNN+DNB
Else
      FD ← α DNB + β DKNN+Ω DSVM
Endif
```

3   Experimental results

The evaluation of the used approach has been carried out on Flavia dataset [4]. Flavia contains 1907 images that are categorized into 32 classes. Fig 5 shows representative simples from this dataset.
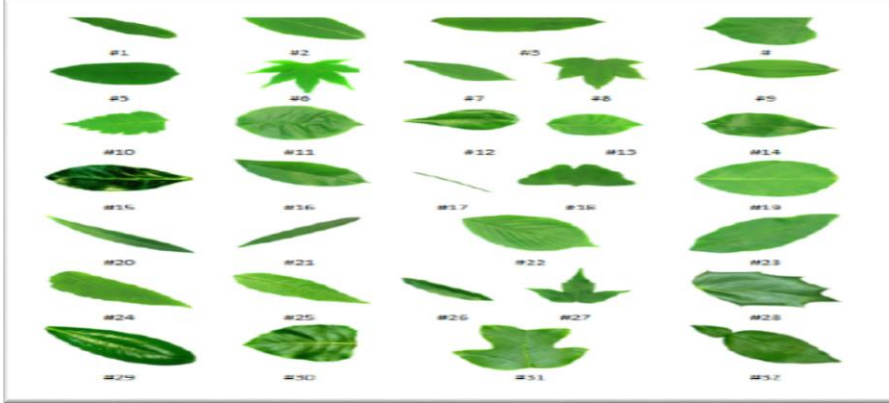

Fig. 5. Representative samples from Flavia dataset.

In our experiments, the dataset has been divided into two subsets. The First one consists of 1284 images (70%) used for training, whereas the rest 623 images (30%) have been used to test the system. We opted for Accuracy to be used as an evaluation metric of our system. It is given by the following formula.

$$\text{Accuracy (\%)} = \frac{Nc}{Nt} * 100 \qquad (4)$$

where Nc is the number of correctly classified images and Nt is the total number test.

In order to prove the performance of our method, we have firstly classify images using independently each classifier. Table 1 shows the obtained results.

Table 1. Accuracy results using each classifier separately.

Classifier	Average accuracy%
NB	72,231
KNN	65,810
SVM	59,390

From the Table 1 we can see that the best result has been yielded by naive-bayes (72%). Because It makes use of all the features contained in the data, and analyses

them individually as though they are equally important and independent of each other. Additionally, we can see that SVM has not performs well compared to the others (59%). This low performance can be attributed to the low-dimensionality of the used data [15].

After evaluating each classifier separately, we will, next, evaluate the combinations of different classifiers. The obtained results are shown in Fig. 6.
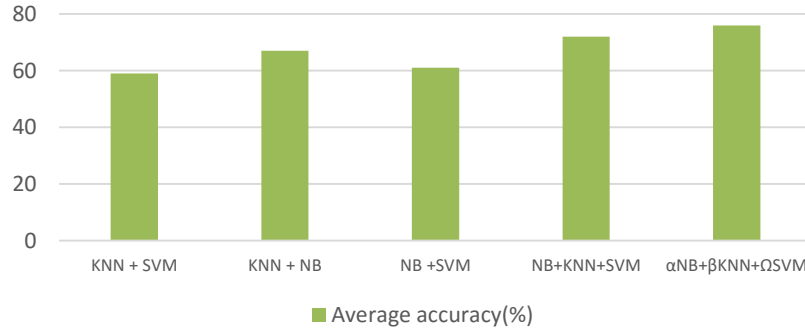

Fig. 6. Comparing the accuarcy yielded by different classifiers.

As Fig. 6 shows, combining classifiers does not improve results but rather decrease them in most cases. Because of suffering from low-dimensionality of data, SVM degrade the results of the different combinations by declaring conflicts in classification. Such issue could be resolved by assigning weights to the different classifiers. Thus and according to Table 1, the best classifiers has been associated with higher weights, combined then evaluated (i.e., $\alpha = 2, \beta = 2, \Omega = 1$). As it is illustrated in Fig.6, combining these three classifiers with their corresponding weighs has yielded better results (76%).

4   Conclusion

In this paper, we have proposed an automatic plant classification system. Our system is based on a parallel combination of three classifiers namely KNN, NB and SVM with two combining modules namely sample majority vote and weighted vote. These classifiers are firstly trained with a set of morphological features that describes the shape of the leaf. Our system has been evaluated using the well-known Flavia dataset. After evaluating each classifier separately, we have found out that NB is the best classifier among all. We, then evaluate the different possible combinations of classifiers. Results indicates that SVM negatively effects the results because of the low-dimensionality of data especially by using simple majority vote module. To solve this issue, we have opted for weighted vote with the following weights 2, 2 and 1 that corresponds respectively to NB, KNN and SVM. By associating the lowest weight to SVM, our system reduces its negative effect to the voting process. The former combination has yielded best results. Future works should consider using more powerful classifiers such as neural networks or discriminate analysis.

5   References

1. Murphy ,G. E., Romanuk ,T. N.: A meta-analysis of declines in local species richness from human disturbances. Journal of Ecology and Evolution, 4(1), 91–103 (2014).
2. Tharwat, A., Gaber, T., Hassanien, A. E.: One-dimensional vs.Two-Dimensional based Features Plant Identification Approach. Journal of Applied logic, 24, 15-31 (2017).
3. Anami, B.S., Suvarna, S.N., Govardhan, A.: A Combined Color, Texture and Edge Features Based Approach for Identification and Classification of Indian Medicinal. International Journal of Computer Applications, 6(12), 45–51 (2010).
4. Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F.: A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network .In: IEEE International Symposium on Signal Processing and Information Technology, Giza.11–16 (2007).
5. Silva, P.F., Marcal, A.R., da Silva, R.M.A.: Evaluation of Features for Leaf Discrimination. Image Analysis and Recognition, In International Conference Image Analysis and Recognition, 197–204, Springer Berlin Heidelberg (2013).
6. Rahman, A.F.R. and Fairhurst, M.C.Es-saady, Y., El Massi, I., El Yassa, M., Mammass, D., Benazoun, A.: Automatic recognition of plant leaves diseases based on serial combination of two SVM classifiers. In: International Conference on Electrical and Information Technologies (ICEIT), 1–13, Morocco (2016).
7. El Massi, I., Es-Saady, Y., El Yassa, M., Mammass, D., Benazoun, A.: Automatic recognition of the damages and symptoms on plant leaves using parallel combination of two classifiers. In: 13th International Conference Computer Graphics, Imaging and Visualization. 131-136, IEEE, Morocco (2016).
8. Asrani , K., Jain, R.: Contour based retrieval for plant species. International Journal of Image, Graphics and Signal Processing 5(9), 29–35 (2013).
9. Casanova, D., de Mesquita Sa Junior, J.J. and Bruno, O.M.: Plant leaf identification using gabor wavelets. International Journal Imaging System Technology 19(3), 236–243(2009).
10. Kadir ,A., Nugroho ,L. E., Susanto ,A., Santosa ,P. I.: Leaf classification using shape, color, and texture features. Computer Vision and Pattern Recognition. (2013).
11. Harish, B, S., Hedge, A., Venkatesh, O., Spoorthy, D. G.: Classification of plant leaves using Morphological features and Zernike moments. In: International Conference Advances in Computing, Communications and Informatics (ICACCI), 1827 – 1831, India (2013).
12. Buttrey, S. E., Karo, C.: Using k-nearest-neighbor classification in the leaves of a tree. Journal of Computational Statistics and Data Analysis, 40(1), 27–37 (2002).
13. Russell, S., Norvig, P.: Artificial Intelligence: A modern approach. Book (3d edition), Englewood Cliffs, (1995).
14. Vapnik, V.: The Nature of Statistical Learning Theory. Book 2d Edition .Springer-Verlag New York. (1995).
15. Chapelle, O., Haffner, P., Vapnik, V. N.: Support vector machines for histogram based image classification. Transactions on Neural Networks, 5(10), 1055–1064 (1999).
16. Ho, T.K., Hull, J.J., Srihari, S.N.: Decision combination in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence. 16(1), 66–75 (1994).
17. Bahler, D., Navarro, L.: Methods for combining heterogeneous sets of classifiers. In 17th Natl. Conf. on Artificial Intelligence (AAAI), Austin-USA (2000).
18. Rahman, A.F.R., Fairhurst, M. C.: Multiple classifier decision combination strategies for character recognition: A review. International Journal on Document Analysis and Recognition (IJDAR), 5, 166-194 (2003).

19. El Massi, I., Es-saady, Y., El Yassa, M., Mammass, D. and Benazoun, A.: A Hybrid Combination of Multiple SVM Classifiers for Automatic Recognition of the Damages and Symptoms on Plant Leaves. International Conference on Image and Signal Processing, 40-50. Springer, Cham (2016).