Automatic recognition of plant leaves using serial combination of classifiers

Lamis Hamrouni
Department of Computer Science and Information Technologies, University of Kasdi Merbah Ouargla
LINATI Laboratory, University Kasdi Merbah
Ouargla, Algeria
lamis0215@gmail.com

Khaldi bilel
Department of Computer Science and Information Technologies, University of Kasdi Merbah Ouargla
Ouargla, Algeria
LINATI Laboratory, University Kasdi Merbah
Ouargla, Algeria
khaldi.belal @univ-ouargla.dz

Mohammed Lamine KHERFI
1. LAMIA Laboratory
Université du Québec à Trois-Rivières, Canada and
LINATI Laboratory, University Kasdi Merbah
Ouargla, Algeria
Mohammedlamine.kherfi@uqtr.ca

***Abstract*——— plants are of great importance in human life, they are useful in many field such as industry, medicine, agriculture, etc. Plant identification is not a trivial task and presents challenges even for specialists. In this paper, we present an automatic leaf classification system based on a serial combination of two classifiers, namely: Linear discriminate analysis and Naïve Bayes. Our system is consisted of two stages, at the first stage, NB classifier attempts to determine, with a reject option, the class that a given sample is belonging to. If the confidence score yielded by NB does not exceed a certain threshold, then the sample will be passed through another classification task using LDA classifier. Our system has been evaluated using the well-known Swedish dataset. Experimental results indicated that the serial combination of the classifiers has shown better performance than those obtained using only one classifier.**

***Keywords*— plant leaves; Morphological features; serial combination; classification.**

I. Introduction

Computer vision is a rapidly growing field. Therefore, Image classification and recognition tasks gained a lot of interest due to the increasing computing capabilities of modern computers. Plant leaf classification became a hot research topic [1] in recent years. Many industry fields benefit from these researches such as medicine, economic agriculture, etc.

The classical plant classification depends mostly on manual recognition using molecular biology and cellular features of leaves. However, the huge number of species that exist in the world makes classification through experts and botanists subjective, slow, difficult, and sometimes not accurate [2]. Subsequently, an automatic plant identification system that facilitate and accelerate the process of identification is highly needed.

Plants are usually recognized 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 2D.

Because of the improvement achieved in machine learning, the automatic identification of plants has become possible. So far, researchers have done a lot of work concerning plant recognition using machine-learning approaches.

Pedro et al. [2] have presented an approach to identify medicinal plants such as herbs, shrubs and trees automatically. As a classifier, they have used Artificial Neural Network (ANN) and Support Vector Machine (SVM). Hariri et al. [3] presented a system based on Random Forest (RF) and Linear Discriminant Analysis (LDA) algorithms for identifying different types of plants using feature combination,

experimental results showed that LDA achieved a classification accuracy of 92.65 %. Wu et al. [4] introduced a system to classify 32 plant species by extracting 12 visual features from the leaf. KNN classifier has been trained with 1800 leaf images and achieved an accuracy of 93 %(1-NN), 92 %(4-NN)

In literature, most works have opted for only one classifier. Thus, some recent works have tried to improve the results by combining more than one classifier.in recent years multiple classifier system have attracted more attention for researchers and have been successfully applicate in several domain such as Arabic handwriting and face recognition and so on. In the domain of plant classification [5], authors present a system based on the serial combination of two SVMs. The 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 plant diseases. The experimental results have reported an 87.7% accuracy.

Multiple classifier system is categorized into parallel and serial approach in the first one all the classifiers are operated independently than the results are fused where in the second one the classifier are operated in the series.

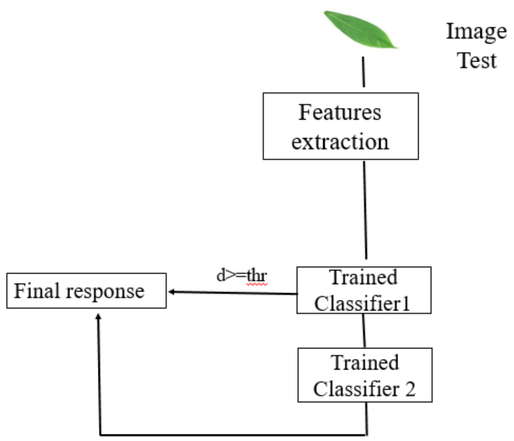
In this paper, we introduce a system that combines two classifiers sequentially, namely LDA an NB. As features, we consider extracting shape and texture features (i.e. Morphological and LBP) from the leaf images. The evaluation of our system has been undergone the well-known Swedish dataset. More details about the proposed system will be given in the next section.

The rest of the paper is organized as follow: section 2 presents the scheme of our system and discusses the used features. In Section 3, we conduct experimentations on the proposed system and report results. Finally, we give some conclusions.

II. Proposed Method

In this section, we explain our proposed method for leaf identification that consists of two main phases : feature extraction and classification. At the first phase we extract o type of features from the leaf image, the first type is texture features (LBP) and the second types shape features (Morphological). At the second phase, we train the first classifier with LBP and the second classifier with Morphological features. Finally, the tested images are passed on a serial combination of the two classifier in order to yield the global response. Fig. 1 illustrates a general scheme of our system.

Fig. 1. Scheme of the proposed system



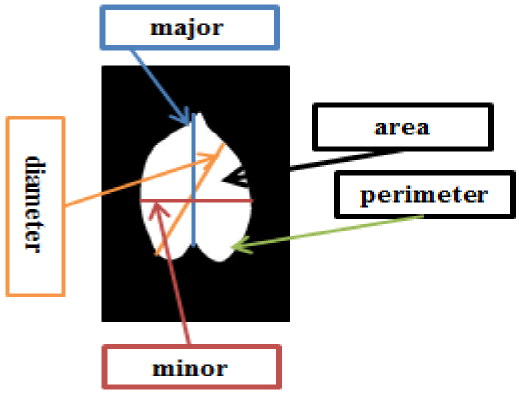
A. *Feature extraction*

The features extraction stage consists in transforming the segmented image into a vector of numeric values. The adopted features include LBP that is regarded as a powerful tool for extracting robust features from texture images, and Morphological features that describe the shape of leaf.

1. *Morphological 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 [6] of the leaf such as: diameter, area, perimeter, major and minor. Fig. 2 shows an example of some geometrical features extracted from a leaf image.

Fig. 2. Basic geometrical features



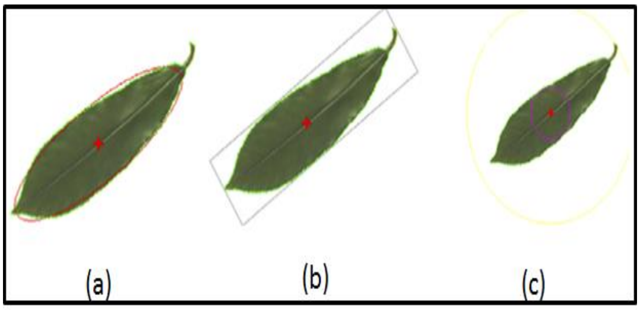
In our method, we extract the following five morphological features:

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 authors of [6] which are:

1. 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.
1. Perimeter Ratio of Physiological length & width: this feature is the ratio of perimeter leaf and the sum of major and minor axis length, given by PRPW=P/(LP+WP).
2. Perimeter Ratio of Diameter: it is the ratio of perimeter to the diameter, given by PRD=P/D.
3. Rectangularity: The similarity between the leaf and a rectangle, given by R=(Lp*Wp)/A.
4. Narrow Factor: the ratio of the diameter D and length Lp (i.e., NF = D/Lp).
5. Circularity: The ratio involving the area A of the leaf and the square of its perimeter P, given by C = 4πA/P2.
6. Solidity: The ratio between A (area) of the leaf and Ach (area of a convex hull), given by S = A/Ach.

Fig. 3 shows some of morphological features.



2. *LBP (local binary pattern):*

Local Binary pattern (LBP) was firstly introduced by Ojala [7] and Pietkain as a statistical approach. A very small local neighborhood (patch) of a pixel is used to calculate a feature vector. The LBP operator labels the pixels of an image by thresholding the local neighborhood around each pixel and considering the result as a binary number. Fig 4 illustrates an example of computing LBP in a 3×3 neighborhood by comparing the intensities of the eight neighbors around each pixel with the intensity of the center pixel.

Fig. 4 example of LBP.

5	13	20
6	10	3
16	70	55

→

0	1	1
0		0
1	1	1

1	2	4
128		8
64	32	16

→

0	2	4
0		0
64	32	16

LBP=0*1+1*2+1*4+0*8+1*16+1*32+1*64=118

B. *Classification*

Image classification is the process in which images are recognized, differentiated, and understood. The classifier, firstly, is trained with a set of labeled data (i.e., training set). Then, test samples are provided to the classifier in order to be assigned to the appropriate class.

In our work, the feature vectors, which are extracted in the previous phase, will be used to train our system. It consists of two stages, in the first stage we adopt NB [8] classifier and in the second one we use LDA [9]. Serial approach [10] is used to combine the classifiers, which consists in positioning the classifiers one after the other. The successive organization of classifiers permit to either improve the decision taken at the previous stage or leave it as it is.

In our work, the recognition process may undergo one or two stages. If the first classifier has yielded a decision that exceeds a certain threshold, then no second classification stage is needed. Otherwise, if the decision is below a certain threshold, then the sample must be fed to the second classifier in order to improve the decision.

The utilized classifiers are presented below:

1. *NB (Naïve Bayes):*

Naive Bayes classifiers [11] are family of simple probabilistic classifiers that apply Bayes' theorem with strong (naive) independence assumptions between the features.

Let x be a data sample whose class label is unknown and let c be some hypothesis, then.

$$P(c/x) = \frac{P(x/c)P(c)}{P(x)} \quad (1)$$

The naive Bayes classifier requires a small amount of training data to estimate the necessary parameters (means and variances of the variables) for classification. Because independent variables are assumed, only the variances of the variables for each class are need to be determined and not the entire covariance matrix.

2. *LDA (Linear Discriminant Analysis):*

Linear Discriminant Analysis (LDA) is a commonly used technique for data classification and dimensionality reduction. Linear Discriminant Analysis easily handles the case where the within class frequencies are unequal and their performances have been examined on randomly generated test data. Its basic idea is to find a linear transformation that best discriminate the classes, and then classification can be performed in transformed space based on some metrics such as Euclidean distance. [3]. Mathematically, LDA implementation is carried out via scatter matrix analysis. For all samples of all classes, two measures have been defined as follows:

Within-class scatter matrix is defined by the following formula :

$$S_w = \sum_{j=1}^{k} \sum_{i=1}^{N_j} (x_i^j - m_k)(x_i^j - m_k)^T \quad (2)$$

Between-class scatter matrix, which is defined by the following formula:

$$S_B = \sum_{j=1}^{k} (m_k - m)(m_k - m)^T \quad (3)$$

C. *Serial combination*

The principal of Serial combination is to put classifier one after other in multi stage way. Hence, at each stage there is only one classifier classifying the patterns. There are two basic categories of serial combination: class set reduction and reevaluation approach.in the first categories, the number of possible classes is reduced continuously in each stage, while the second categories require reevaluation of the patterns, which are rejected at the preceding stage. The decision either accepts or reject it is based on below a pre-defined threshold

III. EXPERIMENT

In order to evaluate our system, we have used the well-known Swedish dataset [12] that contains 15 classes with a total of 1125 images. Figure. 4 shows a representative sample from this dataset.

Fig. 4. Samples of Swedish Dataset

In this experiment, the first 25 images from each class are devoted for the training and the rest are for the test.

In this experiment, the first 25 images from each class are devoted for the training and the rest are for the test.

We opted for Accuracy to be used as an evaluation metric of our system. It is given by the following formula.

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

To prove the performance of our method, we firstly classify images using each classifier independently naïve bayes with normal distribution and linear discriminant analysis (linear). Table 1 shows the obtained results.

Table 1. Accuracy results using each classifier separately.

Classifier	Features	Average accuracy%
NB	LBP	67.73
LDA	Morphological	65.38%

From Table 1 we can see that the results that have been yielded by NB and LDA are approximately the same.

After evaluating each classifier separately, we evaluate the combination of the two classifiers using a threshold of 80%.the values of threshold according to our experiment is fxed during the training phase. The obtained results are shown in Table. 2.

Table 2. Accuracy results using classifier combination

Classifiers	Average accuracy%
LDA+NB	71.33

As Table.2 shows, combining classifiers sequentially has improved the results. This means that errors made by the first classifier have been corrected by the second one.

IV. CONCLUSION

In this paper, we have proposed an automatic plant classification system.in order to improve accuracy Our system is based on a sequential combination of two classifiers namely LDA and NB.

In order to create discrimination of the response between classifiers, the first classifier has been trained with a local binary pattern features and the second with Morphological features that describe the shape of the leaf. Our system has been evaluated using the well-known Swedish dataset.

evaluating each classifier separately, we have found out that LDA and NB has yielded approximately the same results. We have then evaluated a serial combination of classifiers where the result indicated better performance. Other types of classifiers should be sequentially combined and examined for leaf classification.

REFERENCES

.

[1] A. Tharwat,T. Gaber,A. E. Hassanien,:"One-dimensional vs.Two-Dimensional based Features Plant Identification Approach. Journal of Applied logic", vol 24, 2017,pp.15-31.

[2] P.F. Silva,A.R. Marcal,da Silva, R.M.A.: "Evaluation of Features for Leaf Discrimination". Image Analysis and Recognition, In International Conference Image Analysis and Recognition, 2013,pp.197–204.

[3] E. Elhariri, N. El-Bendary, AE. Hassanien.: "Plant classification system based on leaf features". In: 2014 9th international conference on computer engineering and systems (ICCES),(2014),pp 271–276.

[4] SG. Wu, FS. Bao, EY. Xu, Y-X. Wang, Y-F. Chang, Q-L. Xiang. : "A leaf recognition algorithm for plant classification using probabilistic neural network". In 2007 IEEE international symposium on signal processing and information technology, 2007,pp 11–16

[5] A.F.R. Rahman, Fairhurst, M.C. Es-saady,Y. El Massi,I. El Yassa,M. Mammass,D. Benazoun.: Automatic recognition of plant leaves diseases based on serial combination of two SVM classifiers. In: International Conference on Electrical and Information Technologies (ICEIT), 2016, pp1–13, Morocco.

[6] B. S. Harish,A. Hedge,O. Venkatesh,D. G. Spoorthy.: "Classification of plant leaves using Morphological features and Zernike moments". In: International Conference Advances in Computing, Communications and Informatics (ICACCI),2013, pp 1827 – 1831, India .

[7] T. Ojala, M. Pietikäinen, D. Harwood.:"A comparative study of texture measures with classification based on featured distributions",1996,pp 51-59.

[8] A. Muthevi,R.B. Uppu.:" Leaf classification using completed local binary pattern of textures". In: 2017 IEEE 7th International Advance Computing Conference (IACC) ,2017.

[9] S. Russell, P. Norvig.: A modern approach. Book (3d edition), Englewood Cliffs, 1995.

[10] S. Zhang, Y-K . Lei.: "Modified locally linear discriminant embedding for plant leaf recognition",2011,pp 2284–2290.

[11] D. Lin, L. Sun, K-A.Toh.:"Biomedical image classification based on a cascade of an SVM with a reject option and subspace analysis",2018,pp 128-140.

[12] O. Söderkvist, Computer Vision Classification of Leaves from Swedish Trees, Master's Thesis, Linköping University, 2001.