ISPEC Tarım Bilimleri Dergisi
5(2): 393-400, 2021
© Telif hakkı ISPEC'e aittir
Araştırma Makalesi

E-ISSN:2717-7238
www.ispecjournal.com

ISPEC Journal of Agr. Sciences
5(2): 393-400, 2021
Copyright © ISPEC
Research Article

Plant Identification Via Leaf Classification Using Color and Biometric Features

Ümit Çiğdem TURHAL[1*]

[1]Bilecik Şeyh Edebali University, Faculty of Engineering Department of Electrical and Electronics Engineering

[1]**ORCID**: 0000-0003-2387-1637

[*]Corresponding author:

ucigdem.turhal@bilecik.edu.tr

DOI

https://doi.org/10.46291/ISPECJASvol5iss2pp393-400

Alınış (Received): 20/02/2021

Kabul Tarihi (Accepted): 26/03/2021

Abstract

Plants that are of great importance for humans and other living things are an integral part of our ecosystem. In today's world, where many plant species are at risk of disappearance, the identification of plants helps to protect and survive all natural life. There are many studies presented in the literature for plant identification. The most popular of these identification methods is leaf based classification. The reason for choosing leaves in this classification is that they are easier to obtain than other biometric components such as flowers available for a short period of time. Various biometric properties of the leaf must be determined for leaf classifications. In traditionally it is time consuming and expensive to perform this process visually by experts. In this article, various leaf biometric features obtained by digital image processing methods are used as the feature extraction step for automatic leaf classification. As the classification algorithms, Naive Bayes, Linear Regression, Multilayer Perceptron, Decision Tree and Random Forest are used. According to the experimental results using the training set as the test set, 100% recognition rate is obtained for Random Forest classification algorithm and 96% recognition rate is obtained in 30-fold cross validation for Linear Regression classification algorithm.

Keywords
Plant identification, plant venation, image processing, leaf classification, machine learning

INTRODUCTION

Plants convert solar energy into nourishment thus they form the basis to other life forms. Therefore, they are of great importance for both the environment and life. Plants are used as food, as medicine and in many industries. Therefore, while conservation and survival of plant species is very important for all life forms their yield is also has to be increased. For the yield increase it is very important to analyse the plant's physical and enviromental requirements. There can be found many of studies such that in the literature (Coşkun, and Bengisu, 2021; Keten and Tanrıverdi, 2020). These cases have attracted researchers to struggle to classify plants, since ancient times. Classification of plants, can be possible by identifying them. The first work in this field is based on the works of the Swedish botanist Carolus Linnaeus in the eighteenth century. L. R. Hicher was the first scientist to study leaf properties for plant classification in 1973. There have been many developments in this field since then (Beghin et al., 2010). Among the plant species, there are number of plant species whose appearance is similar. This makes it very difficult to classify plants. Plants can be classified using leaves and flowers. However, because of the short existance time of the flowers, plant classification is made using leaves in many studies in the literature. The leaves can be classified in terms of their features such as pattern, shape and arrangement. It is time consuming to make this classification manually. Acquisition of leaf properties by using digital image processing techniques instead of manual eliminates this disadvantage and enables classification with high accuracy in a short time. When studies on plant classification in the literature are examined, it is seen that these studies consist of two main stages. The first of these stages is the extraction of plant leaf features with image processing and the second is the use of these extracted features to create a classification model (Petr and Suk, 2013). Feature extraction phase consists of obtaining visual physical features such as color information and shape information. The color information of the leaf may vary depending on environmental conditions, for example it is very sensitive to the lighting conditions. In contrast, leaf color information was used in the literature as a distinguishing feature for plant species under suitable conditions (Kumar-Saroj et al., 2019). As the shape information, the derivate of the leaf image or edge information of the leaf image (Beghin et al., 2010), texture and shape properties (Kadir et al., 2013), and image segments and image histograms (Chang and Kuo, 1993; Bashish et al., 2010) are used. In the second phase of the plant classification process, as the classification algorithms, it can be seen in the literature that machine learning algorithms such as Artificial Neural Network (ANN) (Chang and Kuo, 1993; Bashish et al., 2010), Wavelet Transform (WT) (Chang and Kuo, 1993), Linear Discriminant Analysis (LDA) (Kadir et al., 2013) and Support Vector Machines (SVM) (Pedro et al., 2013) are used oftenly.

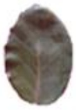
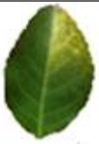
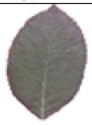
In this study, feature extraction is made by using the leaf color information obtained from digital image processing techniques and some leaf shape information together. Several different classification models have been used to examine the effect of the obtained properties on the plant classification problem. Classification performances of models are given comparatively. Database used in this study is the database obtained in (Weka Hall et al., 2009). While image processing steps are carried out on MATLAB software platform, classification models are created on Weka software. In the rest of the paper while the material and method are given in the second part, Material and Method is explanied, the results of the experimental study are given in the third part. In the fourth part, conclusions are given.

MATERIAL and METHODS

Material

The database used has 40 different plant species. It consists of simple leaves and complex leaves according to their shape. Each leaf specimen's image was accuired over a coloured background using an Apple IPAD 2 device and has a pixel resolution is 720x920 pixels (Pedro et al., 2013). In this study, it is selected 10 plant species each has simple leaves and has 10 different images are selected and classified. Table 1 provides the names and the general aspect of the typical leaves of each plant while Figure 1 provides the different images belong to one specie.

Table1. Different plant species selected from the database

No	Plant species	General view	No	Plant species	General view
1	*Quercus suber*		6	*Ilex perado* ssp. *azorica*	
2	*Magnolia grandiosa*		7	*Buxus sempervirens*	
3	*Corylus avellana*		8	*Urtica dioica*	
4	*Bougainvillea* sp.		9	*Acca sellowiana*	
5	*Euonymus japonicus*		10	*Hydrangea* sp.	

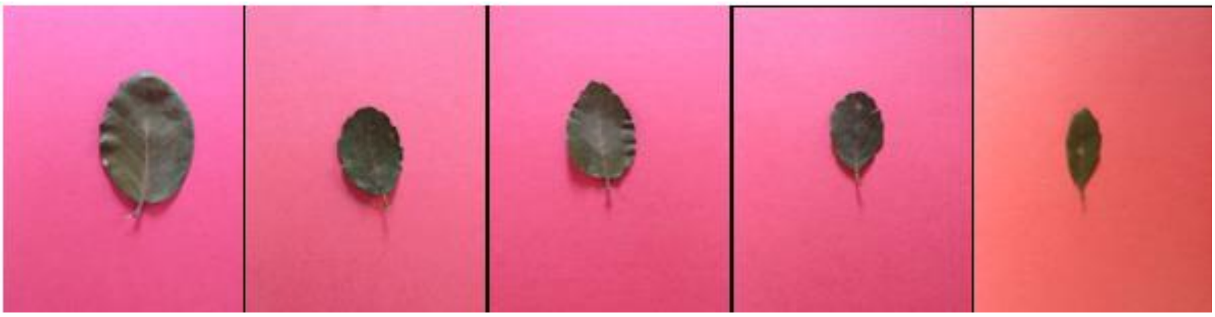


Figure 1. Five image samples belong to one specie in the database

Methods

In this study first of all leaf features such as color information and shape information are obtained using digital image processing techniques in MATLAB platform. Then using the obtained features classification models are constructed for several classification algorithms in Weka software and their performances are evaluated comparatively.

Feature extraction step

In this study, as the image features seven attributes are used such as gray scale image pixel mean and its standard deviation, Hue component pixel mean and its standard deviation, image area, bounding box width and heigt values. The first four attributes of the feature vector for each image contains the color information. The color information is obtained from the gray scale image and from the hue component (H) of HSV color space. The rest of the attributes contain the shape information. In the feature space an image is represented as given in Eq. 1.

$$X = \left\{ \begin{array}{c} \mathit{Hue\ component\ pixel\ mean}, (x_1) \\ \mathit{Hue\ component\ standard\ deviation}, (x_2) \\ \mathit{Gray\ scale\ pixel\ mean}, (x_3) \\ \mathit{Gray\ scale\ image\ standard\ deviation}, (x_4) \\ \mathit{Binary\ image\ area}, (x_5) \\ \mathit{Binary\ image\ bounding\ box\ width}, (x_6) \\ \mathit{Binary\ image\ bounding\ box\ height}, (x_7) \end{array} \right\}^T = \{x_1 \quad . \quad . \quad x_7\}$$

Eq. 1

The process of obtaining feature vector X for an image is explained in the flow chart given in Figure 2. Some image samples belong to the feature extraction step are given in Fiqure 3.

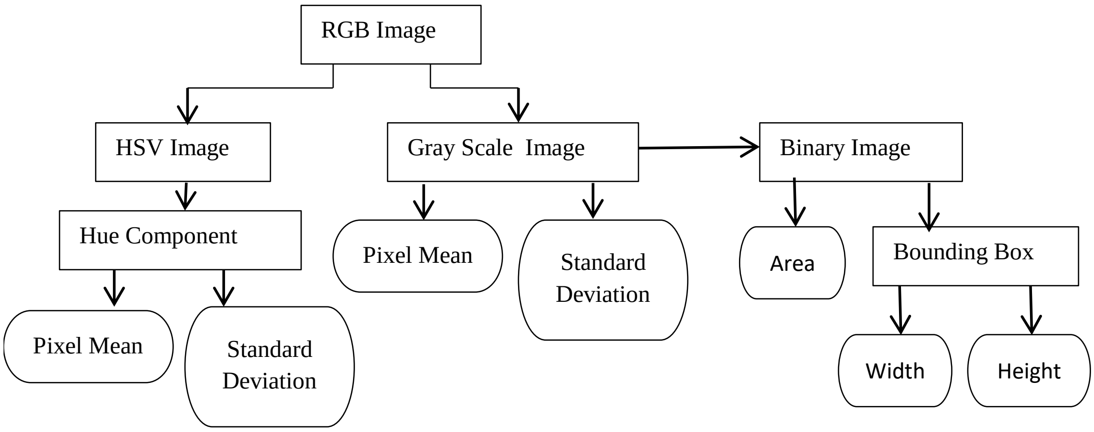


Figure 2. Flow chart of the feature exteraction step

In the feature extraction step, while the terms used as image processing techniques are consisted of RGB image, gray scale image, HSV color space and binary image, the terms used in analysis of the images are consist of pixel mean, image standard deviation, region area, bounding box height and width. The explanations of all these terms are as given in Table 2 and the images are given in Figure 4.

Table 2. Image processing techniques and analysis, that are applied in feature extraction process

Rgb Image	Image with real color pixel values. Each pixel has three components such as red, green and blue. Each component has 256 levels.
Gray Scale Image	Obtained from RGB image using image conversion techniques. Pixel values consist of only one component. Pixel values change between 0-255. 0 corresponds to black while 255 corresponds to white. The values between 0-255 corresponds to the gray levels.
HSV Color Space	Obtained from RGB image using image conversion techniques. Each pixel has three components such as hue, value and saturation. Hue component includes only the color image. Value and saturation components are related to the light components.
Binary Image	Obtained from gray scale image using image conversion techniques. Its pixels have only two value such as 0,1. 0, corresponds to black and 1, corresponds to white.
Image Pixel Mean	Mean of the image pixel values.
Image Standard Deviation	It measures how far a set of random numbers are spread out from their mean value.
Region Area	Pixel number of a region, specified in an image.
Bounding Box	The smallest box that surrounding an image segment.

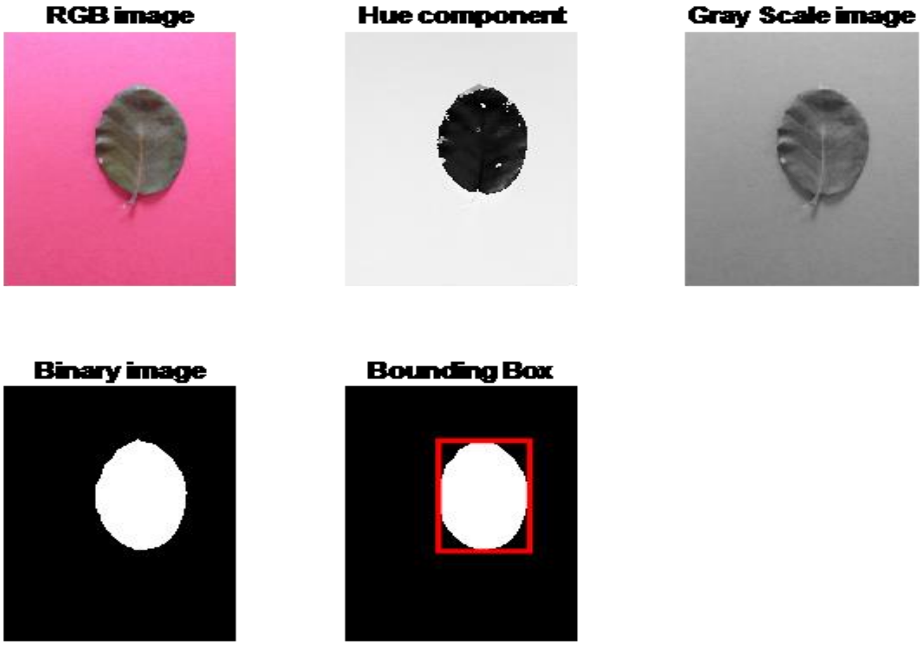


Figure 3. Image output samples of some of the image processin steps

Classification model

In this study, plant identification is performed via leaf claasification. Classification is a supervised learning process. A model is created using the database that its classes are known. Then model's performance is evaluated with the data that are not used in model construction process. The database has a form that is given in Figure 4 and the data has a form given in Eq.1.

Attributes

Objects

Object No	A_1	.	.	A_n
1				
2				
.				
.				
m				

Figure 4. Image output samples of some of the image processin steps

In a classification task one of the attributes in Figure 4 is the class information. The database is first split into two parts, in such a way that one part is training set and the other part is test set. The Training set is used in model construction step and the test set is used in model performance evaluation step. After a model is constructed, which its classification performance is known, a new data that its class is unknown is classified using this model.

In this study five different classification algorithms such as naive bayes, linear regression, multi layer perceptron, desicion trees and random forest are used and their performance is given comparatively. Classification models are constructed using Weka software.

Naive bayes classifier: Naive bayes classifier is a statistical pattern classification algorithm that works on probabilities. Basically works on the principle of bayes theorem which defines the relationship between the conditional probabilities and the prior probabilities.

Linear regresision: Linear regression analysis is a linear approach used to model the relationship between the dependent variable and one or more independent variables. The equation describing this relationship is called the simple/multiple linear regression equation.

Multi layer perceptron: Perceptron Model is a controlled learning algorithm that forms an important basis for today's neural networks. In other words, learning is expected by giving the network both input and output sets. The multiple layer perceptron model is a model that contains multiple neurons and layers that operate in parallel. The information in this model is transferred to the next layer.

Decision tree: It is one of the tree-based learning algorithms and is one of the most used classification algorithms. A decision tree relies on the process of dividing a dataset containing a large number of records into smaller sets by following simple decision-making steps.

Random forest: It is also one of the tree-based learning algorithms such as decision tree. The random forest algorithm works as a community of many individual decision trees. Each tree in the random forest gives a class estimate and the top-rated class is an estimate of the model.

Experimental studies

In the experimental studies five different classification algorithms are used. The objects of the database is constructed as in the form as given in Equation 1. Experimental studies are performed using Weka software. Weka is basically a data mining program developed by Java and open source distributed by Waikato

University. In Weka machine learning algorithms and requirements such as data pre-processing are presented together. The classification process steps for each algorithm are as given in the puseudocode in Figure 5.

```
Classification process using, naive bayes, linear regression, multi layer perceptron, decision tree and random forest.
Step1: Construction of training set and the test set according to cross-validation.
Step2: Model construction as the training process.
Step3: Classification performance evaluation as the testing process via confusion matrix.
```

Figure 5. Puseudocode for the classification process

In the experimental studies four different classifications are performed for each of the method. In the first one, classification performance of each method is evaluated using the training set as the test set. In the second, third and the fourth ones the recognition performances are evaluated by 30-fold cross validation method. 110 data points are randomly divided into 30 slices. Prediction accuracy scores are calculated by using each slice one time as test set in turn and the rest as training set. While in the first one and the second one it is not applied any preprocessing to the database, in the third one and the fourth one a filter is applied to the data as preprocessing. In the third step a filter that select the attributes that are more important is applied. According to this filtering three of the attributes are neglected. And in the fourth one an additional filtering to the filter applied in the third one is applied that is discritization. The recognition results are given in Table 3.

Table 3. Recognition rates of classification methods

Classification Method	Recognition rates (%)			
	Without Preprocessing		With Preprocessing	
			Filter 1	Filter 2
	Training Set	Test Set	Test Set	Test Set
Naive Bayes	92.7273	84.5455	88.1818	91.8182
Linear Regression	96.3636	80.9091	82.72.73	96.3636
Multilayer Perceptron	96.3636	78.1818	84.5455	91,8182
Decision Tree	95.4545	86.3636	86.3636	83.6364
Random Forest	100	88.1818	92.7273	89.0909

RESULTS and DISCUSSION

In this study, it is performed leaf classification automatically for plant identification using naive bayes, linear regression, multilayer perceptron, decision tree and random forest. As can be seen from the applications, 100% accuracy is achieved only in the Random Forest method when using the training set as the test set. The recognition accuracies obtained as a result of 30-fold cross verification are higher for all methods in cases where filter application is applied. Here, the highest accuracies in the filter-free application and the first type filtering application are achieved in the random forest method, while the highest success among all applications was obtained for the linear regression method. Leaf samples of plant varieties used in the database are not very different samples visually. So the similarity between the classes is high. This causes accuracy rates to remain around 90% on average if two filters are applied. However, a high success rate of 96% is obtained with the linear regression method. As can be seen from this study, the identification of plant varieties can be done automatically with high accuracy with the classification of leaves according to its characteristics.

REFERENCES

Bashish, D.A., Braik, M., Bani-Ahmad, S. 2010. A framework for detection and classification of plant leaf and stem diseases. In Signal and Image Processing (ICSIP), International Conference IEEE, 113–118.

Beghin, T., Cope, J.S., Remagnino, P., Barman, S. 2010. Shape and texture based plant leaf classification. In International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, 345–353.

Chang, T., Kuo, C.C. 1993. Texture analysis and classification with tree-structured wavelet transform. IEEE Transactions on Image Processing, 2(4): 429–441.

Coşkun, M., Bengisu, G. 2021. Determine the effects of bacteria ınoculation on yield and yield components of some legume green fertilization crops under organic farming conditions. ISPEC Journal of Agricultural Sciences, 5(1): 10-20.

Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I. 2013. Leaf classification using shape, color, and texture features. arXiv preprint arXiv:1401-4447.

Keten, M., Tanrıverdi, Ç. 2020. The effect of the leonardite dose applied at different rates on the water-yield relationship of amaranth (*Amaranthus cruentus* L.) plants. ISPEC Journal of Agricultural Sciences, 4(4): 823-833.

Kumar-Saroj, S., Oshiro, S., Yadav, P., Pratap-Singh, N. 2019. An efficient approach for plant leaves identification based on texture features. International Journal of Computational Intelligence & IoT 2(3).

Petr, T., Suk, T. 2013. Leaf recognition of woody species in Central Europe. Biosystems Engineering, 115(4): 444-452.

Pedro, F.B., Silva-Andre, R.S., Marcal-Rubim, M., Almeida da Silva. 2013. Evaluation of features for leaf discrimination. Springer Lecture Notes in Computer Science, 79(50): 197-204.

Weka-Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H. 2009. The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11(1).