

# Medicinal and Aromatic Plants Identification Using Machine Learning Methods

G. KAYHAN, E. ERGÜN


**Abstract—** In this study, different machine learning (ML) methods were used to classify medicinal and aromatic plants (MAP) namely St. John's wort (*Hypericum perforatum* L.), Melissa (*Melissa officinalis* L.), Echinacea (*Echinacea purpurea* L.), Thyme (*Thymus* sp.) and Mint (*Mentha angustifolia* L.) based on leaf shape, gray and fractal features. Naive Bayes Classifier (NBC), Classification and Regression Tree (CART), K-Nearest Neighbor (KNN), and Probabilistic Neural Network (PNN) classification were used as methods. The results indicated that plant species were successfully recognized the average of correct classification rate. The best classification rate on the NBC was taken: training data for classification rate 98.39% and test data classification rate for 98.00% are obtained. ML could be accurate tools for MAP classification tasks.

**Index Terms—** classification, feature extraction, image processing, machine learning.


## I. INTRODUCTION

THE AUTOMATIC recognition of plants using computer vision has been commonly utilized in the literature. For example, Kalyoncu et al. used geometric features [1]. Shao et al. have classified the image sets containing the images of the leaves of the same class in each set. [2]. Ruberto et al.'s leaf classification procedure utilizes a high dimensional new features set that consist of the shape, color and texture features [3]. Du et al. [4] used fractal dimension features to recognize of plant leaf image. Xu et al. proposed a Probabilistic Neural Network (PNN) based plant classification [5]. Miao et al. suggested an evidence-theory-based rose classification method, which used many features of roses [6]. Gu et al. recommended the leaf classification method which consists of skeleton segmentation and wavelet transform and Gaussian interpolation [7]. Wang et al.'s leaf recognition methods combined shape features and a hypersphere classifier [8].

GÖKHAN KAYHAN, is with Department of Computer Engineering University of Ondokuz Mayıs University, Samsun, Turkey, (e-mail: [gakayhan@omu.edu.tr](mailto:gakayhan@omu.edu.tr)).

 <https://orcid.org/0000-0003-3391-0097>

ERHAN ERGÜN, is with Department of Computer Engineering University of Ondokuz Mayıs University, Samsun, Turkey, (e-mail: [erergun@omu.edu.tr](mailto:erergun@omu.edu.tr)).

 <https://orcid.org/0000-0003-1446-2428>

Manuscript received November 26, 2019; accepted January 08, 2020.  
DOI: [10.17694/bajece.651286](https://doi.org/10.17694/bajece.651286)

Medicinal and aromatic plants (MAP) harvesting is as old as human history [9]. It is reported that at present, over 60% of the world's population, 80% in developing countries, take advantage of using plants for their medical purposes. In some Asian and African countries, alternative medicine is preferred instead of traditional medicine for healthcare needs. This is due to modern medical inadequacies and cultural preferences. [10]. The active components of the plants (*Melissa officinalis* L. [11], *Mentha piperita* L. [12] and thyme [13]) used for this purpose are concentrated in their leaves. These components are essential oils. Therefore leaves are the most important part of these medicinal and aromatic plants [14]. Studies on plant identification in the literature are widely used databases: the ImageCLEF dataset in 2011 [15] and in 2012 [16, 17], Flavia dataset [18], Leafsnap [19], Swedish leaf dataset [20], Foliage dataset [21], LifeCLEF 2015 [22]. Foliage dataset, was prepared to test leaves with various colors and patterns, contains 60 kinds of leaves [21]. The Flavia dataset contains 1907 scanned images of 32 species [18]. The Swedish leaf dataset comprises 1125 images of 15 kinds of leaves [20]. ImageCLEF dataset in 2011's [15] scanned category includes totally 3070 images and in 2012's [16] scan dataset comprises totally 6630 images. Leafsnap currently covers all 185 tree species from the Northeastern United States [19]. Odabas et al. [23] set up a MAP database, contains 752 scanned images belonging to 5 kinds of medical and aromatic plants. Main objective of this work was to classification of MAP species using computer vision for precision agriculture. MAP leaf classification is original to this work. Firstly, features extractions based digital image processing is taken and secondly ML-based classification was performed. Generally, digital morphological features [1, 18, 6, 7, 8], fractal features [4], and gray/color properties [6] commonly used. So, these powerful feature extraction methods are combined in this paper. Ruberto et al. were used 138 features in the study [3], but we use only 10 features in this study. As plant classifier methods, different ML methods (NBC, CART, PNN, and kNN classification) are used.

## II. MATERIALS AND METHODS

### A. Image acquisition and pre-processing

It should create the database that is needed to recognize the plant species. Obtaining appropriate images of the selected plants to create this database and extracting the features necessary for learning the machine from these images is one of the important steps.

### 1) Acquisition of plant leaf

Color digital images (600 dpi) were taken using a scanner (HP ScanJet G4010) of individual plants of St. John's wort (*Hypericum perforatum* L.), Melissa (*Melissa officinalis* L.), Echinacea (*Echinacea purpurea* L.), Thyme (*Thymus* sp.) and Mint (*Mentha angustifolia* L.). These images are taken from MAP database of 752 images [23]. Leaf samples were randomly taken from in experiment field populations of the MAP on grown season in 2013-2014. In this period, for catching the different phases of leaf morphogenesis, 200 St. John's Wort leaves, 200 Melissa leaves, 88 Echinacea leaves, 112 Mint leaves, and 152 Thyme leaves were collected.

### 2) Image pre-processing

Firstly, leaf images were converted binary images. Fig. 1 shows an example image (a), a normalized image (b), threshold binary image (c) and masked-cropped leaf image (d).

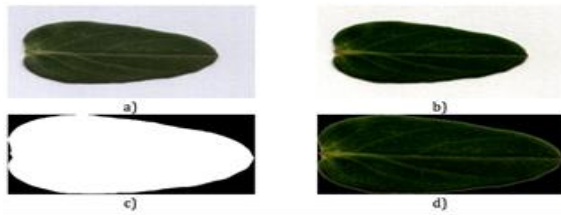


Fig.1. An example image a) raw image, b) normalized image, c) threshold binary image, and d) masked-cropped image

## B. Feature Extraction

In this study, the features are consisted of the shape (3 features), gray (5 features) and fractal (2 features) properties of leaf. The total number of features is 10.

### 1) Shape Features (SF)

The major/minor axis length feature is a scalar that indicates the length of the major/minor axis of the ellipse (ALMaj and ALMin). Fig. 2 shows that major axis length and minor axis descriptors for binary leaf image.

Firstly, the aspect ratio representing the ratio for the ellipse with minor and major axis length, is calculated by

$$SF_{Ar} = \frac{AL_{Min}}{AL_{Maj}} \quad (1)$$

$$SF_{Min} = AR_{Min} \quad (2)$$

Eccentricity is the measure of how much any conic section deviates from being circular.

$$SF_{Sol} = \frac{A_{ROI}}{A_C} \quad (3)$$

Solidity is the measurement of the overall concavity of a particle.

$$SF_{Ecc} = \frac{E_A}{AL_{Maj}} \quad (4)$$

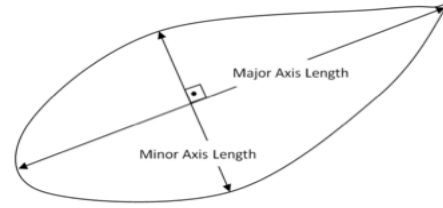


Fig.2. Ellipse's major axis length and minor axis length features for binary leaf image

### 2) Gray Features (GF)

The gray-level co-occurrence matrix (GLCM) is a descriptor that is defined over an image to be the distribution of co-occurring values. Contrast, correlation, energy, and homogeneity descriptors are classically calculated from GLCM. Contrast descriptor extracted the local variations from the GLCM.

$$GF_{Con} = \sum_{i,j} |i - j|^2 p(i,j) \quad (5)$$

Correlation statistic describes the joint probability occurrence of the pixel pairs.

$$GF_{Cor} = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{\sigma_i \sigma_j} \quad (6)$$

Energy feature measures the sum of squared elements in the GLCM.

$$GF_{Ene} = \sum_{i,j} p(i,j)^2 \quad (7)$$

Homogeneity descriptor reveals the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$GF_{Hom} = \sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (8)$$

And another important property is Entropy, which characterizes uncertainty of the gray leaf image.

$$GF_{Ent} = - \sum_{i,j} p(i,j) \log_2 p(i,j) \quad (9)$$

### 3) Fractal Features (FF)

A fractal feature of an image is an important descriptor to measure self-similar pattern that exhibits a repeating pattern that displays at every scale. The box-counting method (BCM), also known as Minkowski dimension, measures the fractal dimension of an image S in a Euclidean space  $R_n$ . In order to calculate the dimension, one should imagine a fractal S lying on an evenly-spaced grid and count the number of boxes that are necessary to cover the set. We make the grid finer by applying a box-counting algorithm to see how this number changes and the box-counting dimension is calculated using

this number. So, the local dimension distribution is calculated as,

$$df(m) = \frac{\lim_{h \rightarrow 0} \frac{N(x+h) - N(x)}{h}}{\lim_{h \rightarrow 0} \frac{r(x+h) - r(x)}{h}} \quad (10)$$

where  $N(c)$  is the number of boxes of side length  $c$  required to cover the image and  $r$  is box size. In this study, we select the expected value ( $FF_{Mea}$ ) of and a measure of dispersion ( $FF_{Std}$ ) of  $df$

$$FF_{Mea} = \frac{1}{M} + \sum_{m=1}^M (df(m) - FF_{Mea})^2 \quad (11)$$

$$FF_{Std} = \sqrt{\frac{1}{M} \sum_{m=1}^M (df(m) - FF_{Mea})^2} \quad (12)$$

### C. ML Methods

The aim of the machine learning is that of letting huge amounts of data dictate algorithms and solutions for complex problems. At the same time, machine learning is to be contrasted with the traditional approach of depend on domain experts for the development of solutions. In this research, NBC, CART, KNN, and PNN machine learning methods were used.

#### 1) NBC

Naive Bayes Classifiers are simple probabilistic classifiers for pattern recognition problems. This classifiers use Bayes theorem. The features which are used Bayes theorem are naive independence assumptions. Naive Bayes classifiers designed with independent predictors for each class are used, when the size of the data is high. Not much training data is needed to figure out the parameters. [24].

Bayes rule :

$p(x|C_j)$  : the probability that  $x$  is a sample in class  $j$

$P(C_j)$  : first probability of class  $j$

$p(x)$  : probability of being  $x$  for any sample.

$P(C_j|x)$  : The probability that sample  $x$  is from class  $j$

$$P(C_j|x) = \frac{P(x|C_j)P(C_j)}{P(x)} \quad (13)$$

Since  $p(x)$  in Equation (13) is the same for all classes, probability can be calculated as given by Equation (14). The maximum value is determined in these possibilities.

$$P(C_j|x) = P(x|C_j)P(C_j) \quad (14)$$

If all properties in Equation (14) are independent, Equation (14) is defined as Equation (15). Thus, the complexity of calculation becomes greatly reduced.

$$P(C_j|x) = \prod_{k=1}^n P(x_k|C_j)P(C_j) \quad (15)$$

$$P(C_j|x) = P(x_1|C_j)P(C_j) \times \dots \times P(x_n|C_j)P(C_j)$$

#### 2) CART

Classification and Regression trees (CART) is a non-parametric rule-based machine learning method [25]. The main objective of CART is to separate the main data matrix (independent variables matrix) into homogeneous subgroups according to the dependent variable. The subgroups are in the form of branches of a tree and they are in a hierarchical order. In the classification tree method, Gini impurity is used to decide the purity of the binary dependent variable. In Gini algorithm, attribute values are divided into two parts.

For each partition,  $G_{left}$  (Equation 16) and  $G_{right}$  (Equation 17) are calculated.

$$G_{left} = 1 - \sum_{i=1}^k \left( \frac{|T_{class_i}|}{|B_{left}|} \right)^2 \quad (16)$$

$$G_{right} = 1 - \sum_{j=1}^k \left( \frac{|T_{class_j}|}{|B_{right}|} \right)^2 \quad (17)$$

$T_{class_i}$ : each class value in the left part

$T_{class_j}$ : each class value in the right part

$B_{left}$ : all number of value in the left part

$B_{right}$ : all number of value in the right part

Gini values ( $G_m$ ) are calculated for each features by using Equation (18). The classification tree is created by selecting the feature with the smallest value from the values found.

$$G_m = \frac{1}{n} (|B_{left}| \cdot G_{left} + |B_{right}| \cdot G_{right}) \quad (18)$$

$n$ : the total class value that belongs to a feature.

$m$ : the number of feature

There are no classes in the regression trees. Therefore, the classification separation rules in the regression tree technique cannot be applied using the Gini index. Differences in the regression tree are performed according to the residual sum of squares. The residual sum of squares is similar to the Gini separation rules [25]. Regression trees are effective in cases of complex and nonlinear interaction with many variables [26].

#### 3) KNN

One of the non-parametric methods used for classification and regression is the K nearest neighbors' algorithm (KNN). The algorithm determines the K nearest neighbors to the training samples of the inputs. The data is considered a member of that class if the class is dominant. K is normally a positive and small integer [27]. In Fig. 3, when the KNN algorithm is

applied (K = 3), the new data belongs to class 2 which is dominant.

KNN Algorithm:

- i. Number of neighborhoods (K) is determined.
- ii. The Euclidean distances are calculated between the input data and all data.
- iii. According to calculated distances, the data is sorted from small to large and first K data is determined.
- iv. The class information of K nearest neighbors is determined.
- v. The class of the estimated data is defined as the most repetitive class.

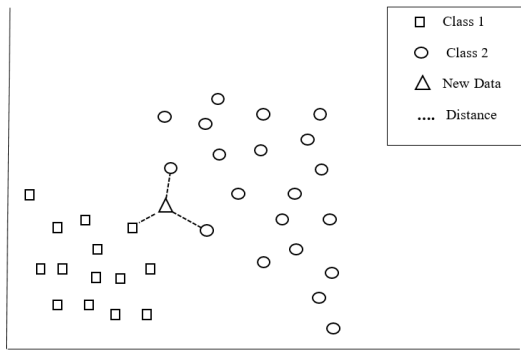


Fig. 3. Class selection for new data in the KNN algorithm

#### 4) PNN

A Probabilistic Neural Network (PNN) is a feed forward neural network. This network was derived from Kernel Fisher discriminant analysis and the Bayesian network. A PNN network consists of four layers. this layers named the Input layer, the Pattern layer, the Summation layer, and the Output layer (Fig. 4) [28].

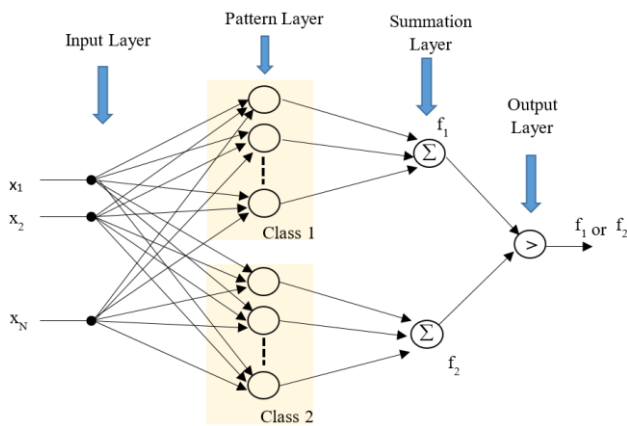


Fig. 4. A Probabilistic Neural Network (PNN)

$$g_1(x) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} e^{-\frac{\|x-\alpha^{(p)}\|^2}{2\sigma^2}} \quad (19)$$

$$g_2(x) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} e^{-\frac{\|x-b^{(q)}\|^2}{2\sigma^2}} \quad (20)$$

N: the dimension of the vectors

p: the number of vector

$\alpha^{(p)}$  : p x N feature vector in class 1.

$b^{(q)}$  : p x N feature vector in class 2.

x: input feature vector

$\sigma_1$  and  $\sigma_2$  :spread parameters

N-dimensional data is passed through the gauss functions that belongs to class 1 and class 2 in the pattern layer (Equation 19 and 20).

The sum of gauss outputs of each class are f1 and f2 respectively.

If  $f_1 > f_2$  then x belongs to class 1.

#### D. Evaluation Strategy

##### 1) k-fold Cross Validation:

The k-fold cross-validation method is widely used to evaluate objectively the classifier performance. This method, by random selection, allows all instances to enter the test set. For the M data sets, k-fold cross validation algorithm comprises the following steps,

Step1. The data set is divided into k-random fold. Thus, each fold M/k will have data.

Step 2. Each i. belonging to the fold (i = 1..k) data consists of test dataset, others consists of the train dataset. ML network was trained with each fold's train set, and then network was evaluated with the test set.

##### 2) Accuracy:

A matrix-shaped special table is used to visualize the performance of a consulting learning algorithm in machine learning applications [29]. Each column and row of the matrix represents the instances in a predicted class and the instances in an actual class respectively. To ISO 5725-1, it is known that the general term accuracy (Equation 21) is used to describe the closeness of measurement to the true value [30]. That is, the accuracy is the proportion of both true positive and true negative results among the total number of cases tested [31].

$$Accuracy(\%) = \frac{\text{numbers of true features}}{\text{total numbers of samples}} \times 100 \quad (21)$$

### III. RESULTS AND DISCUSSION

In this research, k=10 cross validation is used. So total samples are 752, train samples are 677 (90%) and test samples are 75 (10%). NBC, CART, KNN and PNN were used for identification of some medicinal and aromatic plant species. The training accuracy obtained in all classifier is higher (Table I). The highest performance is obtained by NBC in dataset (98.39%, 98.00%). This high performance is directly related to the method followed in the feature extraction.

TABLE I  
PERFORMANCE RESULTS WHICH WERE OBTAINED THE DIFFERENT ML METHODS

ML Methods	Data Set	Accuracy
NBC	Training	98.39%
	Testing	98.00%
CART	Training	99.51%
	Testing	97.61%
KNN	Training	100.00%
	Testing	81.92%
PNN	Training	100.00%
	Testing	67.69%

Each feature (Gray, Shape, and Fractal features) was removed from the model and its contribution to performance was determined. The accuracy decrement when they have removed is shown in Table II.

TABLE II  
CONTRIBUTION OF PROPOSED FEATURES WITH REGARDS TO ACCURACY IN MAP DATABASE.

Item	Contribution (%)
Gray features	5.97
Shape features	3.28
Fractal features	0.04

The high test accuracy (98.00 %) was obtained by NBC, for all folds, total training confusion matrix are given in Fig. 5 and total testing confusion matrix in Fig. 6.

%	Predicted Class				
	St. John'swort	Melissa	Echinacea	Thyme	Mint
St. John'swort	98.50	0.00	1.50	0.00	0.00
Melissa	0.00	99.28	0.11	0.00	0.61
Echinacea	2.53	1.14	96.34	0.00	0.00
Thyme	0.00	0.66	0.00	99.34	0.00
Mint	0.00	2.08	0.99	0.00	96.93

Fig. 5. Confusion Matrix for Classification kNN for train

The 6.659 out of 6.768 train samples were correctly classified (98.39%). In Fig. 5, Echinacea type was classified incorrectly St. John'swort (2.53%) and Melissa (1.14%). Similarly, Thyme type was classified incorrectly Melissa (0.66%). Melissa was confused to mostly Mint (0.6%).

%	Predicted Class				
	St. John'swort	Melissa	Echinacea	Thyme	Mint
St. John'swort	98.50	0.00	1.50	0.00	0.00
Melissa	0.00	98.50	0.50	0.00	1.00
Echinacea	1.14	1.14	95.45	0.00	2.28
Thyme	0.00	0.66	0.00	99.34	0.00
Mint	0.00	2.68	0.89	0.00	96.43

Fig.6. Confusion Matrix for Classification kNN for test

In Fig.6, 737 out of 752 test samples were correctly classified (98.00%). Echinacea type was classified incorrectly St. John's wort (1.14%) , Melissa (1.14%) and Mint (2.28%). Similarly, Thyme type was classified incorrectly Melissa (0.66%). Melissa was confused to mostly Mint (1%). Misclassification style for train dataset is similar to test dataset.

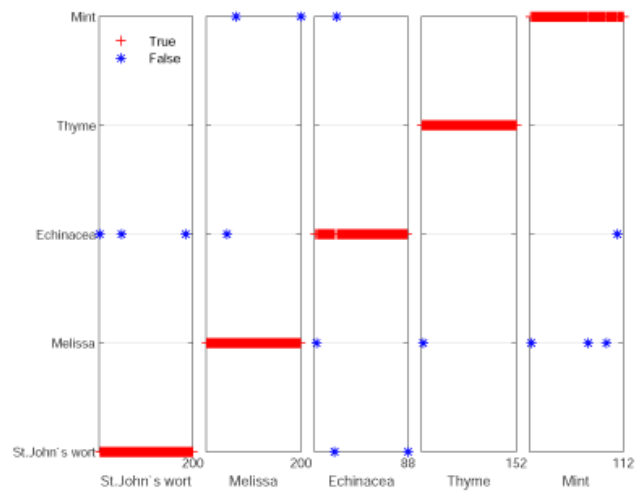


Fig. 7. Result of the classification NBC for test dataset

Wu et al. used 12 features (orthogonalized 5 principal variables) to 32 leaf types and achieved %90.31 accuracy rate with PNN classification [17]. Du et al. used 15 features to 20 leaf types and achieved %93 accuracy rate with 1-NN classification [32]. Du et al. used 20 features to 35 leaf types and achieved %87.14 accuracy rate with k-NN classification [4]. Kadir et al. used 17 features to 32 leaf types and achieved %94.07 accuracy rate with PNN classification [21]. Ruberto et al. used 138 features to 32 leaf types and achieved approximately %100 accuracy rate with SVM classification [3]. Kalyoncu et al. used a lot of features to 32 leaf types and achieved approximately %94 accuracy rate with LDC

classification [1]. In this work, it achieves %98.00 accuracy rate with NBC, which there are 5 leave types and 11 features. Fig.7 shows the correct and incorrect classification in test dataset. For instance, St. John's wort leaves, which have 200 test data, were correctly classified the numbers of 197 data (+ symbol) and incorrectly classified the numbers of 3 data (\* symbol) as Echinacea leaves.

#### IV. CONCLUSION

Due to the difficulty of morphologically similar leaves, plant recognition based on images is difficult processing. Therefore, the performance of many algorithms used in plant identification has low accuracy.

In this study, the leaves of medicinal and aromatic plants were automatically classified according to leaf shape and color using computer vision. For this classification, a new feature extraction approach, which used gray, shape and fractal features of leaf, was proposed. The used feature set is a combination of several known feature sets. Additionally, we proposed the use of NBC that has a higher accuracy compared to KNN, CART, and PNN classifiers. In this research, it has been shown that leaf identification can be developed by combining features complementing each other. In this study, the properties that increase the accuracy of plant recognition are combined. Among the proposed methods, NBC is a better-class classifier. The experimental results on MAP database verify the effectiveness of the proposed method.

The reason is that NBC based on applying Bayes' theorem with naive independence assumptions among the gray, shape and fractal features. This allows the use of a combination of independent leave's features. Additionally, online classification time of NBC is the lowest due to single scan and fast classify. Thus, it is very pertinent for large-scale applications.

#### REFERENCES

- [1] C. Kalyoncu, T. Önsen: "Geometric leaf classification", *Computer Vision and Image Understanding*, 2015, 133, (0), pp. 102–109
- [2] M. Shao, J. Du, J. Wang, C. Zhai: "Recognition of leaf image set based on manifold-manifold distance", In: *Intelligent Computing Theory - 10th International Conference, ICIC 2014, Taiyuan, China, August 3-6, 2014, Proceedings*, pp. 332–337
- [3] C.D. Ruberto, L. Putzu: "A fast leaf recognition algorithm based on SVM classifier and high dimensional feature vector", In: *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Volume 1, Lisbon, Portugal, 5-8 January, 2014*, pp. 601–609
- [4] J. Du, C. Zhai, Q. Wang: "Recognition of plant leaf image based on fractal dimension features", *Neurocomputing*, 2013, 116, pp. 150–156
- [5] S.G. Wu, F.S. Bao, E.Y. Xu, Y. Wang, Y. Chang, Q. Xiang: "A leaf recognition algorithm for plant classification using probabilistic neural network", *CoRR abs/0707.4289*, 2007
- [6] Z. Miao, M. Gandelin, B. Yuan: "An oopr-based rose variety recognition system", *Eng. Appl. of AI*, 2006, 19, (1), pp 79–101
- [7] X. Gu, J.-X. Du, X.-F. Wang: "Leaf recognition based on the combination of wavelet transform and gaussian interpolation", In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) *Advances in Intelligent Computing, Lecture Notes in Computer Science*, Springer, 2005, vol. 3644, pp 253–262
- [8] X. Wang, J. Du, G. Zhang: "Recognition of leaf images based on shape features using a hypersphere classifier", In: *Advances in Intelligent Computing, International Conference on Intelligent Computing, ICIC 2005, Hefei, China, August 23-26, 2005, Proceedings, Part I*, pp 87–96
- [9] S.S. Dhillion, H. Svarstad, C. Amundsen, H. Bugge: "Bioprospecting: Effects on Environment and Development", *Ambio*, 31, (6), 2002, pp:491–493,
- [10] J.R.S. Tabuti, S.S. Dhillion, K.A. Lye: "Traditional medicine in bulamogi county, uganda: its practitioners, users and viability", *Journal of Ethnopharmacology*, 85, (1), 2003, pp 119–129
- [11] H. Sadraei, A. Ghannadi, K. Malekshahi: "Relaxant effect of essential oil of melissa officinalis and citral on rat ileum contractions", *Fitoterapia*, 2003, 74, (5), pp 445–452
- [12] M. Mucciarelli, W. Camusso, C.M. Berteau, S. Bossi, M. Maffei: "Effect of (+)-pulegone and other oil components of menthapiperita on cucumber respiration", *Phytochemistry*, 2001, 57, (1), pp 91–98
- [13] H. Baydar, O. Sağdıç, G. Özkan, T. Karadoğan: "Antibacterial activity and composition of essential oils from *origanum*, *thymbra* and *satureja* species with commercial importance in Turkey", *Food Control*, 2004, 15, (3), pp 169–172
- [14] M. Centritto, F. Loreto, A. Massacci, F. Pietrini, M.C. Villani, M. Zacchini: "Improved growth and water use efficiency of cherry saplings under reduced light intensity", *Ecological Research*, 2000, 15, (4), pp. 385–392
- [15] "The CLEF 2011 plant images classification task", <http://ceur-ws.org/Vol-1177/CLEF2011wn-ImageCLEF-GoeauEt2011a.pdf>, accessed 25 June 2015
- [16] "The imageclef 2012 plant identification task", <http://ceur-ws.org/Vol-1178/CLEF2012wn-ImageCLEF-GoeauEt2012.pdf>, 25 June 2015ww
- [17] B. Yanikoglu, E. Aptoula, C. Tirkaz: *Automatic plant identification from photographs Machine Vision and Applications*, 2014, 25, pp.1369–1383
- [18] S.G. Wu, F.S. Bao, E.Y. Xu, Y. Wang, Y. Chang, Q. Xiang: "A leaf recognition algorithm for plant classification using probabilistic neural network", *CoRR abs/0707.4289*, 2007
- [19] N. Kumar, P.N. Belhumeur, A. Biswas, D.W. Jacobs, W.J. Kress, I.C. Lopez, J.V.B. Soares: "Leafsnap: A computer vision system for automatic plant species identification", In: *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part II*, pp. 502–516
- [20] O.J.O. Söderkvist: "Computer vision classification of leaves from swedish trees". Master's thesis, Linköping University, SE-581 83 Linköping, Sweden, LiTH-ISY-EX-3132, September 2011
- [21] A. Kadir, L.E. Nugroho, A. Susanto, P.I. Santosa: "Neural network application on foliage plant identification", *CoRR abs/1311.5829*, 2013
- [22] M.M. Ghazi, B. Yanikoglu, E. Aptoula: "Plant identification using deep neural networks via optimization of transfer learning parameters", *Neurocomputing*, 2017, 235, pp. 228–235
- [23] M.S. Odabas, N. Senyer, G. Kayhan, E. Ergun: "Estimation of chlorophyll concentration index at leaves using artificial neural networks", *Journal of Circuits, Systems, and Computers*, 2015
- [24] C.D. Manning, P. Raghavan, H. Schütze: "Introduction to Information Retrieval", Cambridge University Press, New York, NY, USA, 2008
- [25] L. Breiman, J.H. Friedman, R. Olshen, A.C.G. Stone: "Classification and regression trees", Wadsworth International Group, Belmont, California, USA, 1984.
- [26] W.-Y. Loh: "Classification and regression trees", *Wiley Interdisc. Rev.: Data Mining and Knowledge Discovery*, 2011, 1, (1), pp. 14–23
- [27] N.S. Altman: "An introduction to kernel and nearest-neighbour nonparametric regression", 1992, 46, (3), pp. 175–185
- [28] D.F. Specht: "Probabilistic neural networks", *Neural Netw.* 1990, 3, (1), pp. 109–118
- [29] S.V. Stehman: "Selecting and Interpreting Measures of Thematic Classification Accuracy", *Remote Sensing of Environment*, 1997, 62, (1), pp. 77–89
- [30] BS ISO 5725-1: "Accuracy (trueness and precision) of measurement methods and results - part 1: General principles and definitions", 1994
- [31] C.E. Metz: "Basic principles of {ROC} analysis", *Seminars in Nuclear Medicine*, 1978, 8, (4), pp 283–298
- [32] J.-X. Du, X.-F. Wang, G.-J. Zhang: "Leaf shape based plant species recognition", *Appl. Math. Comput.*, 2007, 185, (2), pp 883–893

## BIOGRAPHIES



**GÖKHAN KAYHAN** Trabzon, in 1978. He received the B.S. degrees in Electrical and Electronics Engineering from the Ondokuz Mayıs University, Samsun, in 1999 and the M.S. degrees in Electrical and Electronics Engineering from the Ondokuz Mayıs University, Samsun, in 2004 the Ph.D. degree in the B.S. degrees in Electrical and Electronics Engineering from the Ondokuz Mayıs University, Samsun, in 2011.

From 2001 to 2013, he was a Research Assistant in Department of Computer Engineering at Ondokuz Mayıs University. Since 2013, he has been an Assistant Professor Dr. at the Department of Computer Engineering, Ondokuz Mayıs University. His research interests include machine learning artificial intelligence, image processing, signal processing.



**ERHAN ERGÜN** Samsun, in 1970. He received the B.S. degrees in Electrical and Electronics Engineering from the Hacettepe University, Ankara, in 1993; M.S. and Ph.D. degrees in Electrical and Electronics Engineering from the Ondokuz Mayıs University, Samsun, in 1999 and 2008.

From 2009 to 2010, he was a Assistant Professor Dr. at Bafra Vacation School, Ondokuz Mayıs University. Since 2012, he has been an Assistant Professor Dr. at the Department of Computer Engineering, Ondokuz Mayıs University. His research interests include analog electronic circuits, image processing, signal processing.