The Nucleus 55, No. 1 (2018) 1-7

www.thenucleuspak.org.pk

The Nucleus
ISSN 0029-5698 (Print)
ISSN 2306-6539 (Online)

An Ensemble Classifier based Leaf Recognition Approach for Plant Species Classification using Leaf Texture, Morphology and Shape

A. Riaz, S. Farhan*, M. A. Fahiem and H. Tauseef

Department of Computer Science, Lahore College for Women University, Lahore, Pakistan

ARTICLE INFO

Article history :
Received : *06 March, 2017*
Accepted : *15 March, 2018*
Published : *04 April, 2018*

Keywords:
Classification
Morphological features
Texture features
Zernike moments

ABSTRACT

Plant recognition is a main problem for biologists, environmentalists and chemists. Human experts of these fields perform plant recognition manually, which requires more time and is less efficient. Plant recognition systems are used to classify plants into appropriate taxonomies. Such information is useful for botanists, industrialists, food engineers and physicians. Botanists use morphological features of the leaves to identify them. These features are used in terms of automation in identifying the plants. Leaf images of different plants have different characteristics which help in the classification of these species. The proposed approach identifies plant species in three distinct phases: (1) preprocessing, (2) feature extraction and (3) classification. Leaf features like texture, morphology and Zernike moments are extracted and treated as input vector to the four different classifiers. The best accuracy achieved by a combination of textural and morphological features is 87%.

1. Introduction

Plants play a beneficial role in the Earth's ecology by providing foodstuff, fuel, oxygen, shelter, medicines and much more. As there is a large number of plant species available on earth, it is difficult to recognize each plant and its importance. Designing automatic plant classification system is beneficial, as it helps in quick classification of plants and has a varied use in a wide range of scientific and industrial fields. Plant database management for classification and recognition of various plant species is a crucial phase for their preservation. Plant classification is not only of interest to botanists and plant ecologists, it is also significant and of interest to professionals such as landscape planners and designers, arbori-culturists, herbal practitioners and biologists, as well as to nature loving individuals like hikers, eco-tourists and park visitors. Mostly, classification of plants has been done using different parts of plants such as leaves, barks, flowers or fruits. Flowers and fruits make the classification difficult and more time consuming as they are not available throughout the year. Therefore, leaves are considered as an imperative feature to distinguish plant species. Moreover, collecting leaf images for plant classification is low-cost and convenient, and classification can simply be performed using leaf color, texture and shape. The aim of this work is to select stable features having the qualities necessary to classify various types of leaves using highly efficient recognition algorithm. Plant classification is an active research area in recent years. Due to the remarkable development of digital image processing, pattern recognition and machine vision, numerous techniques have been introduced for plant classification using leaves. For automated plant classification, the most frequently used features are texture, color and shape; mostly classifiers used in plant classification are Probabilistic Neural Network (PNN), Linear Discriminant Classifier (LDC), Move Median Centers (MMC) and Support Vector Machine (SVM). The performance of these classifiers is different with various sets of features. The existing techniques perform well and provide acceptable accuracy using a large feature set which is computationally expensive.

In our research work, we propose a classification system using textural features that are calculated using statistical methods of the gray level histogram of the image and concentrate only on first order histogram and co-occurrence based measures; combined with shape features, e.g., aspect ratio, rectangularity etc. and Zernike moments [1]. A feature vector of a small number of features is then created for classification purposes. Initially, classification accuracy is computed for individual descriptors discussed above using ensemble learning classifiers. Then, performance of the same classifiers is measured using feature sets consisting of texture, geometric and Zernike moments collectively. We found that smaller feature vectors can successfully classify the samples of leaves, and the performance of classifiers is improved vividly. Many leaf image datasets are available for the general public such as Flavia dataset, Image CLEF dataset, Leaf snap dataset, the Smithsonian Leaf dataset, Swedish Leaf dataset and many others. The efficiency of the proposed algorithm is evaluated using Flavia dataset.

Many researchers have proposed different methodologies for plant classification. Recently, various leaf features have been used to make a feature set, such as statistical features [2] and Fourier descriptors for leaf identification [3], multi scale fractal dimension [4], leaf

*Corresponding author : saifar79@hotmail.com

color [5], centroid contour distance, vein structures and a combination of morphological and textural features [6]. A fusion of shape features and Hu moments without using any contour information is proposed by [7]. Several features such as aspect ratio (ratio of length and width of the leaf) and rectangularity require length and width of the leaf and preprocessing of the images in order to have extreme points marked manually by the user. A proficient algorithm for plant classification using morphological features and vein features to characterize the leaf is proposed by Wu et al. [8] with a resulting accuracy of 90%, but the features still require length and width of the leaf as an input. Leaf skeletons and veins using a fusion of Wavelet Transform and Gaussian interpolation are used by Gu et al. [9] with k-nearest neighbor (kNN) and radial basis probabilistic neural network (PNN) classifier. Shape features, vein features, leaf contour and leaf centroid are used by Lee et al. [10] for performing plant classification. For acquiring frequency domain data, Fast Fourier Transform has been performed on the values of distance between centroid and contour of leaves. For classification of Indian medicinal plants, Sandeep [11] used leaf area, color and edge features.

Techniques like SVM [12], PNN [9, 13], MMC hyper spheres [5, 8] and ANN with back-propagation [14] have been used for plant classification. Table 1 shows details of some existing classical techniques used for plant classification.

Table 1: An overview of existing classical methods of plant classification

Reference	Preprocessing	Feature extraction	Accuracy	Classifier
[15]	Segmentation, Noise removal, Contour extraction	Moment invariant, Geometrical features, Multi scale distance matrix	70%	LDC
[16]	Rotation, Grayscale conversion, Binarization, Edge detection	Morphological features, Distance map	83.5%	KNN
[17]	Grayscale conversion, Binarization, Edge detection, Corner detection	Leaf tooth feature	76%	Sparse representation classifier
[7]	Grayscale conversion, Background removal	Contour signature, Texture analysis	81.1%	Incremental classification method
[18]	Grayscale conversion, Contour detection	Fractal dimensions	84%	Fractal refinement technique

3. Material and Methods

3.1 Data

The dataset used in this research work is Flavia leaf dataset [8]. It contains 32 plant species, each consisting of 40 to 60 images in RGB color space and white background with no leaf stalk, collected from multiple trees. Filenames of all images are 4-digit numbers, with the file extension .jpg. A sample of leaf images from each of the 32 species is shown in Fig. 1. Table 2 shows some details of Flavia dataset.

Fig. 1: Species of flavia leaf dataset.

3.2 Proposed Approach

The proposed approach consists of three distinct stages, i.e., (i) pre-processing, (ii) feature extraction and (iii) classification, as shown in Fig. 2. The first stage is the preprocessing, which includes RGB to grayscale conversion, segmentation, noise reduction and smoothing. The second stage extracts different types of shape features, textural features and Zernike moments. In the classification stage, these features are used to classify leaf images using ensemble learning classifiers.

3.2.1 Preprocessing

The preprocessing step is required to prepare images in order to extract beneficial features from them. This step is carried out before the actual analysis is performed on the image. Most algorithms in image processing, like feature extraction and segmentation of images, are radically dependent on the quality of the images. Visual reliability of the images is increased by this step. It has a number of techniques to enhance or eradicate some details of the

Table 2: Details of the flavia dataset used in this research.

Attribute	Value
Total number of Species	32
Total number of Leaf Images	1907
Images per Class	40 to 60
Resolution	1600*1200
Color Space	RGB
File Format	jpg

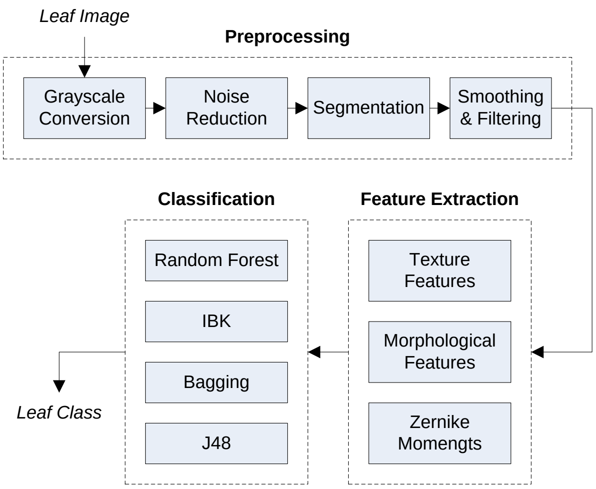

Fig. 2: Detailed working of proposed approach.

image in order to execute further analysis efficiently. In our proposed approach, the step of preprocessing includes RGB to gray conversion, noise reduction, leaf segmentation and smoothing binary images. Details of these steps with the results are shown in Fig. 3.

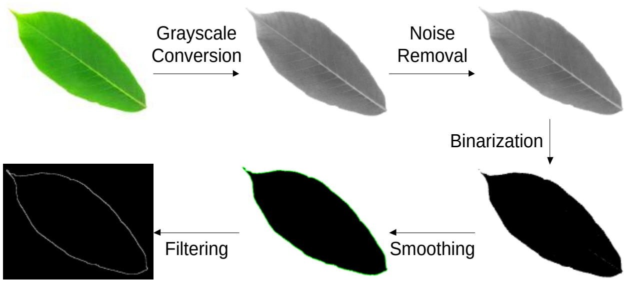

Fig. 3: Preprocessing steps performed on a sample leaf image.

An RGB image I_c *(x,y)* is converted to a grayscale image I_g *(x,y)*. Every leaf image has some noise in the form of small holes and dust particles that are distributed over the image capturing device. It also has undesirable burrs on its periphery, so the noise removing step is necessary to correct the distorted or degraded data using a median filter. Next, leaf segmentation is done. It converts leaf grayscale image I_g *(X, Y)* into a binary image I_b *(X, Y)* using a simple adaptive threshold on the grayscale image. After segmentation, smoothing and filtering is applied. In this step, morphological dilation operation is performed to get rid of non-leaf margin edges retained in the leaf edge image obtained using prewitt edge detection filter in order to get a smoothed leaf image. After dilation, thinning morphological operation is performed to get a leaf contour as thin as one pixel.

3.2.2 Feature extraction

In order to classify images accurately and efficiently, feature extraction is performed to find more relevant and discriminative features, as most of the information residing in images is irrelevant. After the image is segmented, image representation is very important for further processing. Image regions can be represented either using external characteristics or using internal characteristics. External characteristics use the information regarding object boundary and internal characteristic use pixels in a region such as color and texture. In our proposed method, texture, morphological and Zernike moment based features are taken into account. Mostly, leaves of different plants are green in color and have similar texture or shape, therefore, use of a single type of features may not produce valuable results.

3.2.3 Textural features

In the field of image processing and computer vision, texture is considered as an important cue for the analysis of images. Texture represents the tactile quality of the surface and structure of object. It provides information about the spatial arrangement of visual patterns. Textural features can be calculated using different methods. Textural features in our research work are calculated using (i) statistical methods of the gray level histogram of the image and (ii) co-occurrence based measures.

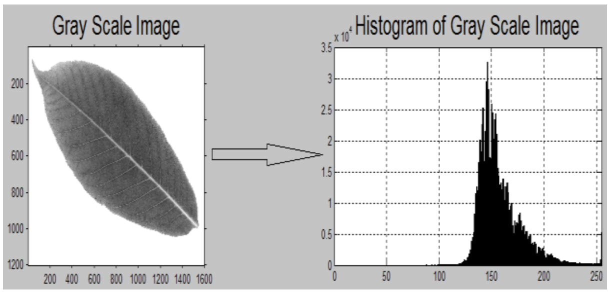

Fig. 4: A gray scale leaf image and its intensity histogram.

Histogram based measures are first order statistics that describe the texture of regions using a gray level intensity histogram *p(i)* of image. The first order histogram statistics include mean, variance, skewness and kurtosis. All of these are calculated using a histogram of grayscale image as shown in Fig. 4.

Mean (μ) is one of the most common features that represents the average of each gray level region. It provides an estimate of intensity in the image using Eq. (1).

$$\mu = \sum_{i=0}^{G-1} iP(i) \quad (1)$$

Variance (σ^2) gives histogram width and measures the fluctuations from mean gray level value as presented in Eq. (2).

$$\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 p(i) \qquad (2)$$

Histogram asymmetry of the gray levels around the sample mean level is described by skewness as shown in Eq. (3). Data is spread more to the left of mean when skewness is negative and to the right of mean when it is positive.

$$Skewness = \sum_{i=0}^{G-1} (i-\mu)^3 p(i) \qquad (3)$$

Kurtosis is a measure of the uniformity of the gray level distribution and is defined as shown in Eq. (4).

$$Kurtosis = \sum_{i=0}^{G-1} (i-\mu)^4 p(i) \qquad (4)$$

The limitation of computed texture using histogram is that they have no information about the positions of pixels with respect to each other. Grey Level Co-occurrence Matrices (GLCM) is one of the most commonly used methods for extracting various textural features. GLCM features consider distribution of pixel intensities and also the positions of pixels with equal, or almost equal, intensity values. GLCM is based on the second order statistics of the grayscale histogram of image describing relation between pixel pair *p*(*i*,*j*). Features that can be calculated based on GLCM are angular second moment (ASM), contrast (C), correlation (Corr), homogeneity (H) and entropy (E).

Angular second moments provide a measure of the homogeneity of an image, as indicated in Eq. (5). The values of *p(i,j)* are relatively high in homogeneous regions and low in inhomogeneous regions.

$$ASM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (P(i,j))^2 \qquad (5)$$

Contrast is a measure that favors contributions from *P(i,j)* away from the main diagonal as mentioned in Eq. (6).

$$C = \sum_{n=0}^{G-1} n^2 \left(\sum_{i=1}^{G} \sum_{j=1}^{G} P(i,j)\right) \qquad (6)$$

Correlation, as presented in Eq. (7), measures linear dependence of gray levels between pixels at specified locations. Here μi, μj and σi, σj denote the mean and standard deviations of the row and column sums of the matrix, respectively.

$$Corr = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \qquad (7)$$

The homogeneity in images is calculated using Eq. (8). Its value for homogeneous regions is high, whereas, it is low for inhomogeneous regions.

$$H = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} p(i,j) \qquad (8)$$

Entropy, as presented in Eq. (9), is the measure of randomness of the gray levels in an image. A homogeneous region has a high entropy value. Maximum value is reached when all probabilities are equal.

$$E = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) \times log(p(i,j)) \qquad (9)$$

3.2.4 Shape features

We defined shape features on the basis of geometrical, morphological and Zernike moment. Geometric features include Leaf's perimeter, length, width, area and diameter as shown in Fig. 5.

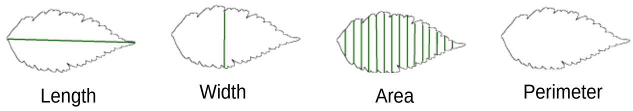

Fig. 5: Leaf geometric features.

Using four basic geometric features, we can define digital morphological features including aspect ratio, rectangularity, convex hull, area of convex hull, perimeter ratio of diameter, form factor, narrow factor, perimeter ratio of length and width, and circularity. Aspect ratio, bounding box and convex hull are shown in Fig. 6.

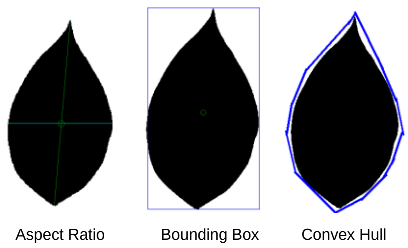

Fig. 6: Leaf morphological features.

3.2.5 Zernike moments

These are shape descriptors belonging to a class of circularly orthogonal moments that are simple, rotation invariant and have higher accuracy for detailed shapes. Zernike moments are calculated using Eq. (10).

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y V_{nm}{}^*(x,y) f(x,y) \qquad (10)$$

Here $x^2+y^2 \leq 1$, $Z_{nm}{}^* = Z_{n\text{-}m}$ and $V_{nm}{}^*(x, y)$ are the complex conjugate of Zernike function.

3.2.6 Classification

Numerous machine learning algorithms are available for the classification of data. Four different types of classifiers, i.e., Random Forest, IBK, J48 and Bagging are used to evaluate the accuracy using different types of features. Random Forest is an ensemble learning classifier that creates a number of decision trees and uses them to make a classification. Each decision tree predicts a class that receives the larger number of votes. K-Nearest Neighbor (IBK) is considered a lazy learning algorithm that classifies data based on its similarity with its neighbors. "K" stands for number of data set items that are considered for the classification. Here, a value of K=1 is used for all experiments. Bagging is also an ensemble method and an application of bootstrap procedure. Bagging provides an improvement in unsteady estimation or classification to solve regression problems and to get an aggregated predictor. Decision tree is a very fundamental technique in classification and is widely used. Its working is based on creating a top-down binary tree recursively and selecting an attribute with the highest information gain as a root. In WEKA [19] decision tree is implemented as J48.

4. Results and Discussions

The present study consists of three types of experiments. In the first experiment, individual sets of features, i.e., Texture (T), Morphological (M) and Zernike (Z) are used with four classifiers. In second experiment different pairs of feature sets are selected, i.e., Morphological + Zernike (M+Z), Texture + Morphological (T+M), Texture + Zernike (T+Z) and classification is performed. In the third and last experimentation, all three feature sets are combined i.e., Texture + Morphological + Zernike (T+M+Z) and classification is done. Ten-fold cross validation strategy is applied throughout the classification process. For all experiments, accuracy, precision and recall values are calculated using Eq. (11), Eq. (12) and Eq. (13), respectively.

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

Where *TP* means 'True Positive', *TN* means 'True Negative, *FP* means 'False Positive' and *FN* means 'False Negative'.

The result was approximately 85% with all three individual feature sets and 87% with Texture + Morphological features using Random Forest algorithm. The results are reported as accuracy (Acc), precision (Prec) and recall (Rec) and can be seen in Table 3. Graphical presentations of results are provided in Fig. 7, Fig. 8 and Fig. 9, respectively.

Table 3: Results of classification.

Feature set	Random forest			IBK			Bagging			J48		
	Acc	Prec	Rec	Acc	Prec	Rec	Acc	Prec	Rec	Acc	Prec	Rec
T	61	62	61	65	67	65	52	53	53	49	52	49
M	78	77	78	72	73	72	67	68	67	65	66	65
Z	74	73	74	80	81	80	63	65	63	57	54	57
M+Z	80	81	80	82	85	82	67	70	67	52	50	52
T+M	87	88	87	83	85	83	73	75	73	70	72	71
T+Z	84	86	84	82	84	82	64	69	64	65	64	65
T+M+Z	85	87	85	85	86	85	67	69	67	65	65	65

The proposed research work is compared with latest research articles. Generalized Relevance Learning Vector Quantization (GRLVQ), which is a competitive based learning algorithm, and is used for integrated feature extraction and classification of plant leaf images [20]. Precision and recall values obtain through this method are 0.9380 and 0.9298, respectively. Although the results seem to be better, the dataset consists of only 15 plant species. The dataset used in our proposed research comprises of 32 classes. Moreover, the feature set is larger as compared to proposed approach. Similarly Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP) and Speeded-Up Robust Features (SURF) with multiclass Support Vector Machine (SVM) classifier have been used for plant leaf recognition [21]. Accuracy achieved through HOG, LBP and SURF is 97%, 97% and 63%, respectively. HOG and LBP are computationally expensive as these techniques work on fine details present in the images. The feature sets hence extracted comprise of hundreds of features which are much more than the number of features in our proposed approach.

From these results, it is concluded that the proposed approach produces better results with smaller feature sets as compared to existing approaches. In Fig. 7, it can be seen that the accuracy produced by random forest and IBK is higher as compared to bagging and J48. Features that provide higher accuracy are Texture + Morphological features or when all features are combined.

4. Conclusion

The algorithm proposed in our research work classifies a plant leaf by incorporating textural features, morphological features and shape features. We investigated the use of ensemble classifiers for plant species

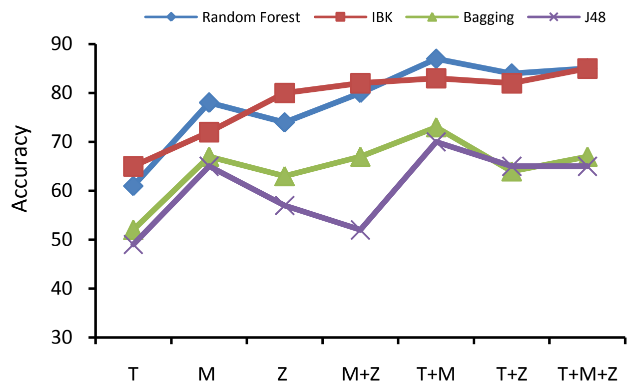

Fig. 7: Comparison of classifier accuracy using different feature sets.

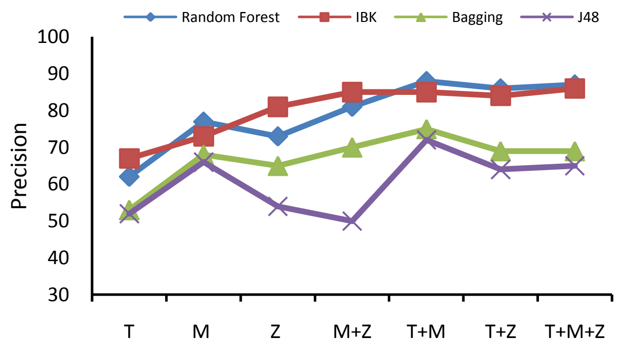

Fig. 8: Comparison of classifier precision using different feature sets.

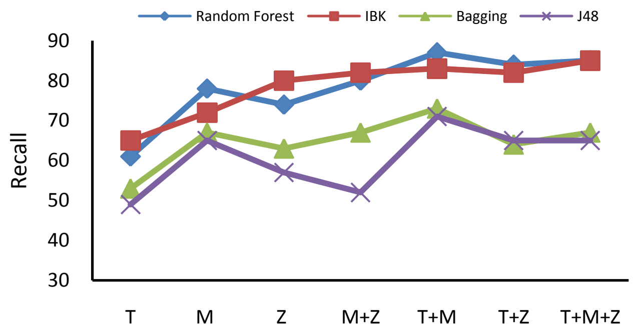

Fig. 9: Comparison of classifier recall using different feature sets.

classification. The classification task is initially performed on a single type of features, but it is observed that single type of features are not sufficient for successful classification, as these do not cover all characteristics of the leaf necessary for its classification. Morphological, textural and shape features are then combined in order to improve classification accuracy. Among all feature types, only the most discriminating features are selected to reduce computational overhead. From the present study, it is found that the proposed approach produced better results with even small feature sets as compared to existing approaches because of appropriate selection between texture features, morphological feature and Zernike moments. The proposed algorithm is tested on Flavia dataset consisting of 1900 samples belonging to 32 plant species and achieved an accuracy of 87%. In future, leaf vein features will be investigated and analyzed for plant species classification.

References

[1]. M.R. Teague, "Image analysis via the general theory of moments", J. Opt. Soc. Am., vol. 70, no. 8, pp. 920-930, 1980.

[2]. A. Shaukat, S. Farhan, M. Tahir, M.A. Fahiem and H. Tauseef, "Automated Classification of Hair Care Plants using Geometrical and Textural Features from Leaf Images: A Pattern Recognition based Approach", Pak J. Sci., vol. 68, no. 4, pp. 419-425, 2016.

[3]. J.C. Neto, G.E. Meyer, D.D. Jones, and A.K. Samal, "Plant species identification using Elliptic Fourier leaf shape analysis", Comput. Electron. Agric., vol. 50, no. 2, pp. 121-134, 2006.

[4]. A.R. Backes and O.M. Bruno, "Plant leaf identification using multi-scale fractal dimension", Int. Conf. on Image Analysis and Processing, September 8-11, 2009, Vietri sul Mare, Italy, Foggia, Pasquale, Sansone, Carlo, Vento, and Mario, Eds. Berlin, Heidelberg: Springer, pp. 143-150, 2009.

[5]. A. Kadir, L.E. Nugroho, A. Susanto and P.I. Santosa, "Leaf classification using shape, color, and texture features", Int. J. Computer Trends and Technology, pp. 256-263, 2011.

[6]. M.Z. Rashad, B.S. El-Desouky and M.S. Khawasik, "Plants images classification based on textural features using combined classifier", Int. J. of Comp. Sci. Inf. Tech., vol. 3, no. 4, pp. 93-100, 2011.

[7]. T. Beghin, JS. Cope, P. Remagnino and S. Barman, "Shape and texture based plant leaf classification", 12th Int. Conf. on Advanced Concepts for Intelligent Vision Systems, Dec. 13-16, 2010, Sydney, Australia, Eds.: J. Blanc-Talon, D. Bone, W. Philips, D. Popescu and P. Scheunders, Berlin, Heidelberg: Springer, pp. 345-353, 2010.

[8]. S.G. Wu, F.S. Bao, E.Y. Xu, Y.X. Wang, Y.F. Chang and Q.L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network", IEEE Int. Symp. on Signal Processing and Information Technology, Dec. 15-18, 2007, Cairo, Egypt, pp. 11-16, 2007.

[9]. X. Gu, J.X. Du, and X.F. Wang, "Leaf recognition based on the combination of wavelet transform and Gaussian interpolation", Int. Conf. on Intelligent Computing, August 23-26, 2005, Eds.: Hefei, China, D.-S. Huang, X.-P. Zhang and G.-B. Huang, Berlin, Heidelberg, Springer, pp. 253–262, 2005.

[10]. K.-B. Lee, K.-W. Chung and K.-S. Hong, "An implementation of leaf recognition system based on leaf contour and centroid for plant classification", Ubiquitous Information Technologies and Applications, Eds.: Y.-H. Han, D.-S. Park, W. Jia and S.-S. Yeo, Springer, pp. 109-116, 2013.

[11]. K. Sandeep, "Leaf Color, area and edge features based approach for identification of Indian medicinal plants", Indian J. Comp. Sci. Engg., vol. 3, no. 3, pp. 436-442, 2012.

[12]. A. Aakif and M.F. Khan, "Automatic classification of plants based on their leaves", Biosystems Engineering, vol. 139, pp. 66-75, 2015.

[13]. H. Zhang, P. Yanne and S. Liang, "Plant species classification using leaf shape and texture", IEEE Int. Conf. on Industrial Control and Electronics Engineering, August 23-25, 2012, Xi'an, China, pp. 2025-2028, 2012.

[14]. B.C. Heymans, J.P. Onema and J.O. Kuti, "A neural network for opuntia leaf-form recognition", Proc. of IEEE Int. Joint Conf. on Neural Network, November 18-21, Singapore, pp. 2116-2121, 1991.

[15]. C. Kalyoncu and O. Toygar, "Geometric leaf classification", Comp. Vis. Image Underst., vol. 133, pp. 102-109, 2015.

[16]. T. Munisami, M. Ramsurn, S. Kishnah and S. Pudaruth, "Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers", Procedia Comp. Sci., vol. 58, pp. 740-747, 2015.

[17]. T. Jin, X. Hou, P. Li, and F. Zhou, "A novel method of automatic plant species identification using sparse representation of leaf tooth features", PloS One, vol. 10, no. 10, pp. e0139482, 2015.

[18]. A. Jobin, M.S. Nair, and R. Tatavarti, "Plant identification based on fractal refinement technique (FRT)", Procedia Tech., vol. 6, pp. 171-179, 2012.

[19]. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA data mining software: an update", ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, 2009.

[20]. E. Imah, Y. Rahayu and A. Wintarti, "Plant leaf recognition using competitive based learning algorithm", 2nd Annual Applied Science and Engineering Conference, IOP Conference Series: Mater. Sci. Engg., January, 2018: IOP Publishing, pp. 012058, 2018.

[21]. Z. Ibrahim, N. Sabri and N.N.A. Mangshor, "Leaf Recognition using Texture Features for Herbal Plant Identification", Indonesian Journal of Electrical Engineering and Computer Science, vol. 9, no. 1, pp. 152-156, 2018.