Tropical Plant Leaf Identification Application

(MYLeaf)

Seif Allah El Kari, Syamsiah Mashohor and Siti Mariam Shafiee

***Abstract*—**Leaf identification is an important aspect for agriculture players and botanists, due to the enormous amount of time, studies and specialized knowledge to be able to identify a particular leaf species. The recognition systems need to be developed locally due to the characteristics of plants which varied according to many factors such as genetic factors and environmental factors. The importance to have a portable identification application is important since the best time to identify a plant is when it is still in the ground or when it is fresh. This usually means the identification should occur in the field. Leaf provides many shape morphological variations which can be observed using image processing to enhance feature extraction performance of the extracted data and Artificial Neural Network (ANN) is used as classifier. This framework was carried out using Opencv library and Multilayer Perceptron (MLP) library for better and optimized leaf identification framework compiled on Microsoft Visual Studio. The obtained results show 90% of classification training accuracy of the nine groups of leaf shape. The testing accuracy of 95% in leaf class 1 and lowest testing rate of 59% in leaf class 9.

Keywords: Multi-layer perceptron, feature extraction, leaf identification, shape features

I. INTRODUCTION

Millions of plants exist on planet earth, which shares an important role in the life of human being, from nutrition, medical use, and environment sustainability. For this reason, botanist spends many years on research and gathering for preserving unknown plants species and endangered from the extension. Traditionally, collected samples were dehydrated, then categorized based on observation. Compared with methods, such as biology analysis like cell and molecules, the study of image processing and the usage of computer vision technologies open up various studies in image-based plant recognition. In this case, the computers can extract features using image processing technique.

Recent studies were investigating all visual aspects of a plant, for example, a flower, leaf, and bark [1]. However, the leaf was considered the most effective scope of area of plant identification because leaf was easy to obtain and to be studied. Leaf provided a set of features that can be explored like color, shape, and texture [2].

Color can vary from leaf to leaf, and it can be distinguished between a set of color histograms but tropical plants are mainly in variations of green color. Texture features typically were the vein structure of a leaf and this feature is very useful to differentiate between species under one generic name. However, the shape feature provided a dominant chance of achieving an accurate leaf identification focused on the leaf regions (contour) using digital morphological analysis [3].

The important area of leaf recognition research and methods were presented in the literature. Researchers used shape features such as aspect ratio, form factor, circularity, smooth factor and convex hull [4]. In fact, these features are derived from primary leaf geometric such as physiological length and physiological width [4].

Typically, after the features are extracted, the classification was considered as part of the studies conducted on leaf recognition such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and k-nearest neighbors' algorithm. The proposed classifier was based on ANN, due to its simplicity and low computation time [5].

This project is aimed to identify tropical plant leaf using an application and classify the leaf based on extracted information from digital images. The main objectives of this project are:

1. To identify the nine categories of shape of tropical leaf and species that falls under them.
2. To design an identification method using leaf's shape.
3. To evaluate the performance of leaf recognition of each categorized species.
4. To develop a web based application for tropical plant leaf identification.

A leaf recognition algorithm using shape features and efficient classification method is implemented in this project. Figure 1 shows the flow diagram of MYleaf application development. The algorithm include data collection, shape feature extraction, MLP classification and web based development. The final output from the application is the leaf shape category and possible species that fall under the respective leaf shape category.

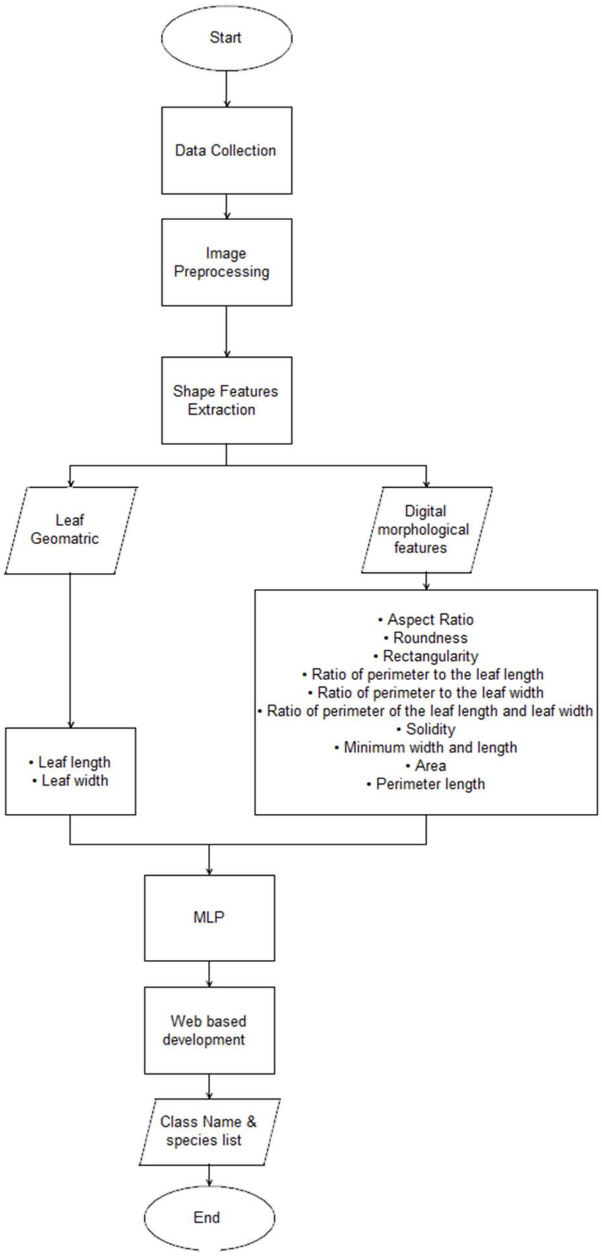


Figure 1: The flowchart of the development of MYleaf application.

II. *Methods*

This project started with meeting the lecturers and students from Forestry Faculty in UPM. From the discussions, we decided to develop the application by categorizing the plant's leafs into 9 shapes of leaf as suggested by the forestry expert. Then, the whole MYleaf application is divided into three main stages which are image preprocessing, feature extraction and classification. All of these are implemented using C++ compiled on Microsoft Visual Studio using OpenCV library and MLP library.

A. Data collection

This project used several datasets that are available online and collected from field exploration, two main data set contributed around 80% of the test data samples, and they are Flavia dataset collected by Stephen Gang Wu [6] and data set collected by the Faculty of Forestry, UPM. The other 20% are from the Putrajaya Botanical garden [7] and Leaf dataset [8]. The total number of collected images are 887 and all the 9 shape groups are distributed according to the sources as in Figure 2. The 9 shape categories for plant leaf are illustrated in Figure 3. This is the first categorization step used by botanists to recognize a plant species.

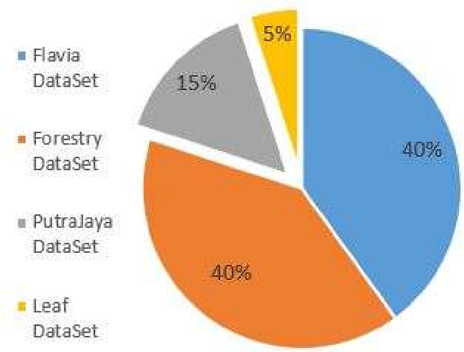


Figure 2: The distribution of images according the data collection sources.

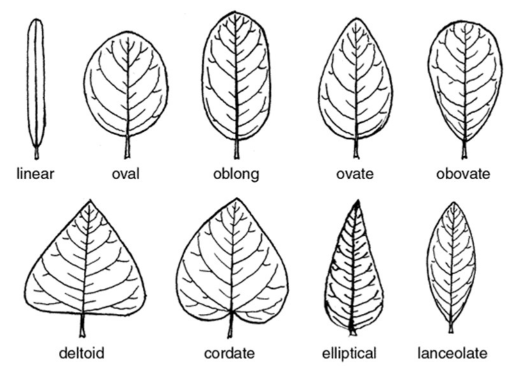


Figure 3: The nine categories of shape for plant leaf

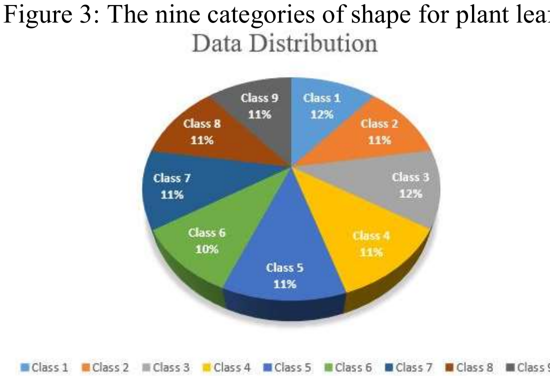
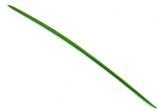
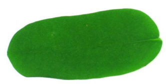
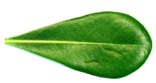


Figure 4: Even Distribution of images according to shape class.

Table 1: The samples of collected images according to the shape categories.

Class No	Shape Name	Sample 1	Sample 2
1	Linear		
2	Oval		
3	Oblong		
4	Ovate		
5	Obovate		
6	Deltoid		
7	Cordate		
8	Elliptical		
9	Lanceolate		

B. *Image preprocessing*

In this stage, a Red -Green-Blue (RGB) acquired image is converted to a grayscale image using the Eq.1. This is performed to reduce the complexity of the algorithm by processing in grayscale images and color features are not used in this framework.

$$Gray = 0.299 * + R0.587 * G + 0.114 * B \quad (1)$$

After the color model conversion, boundary extraction is performed using Canny edge detection method to produce the outline of a leaf [9].

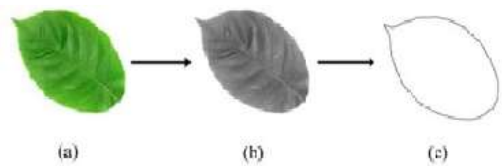


Figure 3: The preprocessing steps comprised of (a) original image, (b) grayscale image and (c) edge detected image.

Canny edge algorithm process the binaries image in 4 steps. Firstly the noise is filtered using smoothing technique where a linear filter filters the output value determined by a weighted summation of the input pixel. Secondly, the intensity gradient of the captured image is obtained. Thirdly, a non-maximum suppression method is applied to remove pixels that are not belong to the edges. Finally, a hysteresis test on pixel gradient of two threshold will be conducted, where the upper threshold would be accepted as an edge [10].

C. *Feature extraction*

In this section, we selected 12 geometric features which has been suggested from literature reviews. The extracted features are obtained from the leaf contour using Canny edge detection. Features extraction include basic geometric features and digital morphological features. Basic geometric feature extraction is illustrated Figure 4 [11].

1. *Basic geometric features*

- Leaf length: defined as the longest distance between leaf centroid line and end margin of the leaf opposite ending donated by LL
- Leaf width: defined as the distance between LL centroid line and the leaf contour margin donated by LW (as shown in Fig.3 (b)).

Both length and width are obtained using bounding box surrounding the extracted leaf edge.

- Leaf Area: calculate the contour area
- Leaf Perimeter: calculate the contour perimeter length (as shown in Fig.3 (a)).

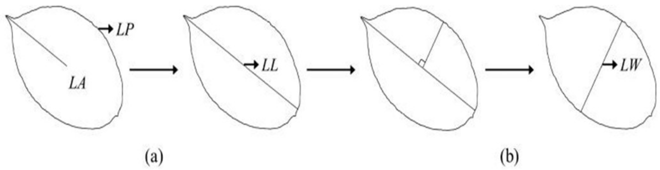


Figure 4: The illustration of geometric features, (a) leaf area, leaf perimeter and leaf length extraction, (b) leaf width [11].

2. Digital morphological features

- Aspect ratio: calculated by using the leaf length, LL and leaf width, LW

$$LL/LW \qquad (2)$$

- Circularity: describe the leaf and a circle or circularity ratio

$$4LA \;/\; LP2 \qquad (3)$$

- Perimeter ratio of height: ratio of the leaf perimeter, LP and leaf length, LL

$$LL/LP \qquad (4)$$

- Perimeter ratio of width: ratio of the leaf perimeter, LP and leaf length, LW

$$LP/LW \qquad (5)$$

- Ratio of perimeter of the leaf length, LL and leaf width, LW

$$LP/(LL + LW) \quad (6)$$

- Solidity: ratio of the area of convex hull and contour area

$$LA/LW * LL \quad (7)$$

- Minimum width and length area: Extracted minimum rectangular area enclosing the points set

D. MLP Classification

Neural network programing structure inspired by the nervous system processing abilities. This network contains artificial neurons called units, this units are arranged in a cascading configuration where it accepts information from an external network. Multilayer perceptron is a feed forward artificial neural network that route a set of data onto a set of its appropriate outputs [12].

A multi-layer neural network consists of large number of units joined together in a configuration of linked units. Units usually separated into three layers: input units, which receive information to be processed; output units, where the results of the processing are found; and units in between known as hidden units. The network is trained on a set of paired data to determine input-output mapping.

The weights of the connections between neurons are fixed, and the network is used to determine the classifications of a new set of data.

ANN depends upon three fundamental aspects, input and activation functions, network architecture and the weight of each input connection. The behavior of the ANN is defined by the current values of the weights [13] (as shown in Fig.5).

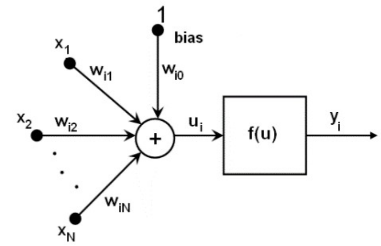


Figure 5: summed value to activation function

The obtained outputs x_j of the layer n, the outputs y_i of the layer $n+1$

$$u_i = \sum_j \left(w_{i,j}^{n+1} * x_j\right) + w_{i,bias}^{n+1} \quad (8)$$

$$y_i = f(u_i) \qquad (9)$$

Backpropagation learning algorithm is used in training this type of network where the output layer of weights are corrected using the following formula [14].

$$w_{ho} = w_{ho} + (n\delta_o o_h) \quad (10)$$

w_{ho}= weight connecting hidden layer unit (h) with output unit (o), n = learning rate, o_h= output at hidden unit (h).

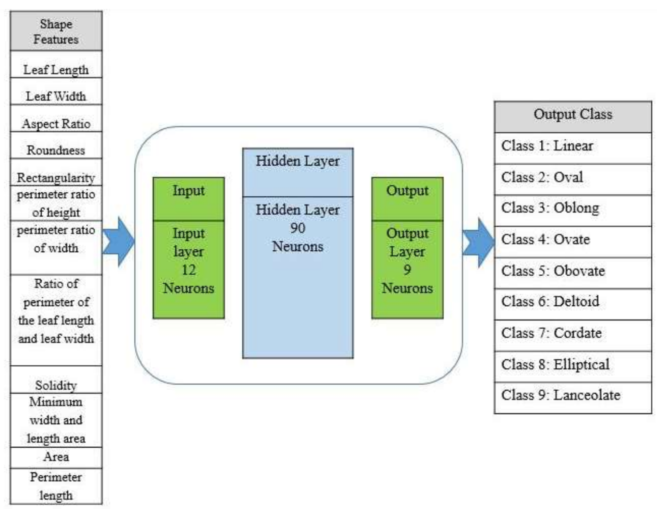


Figure 6: The architecture of the proposed MLP classifier.

The implemented architecture is shown in Fig 6, which is using a 12 feature inputs neurons, 90 hidden layers neurons and 9 output neurons. The parameters used to optimize the classifier are: The maximum number of iteration is 1000, sigmoid activation function default value 0.6 to 1, learning rate of 0.000001 and momentum constant of 0.1.

E. Web based development

The web based development consist of GUI design and integrating the preprocessing, feature extraction and classification code into web based processing.

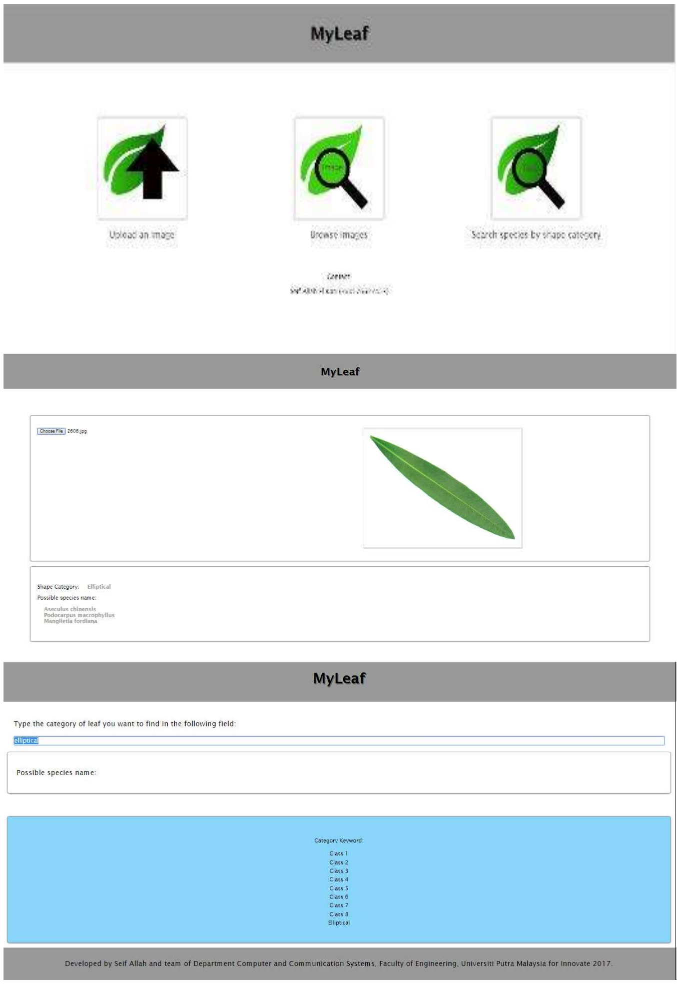


Figure 7: Screenshot of MYleaf web based interface

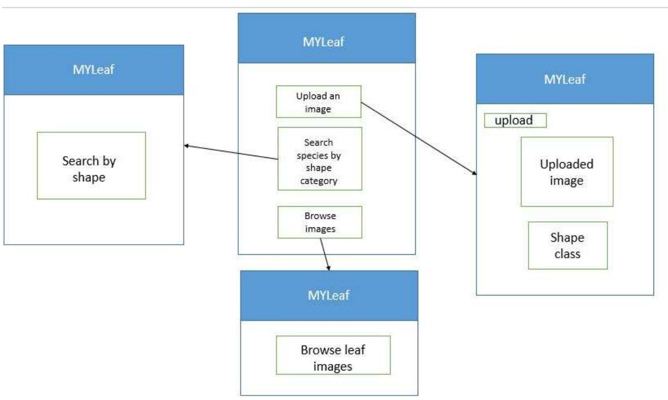


Figure 8: Web based processing flowchart

III. Results and Discussions

the number of data set used were 887 samples some samples shown in table 1.The feature are extracted and classified using multilayer perceptron. The training classification is given in table 2, 3 with 90% accuracy rate and testing confusion matrix accuracy is given in table 4.

Feature combination of the all 12 feature were tested to provide an overview of each possible combination were it can help improve training classification. All 12 combination showed major improvement in training accuracy rate of 91%. As shown in fig 9.

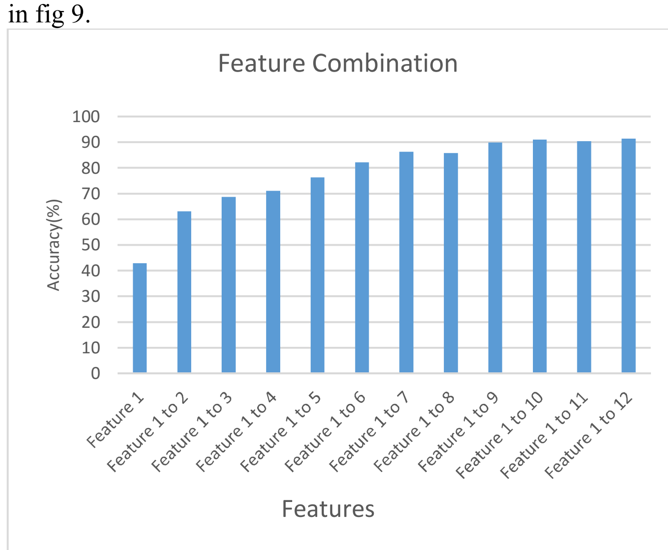


Figure 9: Result of feature combination

The experimental results are divided into MLP classification training accuracy rate and testing by using cross validation leave one out method. The proposed framework has been trained on the whole even distributed datasets. From table 2, it is observed that leaf linear shape gives best classification training accuracy of 100% and Low accuracy classification training rate of 68% for Lanceolate class shape. This due to strong similarity with Elliptical class shape and the wrong segmentation within the Putrajaya dataset of leaf images.

Table 4 describes the obtained results of cross validation which produced the average classification accuracy of 73%. The best shape category that was able to achieve high classification rate is class 1, which is linear shape with accuracy of 95%. This is due to unique leaf shape comparing to other classes. Class 9 get low accuracy rate of 59% with high confusion with class 8. Therefore, further analysis showed that Class 4, 8 and 9 are nearly identical to each other. Even Human observation faces difficulties to categorize these three shapes classes within its correct classes during classification.

Considering the automated leaf feature and proposed fast MLP classification and its parameters. This framework is tolerable to distinguish other 6 classes due to acceptable classification accuracy rates of the real data images and classification performance.

Table 2: Classification training results using the proposed features and MLP classifier.

Class	**Leaf Class Shape**	**No. of sample**	**Retrieved of samples**	**Classification accuracy rate (%)**
1	Linear	100	100	100
2	Oval	99	91	91.9191
3	Oblong	100	97	97
4	Ovate	100	95	95
5	Obovate	100	93	93
6	Deltoid	87	82	94.25
7	Cordate	100	93	93
8	Elliptical	100	87	87
9	Lanceolate	100	68	68

Table 3: Classification training summary using the proposed features and MLP classifier.

Total No. of Sample	**Total Retrieved Samples**	**Average Recognition Rate**
887	807	90.9808

Table 4: Classification testing results using cross validation (leave-one-out)

		Target Class								
		1	2	3	4	5	6	7	8	9
Output Class	1	95 %	0	0	2	0	0	0	1	5
	2	0	75 %	4	14	0	2	2	0	0
	3	0	1	85 %	1	3	2	4	3	2
	4	3	11	3	65 %	10	2	1	1	3
	5	0	1	3	6	64 %	9	7	6	8
	6	0	6	2	3	3	73 %	13	2	4
	7	0	4	0	1	0	7	69 %	0	0
	8	1	2	2	4	10	0	2	70 %	19
	9	1	0	1	4	10	0	2	17	59 %

IV. Conclusion

In this paper, we introduces the use of leaf shape features obtained by extracting the leaf morphological features for automatic leaf shape identification and classification based on the nine shape classes. For Performance Evaluation the results indicates that the algorithm using Multilayer perceptron is efficient with accuracy of training 91% and cross validation of testing 73%. This proposed algorithm showed efficient recognition, time efficiency and easy implementation.

ACKNOWLEDGEMENT

We would like to thank Assoc. Prof. Dr.Mohd Zaki bin Hamzah from Department of Forest management, Faculty of Forestry, UPM for sharing his expertise in recognizing plants with us. Special thanks to Husna Izzati binti Hasanuddin, a final year student from Faculty of Forestry, UPM for helping us in collecting the images.

REFERENCES

[1] M. Kumar, M. Kamble, S. Pawar, P. Patil and N. Bonde, “Survey on Techniques for Plant Leaf Classification”, International Journal of Modern Engineering Research, vol. 1, no. 2, (2011), pp. 538-544.

[2] R. P. S. B. Jyotismita Chaki, "Plant leaf recognition using texture and shape features with neural classifiers," Pattern Recognition Letters, vol. 1, no. 6, pp. 61-68, 2015.

[3] I. Y. Olfa Mzoughi, "Semantic-based automatic structuring of leaf images for advanced plant species identification," Multimed Tools Appl, no. 75, pp. 1615-1646, 2016.

[4] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, “Plant species identification using digital morphometrics: A review,” Expert Systems with Applications, pp. 7562–7573, 2012.

[5] M. Shabanzade, M. Zahedi and S. A. Aghvami, “Combination of Local Descriptors and Global Features for Leaf Classification”, Signal Image Processing: An International Journal, vol. 2, no. 3, (2011), pp. 23-31.

[6]Ahmad Luqman Ahmad Firdaus, Syamsiah Mashohor,” Performance Method of Content Based Image Retrieval (CBIR) For Plant Leaves”, 2011

[8] H. Zhang, P. Yanne, and S. Liang, “Plant species classification using leaf shape and texture,” in International Conference on Industrial Control and Electronics Engineering, 2012, pp. 2025–2028.

[9] M. D. A. Ekshinge Sandip Sambhaji, "Leaf Recognition Algorithm Using MLP Neural Network Based Image Processing," International Journal of Engineering Sciences & Research Technology, vol. 1, no. 3, pp. 778-783, 2014.

[10] G. A. S. J. A. O. JESELLE PETRIZE M. SOSA, "An Application Using Canny Edge Detection and Multilayer Perceptron for Recognizing Leaves of Tropical Plants," Mathematics and Computers in Biology and Biomedical Informatics, 2014.

[11] K.-B. Lee and K.-S. Hong, “An implementation of lea recognition system based on leaf contour and centroid for plant classification,” in International Journal of Bio-Science and Bio-Technology, vol. 5 (2), 2013, pp. 57–66.

[12] P.a. Evanjali Gamit, "Literature Review On Plant Leaf Classification," International Institution for Technological Research and Development, vol. 1, no. 2, 2015.

[13] L. E. Nugroho. A. Santosa. P. S. A. Kadir, "Neural Network Application on Foliage Plant Identification," International Journal of Computer Applications, vol. 29, no. 9, pp. 75-84, 2011.

[14] GAO, L., Lin, X., Zhong, M., and Zeng, J. (2010b). A neural network classifier based on prior evolution and iterative approximation used for leaf recognition. In Natural Computation (ICNC), 2010 Sixth International Conference on, volume 2, pages 1038–1043.