

# Prototype Designing of Computer-aided Classification System for Leaf Images of Medicinal Plants

**Govardhan Jain and Deepti Mittal**

*Electrical and Instrumentation Engineering Department, Thapar University, Patiala (Punjab), India;*  
gjain307@gmail.com; deepti.mittal@thapar.edu

## ABSTRACT

The objective of the present work is to design computer-aided classification (CAC) system to discriminate among various medicinal leaves. This objective is achieved with the designing of a classification system using shape features to obtain high accuracy of end results. In order to evaluate the performance of the presented work six leaf classes were adopted and its shape features were extracted. The role of each shape feature in differentiating the leaves from one another is studied using region of convergence (ROC) curves. The combined effect of both the shape features is studied using neural network classifiers. Two neural network (NN) Classifiers were designed to classify the leaves and their performance is compared. First classifier is designed using radial bases function neural network (RBFNN) and second using feedforward backpropagation neural network (FFBNN) with a single hidden layer. A comparative study for this dataset reveals that RBFNN shows 92% classification accuracy which is 2.7% higher than that of FFBPNN.

**Keywords:** classification, medical leaf, feature extraction, NN classifiers, Otsu segmentation, FFBNN

## 1 Introduction

Ayurvedic medicines are recommended by majority of people all over the world because of less or rather no side effects as compared to the modern medicines. In India there is a huge focus on herbal and indigenous medicines. To make it largely accepted and available for our future generations, the major concern lies in distinguishing the leaves of the herbs.

There are various areas such as medical imaging, disease classification etc. which includes both macroscopic (shape) and microscopic observation (texture) [1-5]. However, classification of leaf images using its visual morphological features like roundness and slimness by macroscopic observation is more promising, time efficient and cost effective than the microscopic observational features like texture. However, it has been observed that the existence of wide morphological varieties through evolution among the leaf cultivars variety complicates the classification process and there are not many experts in the field of identification of medicinal plants. In order to retrieve the information at a faster rate there is an urgent need of making an automated identification system to distinguish leaves of different medicinal plants which requires a great design and development system which we will deal in this paper. So, this

research focuses on developing a CAC system that helps in distinguishing leaves with medicinal properties based on its visual features like slimness and roundness.

Table1 summarizes the work of several researchers in the designing of CAC system to discriminate among different medicinal leaves. *Wang et al.* used a moving median centre hypersphere classifier to classify leaves based on its shape features but they didn't specify the leaf class [6]. Texture features were included by *Zulkifi et al.* to classify leaves [7]. *Zulkifi et al.* also worked on the method of invariant moments and classified the leaves using general regression neural network. *Wu et al.* extracted 12 shape features and classify leaves using probabilistic neural network (PNN) classifier [8]. All the above researchers did not process any colour information. *Kadir et al.* worked on colour, texture and shape feature and classify the leaves using PNN classifier [9]. *Wu. Man et al.* classified the leaves based on their shape, colour and texture using support vector machine classifier [10]. *Rjanani et al.* also worked on the extraction of shape, colour and texture features from the selected medicinal plants and classified the leaves using artificial neural network [11]. They implemented FFBNN with a single hidden layer. *Kumar et al.* classified medicinal leaves based on its shape features using Gaussian distribution [12]. *Rahmani et al.* used K-nearest neighbour classifier to classify leaves using shape, texture and scale margin feature histogram [13]. *Chemburkar et al.* used only morphological features to classify the herbal leaves using FFBNN [14].

**Table1. Brief detail of the studies on leaf classification**

Author (Year)	Leaf Classes (No. of leaves)	Feature extracted	Classifier used
X-F Wang (2005)	-	Shape	MMC Hypersphere
Zulkifi (2007)	-	Shape and texture	General Regression NN
Wu (2007)	Flavia	Shape	Probabilistic NN
Man (2008)	-	Shape Colour and Texture	Support Vector Machine
Kadir (2013)	Flavia	Shape Colour and Texture	Probabilistic NN
Janani (2013)	Thulasi, Neem, Henna, Hibiscus, Curry	Shape colour and texture	Artificial neural network
Kumar (2014)	Medicinal Leaves	Shape	Gaussian distribution
Rahmani (2015)	100 different classes	Shape, Texture, and Scale margin feature histogram	K-NN Classifier
Chemburkar (2014)	Herbal leaves	Shape	Artificial NN

Among all approaches, artificial neural network classifiers (ANN) based approaches has the fastest speed and gives best accuracy for classification. *Du et al.* indicates that artificial neural network classifiers such as multilayer perceptron networks, backpropagation neural network, RBFNN, radial bases probabilistic neural network run faster and achieve higher accuracy than k-NN and move median centre hypersphere classifier [15]. So this research adopts NN classifiers to classify leaves.

In the present work, shape features from the leaf image are extracted after segmentation using Otsu thresholding algorithm [16]. Then the role of each shape feature in differentiating the leaves from one

another is analysed using ROC analysis. The performance of both the shape features are analysed through a comparative study of two classifiers. The classifiers used in this research are radial bases function neural network and feed forward backpropagation neural network with a single hidden layer.

## 2 Materials and Methods

### 2.1 Materials

In this research 255 leaf images of six different classes were taken. These include images of *Castanea Dentata* (CDT) (49), *Asminia triloba* (ATA) (30), *Salix Babylonica* (SBA) (51), *Quercus Velutina* (QVA) (59), *Salix Caroliniana* (SCA) (39), *Starix Obassia* (SOA) (35). The data base has been taken from an online source [19]. Out of 255 leaf images 190 were used for training and 65 were used for testing.



Figure 1. Samples of six different classes of leaves used for classification

### 2.2 Methods

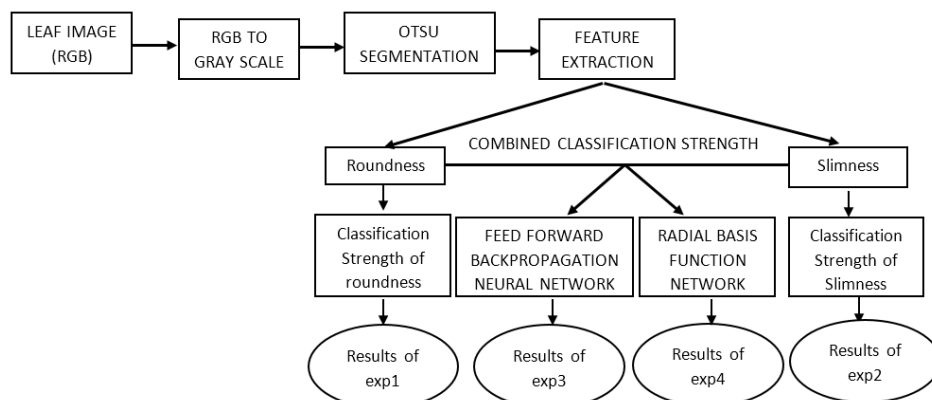


Figure 2. Block diagram of the proposed method

#### 2.2.1 Image Segmentation

Otsu thresholding algorithm is used to convert the grey scale image into the binary image [16]. It is an important technique in image segmentation applications. The basic idea of thresholding is to select an optimal grey level threshold to separate region of interest in an image from the background based on the grey level distribution of pixels of the image. Thresholding creates binary images by turning all the pixels below that threshold value to zero and above or about the threshold value to one. In Otsu's algorithm it is assumed that histograms are bimodal which means that it contains two classes of pixels that is foreground and background. In this algorithm we exhaustively search for the threshold that minimizes the intra-class variance or maximizing the inter-class variance between these two classes of pixels.

### 2.2.2 Feature Extraction Module

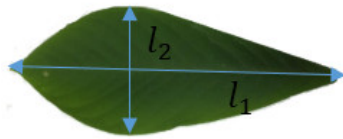
Geometric features which were used to classify leaves are roundness and slimness.

Roundness

Roundness which is also called compactness is defined as:

Roundness =  $\frac{4\pi A}{p^2}$ , Where P is the perimeter of the leaf and A is the area of the leaf.

Slimness



Slimness also called aspect ratio of a leaf is defined as ratio of width to length of the leaf Slimness=  $\frac{l_1}{l_2}$

Where  $l_1$  is the width of the leaf and  $l_2$  is the length of the leaf.

### 2.2.3 Feature Scaling

Feature scaling is a method used to standardize the range of features of data. Since the range of values of raw data varies widely, in some machine learning algorithms, objective functions will not work properly without normalization. The method used in this research is scaling the range of features in the range of [0, 1]. The general formula is given as:

$$\hat{X} = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

In the above formula  $\hat{X}$  represents the new value of the feature,  $X_i$  represents the original feature value,  $X_{min}$  represents the minimum value of the original feature and  $X_{max}$  represents the maximum value of the original feature.

### 2.2.4 Classification Module

The scaled features set as calculated above are given as an input to the artificial neural network classifier. Two classifiers are designed and their performance and accuracy is compared using ROC and confusion matrices. One is designed FFBNN and second is designed using RBFNN.

A two layer FFBNN was used for comparison. The sigmoid function was used in the hidden layer because it is differentiable which makes it possible to characterize the rate of change in the output layer error with respect to a change in a particular weight and reduces computational complexity. Soft max function was used as the output layer function because it gives better results than sigmoid function for multiclass discrimination problem. The network is trained with scaled conjugate gradient transfer function which can train any network as long as weight, net input and transfer function of the network has derivative function































The radial basis function neural network is a type of multilayer network. The basic RBFNN structure consists of three layers i.e. an input layer, a kernel (hidden) layer, and an output layer. In the kernel layer, it consists of a set of kernel basis functions called radial basis functions. The output of the RBFNN is a linear combination (weighted sum) of the radial basis function calculated by the kernel units. In RBF networks, the outputs of the input layer are determined by calculating the distance between the network inputs and hidden layer centres. The second layer is the linear hidden layer and outputs of this layer are weighted forms of the input layer outputs. Each neuron of the hidden layer has a parameter vector called centre. RBFNN can overcome some of the limitations of FFBPNN. RBFNN can use a single hidden layer for modelling any nonlinear function therefore, it is able to train data faster than FFBPNN.

### **3 Results and Discussion**

#### **3.1 ROC results**

In this section the role of each shape feature in identifying a leaf class is studied using ROC. A Receiver Operating Characteristics plots between Sensitivity vs Specificity across a range of values and its area represents the predictability of the classifier.

**Table2. ROC results of individual feature for binary classification of medicinal leaf**

SNO.	Leaf Class 1		Leaf Class 2		Roundness (Az)		Slimness (Az)	
					Train	Test	Train	Test
1	ATO		SBA		1	1	1	1
2	CDA		SBA		1	1	1	1
3	CDA		SOA		1	1	1	1
4	QVA		SOA		1	1	1	1
5	ATO		QVA		1	1	1	0.90
6	SBA		SCA		1	1	0.99	0.98
7	SCA		SOA		1	1	0.82	0.77
8	ATO		CDA		1	1	0.64	0.70
9	ATO		SOA		1	0.95	1	1
10	ATO		SCA		1	0.93	1	1
11	SBA		SOA		0.90	0.80	1	1
12	CDA		QVA		0.96	0.90	1	0.91
13	SBA		QVA		0.90	0.80	1	0.92
14	QVA		SCA		0.80	0.85	1	0.91
15	CDA		SCA		0.76	0.61	1	0.91

In Table2 the role of each feature in binary classification of leaves has been shown for both training and test data set by computing area (Az) under the roc curves. The area under the roc curve is a measure of how well a particular parameter can distinguish between two classes. The Az values are 1 for ATO vs SBA, CDA vs SBA, CDA vs SOA, QVA vs SOA with both the features roundness and Slimness. It can be depicted

that any one either roundness or slimness can be chosen to classify these classes. For the classes ATO vs QVA, SBA vs SCA, SCA vs SOA, ATO vs CDA the Az values are 1 for roundness and 0.90, 0.98, 0.77, 0.70 for slimness respectively. These results shows that for these classes roundness is considered to be the powerful feature in differentiating these classes. The Az values for ATO vs SOA, ATO vs SCA, SBA vs SOA are 1 for slimness and 0.95, 0.93, 0.80 for roundness. For these classes only slimness feature is sufficient and shows exciting results in classifying the leaves. For rest of the classes the results are promising for both training as well as test data set but are not able to discriminate among leaf classes individually. The Az values for CDA vs QVA, SBA vs QVA, QVA vs SCA, CDA vs SCA are 0.90, 0.80, 0.85, 0.61 for roundness and 0.91, 0.92, 0.91, 0.91 for slimness respectively. So in order to classify these classes both the features are needed together. The combined effect of both the features is studied by using CAC system. Two neural network classifiers RBFNN AND FFBPNN are designed and there performance is compared using confusion matrices.

### 3.2 Classifier Results

The results of designed CAC systems are discussed in this section. First, the results of designed CAC system using FFBNN is studied. Then the results of RBFNN are studied and are compared with FFBPNN. A total of 255 leaf images were used for the classification, out of which 190 were used for training, 65 were used for testing. Both the classifiers are trained to 100% accuracy among the six leaf classes for the training data set and its performance is evaluated for the test data set.

**Table3. Comparison of confusion matrices of RBFNN and FFBNN**

Comparison of test results for FFBPNN and RBFNN												
CLASS CODE	ATO		CDA		SBA		QVA		SCA		SOA	
	BP	RB	BP	RB	BP	RB	BP	RB	BP	RB	BP	RB
ATO	17	18	0	0	0	0	0	0	2	2	1	0
CDA	1	1	10	10	0	0	0	0	1	1	0	0
SBA	0	0	0	0	11	11	1	1	0	0	0	0
QVA	0	0	0	0	0	0	10	10	0	0	0	0
SCA	0	0	0	0	0	0	0	0	8	8	0	0
SOA	2	1	0	0	0	0	0	0	0	0	11	12
Class Accuracy (%)	85	90	100	100	100	100	90.9	90.9	72.7	72.7	91.7	100
Overall Accuracy (%)	89.3	92.0										

Table 3 shows the comparison of test results in the form of confusion matrix when above mentioned classifiers are used for multiclass classification using FFBNN and RBFNN. Classification results of CDA, SBA

are best with the classification accuracy of 100% for both the classifiers. The classification results of ATO, QVA and SOA are highly exciting with the classification accuracy of 85%, 90.9%, 91.7% for FFBPNN and 90%, 90.9%, 100% for RBFNN respectively. The classification results of SCA were promising with the classification accuracy of 72.7% for both FFBNN and RBFNN. Finally, the overall classification strength is 92% for RBFNN and 89.3% for BPNN which shows that RBFNN gives a 2.7% higher accuracy than FFBPNN.

#### 4 Conclusion

The results lead to a conclusion that only visual shape features that is roundness and slimness can differentiate a particular leaf class from six different classes to a higher extent with less computational time and complexity. Therefore, this preliminary work provides the exciting results in medicinal leaf classification with two features only. The results shows almost 100% discrimination capability in most of the binary classification and very promising results in multi-class classification using neural network classifiers. However, RBFNN outperforms FFBPNN by showing 92% accuracy whereas BPNN shows 89.3% accuracy which is 2.7% higher than that of FFBPNN. The proposed CAC system designed with two features only saves computational time and enhances the possibility of implementation the same methodology with comparatively large set of medicinal leaf classification.

#### REFERNCES

- [1] Mittal, D., Kumar, V., Saxena, S.C., Khandelwal, N., Kalara, N. (2011, January). Neural network based focal liver lesion diagnosis using ultrasound images, *Computerized Medical Imaging and Graphics*, 35(4), 315-323.
- [2] Mittal, D., Rani, A. Ritambhara (2016, February). Detection and classification of focal liver lesions using support vector machine classifiers, *Journal of Biomedical Engineering and Medical Imaging*, 3(1), 21-34.
- [3] Mittal, D., Kumari, K., (2015, September). Automated detection and segmentation of drusen in retinal fundus images, *Computers and Electrical Engineering*, 47, 82–95.
- [4] Mittal, D. (2017). Impact of modified anisotropic diffusion–based enhancement method in computer-aided classification of focal liver lesions, *Ultrasonic Imaging*, 39(2), 96-107.
- [5] Kumari, K., Mittal, D (2015, June). Drusen quantification for early identification of age related macular degeneration, *Advances in Image and Video Processing*, 3(3), 28-40.
- [6] Zhang, G. J., Wang, X. F., Huang, D. S., Chi, Z., Cheung, Y. M., Du, J. X., & Wan, Y. Y. (2004, October). A hypersphere method for plant leaves classification. In *Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on* (pp. 165-168). IEEE.
- [7] Zulkifli, Z., Saad, P., & Mohtar, I. A. (2011, December). Plant leaf identification using moment invariants & general regression neural network. In *Hybrid Intelligent Systems (HIS), 2011 11th International Conference on* (pp. 430-435). IEEE.



- [8] Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007, December). A leaf recognition algorithm for plant classification using probabilistic neural network. In *Signal Processing and Information Technology, 2007 IEEE International Symposium on* (pp. 11-16). IEEE.
- [9] Kadir, A., Nugroho, L. E., Susanto, A., & Santosa, P. I. (2013). Leaf classification using shape, color, and texture features. *arXiv preprint arXiv:1401.4447*.
- [10] Man, Q. K., Zheng, C. H., Wang, X. F., & Lin, F. Y. (2008, September). Recognition of plant leaves using support vector machine. In *International Conference on Intelligent Computing* (pp. 192-199). Springer Berlin Heidelberg.
- [11] Janani, R., & Gopal, A. (2013, September). Identification of selected medicinal plant leaves using image features and ANN. In *Advanced Electronic Systems (ICAES), 2013 International Conference on* (pp. 238-242). IEEE.
- [12] Kumar, E. S., & Talasila, V. (2014, April). Leaf features based approach for automated identification of medicinal plants. In *Communications and Signal Processing (ICCSP), 2014 International Conference on* (pp. 210-214). IEEE.
- [13] Rahmani, M. E., Amine, A., & Hamou, M. R. (2015). Plant Leaves Classification. *ALLDATA 2015*, 82.
- [14] Chemburkar, A., Sartape, A., Gawade, A., Somawanshi, P., & Ghorpade, J. (2014). Automated Tool for Plant Leaf Classification Using Morphological Features. *International Journal Of Engineering And Computer Science*, 3(11), 9098-9102.
- [15] J. Du, D. Huang, X. Wang, and X. Gu, "Shape recognition based on radial basis probabilistic neural network and application to plant species identification," in *Proceedings of 2005 International Symposium of Neural Networks*, ser. LNCS 3497. Springer, 2005.
- [16] Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296), 23-27.
- [17] <http://leafsnap.com/dataset/>