

Leaf Identification Using Feature Extraction and Neural Network

Satnam Singh¹, Dr. Manjit Singh Bhamrah²

¹(Pursuing M.tech in Department of Electronics and Communication Engineering, Punjabi University, Patiala, India)

²(Professor in Department of Electronics and Communication Engineering, Punjabi University, Patiala, India)

Abstract: In the world, some important species of plants are going extinct day by day. To control the phenomena, steps are identification, restoring and protecting of the plants. Among them, identification of proper medicinal plants is quite challenging. Usually, plants are identified from their leaves. In this paper, a method is proposed for the extraction of shape features from leaf images. A classifier named as an Artificial Neural Network (ANN) is trained to identify the exact leaf class. It is done to attain high efficiency with less computational complexity. This work has been tested for the accuracy of network with different combination of image features. The results are tested on 80 leaf images. It is revealed that this method give 98.8% accuracy with a minimum of seven input features. This approach is more promising for leaf identification systems that have minimum input and demand less computation time. Neural Network also developed and trained to classify and recognize plants from leaf images. This work is implemented using the image processing and neural network toolboxes in MATLAB.

Keywords: Image Processing, Leaf Classification, Shape Features, Artificial Neural Network

I. Introduction

Plant is one of the most important forms of life on earth. Plants control the balance of carbon dioxide and oxygen of earth's atmosphere. The relationship between human beings and plants are also very close. In addition, plants are important means of circumstances and production of human beings [1]. Regrettably, the amazing development of human civilization has disturbed this balance to a greater extent than realized. It is one of the biggest duties of human beings to save the plants from various dangers. So, the diverseness of the plant community should be restore and put everything back to balance.

A computer based plant identification system can be used. Different characteristics of the plants are used, starting at very simple level such as color and shape of the leaf to very complex such as tissue and cell structure. Now-a-days mobile phones can take very high quality images, as they are integrated with digital cameras. This makes the use of such a system even wider. Plant leaves are two dimensional in nature. They hold important features which are useful for classification of plant species. There are set of appropriate numerical attributes of features to be taken out from the object of interest. This is done for the purpose of classification.

Research has received considerable attention in recent years [2, 3] on the utilization of moments for object characterization in in-variant tasks and non-invariant tasks. The mathematical concept of moments has been around for many years. It has been used in various fields varying from mechanics and statistics to pattern recognition and image understanding. This paper focuses on automation through computer vision. Computer vision [4] is concerned with the theory behind artificial systems that extract information from images. This work was done to extracts different features from a leaf image and classifies different classes of leaves based on the extracted features. Furthermore, the system used the results of the classification scheme in identifying the class of new leaf images.

II. Related Works

Jyotismita Chaki et al. [5] worked on a complete plant leaf recognition system which had two stages. The first stage was the extraction of leaf features. Moment invariants technique was often used as features for shape recognition and classification. The second stage involved classification of plant leaf images based on the derived feature obtained in the previous stage. Classification was performed by comparing descriptors of the unknown object. The features were given as input to the neural network which was based on supervised learning algorithm having multilayer preceptor with feed-forward back propagation architecture. The result of this ranged from 90-94% accuracy having mean square error in the range of 0-0.9.

Stephen Gang Wu et al. [6] worked using PNN (Probabilistic Neural Network) for classification of leaf images based on 12 leaf features. A principal component analysis was used for reducing the 12 dimensions into five dimensions for faster processing. The 12 features used were physiological length, physiological width, leaf area, and leaf perimeter, smooth factor, aspect ratio, form factor, rectangularity, narrow factor, perimeter ratio of diameter, perimeter ratio of physiological length and physiological width, and vein features. If a Standard size

image was not used then the features of the same leaf vary with different sizes images. Since, it was the same leaf, the value of all the features were expected to be same. The values for these features vary for different scaled versions of the same leaf image. In this research, features used were scaling, translation and rotation invariant.

Y. Nam et al. [7] studied a shape-based leaf image retrieval system using two novel approaches. These were named as, an improved maximum power point algorithm and a revised dynamic matching method. In his approach, they used a region-based shape representation technique to define the shapes of the leaf. Certain points of interest were evaluated and the distances between the points were taken as the features for image matching.

Bivashrestha et al. [8] studied various features related to the shape of the leaves for leaf image-based plant classification. The features selected in this automated plant identification were isoperimetric quotient, convex hull ratio, aspect ratio and eccentricity. These features were vital in identifying leaf shapes, number of lobes and margin types. Leaves of nine classes were used and a classifier was implemented and tested using 105 leaves of 38 different species. K-means clustering was used to classify 105 leaf images into nine clusters.

R. Janani et al. [9] proposed a method for the extraction of shape, color and texture features from leaf images and training an artificial neural network (ANN) classifier to identify the exact leaf class. The key issue lies in the selection of proper image input features to attain high efficiency with less computational complexity. They tested the accuracy of network with different combination of image features. The tested result on 63 leaf images revealed that their method gave the accuracy of 94.4% with a minimum of eight input features. Image processing and neural network toolboxes were used in MATLAB to implement the work.

In 2000, Bivashrestha et al. [8] used K-mean clustering on 105 leaves and less features were used. In 2005, Y. Nam et al. [7] used region based shape representation technique for leaf identification. In 2007, Stephen Gang Wu et al. [6] did the work on leaf images with PNN with some extracting features. The size of leaf image was almost same. So the features were almost same. In 2013, R. Janani [9] et al. used ANN classifier using only 63 leaf images and with accuracy 94.4%.

In this paper, the salient features of leaves were studied. Recognition was studied to develop an approach that produced the better classification algorithm. This work was done on leaf identification using Artificial Neural Network by first extracting the shape features. It was based on local region leaves and was tried to get better accuracy results as compared to previous works.

III. Methodology

The software implementation of the project has been done using MATLAB (MATRIX LABORATORY), software developed by Math Works (www.mathworks.com) in USA. Proposed algorithm is applied on different datasets collected by digital camera. In order to verify the algorithm, it is implemented on variety of leaves. All images used for the experimental results are shown below in fig. 1. Images are chosen with different lightning conditions, shapes and using different locations of the camera from the objects.



i.

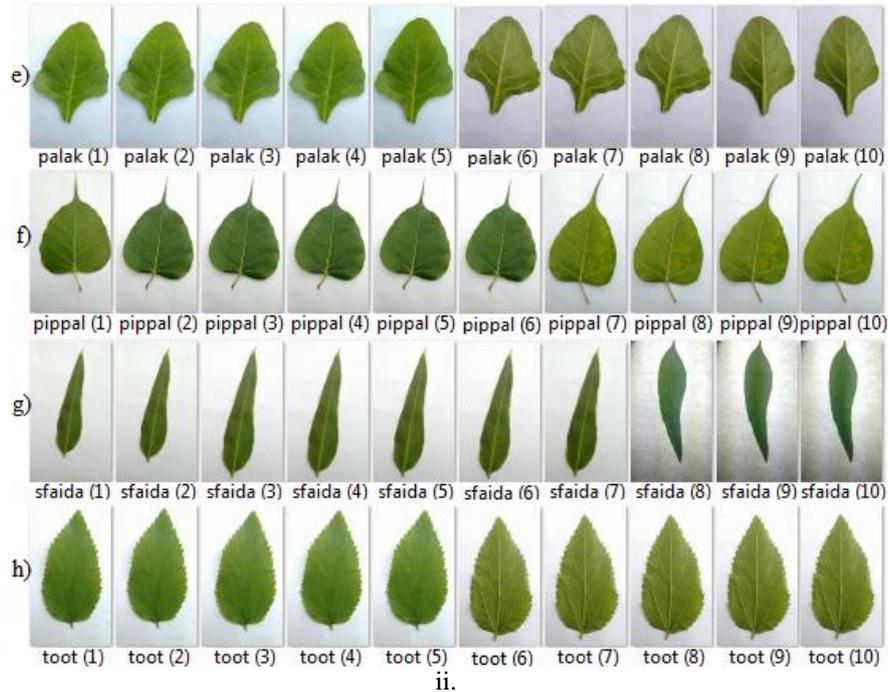


Fig. 1: All training dataset (i. and ii.); ten samples of each leaf category; a) Amrood, b) Jaman, c) Kathal, d) Nimbu, e) Palak, f) Pippal, g) Sfaida and h) Toot

The proposed method has four major steps as shown in fig. 2 and is explained below:

- 1.1 Image Acquisition and Preprocessing
- 1.2 Segmentation
- 1.3 Feature Extraction
- 1.4 Classification and Recognition

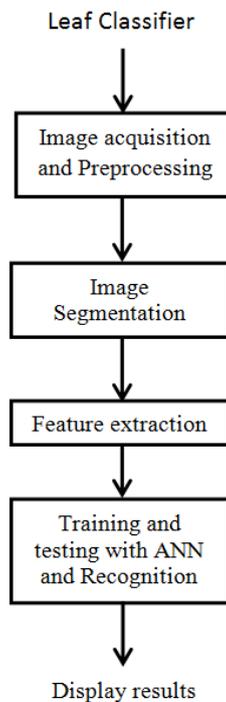


Fig. 2: Flow chart of proposed method for identification of leaf images

3.1 Image Acquisition and Preprocessing

Leaf images are collected from variety of plants with a digital camera. In this work, 8 species of different plants are taken. Each category of plant included 10 samples of leaf images. These leaf images come in with different of size, shape and class. Pre-processing contains sequential operations. It included prescribing the image size, conversion of gray-scale images to binary images. In addition, modifying the scaling and rotation factors of the image and de-noising the image using wavelets are also included.

3.2 Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments. It includes image enhancement, extraction of leaves from background and aligning the leaf image horizontally. The aim of image enhancement [11] is to improve the interpretability or perception of information in images for human viewers, or to provide better input for other automated image processing techniques. These enhancement operations are performed in order to modify the image brightness, contrast or the distribution of the grey levels. After this the leaf image is extracted from background. Then the boundaries are detected around object region using active contours. This is done after applying wavelets. It is aligned according to center of mass of the leaf after getting the leaf object. Then alignment is done with respect to x-axis. The imrotate command is used to rotate the image with respect to horizontal. After this step, leaf is ready for feature extraction process.

3.3 Feature Extraction

Feature extraction is used to extract relevant features for recognition of plant leaves. The redundancy is removed from the image and the leaf images are represented by a set of numerical features. These features are used to classify the data. The following features are used for doing leaves classification:-

- Extraction of Eccentricity: Region-based method is used to estimate the best fitting ellipse for the extraction of eccentricity. fig. 3 illustrates a best fitting ellipse to a polygon. It also showed the major axis, M_a and the minor axis, M_i of the best fitting ellipse.

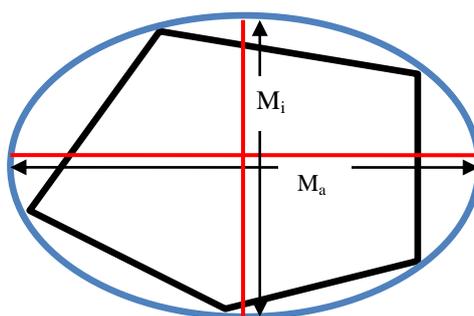


Fig. 3: The best fitting ellipse to a polygon with its major and minor axes; M_i i.e. minor axis and M_a i.e. major axis are shown by red lines

- Extraction of Aspect Ratio: For a rectangular shape, aspect ratio is defined as the ratio of its breadth by length. For leaves having irregular shapes, the aspect ratio is defined as the ratio of the minor axis to the major axis of the best fitting ellipse.
- Leaf Area: The value of leaf area is evaluated by counting the number of pixels having binary value 1 on smoothed leaf image. It is denoted as A.
- Leaf Perimeter: Leaf perimeter is defined as distance around the boundary of the region. It is evaluated by counting the number of pixels containing leaf margin. It is denoted by P.
- Area ratio of perimeter: Ratio of Area to perimeter is defined as the ratio of leaf Area A and leaf perimeter P, is calculated by A/P .
- Length of the major and minor axes: Area of the ellipse that just encloses the region.
- Solidity: Scalar specifying the proportion of the pixels in the convex hull that are also in the region computed as $\text{Area}/\text{Convex Area}$. This property is supported only for 2-D input label matrices.

3.4 Classification using ANN

Artificial Neural Network (ANN) [10] has been the successfully used classifier in numerous fields. It can be modeled on a human brain. The basic processing unit of brain is neuron which works identically in ANN. The neural network is formed by a set of neurons interconnected with each other through the synaptic weights. It was used to acquire knowledge in the learning phase. The number of neurons and synaptic weights can be changed according to desired design perspective. The basic neural network consists of 3 layers.

- 1) Input layer
- 2) Hidden layer
- 3) Output layer

The input layer consists of source nodes. This layer captures the features pattern for classification. The number of nodes in this layer depends upon the dimension of feature vector used at the input. The hidden layer lies between the input and output layer. The number of hidden layers can be one or more. Each hidden layer has a specific number of nodes (neurons) called as hidden nodes or hidden neurons. The hidden nodes can be varying to get the desired performance. These hidden neurons play a significant role in performing higher order computations. The output of this layer is supplied to the next layer. The output layer is the end layer of neural network. It results the output after features are passed through neural network. The set of outputs in output layer decides the overall response of the neural network for a supplied input features.

IV. Results and Discussion

Proposed algorithm is applied on different datasets collected by digital camera. In order to verify the algorithm, it is implemented on variety of leaves. Images are chosen with different lightning conditions.

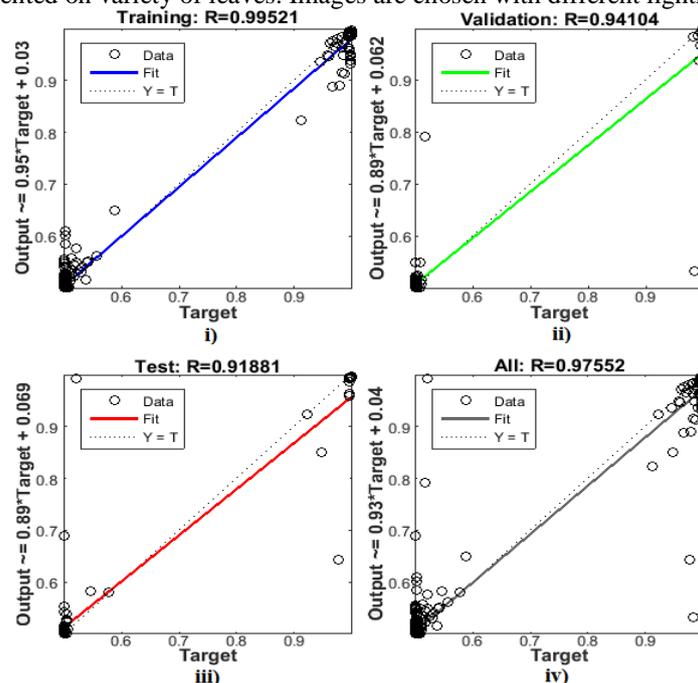


Fig. 4: Neural Network Regression Rate; the graphs of output class vs. target characteristics of i) Training, ii) Validation, iii) Testing and iv) Overall phase. Circles indicate data used in respective phases; colored line indicates the fit line; dotted line is drawn where output(Y) is equal to (T).

Fig. 4 indicates the regression graphs for training, validation, testing and overall phases. In these graphs, the data trend analysis is checked whether it (best fit line) is proper or not. It is determined by coefficient of determinant R. The blue, green, red and dark line is indicated to fit line of training, validation, testing and overall regression graph. In training graph it is shown that how much result is trained correctly. The linear dotted line is denoted for output-input relation, where it is explained that output equals to input. The circles are for trained data samples. Similarly, it is for the graphs of validation, testing and overall graphs. The value of R in training phase is 0.99521 which means that it is fitted and classified in a correct manner. In the case of validation, R is 0.94104 which is an acceptable value for validation. The value of R in the case of testing is 0.91881. It means testing is done properly. It is shown that when all samples are considered after all three training, validation and testing phases, the value of R is 0.97552. This means data has been fitted and classified in a correct manner which is the ideal condition for goodness of classifier.

The objective of Neural Network is to run simulation until it reached the minimum possible mean square error. There is not much variation in mean square error by simulating the conditions in each epoch. It is more or less steady graph but as the epoch increased in each phase shown by red, green and blue lines the value reached to best validation value of 0.003035 which is very close to zero in ideal situation.

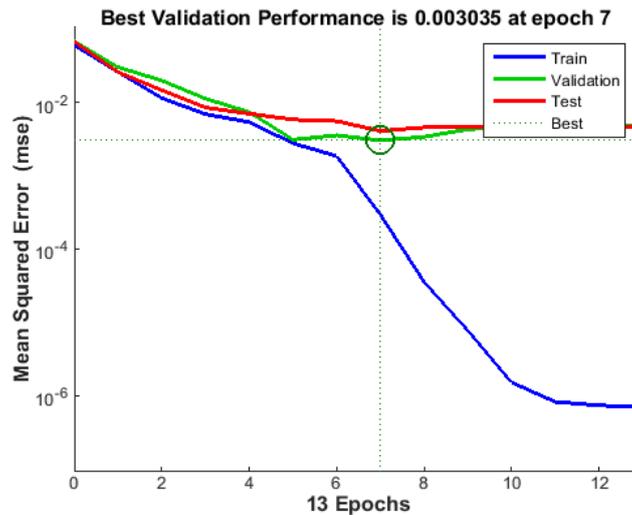


Fig. 5: Performance Graph: the graph of mean square error vs. epochs; blue line indicates the performance of trained data, green for validation data, red for test data. Best Validation Performance is 0.003035 at epoch 7.

The recognition system is implemented using MATLAB. The image is taken as dataset input and feed-forward architecture is used. Firstly neural network is trained using known dataset of different plant leaves. After training, system is tested using several unknown dataset and result is obtained. The desired performance goal is achieved in 7 epochs. Fig. 5 shows performance graph. It specifies the gradient at epoch 7, validation checks and also learning rate. The recognition rate is defined as ratio of number of leaves identified to total number of leaves, i.e.

$$\text{Recognition rate} = \frac{\text{Number of leaves identified correctly}}{\text{Total number of leaves}} * 100$$

An input vector is created by placing first 60 values out of 80 of segmented leaf images. 8 leaf images are selected for validation. For testing phase, 12 images are taken randomly.

The following parameters are used for creating network for training:

No. of neurons in input layer: 10

No. of layers: 2

No. of epochs: 13

Transfer Function: “transig”

Adaptive Learning Function: “LEARNGDM”

Validation checks = 6

Performance Function: “MSE”

| Output Class \ Target Class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Accuracy |
|-----------------------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|
| 1 | 9 11.3% | 0 0.0% | 100% 0.0% |
| 2 | 0 0.0% | 10 12.5% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 3 | 0 0.0% | 0 0.0% | 10 12.5% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 4 | 1 1.3% | 0 0.0% | 0 0.0% | 10 12.5% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 90.9% 9.1% |
| 5 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 10 12.5% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 6 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 10 12.5% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 7 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 10 12.5% | 0 0.0% | 100% 0.0% |
| 8 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 10 12.5% | 100% 0.0% |
| Overall | 90.0% 10.0% | 100% 0.0% | 98.8% 1.2% |

Fig. 6: Overall Confusion Matrix: matrix of output and target class; trained result gives accuracy of 98.8% where class 1 is one time misclassified as class 4 by 1.3%. Blue box indicates accuracy result.

Fig. 6 shows the summary of results of plant species classification for the 8 classes. The accuracy is improved and reaches to 98.8% which is a good amount as shown in blue box. Here, one exception is observed that classes 2, 3, 4, 5, 6, 7, 8 are correctly classified. But class 1 is one time misclassified as class 4 by 1.3%.

V. Conclusion

In this paper, method is implemented to identify the correct species of the leaf from eight different classes. The combination of shape feature is resulted in a correct leaf identification rate of 98.8% with a minimum of seven features. This is the better result than previous work done by R. Janani et al. [9]. They worked on less number (63) of leaf images and with accuracy of 94.4% only.

This work is performed on the regional leaves. In future it can be extended to a larger number of leaf classes with improved efficiency. More extraction features can be used using increased number of leaf dataset in future work.

Acknowledgement

I am greatly indebted to Gurjeet Kaur, who has graciously applied herself to the task of helping me with ample support and valuable suggestions. She has completed her dissertation work on character recognition using shape descriptor feature extraction method and gradient method using nntool for classification. Same work is carried in my dissertation on leaves classification using shape descriptor feature extraction method and nntool for classification.

References

- [1] Hang Zhang; Shansong Liang, "Plant species classification using leaf shape and texture" Industrial Control and Electronics Engineering, 2012 International conference, Page(s): 2025-2028, Year 2012.
- [2] M K HU, "Visual pattern recognition by moment invariants", IRE Trans. Info Theory, vol. 8, pp. 179-187, Year. 1962.
- [3] T. H. Reiss, "The Revised Fundamental Theorem of Moment Invariants", IEEE Transaction Pattern Anal. Machine Intelligence, Vol. 13, No. 8, Page 830-834, Year. 1991.
- [4] J. Du, D. Huang, X. Wang, and X. Gu, "Computer-aided plant species identification (CAPSI) based on leaf shape matching technique," Transactions of the Institute of Measurement and Control. Vol. 28, No.3 Page. 275-284, Year 2006.
- [5] Jyotismita Chaki, Ranjan Parekh, "Plant Leaf Recognition using Shape based Features and Network classifiers" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, Year 2011.
- [6] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang and Qiao-Liang Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", arXiv:0707.4289v1 [cs.AI], Year 2007.
- [7] Y. Nam, "Elis: An efficient leaf image retrieval system", in Proc. Advances in Pattern Recognition Int. Conf., Kolkata, India. Year 2005
- [8] Bivashrestha, "Classification of plants using images of their leaves" ICPR, vol. 2, pp. 2507-2510, Year 2000.
- [9] R. Janani, A. Gopal, "Identification of selected medicinal plant leaves using image features and ANN", Advanced Electronic Systems (ICAES), 2013 International Conference, Page(s): 238 – 242, Year 2013.
- [10] N. K Bose, P. Liang, "Neural network fundamentals with graphs, algorithms and applications", Year 1994.
- [11] R. E. Gonzalez and R. E. Wood, "Digital Image Processing" Addison-Wesley, Year 1993.