Recent Advances in Technology and Engineering (RATE-2017)
6th National Conference by TJIT, Bangalore

International Journal of Science, Engineering and Technology
An Open Access Journal

Pattern Recognition and Classification of Indian Herbal Leaves

[1]Kashif Akhtar, [2]Kiran Singri, [3]Divyashree P, [4]Monika Banerjee, [5]Roopashree S

Abstract
In the view of building awareness towards herbal remedies which has no side effects and is cost effective when compared to modern drugs. Ayurveda (Traditional Indian Medicine) plays a vital role in the herbal medicines. Herbal medicines are manufactured through extensive usage of medicinal plants. It's important to conserve the plant habitats to access these medicinal plants. We build the leaf dataset consisting of medicinal leaves about different plants and develop automated plant identification tools through aid of Computer Vision, Pattern recognition and Image processing methods. The proposed image recognition system process the input image converted into suitable format for feature extraction and classification of the leaf image. Key point descriptor features are extracted using Scale Invariant Feature Transform algorithm. Classifiers such as Binary and Multi class Support Vector Machine and Probabilistic Neural Network are used to categorize the leaf image. The multi class SVM emerges as the winner in providing the best classification accuracy. This classification system achieves more than 94% accuracy, and utilizes limited computational time to perform this task. Furthermore, this system will aid in the conservation of ecology.
Keywords: Herbs, Ayurveda, Feature Extraction, Classifier, SIFT, SVM, Multi SVM, PNN, Image Processing.

1. Introduction

In comparison with modern medicines, floral medicines possess the ability to upgrade common physical condition at cheap rates which will result in the reduction of unwanted reactions. There is a growing scientific concord that plant habitats have been altered and species are disappearing at rates which were never witnessed before. Various plant varieties of data are required to correct this situation, so that for future generation they could be monitored and protected. Handbook or expert is needed to identify valued of such plants.

But, there is a scarcity of botanists and the latter process is very difficult to operate and time consuming. So, to retrieve faster information there is a need to automate the process. Computer vision based image processing is one of the emerging techniques to computerize such tasks by imitating the human visual system. SVM is built originally for binary classification by providing either -1 or +1 values to the variables. SVM was further extended for more than 2 class classifiers which is termed as multi class SVM. Many models were invented.

This paper overcomes the disadvantages of Binary SVM by using two multi class classifiers and hence gives the graphical comparison between multi SVM and PNN. Using these multi class classifiers we can also depict the class to which particular specie belongs.

Bayesian classification is used in probabilistic neural network (PNN). The approximation of probability density function (PDF) is also used in the above classifier. PNN was invented by Donald F. Specht in 1990. PNN is a successful device for deciphering numerous classification problems. PNN can compute non-linear decision boundaries

2. Existing System

Existing system can be used to recognize plants based on the features extracted from their leaves to specify it is herbal or non herbal. The feature extraction mechanism used Scale Invariant Feature Transform (SIFT) and method for classification was Support Vector Machine (SVM). Images were partitioned into training and testing sets where 70% were used for training and 30% for testing purpose.

Leaf features extracted were used for storing incase of training images and compared incase of test images. Accuracy was achieved by extracting large number of features for an approximate match. Classification using SVM was done to categorize leaves as herbal or non herbal class. This model is robust and it models even non-linear relations [1]. Support Vector Machines (SVMs) uses a machine learning task of deducing a function from labeled training data. It uses learning algorithms that examine data used for classification analysis [3].

3. Proposed System

Szedmak et al. [8] proposed a multi class design which uses One-versus-rest framework in their process of formulation. Chawla et al [9] along with Tanget al and Wang and Shen have proposed an idea that overcomes the problems of the One-versus-rest framework model of classifying leaves into different species. We use various data modeling techniques that include SIFT, SVM multiclass and PNN for the classification of leaf images to recognize herbal plant species [2]. Features from different layers will be given as an input to traditional classifiers for a more robust recognition [2]. Few of the characteristics which we will extract include optical character recognition [4], intrusion detection [5], speech recognition [6], and bioinformatics [7]. We will discriminate leaf species for more than 10 categories. Those species comprise curry, neem, tulsi, beetel, amaranth, basil, henna, Indian borage, paijatha and Vincarosea.

4. Methodology

System is arranged to perform prettified functionality at every stage in the program. It includes:

1. Acquisition
2. Pre-processing
3. Extraction using SIFT
4. Classification using SVM and PNN
5. Display Result

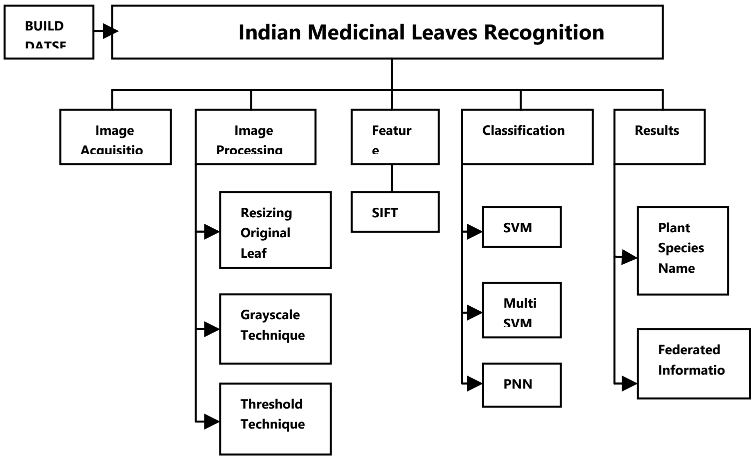

Figure 1: System Architecture

The medicinal plants are collected and captured using 16.0 megapixel cameras and stored in Matlab in the form of JPEG format. In this technique, we will be using a computer algorithm to process the leaf images. The scale invariant feature transform (SIFT) method is used for the feature extraction in leaves. The next step is classification of herbal leaves using SVM and PNN classifiers. Depending on the results, the output is shown which results in the comparison of accuracy between the two classifiers.

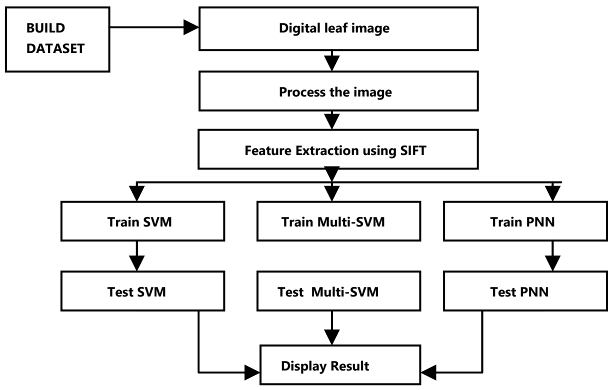

Figure 2: System flow diagram

5. FEATURE EXTRACTION USING SIFT

David Lowe[10] preferred Scale-Invariant Feature Transform (SIFT) to discover and report the characteristics of images. SIFT keypoints of items are initially taken out from a source and saved in the dataset. Features matching are found using euclidean distance of their feature vectors which are consistent to any modifications in illumination or affine transformations such as uniform scaling and rotation. Co-linearity relation between the keypoints and distance ratio along a line are maintained in affine transformations.

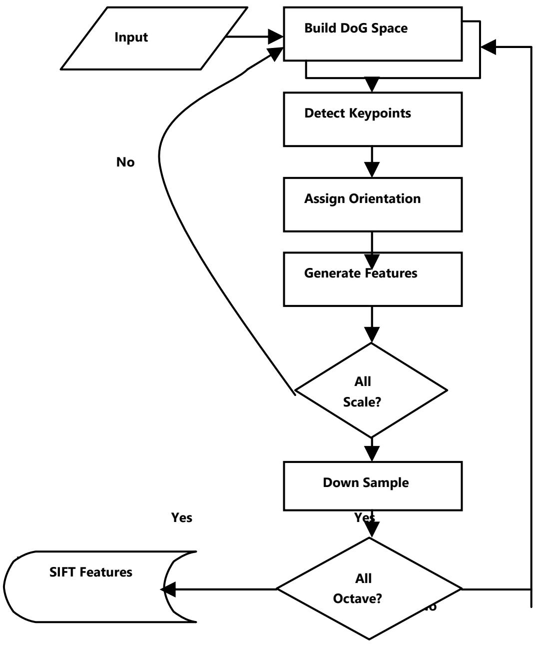

Figure 3: Steps in SIFT process

This method is simpler, robust and accurate in the process of extracting features compared to texture analysis. The features are the key points that can be stored and images can be compared against it. SIFT can be classified into the following four stages:

1. **Scale-space extrema detection**:-It involves finding of all scales in distinct views of the input image and in different location. Continuous function of scale is used to find stable features which are called as scale space. Mathematically, it can be executed using Difference-of-Gaussian, result is point of interests that remain constant to rotation and scaling transformations.

2. **Key point localization and filtering**:-For every candidate location there is a detailed model which is secured that determines scale and location of it. Selection of a key-point is found on the measure of its stability, thus removing points on the edge which are not stable.

3. **Orientation assignment**:- Based on gradient directions of the local image, further orientation is assigned to each key point called beans to make it invariable to rotation location. And all of the future operations to be executed on the data are done using this image.

4. **Key point descriptor**:-Initially orientation histograms are constituted on 4x4 pixel area for every 8 bins. Then histograms are computed by considering their magnitude and orientation values of samples in a 16 x 16 area around the keypoint. Matching accuracy is also calculated across changing number of keypoints from the dataset.

6. CLASSIFICATION USING SVM

Support vector machine (SVM) is a managed machine learning technique among associated learning algorithms in order to examine data used for classification and regression analysis. The SVM proceeds with statistical learning assumption where the learning complication was developed as problem of quadratic optimization. The plane had global optimum and was free of local minimum. The objective here is to set a Hyperplane or an optimal separating hyper plane (OSH) that separates two sets of data (+1 and -1).

Detecting OSH by SVM is done by augmenting the margin between those two classes. The input data is transformed by using kernel functions and then by constructing OSH in the given transformation space. Data vectors which are adjacent to the line constructed are called the Support Vectors. SVM then classifies data where the partition is linear, into the high dimensional feature space. Input vectors are located by using a non-linear transformation that is based on a regularization factor.

A hyperplane can be accepted best for any SVM classifier if it has the peak margin amid the two classes. Here, the border is the maximum width of slab that is parallel to the hyper planes which there are no inner data points.

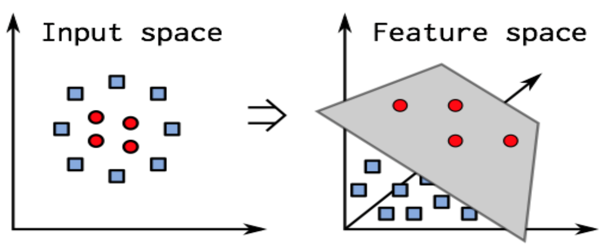

Figure 4: Support Vector Machine classifier

Steps in SVM classification are similar to any supervised learning model. First step is to train SVM, and then validate the SVM classifier. We use the trained system to classify new data. Also, to obtain better results, we can use any SVM kernel function by tuning the parameter to match the particular kernel function, as in below steps:

- ✓ Training using SVM classifier.
- ✓ Testing of new data using SVM classifier.
- ✓ Tuning of an SVM classifier as required.

The inner product (a, b) is retrieved by using kernel function K (a, b) to create a nonlinear support vector classifier.

$$f(a) = sgn(\sum \alpha_i y_i K(a_i, a) + y)$$

Where, f (a) is a function that determines the membership of a, i = 1, 2... n. The desired class was given as -1 and the others as +1. There are two layers to the functioning of SVM. First of which selects the basis K (ai, a) during the learning process, from the agreed kernels. And the next layer is responsible for the construction of space linear function. It is thus comparable of knowing an optimal hyperplane in the given feature space. Support Vector machines can thus be used to build a range of learning machines that employs dissimilar kernel function.

7. CLASSIFICATION USING MULTI-CLASS SVM

Multi class Support Vector Machine is a classifier which has more than two classes which is used in many real world problems. One of those are classification of instances into more than one class in addition to the Binary classification. These are essentially two-class classifiers. Two or more binary classifiers are used to form a multi class classifier. There can be several approaches to multi class SVM. First by building and merging various classifiers that consist of two levels. Second by straightaway taking into consideration each statistic in one optimization articulation. Dealing with multi class problems are way more expensive when compared to the binary class problems with the same data set. Till now, the experiments are limited to dataset comprising ten species. The figure below shows the One-vs-all process of classifying problems in PNN. It constructs many SVM models let's say n where n is the number of classes. Then the n classes are given as an input to the SVM training algorithm. Second method is the one-against-one method which is depicted in the diagram below. This builds n(n−1)/2classifiers where each classifier is trained on data values from two classes.

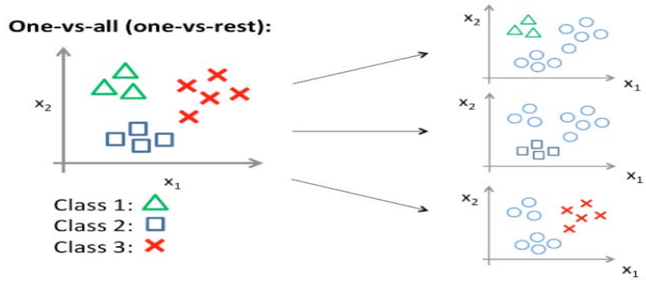

Fig. 5: One-vs-rest method

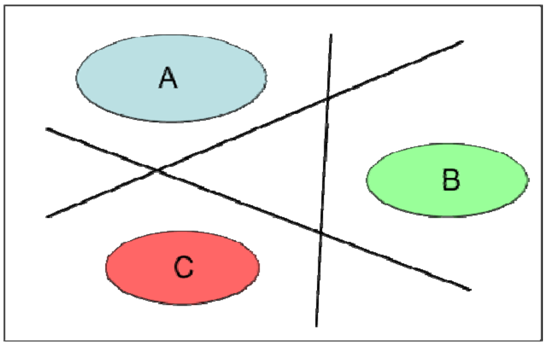

Fig. 6: One-across-one method

SVM is efficient as compared to PNN because it uses a convex optimization method to solve problems which results in faster computation of results.

8. CLASSIFICATION USING PROBABILISTIC NEURAL NETWORK (PNN)

Probabilistic neural network (PNN) was first introduced by D.F Specht in the 1966. PNN is also known as feed forward neural network. This algorithm is broadly used for classification and pattern recognition in image processing. In view of this, parent probability distribution function (PDF) is approximated through a non-parametric function and a Parzen window. Subsequently a new input data is calculated using PDF of every class. The highest posterior probability for new input data is allocated by utilizing Baye's rule. PNN operations are classified into four layers namely Input layer, Hidden layer, Pattern layer/Summation layer and output layer.

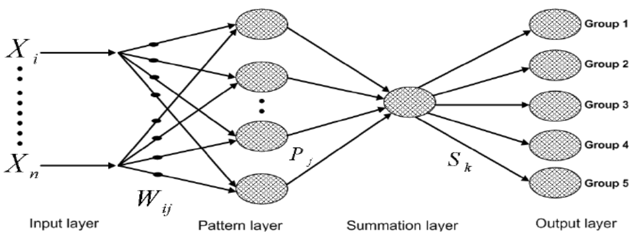

Fig. 7: PNN Structure

- ✓ Input layer: In this layer it represents a seer variable where N-1 neurons are utilized for N number of categories. And values for every neuron in hidden layer are fed by input neurons

- ✓ Pattern layer: Stores the values of the seer variables with the destination value. For the given test case a hidden neuron calculates the Euclidean distance from the neurons center point and then kernel function is applied by utilizing the sigma values.

- ✓ Summation layer: The category of interest for every training case is stored along with each hidden neuron. The values coming out of a hidden neuron are then fed to the pattern neuron. And then the values for the class are added by the pattern neurons they represent.

- ✓ Output layer: Category accumulated in the pattern layer uses the largest vote to estimate the aimed category by comparing the weighted votes.

PNN is slower at classifying new cases and consumes more memory space.

9. IMPLEMENTATION

Algorithm for Training

Step 1: Accept herbal metadata and non-herbal metadata

Step 2: Find number of feature vector present in herbal metadata, p

Step 3: Find number of feature vector present in non-herbal metadata, q

Step 4: Get the minimum of both vectors, x = min (p, q)

Step 5: for i from 1 to x do

5.1: Select i^th feature vector from herbal metadata and assign to m

5.2: Select i^th feature vector from non-herbal metadata and assign to n

5.3: Merge m and n, assign to f

5.4: Use SVM training method and pass f with associated label

Step 6: Display appropriate message

Algorithm for Testing

Step 1: Take the input from the user, the input is an image of plant leaf in a white background.

Step 2: Next step is to take the input image for processing. The pre-processing part resizes the image as well as converts it into grey scale.

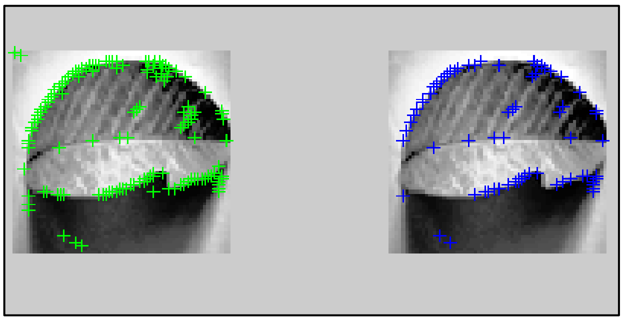

Fig. 8: Preprocessing

Step 3: The pre-processed image is then passed to the extraction module.

Step 4: SIFT extraction module performs a series of steps to extract set of features called key points. These key points are locations in the image space that remains invariant to rotation, scaling, illumination and noise.

Step 5: These key points are passed as a descriptor to the next module.

Step 6: Key points descriptor along with the kernel function, parameters and stored features are used by SVM and PNN classifier to elegantly classify so as to predict the features to be of either herbal or non-herbal species and also specifying specie's name.

9. Results

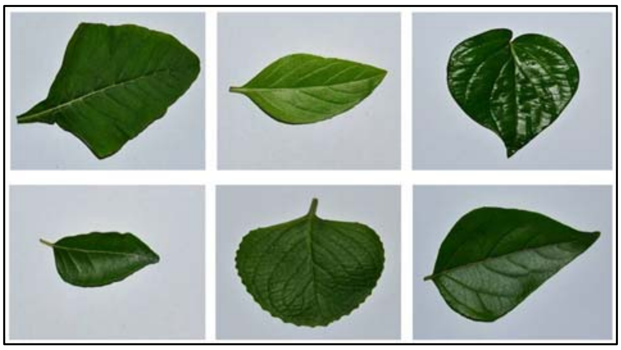

Fig. 9: Sample Leaves – Data Set

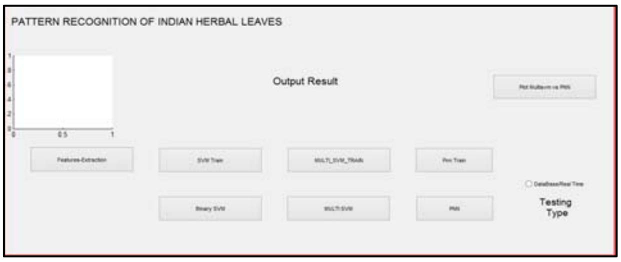

Fig. 10: User Interface

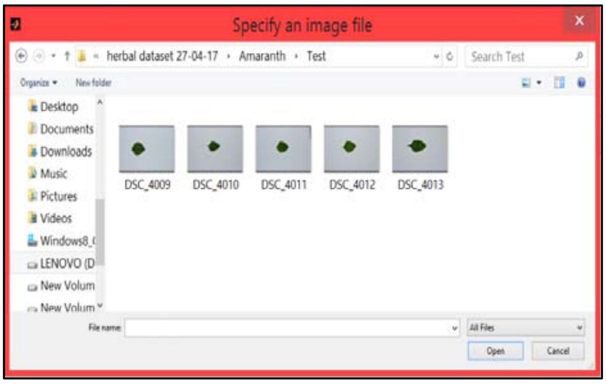

Fig. 11: Selecting an image

Fig. 12: Classification Result

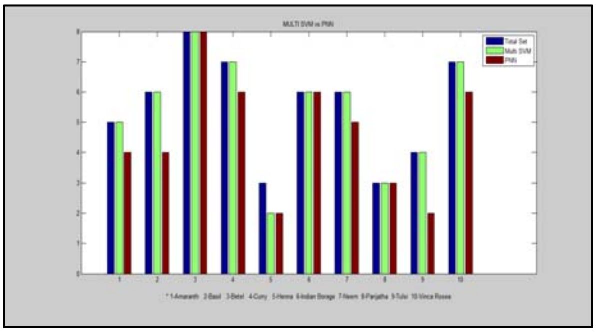

Fig. 13: SVM and PNN Comparision

10. Conclusion

The project builds an automated system for herbal medicines. Since manual identification is very time consuming, the pattern recognition and Image processing techniques have led to build the system which is efficient. The multi class SVM emerges as the winner among all classifiers with 94% accuracy. The system helps in safeguarding the Indian traditional medicines (Ayurveda). The system works real time with very good accuracy rate. Future enhancements are:

1) The dataset can be enlarged and can be used to classify a large variety of herbal leaves and their classes.

2) The work could also be brought into mobile platforms such as Android and iOS as such it can be portability and reliability.

References

[1]. Soumitro Swapan Auddy, Kanishka Tyagi, Son Nguyen and Michael Manry ,"Discriminant Vector Tranformations in Neural Network Classifiers", International Joint Conference on Neural Networks (IJCNN), 2016.

[2]. Chaoyun Zhang, Pan Zhou, Chenghua Li, Lijun Liu "A Convolutional Neural Network for Leaves Recognition Using Data Augmentation IEEE International Conference on Computer and Information Technology", 2015.

[3]. Sapna Sharma , Dr. Chitvan Gupta, "Recognition of Plant Species based on leaf images using Multilayer Feed Forward Neural Network", International Journal of Innovative Research in Advanced Engineering (IJIRAE) Issue 6, Volume 2, June 2015.

[4]. E. Sandeep Kumar and Viswanath Talasila, "Leaf features based approach for automated identification of medicinal plants", Communications and Signal Processing (ICCSP), 2014 International Conference, April 2014.

[5]. T. Sathwik, R. Yasaswini, Roshini Venkatesh and A. Gopal, "Classification of selected medicinal plant leaves using texture analysis " Computing, Communications and Networking Technologies (ICCCNT), Fourth International Conference, July 2014.

[6]. C H Arun, W R Sam Emmanuel and D. Christopher Durairaj, "Texture Feature Extraction for Identification of Medicinal Plants and Comparison of Different Classifiers", International Journal of Computer Applications (0975 - 8887) Volume 62 - No. 12, January 2013.

[7]. Jyotismita Chaki and Ranjan Parekh, "Plant Leaf Recognition using Gabor Filte", International Journal of Computer Applications (0975 – 8887) Volume 56 – No.10, October 2012.

[8]. Liwen Gao, Xiaohua Lin, Mi Zhong, Junmin Zeng, "A neural network classifier based on prior evolution and iterative approximation used for leaf recognition" Natural Computation (ICNC), 2010 Sixth International Conference, August 2010.

[9]. Krishna Singh, Indhra Gupta and Sangeeta Gupta, "SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape", International Journal of Signal Processing, Image Processing and Pattern Recognition rocessing and Pattern Recognition Vol. 3, No. 4, December, 2010.

[10]. David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", Computer Science Department University of British Columbia Vancouver, B.C., Canada, January 5, 2004.

Author's details

1. Mr. Kashif Akhtar, student, Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, Karnataka, India, akhtar.kashif95@gmail.com

2. Ms. Kiran Singri, student, Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, Karnataka, India, kiransingri1@gmail.com

3. Ms. Divyashree P, student, Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, Karnataka, India, divyahemade@gmail.com

4. Ms. Monika Banerjee, student, Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, Karnataka, India,monikabanerjee44@gmail.com

5. Ms. Roopashree S, Assistant Professor, Department of Computer Science and Engineering, T. John Institute of Technology,Bengaluru,Karnataka,India,roopashaily@gmail.com