International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

Identification of Ayurvedic Medicinal Plants by Image Processing of leaf samples

Tejas D. Dahigaonkar[1], Rasika T. Kalyane[2]

[1,2] Department of Computer Science and Engineering, SGGSIE&T, Nanded

---***---

Abstract - *Ayurveda is being practiced in India since the Vedic times for the well-being of the people. It is one of the oldest systems of Medicinal science that is used even today. Being used extensively in medicine, proper identification of the appropriate medicinal plants used for the manufacture of medicine is very important in Ayurvedic medicinal industry. Identification of the appropriate medicinal plants is usually done by the skilled practitioners who have achieved expertise in this field. The knowledge of identifying these medicinal plants is being passed from generation to generation. However, as the identification is completely based on human perception, there can be scope for error or misjudgment. Incorrect or substituted raw materials make the medicine ineffective. It can even result in certain side-effects too. This makes it necessary for an efficient system to be developed that can properly identify the Ayurvedic plant based on the leaf samples.*

The main features required to identify a medicinal plant are shape, color and leaf texture. Color and texture of the leaf contain vital parameters that can uniquely identify to which species a certain plant belongs. This will eventually improve the effectiveness of the medicine. This project focuses on creating an intelligent system that can identify the plant species based on the input leaf sample. This project explores feature vectors of a green leaf and obtains a unique set of feature parameters responsible to maximize the accuracy of identification. The classification of the leaf samples is based on this unique feature set. Accuracy up to 96.66% has been achieved when the samples were tested using Support Vector Machine.

1.INTRODUCTION

Ayurveda is an ancient medicinal system that is practiced in India and has its origins in the Vedic times, approximately 5000 years ago. Ayurveda is considered to be the oldest healing science. In Sanskrit language, Ayurveda means "The Science of Life". In spite of being suppressed during years of foreign occupation, Ayurveda is being widely used in its native land and throughout the world.

The main ingredients of ayurvedic medicines are plant leaves, root, bark, fruits, seeds etc. It is said that about 8000 plants of Indian origin are known to possess medicinal attributes. The fundamentals on which this Ayurvedic system is built upon are always true for all ages and can be easily adopted generation after generation. This, in turn, makes Ayurveda one of the very few medicinal systems that were developed in ancient times and is still implemented to date. Ancient Tibetan and Traditional Chinese Medicine and Early Greek medicine have accepted many concepts that were described in the ancient Ayurvedic literature dating back to several thousands of years. This widespread and extensive use in the field of medicine makes Ayurveda the 'mother of healing'.These Ayurvedic plants are used for preparing medicines on a commercial basis. This has resulted in the production and marketing of Ayurvedic medicines to become a thriving industry with its annual turnover exceeding Rs 4000 crores. Because of this, the number of licensed Ayurvedic medicine manufacturers in India now exceeds 8500.

Because of this increased commercialization of Ayurvedic sector, several issues regarding the raw material quality used for their preparation need to be focused. These plants are usually collected by tribal masses that are not professionally trained in the work of identifying the correct plants. Even the manufacturing units, at times, receive improper or substituted medicinal plants. Most of these manufacturing units do not have proper quality control mechanisms that can screen these plants. In addition to this, confusion due to different local names of these plants makes the matters worse.

Because of the usage of improper raw materials, the Ayurvedic medicine becomes inefficient. It may result in certain unpredictable side-effects too. As a result, an intelligent system needs to be developed that can properly identify the ayurvedic plant based on the leaf samples. This will, in turn, improve the quality of the medicine and will also maintain their credibility.

2. RELATED WORKS

This paper is inspired by the research paper written by Dr. Varun P. Gopi, Prof C.M. Surya and Prof Manojkumar P. Arur [1] of Government engineering College Wayanad, Kerala, India. They used different combinations of plant features like Zernike moments, texture, color, centroid-radius distance, HU Invariant moments and geometric features with different classification techniques like SVM, MLP, K-NN K-means clustering and so on.

T. Sathwik, R. Yasaswini [2] developed a plant identification method by analyzing the texture features only from the Gray Level Co-occurrence Matrix of the leaf sample. They used the least dissimilarity method for classification. The system had achieved an accuracy of 95%.

Nuril Aslina and Nursuriati Jamil [3] used Scale Invariant Feature Transform (SIFT) as a shape descriptor and Grid-Based Color Moments (GBCM) as color moments. The image was decomposed into HSV planes and every plane was divided into 9 grids. Color moments were computed for each grid and were used as a feature vector. The identification was done by calculating the least Euclidean distance between the train and test sets. An accuracy of 87.5% was obtained independently of scaling and rotation of images.

Parag Bhandarkar, Rizwan Ahmed [4] decomposed the morphology of leaf edges by using predefined structural elements and obtained a unique structural signature that quantifies the shape of the leaf. The identity was computed by using root mean square errors between the feature values of training and test samples. They achieved an accuracy of 67.5% due to a small dataset being used.

Itheri Yahiaoui, Olfa Mzoughi [5] used 5 geometrical features and a boundary descriptor called Directional Fragment Histogram for identification. All the results obtained were expressed using the Mean Average Precision (MAP). They achieved an accuracy percentage of 77.83%.

Pavan Kumar Mishra, Sanjay Kumar Maurya [6] used four geometric features i.e. solidity, convexity, circularity, and eccentricity along with three RGB color features in the form of red, green and blue indices for the experiment. They used a three-stage comparison of the obtained feature vectors for a faster rate of the identification procedure. They achieved an overall identification rate of 85%.

A. Gopal, S. Prudhveeswar Reddy [7] used four geometric features, HU's invariant moments and Polar Fourier transform coefficients as shape descriptors for identification. The color descriptors they employed were standard deviation, mean, kurtosis and skewness of the RGB planes. Least dissimilarity method was used for the identification process. Overall classification efficiency of 92% was achieved.

B. B.S. Harish, Aditi Hedge [8] used geometrical features like the form factor, aspect ratio, perimeter ratio of diameter, rectangularity, solidity, irregularity, and convexity and Zernike moments as the shape descriptor. The experiments were conducted on Flavia database and classified the images using SVM algorithm and achieved an overall accuracy of 89%.

3. PROCEDURE

3.1 Dataset

There is no standard dataset of Ayurvedic plant leaves was available for conducting the experiments. So a dummy dataset having 32 different species of plant leaf images were selected and was augmented to obtain a greater no of images. The images of individual leaves were of 1600×1200 resolution and were arranged in a way to separate sets of training and testing samples.

3.2 Proposed Work

The first step of training the data is preprocessing the input data one by one. The main purpose of pre-processing the data is to enhance the visual appearance of the image and to improve the manipulation of datasets. This is achieved by removing unwanted noise, reconstructing the image, enhancing its quality etc. The first step in pre-processing is by sharpening the RGB image. This is done by employing unsharp masking. This sharpening of the image improves the image appearance. Also, the edge or boundary points of an image are sharpened. The next step is Segmentation [9]. Image segmentation is the process where image is divided into multiple parts. This is mainly employed to extract relevant information from a digital image. Segmentation involves binarization. This binary image then passes through a morphological erosion and dilation process so that small imperfections like dots are removed. The largest component of the binary image is selected for determining the different morphological features of the image i.e. geometric features, shape, color features, texture analysis.

FEATURE EXTRACTION

- Geometric Features:

$$\text{Equivalent diameter} = \sqrt{\frac{4 * area}{\pi}}$$

Convex area specifies the total number of white pixels that are present in the Convex Image. A convex image is a binary image that specifies the smallest convex polygon that contains the region in which all the pixels are filled in within the polygon. Solidity can be calculated as

$$\text{Solidity} = \frac{Convex\ area}{Original\ area}$$

Eccentricity is a characteristic of a conic section. It can be calculated as

$$\text{Eccentricity} = \sqrt{1 - \left(\frac{Minor\ axis\ length^2}{Major\ axis\ length^2}\right)}$$

Extent is the ratio of the total number of pixels present in the region to the number of pixels present in the bounding box. It can be calculated by the formula.

$$\text{Extent} = \frac{No.of\ pixels\ in\ region}{No.of\ pixels\ in\ the\ bounding\ box}$$

- Color Feature:

Color is an important feature in the representation of an image. The fact that it is invariant to rotation, translation and scaling makes this feature even more significant. Color space, similarity measurement, and quantification are the key components involved in Color Feature Extraction. Color moments are mainly defined by mean and standard deviation.

$$\text{Mean } (\mu) = \sum_{i=0}^{M}\sum_{j=0}^{N}\frac{P(i,j)}{MN}$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\sum_{i=0}^{M}\sum_{j=0}^{N}\frac{(P(I,j)-\mu)^2}{MN}}$$

Where MN ⇒ image dimensions
P(i, j)⇒color value on column i and row j

- Texture Features:

A Gray Level Co-occurrence Matrix [2] is a square matrix that has an equal number of rows and columns as the number of gray levels present in the grayscale image. From the Gray Level Co-occurrence Matrix (GLCM) following features were calculated.

Correlation

$$CORR1 = \frac{H_{XY} - H_{XY1}}{\max\,[P_x\ (i)\log P_x(i)\ ,(j)\log P_y(j)]}$$

$$CORR2 = \sqrt{1 - e^{-2(H_{XY2}-H_{XY})}}$$

$$\text{Contrast} = \sum_{n=0}^{G-1}\left\{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}P(i,j)\right\}$$

Where

$$H_{XY} = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}P(i,j)\log P(i,j)$$

$$H_{XY1} = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}P(i,j)\log((P_x(i).\log P_y(j))$$

$$H_{XY2} = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}P_x(i).\log P_y(j)\log((P_x(i).\log P_y(j))$$

$$\text{Entropy} = \sum_{i=0}^{2G-2}P_{x+y}(i).\log P_{x+y}(i)$$

G-No. Of gray levels used
P-Probability distribution of GLCM

- Shape Feature:

Leaf shape is determined by employing the erosion technique. Using this technique, the *binary erosion* of A by B [12] is denoted by ABQ and defined as

$$B\{z|\ (Bz \subseteq A\}$$

In other words, it can be stated as the set of pixel locations *z*, in which the structuring element is translated to location z overlaps only with foreground pixels in A.

How the shape of an image is detected is shown in .

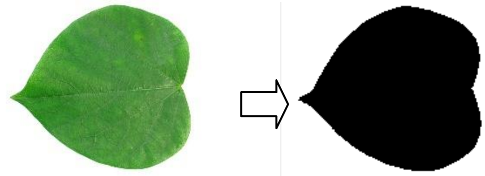

Fig -1: Shape detection using Erosion method

A sample value of feature set used for classification is shown in.These values are computed during feature extraction for all image samples.

Table -1: Feature set for image sample

Entropy	1.6253
Contrast	1.2565
Correlation	0.8365
Solidity	0.8571
Eccentricity	0.3794
Extent	0.8571
Equivalent diameter	58.5019
Standard Deviation	12.9206
Mean	12.1551
Class	1

Classification Techniques:

Human beings often tend to make mistakes while analyzing large amounts of data. Such mistakes are prevalent in the situations when we need to consider the relationships between multiple features of a data set. Thus manual classification of data becomes a herculean effort. To circumvent this issue, different machine learning algorithms prove to be handy. Also, these algorithms improve the efficiency of the systems.

The results of classification can also be shown by using Weka [10]. Weka is a collection of different machine learning algorithms that are written in Java. It is open source software that is developed by the University of Waikato. The algorithms can be applied directly to a dataset. Weka contains different tools that can be used for data pre-processing, classification, clustering, regression, feature selection, association rules, and visualization. It is also well-suited to develop new machine learning schemes. These functions can be accessed by means of a user-friendly graphical user interface. The attribute selection panel provides different algorithms by which the most predictive attributes in a dataset can be identified. The visualization panel shows a scatter plot matrix, where individual scatter plots can be displayed, enlarged and analyzed further by making the use of various selection operators. By evaluating correctly classified instances on the total no. of instances, an accuracy of the algorithm can be determined.

The algorithms can also be implemented in Matlab or Python.

TABLE I shows the example of feature values in a feature matrix using which classification is performed.

Some of the classifiers that can be used for identification are Support Vector Machines*:*

A Support Vector Machine (SVM) is a classifier that is defined by a separating hyperplane that classifies data. In other words, from the given labeled training data, the algorithm outputs an optimal hyperplane which categorizes data to classes. It is a supervised learning algorithm. The hyperplane can be calculated by the equation

$$f(x) = \beta_0 + \beta^T x$$

where

β is a weight vector
β_0 is fixed bias
x is the training sample closest to the hyperplane

The optimal hyperplane is selected that satisfies the equation

$$|\beta_0 + \beta^T x| = 1$$

The margin between hyperplanes can be calculated as

$$M = \frac{2}{|\beta|}$$

This value of M can be maximized by

$$\min_{\beta,\beta_0} L(\beta) = \frac{1}{2}||\beta^2||$$

The **Error! Reference source not found.**describes how data is classified by using SVM classifier by a hyperplane

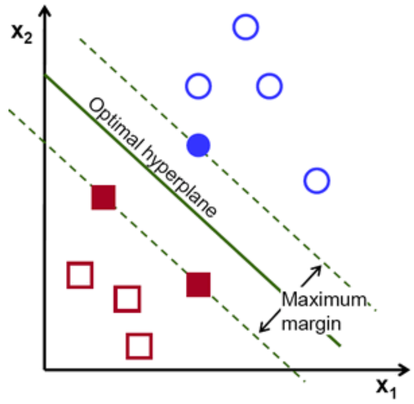

Fig -2: SVM classification

4. PROPOSED SYSTEM

4.1 Overview

This proposed system is inspired by the work of Dr. Varun P. Gopi, Prof. C.M Surya and Prof. ManojKumar P. Arur on the same topic. As per the work was done by T. Sathwik, R. Yasaswini, the use of texture features provides greater accuracy. So in this work, texture features are selected.Nuril Aslina and Nursuriati Jamil used color and shape feature for classification. However, a bit complex techniques were implemented to obtain the parameter values. So a relatively simple approach is used to obtain these values. Itheri Yahiaoui, Olfa Mzoughi used geometric features and shape descriptors for classification. However complex method to calculate shape cost them accuracy. Parag Bhandarkar and Rizwan Ahmed used fewer samples for training and testing. This was the reason behind their reduced accuracy. Pavan Kumar Mishra and Sanjay Kumar Maurya used four geometric features and RGB color features. However, a complex method for classification reduced the accuracy. A. Gopal, S. Prudhveeswar Reddy used 4 geometric features, HU invariant moments as shape descriptor for identification and achieved a good accuracy. B.S. Harish and Aditi Hedge used geometrical features and Zernike moments as shape descriptor for classification on a standard dataset by using SVM classifier. However, the use of Zernike moments created complexity in shape calculation and reduced accuracy.

Thus, looking at the advantages and disadvantages of the work done previously that were inferred from the brief literature survey, Geometric features, texture features, Color features and shape features were selected for classification. The classification was done by using SVM classifier.

4.2 Analysis

The process would go as follows. The given image is first preprocessed. It involves sharpening that was done by using a 2-D spatial unsharp filter and segmentation by thresholding technique. After preprocessing, the image will be ready for feature extraction.

From this preprocessed image, parameter values of different features will be extracted. After preprocessing is done, the geometric features namely Extent, Solidity, Eccentricity, Equivalent Diameter are calculated. After this color features namely mean and standard deviation are calculated. Followed by that, texture analysis is done from the Gray Level Co-occurrence matrix of the image and contrast, correlation and entropy are calculated. Lastly, the shape of the image is calculated by using erosion technique.

After feature extraction is completed, the values of these parameters will be stored in a temporary database as shown in Table II. Based on these feature values, the image samples are classified into different classes and are ready to be tested for identification. The classification is done by using SVM algorithm.

Based on the correctly classified testing instances, accuracy is calculated as a measure of the efficiency of the algorithm.

The **Error! Reference source not found.** below shows how the algorithm will proceed.

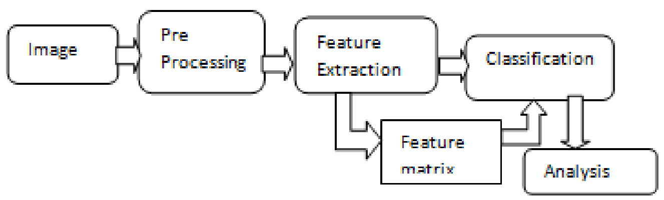

Fig -3: Block diagram of proposed system

5. RESULTS

In this system, we used 64 samples for training and 64 for testing. This training and testing are based on the features namely entropy, contrast, correlation, solidity, eccentricity, extent, equivalent diameter, mean and standard deviation whose values are obtained from the images.

This combination of features along with SVM classifier provides an accuracy of 96.6677%.

6. CONCLUSIONS

Thus, a unique methodology for identification of Ayurvedic medicinal plants from images has been proposed. A unique combination of geometric, color and texture feature have been identified to maximizes the accuracy of identification.

ACKNOWLEDGEMENT

This experiment was conducted at CSE Department, Shri Guru Gobind Singhji Institute of Engineering and Technology, Vishnupuri, Nanded. We would like to thank Prof. S.M. Bansode and Dr. U.V. Kulkarni for their assistance and support throughout the course of the paper. We would also thank the original author of the paper, Dr. Varun P. Gopi and his team for being in touch regarding the details of the experiment.

REFERENCES:

[1] Manojkumar P., Surya C. M., Varun P. Gopi "Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples", Third International Conference on Research in Computational Intelligence and Communication Networks(ICRCIC), 2017.

[2] T. Sathwik, R. Yasaswini, "Classification of Selected Medicinal Plant Leaves Using Texture Analysis", 4th ICCCNT - 2013 July 4 -6, 2013, Tiruchengode, India.

[3] Nuril Aslina, Nursuriati Jamil, "Plant Species Identification by using Scale Invariant Feature Transform (SIFT) and Grid-Based Color Moment (GBCM)", 2013 IEEE Conference on Open Systems (ICOS), December 2 - 4, 2013, Sarawak, Malaysia

[4] Parag Bhandarkar, Rizwan Ahmed, "Leaf Identification using Morphology and Structural Decomposition", International Conference on Signal Processing and Integrated Networks (SPIN), 2014.

[5] Itheri Yahiaoui, Olfa Mzoughi, "ADVANCED TREE SPECIES IDENTIFICATION USING MULTIPLE LEAF PARTS IMAGE QUERIES", 1-INRIA FRANCE 2-SIIVA/RIADI, Intitut Sup´erieur d'Informatique, Universit´e de Tunis El Manar, Tunisia 3- CReSTIC Universit´e de Reims, FRANCE

[6] Pavan Kumar Mishra, Sanjay Kumar Maurya, "A semi-automatic plant identification based on digital leaf and flower Images," IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012

[7] A.Gopal, S. Prudhveeswar Reddy, "Classification of Selected Medicinal Plants Leaf Using Image Processing" 2012 IEEE.

[8] B.S. Harish, Aditi Hedge, "CLASSIFICATION OF PLANT LEAVES USING MORPHOLOGICAL FEATURES AND ZERNIKE MOMENTS", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013.

[9] R.C. Gonzales, R.E. Woods, "Digital Image Processing", PEARSON, 2016

[10] Shweta Srivastava, "Weka: A Tool for Data preprocessing, Classification, Ensemble, Clustering and Association Rule Mining", International Journal of Computer Applications (0975 - 8887), vol. 88 no.10, Feb. 2014

[11] MathWorks Inc® [online]. Available: https://in.mathworks.com

[12] R.C. Gonzales, R.E. Woods, S.L. Eddins, "Digital Image Processing using Matlab", Gatesmark Publishing, 2009.