Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images

Pushpa BR, Anand C and Mithun Nambiar P
Department of Computer Science
Amrita Vishwa Vidyapeetham, Mysuru Campus
Amrita University, India

Abstract
Automatic recognition of plant species recognition is a challenging problem in the area of pattern recognition and computer vision. An efficient plant recognition system will be beneficial to many sectors of society which includes medical field, botanic researches and plant taxonomy study. Manual identification process requires prior knowledge and also it is a lengthy process. This paper proposes a simple and efficient methodology for Ayurvedic plant classification using digital image processing and machine vision technology. The three major phases in proposed methodology are pre-processing, feature extraction and classification. Pre-processing is done in order to highlight the relevant features to be used in the proposed methodology as well as to reduce unwanted noise from the input image, which reduces the chance of getting optimal feature values. In feature extraction phase, different morphologic features such as mean, standard deviation, convex hull ratio, isoperimetric quotient, eccentricity and entropy are extracted from the pre-processed leaf image. In the third phase, a new approach to classify ayurvedic plant species is adopted to recognize plant species by calculating the leaf factor of the input leaf using the extracted feature values and it is compared with the trained values that are stored in the database. An accuracy of 93.75% is obtained for the proposed methodology.

Keywords: Feature extraction, Leaf factor, Ayurvedic Leaf Classification

INTRODUCTION
Ayurveda is an ancient medicinal system evolved in India around thousands of years ago, still followed by many people as it is purely natural and has no side effects. It is very relevant from ancient to this most modern time because of its power to cure chronic diseases. The parts like leaf, flower, root, bark and fruit are mainly used in the preparation of medicines in Ayurveda. At present, the plants are identified manually by experienced physician or taxonomists, which are prone to human errors in many cases. In order to avoid theses human errors, this paper proposes an automated methodology for the identification of medicinal plants which make use of a plant leaf as an input and the classification of plants and its medicinal values as output.

According to Ayurveda every plant on earth has some medicinal value, so it is important to protect the plant and identify its medicinal values. Studies have proved that consuming so much of allopathic medicines may lead to side effects as it carries out many chemical reactions within the body. A general fact about Allopathy is that once it is taken, it requires taking another medicine to cure the side effects which has happened due to the previous medicine. In general, process of consuming medicines will not end. Allopathic treatments are meant to treat the Symptoms of a disease whereas Ayurveda treats the root of the disease.

One of the major advantages of Ayurveda is that it does not have any side effects as it is purely natural, that is relevant in this most modern time as new diseases evolve due to changed life style and changed diet. So it is important for every human being to return back to Ayurveda. Almost all general diseases can be cured through Ayurveda using the shrubs and herbs that are around us. Ayurveda also brings lots of foreign money to the country since many foreign countries are inclining towards it.

Plants are the basic building blocks of life on earth and it is complex to identify a plant species through a photo graph because of its complex three-dimensional structure which cannot be captured through cameras, but it is possible if the leaf can be identified. Fortunately most of the leaves are two-dimensional and it is possible to automate the identification of a plant species through its leaf morphology exploitation.

It is essential for everyone to protect the plants as in today’s world depletion of plants and trees are happening faster due to urbanization. Medicinal plants can be classified based on their internal as well as external features. The external features such as colour, shape, textures and edge histogram are used as their identification parameters. In this paper, an automated system is proposed to recognize the taxonomy of Ayurvedic plants by extracting features from their leaf images and applying mathematical operation to obtain the leaf factor, which can be compared with the trained leaf factor that is in the database to match the input leaf image and classify the Ayurvedic plant species.

RELATED WORK
Researchers have tried many methodologies to extract the features and identify the plant species automatically. Most of these methods make use of combination of many parameters like colour, shape and texture features.

Abdul Kadir et.al [1] proposed a method to identify the plants and various features like texture, vein, shape and colour of the leaves are extracted. The vein feature is extracted using the morphological opening operation and probabilistic neural network is used for classification. The paper limits in achieving reliability with respect to colour feature.

5142
Abdolvahab Ehsanirad et.al [2] proposed a methodology to extract the texture feature of the leaf image and classification. Two different algorithms namely Principal Component Analysis Algorithms (PCA) and Gray Level Co-occurrence Matrix (GLCM) are used to achieve an accuracy of 78% in extracting the texture feature.

A.J. Pérezet.al [3] used the colour and shape feature of leaf image to discriminate soil, weed, and crops. For the proposed methodology different shape features like ratio of the major axis length squared to the area, first invariant central moment, major axis length, ratio of the perimeter squared to the area, minor axis relation, distance to the crop row are used to discriminate soil and weeds. K-Nearest Neighbor (KNN), bayes rule and heuristic approaches are used in classifying the leaf image. An accuracy of 89.7% is acquired from the proposed method.

Pande Ankita Vet.al [4] proposed a methodology for fruit tree recognition using the chain code method. The paper introduced a method called Computer-Aided Plant Species Identification Technique (CAPSI), which is based on the image matching technique of leaf shape. Different biometric features like width factor, diameter, major axis, minor axis, area, perimeter and aspect ratio of the leaf image are extracted. Artificial neural network (ANN) classifier is used for classification of leaf image.

Colhong Lm et.al [5] proposed a system which uses contour of leaves as the main feature to classify the leaf image. In the first step teeth of the leaf and secondly the global structure of the leaf is calculated. Polygonal approximation is done to extract the contour of leaves and also used to detect the mid-most piece of leaf. For classification, a hierarchical method is used. Based on the parameter like apex numbers and similarity measure, the leaf is detected.

Neto.J et.al [6] carried out an experiment with four different plant species namely young soybean, sunflower, red root pig weed, velvet leaf. Elliptic Fourier (EF) and harmonic functions were generated based on the contour of the leaf. Based on the variations between consecutive EF functions and a complexity, index of the leaf shape is calculated. This methodology achieves an accuracy of 89.4%.

James S Cope et.al [7] proposed an efficient methodology for plant classification based on the texture of leaves. Joint distributions for the responses from applying different scales of the gaborfilterare calculated and the difference in leaf structure is calculated by the Jeffrey- divergence measure of corresponding distributions. This methodology worked well in terms of accuracy.

Kue-Bum Lee et.al [8] proposed a Methodology to extract the features of the leaf vein. This work is carried out by finding the contour of the leaf. Firstly by converting the colour image into a gray scale and then to binary image and hence outline of the leaf is extracted. To extract the veins of the leaf, opening operation is done on the grayscale image and the difference in the final image and gray scale image is obtained to get the features extracted.

Sanjay B et.al [9] proposed a methodology in which the input image is converted from RGB to Hue Saturation Value (HSV) and the green pixels are masked before removing and the components are segmented then the useful segments are obtained and finally the colour co-occurrence methodology is used for classification.

Sanjeev S Sannakki et.al [10] presented a paper on comparison between different edge detection methods for leaf images. Fuzzy mathematical morphology is used to carry out opening, closing, erosion and dilation on an image. The dilation and erosion operation can be carried out by the way that the mask is shifted over the image which has been processed by a membership function.

Vijay Satti et.al [11] describes how features are extracted after pre-processing. The procedure involved are pre-processing, RGB to Gray scale and then Gray scale to binary followed by smoothing and filtering. Finally the colour shape and geometric features are extracted. The paper deals with the disease detection in paddy leaves by the approach of histogram processing mechanism. The original disease free leaf is stored in the database and whenever a disease affected leaf image is given as input to the system, it predicts the amount of disease infected in the leaf by analyzing the histogram.

Sanjeev Kumar. E[12] proposed a system with devised methodology which gives the identification of medicinal plants based on its edge features. The colour image is converted to its gray scale equivalent image. From this gray scale image edge histogram is calculated. Canny edge detection algorithm is implemented in this work. The process includes the stages of Image acquisition, feature extraction and comparing the image with those images that are previously stored on the database and the area of leaf is determined by taking one Rupee coin’s area as the reference, which is comparatively effective since the photograph taken may vary from person to person. This work is limited to detecting only the mature leaves since the tender leaves changes slightly when it became mature.

Ji-Xiang Du et.al [13] proposed a new classification method called Move Median Centres (MMC) hyper sphere classifier. From the experimental results of this paper, the methodology save both storage space and reduces the classification time.

From the above review of literature it is very clear that no effective methods were proposed for Ayurvedic plant species recognition and hence this proposed system addresses the Ayurvedic plant recognition.

**METHODOLOGY**

The proposed methodology consists of five steps namely, Image acquisition, pre-processing, feature extraction, classification training and testing. The flow of the system is depicted in Fig. 1.
The leaf recognition is carried out through image processing techniques. The leaf image of a particular plant is fed into the system and the system will pre-process the image in order to reduce the noise present in it and to obtain gray scale, binary and edge for future extraction. In the feature extraction phase, arithmetic mean on colour image, standard deviation on colour as well as convex hull of the leaf in the image, Entropy on gray scale image, convex hull ratio, isoperimetric quotient and eccentricity is calculated. After the feature extraction phase, the leaf factor of the particular leaf is calculated on eight different samples of that plant type and the average leaf factor is calculated which is unique for a particular leaf type and its value is stored in the database. When a new leaf is fed into the system for recognition, the leaf factor of that particular leaf is calculated and it is compared with the leaf factor which is stored in the database and the most matching leaf is returned as the output.

Image Acquisition
Datasets are collected in such a way that leaves are captured against white background using a digital camera or through a scanner. For the proposed work, the dataset of 208 leaf images of 26 different species are collected, which are in different angles and orientation. The experiments are carried over the datasets collected.

Pre-Processing
Pre-processing the images is an important step as it increases the probability of getting desired output in the future steps of image processing. In order to extract the colour feature of the input leaf, global threshold value of the input leaf image is calculated which is useful in enhancing the image and thereby makes the feature extraction phase easier. Then the input image is converted to gray scale and to binary image so as to keep the pixel values as either 1 or 0, so as the feature extraction operations can be made simpler as well as the image gets stored as lower sized binary images, without losing any of its morphological features. Input image is smoothened in order to reduce the noise in the image. Smoothening reduces the number of pixels in the image and it helps in detecting the edges in an image. There are low pass filters as well as high pass filter to make an image smoother. In the proposed methodology Laplacian filtering is applied for edge detection, which computes the second order derivatives of an Image. The pre-processing steps are shown in Fig. 2.

Feature Extraction
Image processing techniques are used to extract a set of features that characterize or represent the image. The values of the extracted features represent the information in the image.

Arithmetic Mean:
Also known as Averaging filter used to find the mean of the pixels in an image through moving an M*N matrix over the pixels of the image and it finds the mean value of the pixels inside the mask, in the next step it replaces the center pixel value with the mean value. Process is repeated till all the mean pixel value are calculated and assigned to the restored image. From this the individual intensity contribution of pixels is extracted. In the proposed methodology the use of arithmetic mean is to extract the colour feature from the input Image. The result obtained after testing the mean value of different species shows slight variation from species to species. Mean value of the images are given by the equation 1.

$$A = \frac{1}{n} \sum_{i=1}^{n} x_i$$  \hspace{1cm} (1)
In which, \( A \) is the average or arithmetic mean obtained, \( n \) is the number of terms (e.g. the number of items or numbers being averaged) and \( x_i \) is the value of each individual item in the list of numbers being averaged.

**Standard deviation:**
It is one of the most used measures to find diversity in statistics. In Image processing it is possible to find the variation by calculating the mean value obtained through standard deviation. Low in standard deviation indicates that data points are very close to the mean and a higher deviation from the mean indicates that data points spread in very large amount of area. This is calculated using the formula 2.

\[
S = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (A_i - \mu)^2}
\]

(2)

For a random variable vector \( A \) made up of \( N \) scalar observations. Standard deviation on colour image and convex hull of the leaf are calculated.

**Convex Hull Ratio:**
Calculation of convex hull requires two spatial parameters. First the area of the leaf and the second area of the convex hull on the leaf. The area of the leaf is calculated as the number of pixels in the foreground and the area of the convex hull is calculated as the number of pixels in the hull. The extraction of binary image and construction of convex hull from its binary image is shown in Fig.3.

\[
\text{Convex Hull Ratio} = \frac{a_{\text{leaf}}}{a_{\text{hull}}}
\]

(3)

Where \( a_{\text{leaf}} \) is the area of the leaf and \( a_{\text{hull}} \) is the area of the convex hull. Equation 3 is applied to get the convex hull ratio.

**Isoperimetric quotient:**
Extraction of the isoperimetric quotient required extraction of the area and the perimeter of leaves which is shown in Fig. 4. Given a binary leaf image, the area is calculated as the number of pixels in the foreground. Laplacian filter is used to get the edge image for the foreground region and perimeter is calculated as the number of pixels in the boundary of the foreground region.

\[
\text{Isoperimetric Quotient} = \frac{4\pi A}{P^2}
\]

(4)

Where \( A \) is the area and \( P \) is the perimeter of the leaf. The result can be used to predict the roundness of the leaf. Equation 4 is applied to get the isoperimetric quotient.

**Eccentricity:**
Eccentricity for an ellipse with major axis \( M_a \) and minor axis \( M_i \) is defined by the equation 5.

\[
\text{Eccentricity} = \frac{M_a^2 + M_i^2}{M_a}
\]

(5)

In order to extract the eccentricity of an object in a digital image, the best fitting ellipse is first evaluated. The best fitting ellipse is the ellipse for which the sum of the squares of the distances to the given points is minimal. In other words, the best fitting ellipse is an ellipse that best fits to the data points contained in the region of interest. Fig. 5 illustrates a best fitting ellipse to a polygon.

**Entropy:**
Entropy is the statistical measure of randomness that can be used to characterize the texture of an image. It helps in identifying the similarity of two images. Here entropy is calculated on the grayscale image. Fig. 6 illustrates the grayscale conversion of the colour image.
Figure 6: Gray Scale Conversion for Entropy Calculation

\[ H = \sum p_i (\log_2 p_i) \]  
(6)

Where \( p \) is the histogram count returned from imhist function in MATLAB. The result shows notable difference in different plant species.

Classification

To classify a leaf according to their species, the values that are extracted from pre-processing stage are considered such as Mean, Standard Deviation, Convex hull ratio, Isoperimetric Quotient, Eccentricity and entropy.

Let \( P_i \) be the various parameters extracted from the leaf image. Leaf factor of a leaf is calculated using the formula

\[
\text{factor}_{\text{leaf}} = \left[ \frac{\frac{1}{N} \sum P_i^2 (\mu_p + K_p)}{\mu_p} \right] \times [P_1 + P_2] \times P_3
\]  
(7)

Where \( P_1 \) denotes eccentricity and \( P_2 \) denotes entropy and \( P_3 \) denotes the standard deviation on colour Image. \( \mu_p \) and \( K_p \) are given by equation 8 and equation 9.

\[
\mu_p = \frac{1}{N} \sum_{i=1}^{N} P_i
\]  
(8)

\[
K_p = \frac{\sum_{i=1}^{N} (P_i - \mu_p)^2}{\sigma^4}
\]  
(9)

Where \( \mu_p \) is the mean of the parameters and \( K_p \) is the Kurtosis of parameters, where \( \sigma \) is given by equation 10.

\[
\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} P_i - \mu_p}^2
\]  
(10)

Proposed Algorithm for Leaf Detection

**Step 1:** Read Image

**Step 2:** Derive Standard Deviation (Sd) and Mean (A) of

**Step 3:** Convert to Gray scale

**Step 4:** Calculate entropy (En) of

**Step 5:** Convert gray to binary

**Step 6:** Construct a convex hull (β) for

**Step 7:** Calculate Eccentricity (Ecc) & Isoperimetric Quotient (IQ) and Convex hull ratio (CHR)

**Step 8:** Calculate Standard Deviation (Sc) of convex hull.

**Step 9:** Calculate Leaf Factor (Equation 7)

**Step 10:** Find out Minimum difference Leaf Factor from Database.

IMPLEMENTATION & EXPERIMENTAL RESULTS

Leaf factor for different images of each species is separately calculated and stored in the database. When a new image is given as input, the system find the leaf factor for the leaf and find the most matching case from the database. The proposed algorithm yield 93.7% of accuracy which is calculated from Table 1 and Visualized in Fig 5.
Table 1: Leaf Detection Accuracy

<table>
<thead>
<tr>
<th>Number of Species</th>
<th>Number of Samples</th>
<th>Correctly detected Samples</th>
<th>Wrongly Detected Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>208</td>
<td>195</td>
<td>13</td>
</tr>
</tbody>
</table>

Figure 5: Test Result

CONCLUSION
The proposed methodology is tested with 208 different sample leaf images of 26 different species and noticed positive response in most cases. To identify a medicinal plant, there is a need for experienced taxonomists or a trained medicine practitioner. Through this work, manual labor needed and time required to perform Ayurvedic species recognition can be reduced. The proposed work can be extended to find the defected leaves to increase the accuracy.

REFERENCES


AUTHOR PROFILE
Pushpa B R has completed Master’s degree in Computer Applications at Visvesvarya Technological University, Belgaum, Karnataka and currently working as a Faculty in the Department of Computer Science at Amrita Vishwa Vidyapeeth University, Mysuru Campus. Her areas of interests are Cryptography and Network Security. Anand C is currently pursuing Master’s degree in Computer Applications in Department of Computer science at Amrita Vishwa Vidyapeeth University, Mysuru campus and completed Bachelor’s degree in Computer Applications at Amrita Vishwa Vidyapeeth University, Mysuru campus. His areas of interest are Computer vision and Pattern Recognition.
Mithun Nambiar P is currently pursuing Master’s degree in Computer Applications in Department of Computer Science at Amrita Vishwa Vidyapeeth University, Mysuru campus and completed Bachelor’s degree in Computer Applications at Adithya Kiran College of Applied Studies, Kannur University. His area of interest are Computer Vision and Medical Image Processing.