Ayurvedic Plant Identification using Image Processing and Artificial Intelligence

Amey Sunil Deshmukh*, Pushppavisha Mani Mudhaliar, Dr. Surabhi Thorat

Department of Computer Science, S K Somaiya College, Mumbai, Maharashtra, India

ABSTRACT

Plants play an important role in Earth's ecology by providing sustenance, shelter and maintaining a healthy atmosphere. Some of these plants have important medicinal properties. Automatic recognition of plant leaf is a challenging problem in the area of computer vision. An efficient Ayurvedic plant leaf recognition system will beneficial to many sectors of society which include medicinal field, botanic research etc. With the help of image processing and artificial intelligence, we can easily recognize the leaf images.

Keywords: ANN(Artificial Neural Network), KNN(k-nearest neighbors), PNN(Probablistic Neural Network),SVM(Support Vector Machine)

I. INTRODUCTION

Ayurveda medicinal system is a very huge and better medicinal technique, apart from western medicine methods. Since the Vedic times Ayurveda is being practiced in India. Ayurveda is one of the oldest systems of Medicinal science that is even used today[3]. Ayurvedic plant identification has become a challenging process and an active area of research, it also plays a vital role in preserving physical and mental health of human beings[5]. As oxygen are the source of plants, they release it by the process of photosynthesis. Apart from this, the plants are used in variety of industrial applications such as herbs and ingredients in ayurvedic medicines[2]. People have been using plant as a traditional medicine and identification of the correct ayurvedic plants that goes into the preparation of a medicine is very important in ayurvedic medicinal industry. Cardiac disorders, Respiratory diseases, Fertility issues, etc can be cured using ayurvedic medicine[7]. So a precise identification of ayurvedic plant is crucial for proper treatment but there is a huge problem that most people cannot recognize these ayurvedic plants and thus are not able to take advantage of herbal power to cure disease[8]. Every plant on earth has some medicinal value, according to Ayurveda. Medicinal plant are those plants that are used in treating and preventing specific aliments and diseases that affect human beings. It is considered a form of alternative to allopathic medicine in the world. One of the major advantages of ayurvedic plant is that it does not have any side effects[2]. Many researchers have made an attempt for ayurvedic plant identification using the image of leaves. Researchers are using different techniques for identification. Some researchers are

212

using shape feature while others are using texture features. Classification of plants into their appropriate species can be done using classifiers like KNN, ANN, PNN, SVM, etc. The Reason behind selecting this topic because Ayurveda medicines are easily available in our nature or surroundings, also it is cheaper. People are not aware of the medical knowledge of the plants which are present in surroundings. So, our purpose is to make a well-defined system which is based on Artificial intelligent to identify ayurvedic plants and gives its medical information of the particular ayurvedic plant's by just giving the information through the image processing method and machine learning technique is used to automatically classify leaf types. The shape of leaves often vary from species to species. Classify the plants into their appropriate species using classifiers like SVM, KNN, ANN, PNN, etc[5].

II. LITERATURE REVIEW

Manojkumar P. et al. [1] explored feature vectors from both the front and back side of a green leaf along with morphological features to arrive at the unique optimum combination of features that maximizes the identification rate. Different classifiers were used to identify the plants using dry leaves. Among this MLP yields highest ID rate of 94.5%. Unique combinations of morphological, colour and texture feature have been identified that maximise identification rate of green leaves.

Dileep M.R. et al. [2] proposed AyurLeaf, a Deep Learning based Convolutional Neural Network (CNN) model, to classify medicinal plants using leaf features such as shape, size, color, texture, etc. A five-fold cross validation is performed on AyurLeaf dataset, maximum accuracy achieved for a single run out of five consecutive runs is 98.46%. AyurLeaf is a CNN-based classification model which was trained and tested on its own dataset and Dleaf dataset. SVM

classifier brought out the best classification accuracy on AyruLeaf dataset.

Amrutha M Raghukumar et al. [3] developed an automatic recognition for medicinal plants. The features were extracted from images of plant leaves and then classified. The features considered are the shape, textural and colour features. When simulated using MATLAB R2019a, the accuracies for KNN was obtained as 100% and SVM was about 93.23%. The KNN is preferrable over SVM for this application.

Amala Sabu et al. [4] proposed a computer vision approach for the recognition of ayurvedic medicinal plant species found in Western Ghats of India and developed an automated system to recognize species of plants by analysing the digital images of their leaf. Their proposed system used a blend of SURF(Speeded Up Robost Features - based on concept of integral image) and HOG(Histogram of Oriented Gradients) features which provides nearly 100% of accuracy when experimented with KNN classifier. Implemented and proposed a system for automatic identification of medicinal plants from their leaves also proposed system that makes use of computer vision and machine learning approaches to identify a pre-trained medicinal plant from its leaf.

Amala Sabu et al. [5] gave a survey on different leaf recognition methods and classification using Artificial Neural Network (ANN), KNN (K- Nearest Network, Probabilistic Neural Network (PNN), Support Vector Machine (SVM), Decision Tree. In this comparative study, the classification techniques they mentioned according to their study they stated KNN method is the simplest classification technique.

T.Sathwik et al. [6] developed a method to identify and classify medicinal plants from their leaf images using texture analysis of the images as a basis for classification. A total of 33 leaves outside the database was tested. 31 out of these 33 leaves were identified correctly giving a success rate of 94% to the program.

The texture features are used for the comparison of the leaves. Their automated classification is efficient and fairly reliable.

A.D.A.D.S. Jayalath et al. [7] proposed Ayurveda methods identification for indigenous plants to predict the medicine. A potent technique using CNN for identification rare medicinal plant was proposed in this article. CNN gives 100% accuracy for training the images and when tested on 5000 images the accuracy was between 95%-99%. A potent technique using CNN for identification rare medicinal plant was proposed in this article.

E. Sandeep Kumar et al. [8] focused on the automatic identification of medicinal plants that are commonly used in Ayurveda. The main idea behind this work was to generate a unique identity number for each plant by calculating the standard Deviation, mean and variance of the plant species. Attempt to perform automatic recognition of medicinal plants, as well as to analyse the statistical nature of the image features used for recognition.

Anu Paulson et al. [9] identified indigenous ayurvedic medicinal plant species using the deep learning techniques. Results show that VGG16 achieved the highest classification accuracy, which is 97.8%, VGG19 achieved 97.6% accuracy and basic CNN achieved 95.79% accuracy. From the results obtained, it shows that VGG16 outperformed VGG19 and basic CNN. A comparison between basic CNN, VGG16 and VGG19 has been performed concerning accuracy.

Sujith A. et al. [10] proposed a set of optimal features for plant classification using the combination of feature extraction techniques. After the Analysis of LBP, HOG and GLCM feature Extraction techniques, they found that LBP has better result among these i.e 97.63% in 291.24 seconds. The combined feature extraction techniques give better result in accuracy and computational time than single method. This

model has achieved good results in Swedish leaves than Flavia and D-Leaf.

III.PROPOSED SYSTEM
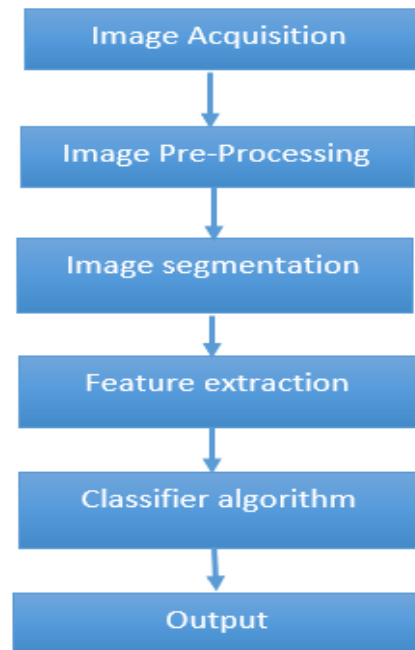
Basic step for ayurvedic plant recognition shown below.

Figure. 1. Basic steps for leaf recognition

The images of the plant leaf are captured through the camera. To remove noise in image or other object removal, different pre-processing techniques is considered (E.g. Image clipping,

Image smoothing). Segmentation means partitioning of image into various parts of same features or having some similarity. Feature extraction plays an important role for identification of an object. In many application of image processing feature extraction is used. Morphology, color, texture, edges etc, are the features used for leaf recognition. After feature extraction is done, the database images are classified. From these methods, we can accurately identify the leaves.

Review of Pre-Processing

The objective of the pre-processing step is to standardize the scale and orientation of the image before feature computation. The raw image is

typically a color image oriented at a random angle and having a random size. The image is first converted to binary and gray scale forms. To make features rotation-invariant, the angle of the major axis of the leaf is extracted from the image and used to rotate it so that the major axis is aligned with the horizontal line[5].

- Gray scale conversion: The image converting to gray scale. The gray scaled images were subjected to the process of image contrast and intensity enhancement techniques and then stacked together as slices for further processing.
- Binary conversion: Thresholding can be used to create binary images from a gray scale image. Binary image is a digital image that has only two possible values for each pixel. Commonly, the two colors used for a binary image are black and white.
- Noise Removal: Digital images are prone to a variety of types of noise. Noise is the result of errors in the digital image acquisition process that result in pixel values We can use linear filtering to remove certain types of noise. Certain filters, such as averaging, low pass filter or Gaussian filters, are appropriate for this purpose. For example, an averaging filter is useful for removing grain noise from a photograph. Salt and pepper noise from an image remove using an averaging filter and a median filter.
- Contrast stretching: Also called Normalization. Some images are homogeneous, they do not have much change in their levels. In terms of histogram representation, they are characterized as the occurrence of very narrow peaks. The homogeneity can also be due to the incorrect illumination of the scene. Ultimately the images hence obtained are not easily interpretable due to poor human perceptibility. This is because there exists only a narrow range of gray-levels in the image having provision for wider range of gray-levels. The contrast stretching methods are designed exclusively for frequently encountered situations. Different stretching techniques have been developed to stretch the narrow range to the whole of the available dynamic range.

- Histogram Modification: Histogram has a lot of importance in image enhancement. It reflects the characteristics of image. By modifying the histogram, image characteristics can be modified. One such example is Histogram Equalization. Histogram equalization is a nonlinear stretch that redistributes pixel values so that there is approximately the same number of pixels with each value within a range. The result approximates a flat histogram. Therefore, contrast is increased at the peaks and lessened at the tails.

IV.Review of Features

A. Shape Features

- Length of leaf: This is the distance between the two ends of the main vein of the leaf[1].
- Breadth of leaf: This is the distance from the left most point in a leaf to the right most point in the leaf.
- Aspect Ratio: The aspect ratio of a leaf is the ratio of the length to its breadth.
- Diameter: The diameter of a leaf is the maximum distance between any two points which lie inside the area covered by the leaf.
- Shape of leaf (Convex Hull): Using the Convex Hull algorithm we can find the coordinates of the points under which the entire area of the leaf is covered.
- Leaf perimeter: Leaf Perimeter is calculated by counting the number of pixels consisting leaf margin.
- Rectangularity (R): The process includes drawing a rectangle outside the image, so that the image just fits in the rectangle. R= LpWp/A, where Lp is length, A is leaf area and Wp is breadth.
- Leaf Area: Smoothed leaf image is consider to find out leaf area. Number of pixels having binary value 1 is termed as leaf area.
- Compactness: It is defined as the ratio of the product of area with 4S to the square of perimeter. It is also referred to as roundness[2].
- Eccentricity: It is a characteristic feature of any conic section, where 'b' and' a 'refer to minimum and maximum axial length respectively for an ellipse.

$$Eccentricity = \sqrt{1-(b/a)^2}$$

B. Texture Features

Energy:

$$E = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (P(i,j))^2$$

Contrast:

$$C = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{j=1}^{G} P(i,j) \right\} \; ; |i-j|=n$$

Entropy:

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

Where G is the number of gray levels, P(i,j) is the probability distribution in the GLCM[5].

Color Features

Arithmetic Mean:

$$mean = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} P(i,j)$$

Standard Deviation:

$$SD = \left[\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (P(i,j) - mean) \right]^{1/2}$$

Skewness(θ):

$$\Theta = \frac{1}{MN\sigma^3} \sum_{i=1}^{M} \sum_{j=1}^{N} (P(i,j) - mean)^3$$

Kurtosis:

$$\Upsilon = \frac{1}{MN\sigma^4} \sum_{i=1}^{M} \sum_{j=1}^{N} (P(i,j) - mean)^4$$

Where M and N are the dimensions of the image, P(i,j) are values of the color on column i and row j[7].

V. Review of Classifiers

A. K-Nearest Neighbor(KNN)

k-nearest neighbor algorithm is a method for classifying objects based on closest training examples in the feature space. k-nearest neighbor algorithm is among the simplest of all machine learning algorithms. Training process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabelled query point is simply assigned to the label of its k nearest neighbors. Typically the object is classified based on the labels of its k nearest neighbors by majority vote. If k=1, the object is simply classified as the class of the object nearest to it. When there are only two classes, k must be a odd integer.

However, there can still be ties when k is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance[4].

B. Probabilistic Neural Network (PNN)

A probabilistic neural network (PNN) is a feed forward neural network . PNN is often used in classification problems. When an input is present, the first layer computes the distance from the input vector to the training input vectors. This produces a vector where its elements indicate how close the input is to the training input. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 (positive identification) for that class and a 0 (negative identification) for non-targeted classes. In a PNN, the operations are organized into a multilayered feed forward network with four layers. Four layers are Input layer, Hidden layer, Pattern layer/Summation layer and Output layer[1].

C. Support Vector Machine(SVM)

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships[12]. An SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. It can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier[3].

D. Decision Tree Classifier

Decision tree is an approach for predictive modeling used in data mining and machine learning, by which the data in data sets can be classified into classes. By using decision trees, the system eliminates half of the cases at each step. This is extremely beneficial because in a system like this, where the dataset is huge in number, the processing takes a chink of the total time. Hence, it has to be minimized wherever possible. The decision tree classifiers organized a series of test questions in a tree structure. The final result is a tree with decision nodes and leaf nodes. Each internal node represents a "test" on an attribute. Each branch represents the outcome of the test. Follow the appropriate branch based on the outcome of the test. It then leads us either to another internal node or to a leaf node[5].

E. Artificial Neural Network (ANN)

Artificial Neural Network have proven themselves as proficient classifiers and are particularly well suited for addressing non-linear problems like leaf classification. ANN is an interconnected group of nodes. Neural networks consist of multiple layers and the signal path traverses from front to back. Back propagation is where the forward stimulation is used to reset weights on the "front" neural units and this is

sometimes done in combination with training where the correct result is known[7].

VI. COMPARATIVE STUDY

Classification Techniques	Advantages	Disadvantages
KNN Classifier	1. Simplest 2. Robust with regard to search space 3. No training is required confidence level can be obtained	1. Expensive testing of each instance 2. Sensitiveness to noisy or irrelevant inputs 3. Lazy Learning
Probabilistic Neural Network	1. Tolerant of noisy inputs 2. Instances can be classified by more than one output 3. Adaptive to changing data	1. Long training time 2. Large complexity of network structure 3. too many attributes can result in over fitting
Support Vector Machine	1. Good generalization capability 2. Sparseness of the solution and the capacity control obtained by optimizing the margin 3. SVMs can be robust, even when the training sample has some bias	1. Slow training 2. Difficult to understand structure of algorithm 3. limitations in speed and size, both in training and testing
Decision Tree	1. Decision trees are simple to use, easy to understand 2. System eliminates half of the cases at each step 3. Fast prediction	1. A small change in input data can at times, cause large changes in the tree.
Artificial Neural Network	1. Requiring less formal statistical training 2. Ability to implicitly detect complex nonlinear relationships between dependent and independent variables	1. Greater computational burden 2. Proneness to overfitting

Table I. Comparision table

VII. CONCLUSION

This work is based on identification of leaf images from the dataset and to identify the image and give its medical use. There are many techniques relevant for the purpose automated recognition for plant leaf recognition. This survey is used to identify different features and classifiers which are used for medicinal plant leaf recognition. From study of above classification techniques we come up with a comparison table. The k nearest-neighbor method is the simplest classification technique. The disadvantage of the KNN method is the time complexity of making predictions.

VIII. REFERENCES

[1] Manojkumar P., Surya C. M., and Varun P. Gopi. "Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples." In 2017 third international confrence on Research in computational intelligence and communication networks (ICRCICN), pp. 231-238, 2017.

[2] Dileep M.R., and Pournami P.N. "AyurLeaf: A Deep Learning Approach for classification of Medicinal Plants." In 2019 IEEE Region 10 confrence (TENCON), pp. 321– 325, 2019.

[3] Amrutha M Raghukumar, and Gayathri Narayanan. "Comparison of Machine Learning Algorithms for detection of Medicinal Plants." In fourth international conference on computing methodologies and communication (ICCMC), pp. 56– 60, 2020.

[4] Amala Sabu, Sreekumar K, and Rahul R Nair. "Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach." In 2017 foutrh International conference on image information processing (ICIIP), pp. 574-578, 2017.

[5] Amala Sabu, and Sreekumar K. "Literature Review of Image Features and Classifiers used in leaf based plant recognition through Image Analysis Approach." In international conference on inventive communication and computational technologies (ICICCT), 2017.

[6] T.Sathwik, R.Yasaswini, Roshini Venkatesh, and A.Gopal. "Classification of selected medicinal plant leaves using Texture Analysis." In 4th ICCNT, 2013.

[7] A.D.A.D.S. Jayalath, P.V.D. Nadeeshan, T.G.A.G.D. Amarawansh, H.P. Jayasuriya, and D.P. Nawinna. "Ayurvedic Knowledge Sharing Platform with Sinhala Virtual Assistant." In 2019 international conference on advancements in computing (ICAC), pp. 220-225, 2019.

[8] E. Sandeep Kumar, and Viswanath Talasila. "Leaf Features based approach for Automated Identification of Medicinal Plants." In International conference on communication and signal processing, pp. 210-214, 2014.

[9] Anu Paulson, and Ravishankar S. "AI based Indigenous Medicinal Plant identification." 2020.

[10] Sujith A., and Aji S. "An optimal Feature Set with LBP for leaf image classification." In fourth international conference on computing methodologies and communication (ICCMC), pp. 220-225, 2020.