International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 01 | Jan 2022 www.irjet.net p-ISSN: 2395-0072

IDENTIFICATION AND CLASSIFICATION OF RARE MEDICINAL PLANTS USING MACHINE LEARNING TECHNIQUES

Shalma Preethi[1], Varshitha[2], Jane Princy[3], Anusha Gowda[4], Mrs. Archana Priyadarshini[5]

[1-4]Student, Dept. of Information Science and Engineering, CEC, Karnataka, India
[5]Asst. Professor, Dept, of Information Science and Engineering, CEC, Karnataka, India

---

Abstract - *As the demand for Ayurvedic medicine increases, demand for medicinal leaves will also increase. Plants are usually identified by human experts using their visual features and it takes a lot of time. In medicinal field, Incorrect identification or error in identification of medicinal plants can lead to unfavorable results. Plant identification can be automated using visual morphological characteristics such as the shape, colour, and texture of the leaves and flowers. This project focuses on machine learning algorithms to get accurate results. Random forest algorithm, k-nearest neighbor algorithm and Logistic regression algorithm are applied to classify the rare medicinal plants from non-medicinal plants. Rare medicinal plants are rarely found in the local gardens and the knowledge of these medicinal plants is in the verge of extinction. This project helps to identify rare medicinal plants using different machine learning techniques. The leaves are identified and classified based on the unique feature combination. Maximum accuracy is expected using various machine learning techniques.*

Key Words: **Rare medicinal plants, Segmentation, Machine learning algorithms, Accuracy**.

1. INTRODUCTION

Medicinal plants can be identified based on their different parts such as leaves, flowers, fruits are planted as a whole. Among all the modalities, leaves are considered as a major and promising modality for the effective classification of medicinal plants. Leaves are obtained in all seasons hence these are identified using machine learning and appropriate names are detected. Both the front andback sides. Maximum accuracy is expected using various machine learning of leaves are captured and detected. Therefore this project contributes in faster identification of medicinal plants without any manual support the leaves are classified based on the unique feature combination. This project can be widely applied in the field of Ayurveda to classify medicinal plants from non-medicinal plants on day to day life basis.

2. RELATED WORKS

In paper[1] the author proposed a method where different features are tested using MLP(multi-layer percepton) and SVM(support vector machine) . MLP has used a Geometrical colour texture combination with 38 different features of leaves where both the sides of a leaf were scanned and features were extracted in order to achieve the higher identification rate. Features such as geometric features, colour features, texture features, HU invariants and Zernike moments are extracted. This method has provided accuracies based on SVM classifiers and MLP classifiers and overall accuracies have been noted. These selected feature combinations are tested using 14 different classifiers and maximum accuracy is noted. These selected features can be set for both dry and green leaves. Since various machine learning languages and neural network algorithms are used, the highest accuracy has been achieved by using MLP neural network classifier i.e 99% of accuracy.

In paper[2] the author proposed a straightforward method for medicinal leaf identification and classification using image processing. It basically had 3 basic phases, Image Acquisition Phase, Image Pre-processing Phase and Feature extraction Phase. Image Acquisition means capturing the image of the leaf using a camera or any other suitable gadget. In Image Pre-processing Phase the background noise and other irregularities are removed and in the last step the features are extracted namely morphological features such as size, area and thickness are acquired. It uses a reference table for comparing these above-mentioned features. Various software methods are implemented here such as ANN is used for classification, Python programming is used for maintaining a dataset and MATLAB used for testing and comparison. In this process, the image is converted into a grayscale and then into a black and white pixel layout. In the next step that is classification various algorithms are used such as Classification and Regression Trees(CART), K-Nearest Neighbors, Logistic Regression, Naive Bayes, State vector machine, Linear discriminant analysis(LDA) and Neural network. These algorithms were applied on 12 herbal medicines with 50 samples. This method has resulted in 98.61% accuracy.

Paper[3] mainly contributes in explaining about the feature extraction from outdoor images. Here the images of leaves are captured along with its background. It uses two segmentation approaches watershed approach and graphcut approach to extract the features from leaves. The Watershed algorithm is an image processing algorithm. The main purpose of this algorithm is to separate the object from its background noise. Here it separates the leaf image from its background or the noise. Using Graph cut we can segment an image into foreground and background elements. It is a is a semi-automatic segmentation technique. In graph cut method, In order to identify what we want in the foreground and what we want in the background we can draw scribbles. Scribbles imply that we can draw lines on the image.

Markers have a common relationship with a specific watershed region that is one to one relationship. we create two markers one for the target leaf and the other for the background. These two markers are also used as the sink and the source seeds in the graph-cut method. We can also use Primary and secondary vein detection to detect the names of the medicinal plants. To obtain the final result they have proposed a leaf refinement using leaf characteristics – colours, shape and other morphological features. Here the results obtained have high accuracy rate than the previously explained methods.

In paper[4] the author identifies the medicinal plants based on their colour, edge detection and GCLM(grey level concurrence matrix)etc. The Artificial neural network has been used as the classifier. The features are extracted from 50 different images and the colour and edge histogram features are extracted from plant leaf and trained with ANN classifier. The colour image is converted to greyscale image and its edges are detected and the medicinal plants are identified from the trained data sets .It can be applied on limited species because there might be similarities between the features that are being extracted here. Hence it may lead to confusion in the result but this method can be applied on the limited number of datasets and can identify the appropriate medicinal plant.

In paper[5] the author proposed a system for training and testing total 64 samples. The main classifier used here is SVM(support vector machine) and the accuracy of around 96.6677% was identified. The image is pre-processed by sharpening the RGB image and converting it to a greyscale image. The next step was segmentation. Segmentation helps to identify the morphological features. This step also involves binarization of the digital image. The largest component of the binary image is selected for determining the different morphological features of the image that is geometric features, colour features and texture features, solidity, eccentricity Entropy, contrast, correlation, solidity, extent, equivalent, extent, Equivalent diameter, standard deviation, mean and class etc are involved in the feature set of image sample for training and testing the sample. Weka is a collection of different machine learning algorithms that are written in Java. It is an 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. these algorithms can be used in MATLAB or python.

In paper[6] the author proposed a leaf detection application that has been developed for Android systems with the help of Open CV libraries which provides the basic tools for computer vision, whose main programming language is C++, but can also be used with other languages such as C, Python and Java, and consequently also in the Android environment. Here they have used saliency maps methods to identify the medicinal plants. Saliency algorithms used by the humans so that they can easily segment image regions that are distinct from their neighbourhood for brightness, colour, shape and movement and therefore are defined salient.. Saliency methods have been applied to numerous vision problems included image segmentation. The first method for the saliency extraction is the algorithm of Itti, Koch and Niebur, based on the extraction of the Gaussian pyramids The other method for the saliency extraction is the Visual Saliency Feature (VSF) method. other algorithms such as Simple Linear Iterative Clustering (SLIC) algorithm was also applied in this approach.

This paper [7] has 5 different classifiers k-Nearest Neighbor algorithm, random forest algorithm, SVM, naive Bayes and Multilayer Perceptron Neural Network .It uses different morphological features like structure, shape, colour, different maps etc. Different accuracies are achieved using these methods. The highest accuracy of 90.1% was obtained from the random forest classifier and second-highest accuracy has been achieved by Multilayer Perceptron Neural Network. This excellent performance indicates the viability of such computer-aided approaches in the classification of biological specimens and its potential applicability in combating the 'taxonomic crisis'. It is designed as a mobile application where one can just scan the image of medicinal plant and if that leaf is trained by the dataset, we can identify the required medicinal leaf.

In paper[8] the author used texture, shape, colour and edge features of leaves, fruits and seeds. Here the image was captured and pre processing was done using K means cluster, Gausian filter, graphcut and grabcut. Various machine learning classifiers are used to identify the types of medicinal plants. classifier algorithms such as Binary classification tree, Multi boost classifier, Random forest algorithm, K nearest neighbor algorithm and PNN(problamistic neural network) and ANN(artificial neural network) are applied on 30 different species. In this method all the features of plants such as flowers, fruits and leaves are extracted hence it will lead to a higher identification rate but it will be time-consuming as it as to access all features of all the parts of plant.

In paper[9] the author proposed a simple method to classify the medicinal plants using various SVM classifiers. Features of the leaf are extracted using convnets and traditional SIFT +bag of features. These methods were used on different datasets such as Swedish dataset, Flavia dataset, CLEF_Uniform and CLEF_Natural. But according to this paper it gives pretty good outcomes for swedish datasets followed by flavia dataset but not really good output in case of other datasets. This method was applied over 100 training samples. As for our expectations, CNN algorithms with pre-trained weights gives similar or better accuracy compared to SIFT+BoF. Particularly traditional method suffers on noisy datasets. The accuracy of the datasets are noted down, where the accuracy of Swedish dataset is obtained around 92%, with a test accuracy of 90.46%, but it doesn't give

expected outcomes with other datasets especially Image CLEF dataset, which is almost equal to random guessing.

In paper[10] the author has described about different machine level approaches. The features like venation, curvature and morphometric features were extracted in order to identify the medicinal plants. SIFT and SURF methods are used to extract the above mentioned features. The SIFT and grid based colour moment is used to classify plant species and it was applied on 40 different species. It achieved an accuracy around 87.5%. The classifier used was SVM support vector regression.The regression can be linear or nonlinear .SVM classifier is a supervised computing device learning algorithm which can be used for classification and regression problems.SVM classifier classifies the images based on its unique features. This classification is used for each analysis and the trying out phase.

3. PROPOSED SYSTEM

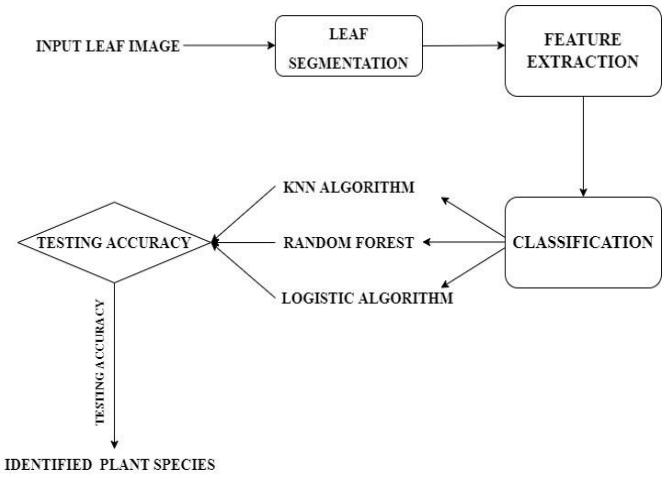

Fig-1.1 : Methodology

Step 1: Input Leaf Image

• Pictures of different leaves are photographed using a camera or comparative gadgets.

• Image acquisition means to collect different type of samples for the formation of the input dataset. Dataset images further go through various processes. In order to provide best solution to any problem it is necessary that dataset cover majority of the different type of inputs.

• One of the most common pre-processing practices is the conversion of the RGB image to a grayscale image. A grayscale image consists of shades of grey. This means that every pixel represents the intensity value at that pixel without showing any colour. Also in a grayscale image the intensity values at a pixel are not absolute and can be infractions. Gray scaling is very significant as it provides a more genuine colour information which plays a major role in segmentation.

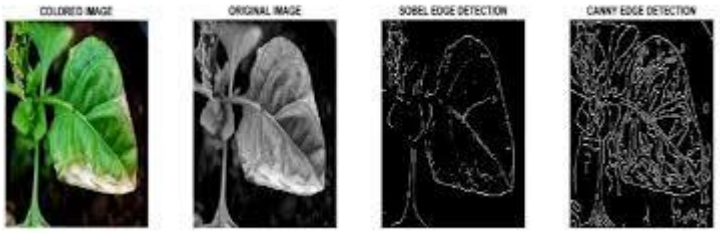

Fig 1.2: Edge detection

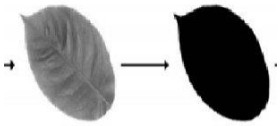

Fig 1.3 : Grayscale image to binary image

Step 2: Image Segmentation

•Image segmentation is a part of digital image processing and computer vision, segmentation is the process of dividing a digital image into a number of segments.

• The main aim of segmentation is to change the representation of an image into something that is more meaningful and more simpler so that it is more easier to analyze.

• Because of colour or shape similarities, the foreground is separated from the background or densely packed groups of pixels.

• Image segmentation is done using Otsu algorithm. Otsu algorithm is used to provide automatic image thresholding.

Step 3: Feature Extraction

• Plants consists of unique features such as different shapes, sizes, texture and edges. These features vary from one plants to another.

• These features are easy to process, but still able to describe the actual data set with the accuracy and originality.

• In this step we extract parameters or unique features from an image. The feature set is extracted by the MATLAB region props method[2]. Many features like area, eccentricity, major axis, minor axis, solidity, inertia tensor, image orientation, bbox, mean intensity, moments central, euler number are collected.

Fig 1.4: Medicinal plants with unique features

Step 4: Classification

• After feature extraction the next step is using the data and classifying different types of leaves.

• Classification is the process of grouping data based on similarities or unique features of the leaf image.

• Classification is done using 3 machine learning algorithms:

• Random forest algorithm

- Random Forest is one of the efficient machine learning classification methods.
- A Large number of leaf datasets can be classified using this technique.
- Random forest mechanism builds many decision trees on various leaf samples the result is based on the majority votes obtained.

• K-Nearest Neighbor algorithm

- Similar and dissimilar data are classified to more than one class using KNN classification technique.
- K-NN algorithm does not provide any output based on assumptions.
- It is also called a non – parametric , instance based learning and lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

• Logistic algorithm

- Logistic regression is one of the notable Machine Learning algorithms, which comes under the machine Learning technique.
- It predicts the output of a categorical dependent variable. Therefore the outcome will have only one value. The output can be either Yes or No, 0 or 1, true or False, etc. But it gives the probabilistic values which lie between 0 and 1 , instead of exact value 0 and 1.

Step 5: Identification

• In this step the names of the rare medicinal plants are obtained. The accuracy obtained by these three algorithms- Random forest, K- nearest neighbour and Logistic regression will be compared and displayed in the form of Graph. As the user inputs a raw image of the medicinal plant, the output will be displayed to the user along with its data.

4. CONCLUSIONS

In this work, we proposed a web based method to identify the rare medicinal plants using their unique morphological features. Otsu method was used for removing the background noise or irregularities of the image. After the completion of segmentation, Feature extraction was done using MATLAB region props table. Based on feature extraction and classifiers the model will be trained and tested. After extracting the features, three algorithms that is random forest, K-nearest Neighbor and logistic regression are used for classification. Given the input image the system will detect whether the leaf is medicinal or not. It displays the compared accuracies of three algorithms in the form of graph. For future work, we can try to include variety of images of leaves and try to include other parts of plants such as fruits, flowers and stem or as a whole plant and soil characteristics as well. We can also use Artificial neural networks and other advanced methods to identify the medicinal plants.

REFERENCES

[1] Kumar, P. M., Surya, C. M. and Gopi, V. P , "Identification of Ayurvedic Medicinal Plants by Image Processing of leaf samples." 2017 Third international Conferences on Research in Computational Interllence and Communication Networks(ICRCICN), Kolkata ,2017,pp. 231-238.

[2]Robert G. de Luna, Renann G. Baldovino, Ezekiel A. Cotoco, Anton Louise P. de Ocampo, Ira C. Valenzuela, Alvin B. Culaba, Elmer P. Dadios Gokongwei College of Engineering, De La Salle University Manila, Philippines "Identification of Philippine Herbal Medicine Plant Leaf Using Artificial Neural Network" 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)

[3] Anantrasirichai, Nantheera & Hannuna, Sion & Canagarajah, Nishan. (2017). "Automatic Leaf Extraction from Outdoor Images".

[4]Aitwadkar P.P., Deshpande S.C., Savant A.V. "Identification of Indian Medicinal Plant by using Artificial Neural Network" International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 04 | Apr-2018

[5]Tejas D. Dahigaonkar1, Rasika T. Kalyane "Identification of Ayurvedic Medicinal Plants by Image Processing of leaf samples" Volume: 05 Issue: 05 | May-2018

[6]Putzu, Lorenzo & Di Ruberto, Cecilia & Fenu, Gianni. (2016). "A Mobile Application for Leaf Detection in Complex Background Using Saliency Maps". 10016. 570-581. 10.1007/978-3-319-48680-2_50.

[7]Adams Begue, Venitha Kowlessur Fawzi Mahomoodally, Upasana Singh, Sameerchand Pudaruth* "Automatic Recognition of Medicinal Plants using Machine Learning Techniques" IJACSA) International Journal of Advanced Computer Science and ApplicationsVol. 8, No. 4, 2017

[8]RAJANI S, VEENA M.N "STUDY ON IDENTIFICATION AND CLASSIFICATION OF MEDICINAL PLANTS" International Journal of Advances in Science Engineering and Technology, ISSN(p): 2321 –8991Vol-6, Iss-2, Spl. Issue-2 Jun.-2018

[9]Albert Liu, albertpl@stanford.edu Yangming Huang, yangming@standford "Plant Leaf Recognition"

[10]Samrity, Dr. Sandeep Kumar PLANT SPECIES RECOGNITION METHOD USING VARIOUS ML-APPROACHES International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 4, April 2017