International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1802-1809 www.ijaem.net ISSN: 2395-5252

Identification and Classification of Rare Medicinal Plants Using Ml Techniques

[1]Student, Anusha B Gowda, [2]Student, Jane Princy, [3]Student, Shalma Preethi D'souza, [4]Student, Varshitha Suvarna, [5]Professor, Mrs. Archana Priyadarshini

Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Dept.of ISE, Canara Engineering College Mangalore, Karnataka

Submitted: 05-06-2022 Revised: 17-06-2022 Accepted: 20-06-2022

ABSTRACT: Ayurveda, often known as the mother of healing arts, is an ancient Indian method of medicine that uses medicinal herbs found natively in the Indian subcontinent. As Ayurvedic medicine is in high demand, demand for medicinal leaves also will also increase. Plants are usually identified by human experts using their visual features and it takes lot of time. The field of medicine incorrect identification of medicinal plants can result in unfavourable outcomes. Plants' morphological properties, such as the form, colour, and texture of their leaves and blooms, can be used to automate visual identification. 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 local gardens, and knowledge of these medical plants is rapidly vanishing. Using several machine learning approaches, this study aids in the identification of rare medicinal plants. Based on the unique feature combination, the leaves are identified and categorised. Various machine learning algorithms are intended to achieve maximum accuracy.

KEYWORDS: Rare medicinal plants, Segmentation, Machine learning algorithms, Accuracy.

I. INTRODUCTION

Among all the modalities, leaves are considered as a major and promising modality for productive categorization of medicinal plants. Medicinalplantscanbelabeled basedontheir various parts such as leaves, flowers, fruits are planted as a whole. Leaves are acquired in all seasons hence these identified using machine learning and appropriate names are discovered. Both the front and back sides.Maximum accuracyisexpectedusingvariousmachinelearningofleavesarecapturedanddetected.Theleavesareclassifiedbasedontheuniquefeature combination.This project can be widely applied in the field of ayurveda to classify medicinal plants from non-medicinal plants on day to day lifebasis.
used, highest accuracy has been achieved by using MLP neural network classifier i.e 99% of accuracy.

Using image processing, the author proposes a simple method for recognising and classifying medicinal plants. The image acquisition phase consisted primarily of clicking on images of the leaf of the plant. The image pre-processing phase did some basic adjustments, like brightness and contrast, and then the feature extraction phase extracted specific information from the images. 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 gray scale and then into a black and white pixel layout. Different methods, such as Classification and Regression Trees (CART), K-Nearest Neighbors, Logistic Regression, Naive

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1802-1809 www.ijaem.net ISSN: 2395-5252

Bayes, SVM method, Linear discriminant analysis(LDA), and Neural network, are employed in the next phase, which is separation or classification. These algorithms were tested on 50 samples of 12 natural medications. 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. Watershed algorithm is a image processing algorithm . This algorithm's main goal is to isolate the item from the background noise. It distinguishes the leaf picture from the background or noise in this case. In the graph cut method, we can draw scribbles, or lines, on the image to define what we want in the foreground and what we want in the background. 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 . Primary and secondary vein detection is also used to identify the medicinal plant. 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 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 limited number of datasets and can identify the appropriate medicinal plant.

In paper[5] the author proposed a system for training and testing 64 samples. The main classifier used here is SVM(support vector machine) and the accuracy around 96.6677% was identified. The image is pre-processed by sharpening the RGB image and converting it to greyscale image. The next step was segmentation. Segmentation is the process where image is divided into different segments and each part is treated as separate image. It 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 i.e 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. we have a library which has different machine learning algorithms written in Java that is called Weka. Waikato university developed this open source software. The algorithms can be applied on a dataset immediately. Weka has a wide range of tools for pre-processing data, classification, clustering, regression, characteristic selection, affiliation rules, and visualisation. These algorithms work in both MATLAB and Python.

The author of paper[6] offered a leaf identification application for Android platforms that was developed with the help of Open CV libraries, which give 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 Visual Saliency Feature (VSF) approach is another way to extract saliency. Other methods were used in this approach, such as the Simple Linear Iterative Clustering (SLIC) algorithm.

The k-Nearest Neighbor technique, random forest algorithm, SVM, naive bayes, and Multilayer Perceptron Neural Network are all used in this paper [7]. .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

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1802-1809 www.ijaem.net ISSN: 2395-5252

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.

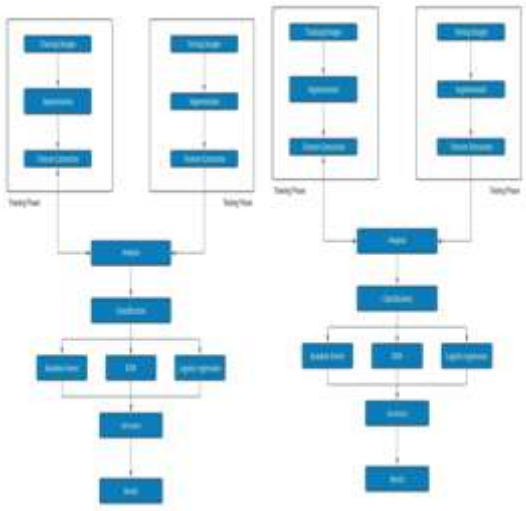
In paper[8] the author used texture, shape, colour and edge features of leaves, fruits and seeds. Here the image was captured and pre processing was done using K means cluster, Gausian filter, graphcut and grabcut. Various machine learning algorithm of classifiers are used to pick out the types of medicinal plants. classifier algorithms consisting of Binary type tree, Multi increase classifier, Random forest set of rules, K nearest neighborset of rules and PNN(problamistic neural network) and ANN(artificial neural network) are carried out on 30 exclusive species. IPNN(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 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 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 a accuracy around 87.5% . The classifier used was SVM support vector regression. The regression can be linear or non-linear.. The SVM classifier classifies the images based on its unique characteristics.This classification is used for each analysis and the trying outphase.

II. IMPLEMENTATION

1. ARCHITECTURAL DESIGN



Figi. ArchitecturalDesign

An architectural diagram is a system diagram that depicts a software system's general structure as well as the relationships, restrictions, and boundaries that exist between its many components. It's a powerful tool since it gives you a bird's-eye view of the software system's physical deployment and evolution path. It contains two phases training phase and testing phase. In training phase the dataset is trained from the features that are extracted and in testing phase the input is segmented and finally both images are analysed.Based on the analysis the results will bedisplayed

2. FUNCTIONAL DESIGN

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1802-1809 www.ijaem.net ISSN: 2395-5252

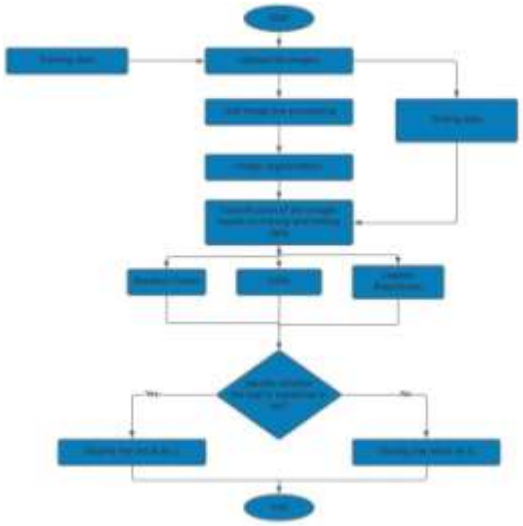


Fig ii.Functional design

Functional flow diagram usually shows the process, workflow or represents a step by step process in solving a task or finding solution to a problem is said to be a flowchart. Flowcharts use boxes and arrows to show the steps taken to solve the problem. These boxes will be connected to each other by using the arrows in the order in which they will be evaluated. Flowcharts are very helpful when it comes to documenting or designing the program or a process. It is also used to analyze a particular problem and find a solution for it in steps. It can be used in various fields.

3. DATA FLOW DESIGN

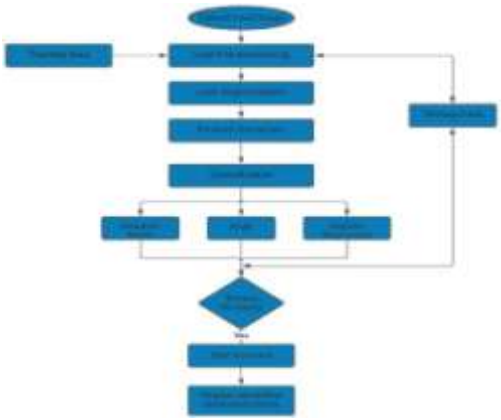


Fig iii. Data flow design

A dataflow diagram depicts the flow of data through an information system graphically. The user will first send an image to the system, which will be segmented, next the feature will be extracted, and finally the image will be classed.

4. USE CASE DESIGN

A use case is a type of interaction between users and the system.It expresses the users' goals as well as the systems duty to them. A use case diagram is made up of a group of actors a collectionof cases encompassed by a system boundary, Communication between actors and use cases, as well as generalisation among use cases. If you're talking to someone or anything.

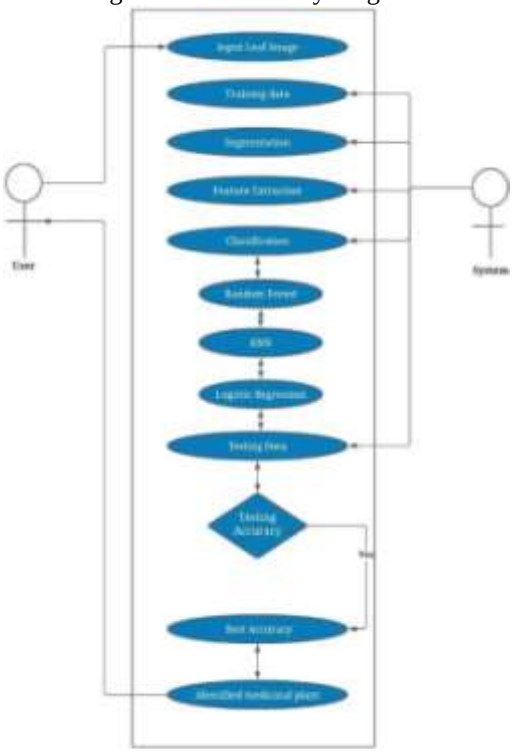


Fig iv. Use case diagram

5. SEQUENCE DESIGN

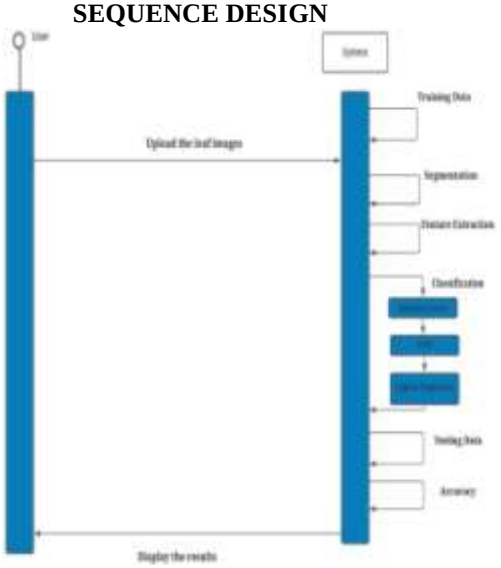
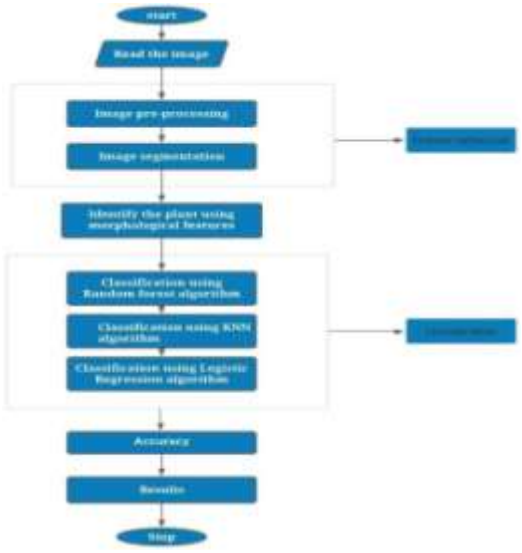


Fig v. Sequence design

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1802-1809 www.ijaem.net ISSN: 2395-5252

A sequence diagram is type of interaction diagram in the Uni field Modelling Language (UML) that describes how processes interact with one another and in what order. It is a component of a Message Sequence Chart. A sequence diagram depicts different processes or items that exists at the same time as parallel vertical lines and horizontal arrows. This enables the graphical specification of simple run-time scenarios transmission. Diagram showsthat the user will input the image to the system. The image of the leaf will be sent by user to the system is segmented feature will be extracted and it will be classified.

6. ARCHITECTURE



Following feature extraction, the data is used to classify distinct types of leaves. The practise of classifying data based on similarities is known as classification. Three machine learning algorithms are used to classify the data:

- Random forest algorithm

One of the most effective machine learning categorization methods is Random Forest. This technique can be used to classify a large number of datasets.

- K-Nearest Neighboralgorithm

Similar and dissimilar data are classified to more than one class using KNN.

- Logistic Regressionalgorithm

Under the Supervised Learning technique, one of the most often used Machine Learning algorithms is logistic regression. It's utilised to forecast the categorical dependent variable from a series of independent variables.

III. PROPOSED SYSTEM

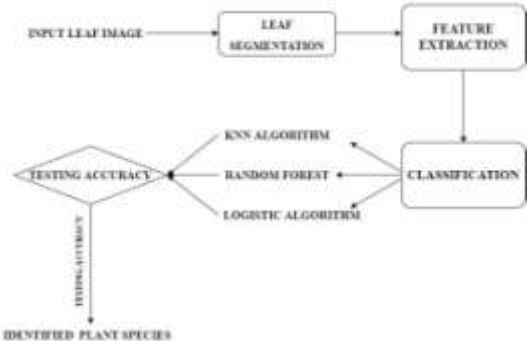


Fig-i : Methodology

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 gray scale image. A gray scale image consists of shades of grey . This means that every pixel represents the intensity valueat that pixel without showing any colour. Also in a gray scale mage 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.

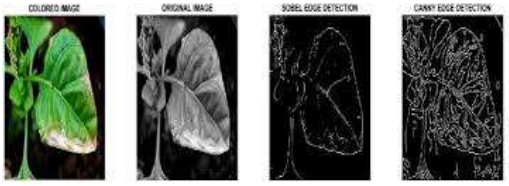


Fig ii: Edge detection

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1802-1809 www.ijaem.net ISSN: 2395-5252

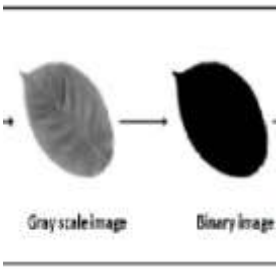


Fig iii :Gray scale image to binary image

Step 2: Image Segmentation

•Image segmentation is a method that divides a digital image into numerous segments. It is a part of digital image processing and computer vision.

• 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.

• Image segmentation is part of digital image processing and computer vision, and how many segmentations are Digital image to that segment

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.

• 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 iv: 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 method.
- 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 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

- One of the most well-known Machine Learning algorithms, which falls under the Machine Learning approach, is logistic regression. It anticipates a categorical dependent variable's outcome.
- As a result, the outcome will only have one value. The result can be Yes or No, 0 or 1, true or false, and so on. However, instead of the exact number 0 and 1, it delivers the probabilistic principle that sits between 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.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1802-1809 www.ijaem.net ISSN: 2395-5252

IV. RESULTS

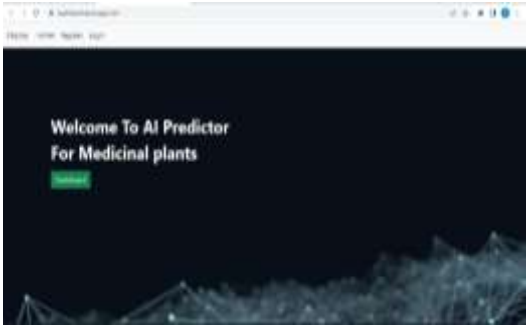


Fig i. HOME PAGE

This page allows users to select different options such as register, login, and admin. This is the home page of the website.

Fig ii. Registration page

Users must first register in order to enjoy the services offered. A username, password, email address, phone number, and profile image must be provided by the user. The user will be registered if they enter valid values when registration.

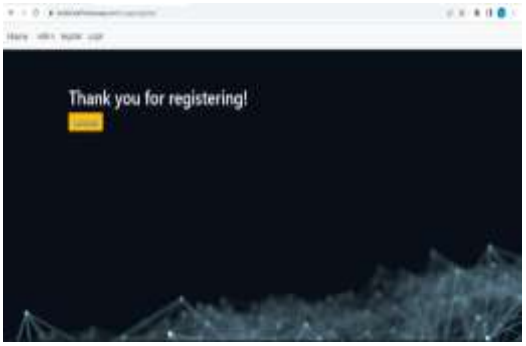


Fig iii. Registration successful page

The user is provided with above message as he finishes registration process. The user should click on continue button to proceed further.

Fig iv. Login page

This is the authentication page for the user. The user can access the system if his or her User Name and Password match.

Fig 6.5 web page for uploading image

In this page the interface provides option for uploading the image of the medicinal leaf. Once uploaded this will be proccessed and compared with the database. Here we need to enter the ph value and temperature of the leaf too.after uploading the system will detect whether it's a medicinal leaf or non-medicinal leaf.

Fig vi. Output of the leaf image

This page provides the details whether the detected plant is medicinal or not

V. CONCLUSION

We suggested a strategy for identifying and classifying uncommon medicinal plants based on their distinctive morphological traits in this research. To remove background noise or imperfections from an image, we employ the Otsu method. Using a MATLAB region props table, feature extraction was conducted after the segmentation phase was completed. Feature extraction and classifiers are used to train and test models. Following the extraction of the features, use the three algorithms of random forest, K-nearest neighbor and logistic regression for

classification, and use the one with the highest accuracy. Given the input image the system will detect whether the leaf is medicinal or not. For future work, we can try to include variety of images of leaves and try to include 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 rllence and Communication Networks(ICRCICN), olkata ,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. DadiosGokongwei 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, VenithaKowlessurFawziMahomoodally, Upasana Singh, SameerchandPudaruth* "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