

A Comparative Analysis on Machine Learning Models for Accurate Identification of Medical Plants

Prabhat Kumar Thella¹; V. Ulagamuthalvi²

¹Research Scholar, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India.

¹thellaprabhatkumar@gmail.com

²Associate Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India.

Abstract

Plants are essential for human life. They help people breathe, provide food, clothing, medicine, and fuel, and also safeguard the environment. Plants can be loaded with medicinal properties and possess active substances that can be used for medical purposes. Several beneficial plant species are disappearing as a result of such factors as global warming, increasing population, professional secrecy, insufficient government support for research efforts, and the lack of public understanding of medicinal plants. It takes time to identify medicinal plants, therefore use professionals to assist you. For better benefit to humankind, a new method to identify and classify therapeutic plants must be developed. Because of the advanced technology in our day and age, medicinal plant identification and classification is an important subject of research in the field of image processing. Feature extraction and classification are the most important components in the process of identifying medicinal plants and classifying them. This research examines methods used in identifying and classifying medicinal plants as well as the medicinal properties of plants that have become increasingly relevant in the recent past. There is a vital importance placed on identifying the suitable medicinal plants in the creation of an ayurvedic medication. In order to identify a medicinal plant, look for these three features: leaf form, colour, and texture. From the both sides of the leaf, there are both deterministic and nondeterministic factors that identify the species. In this study, a combination of traits is designed that is said to identify a single tree the most effectively while minimising errors. The database is made up of scanned photos of both the front and back side of ayurvedic medicinal plant leaves, which is an ayurvedic medicinal plant identification database. In leaf identification, rates as high as 99% have been found when tested on a wide range of classifiers. Extending the prior work by using dried leaves and feature vectors results in identification using which identification rates of 94% are possible.

Identification of the correct medicinal plants that goes in to the preparation of a medicine is very important in ayurvedic medicinal industry. The main features required to identify a medicinal plant is its leaf shape, colour and texture. Colour and texture from both sides of the leaf contain deterministic parameters to identify the species. This paper explores feature vectors from both the front and back

side of a green leaf along with morphological features to arrive at a unique optimum combination of features that maximizes the identification rate. A database of medicinal plant leaves is created from scanned images of front and back side of leaves of commonly used ayurvedic medicinal plants. The leaves are classified based on the unique feature combination. Identification rates up to 99% have been obtained when tested over a wide spectrum of classifiers. The above work has been extended to include identification by dry leaves and a combination of feature vectors is obtained, using which, identification rates exceeding 94% have been achieved.

Key-words: Medical Plant Identification, Performance Analysis, Feature Extraction, Classification, Clustering, Image Processing.

1. Introduction

Ayurveda is an ancient system of medicine dating back to roughly 5000 years ago that is prevalent in India. Ayurvedic remedies consist of plant leaves and other plant parts, such as roots, bark, and flowers. Over 8000 native Indian plants have been identified as having therapeutic properties. 1500 of these plants are employed in the many systems of Indian herbal medicine. This holds true for all Ayurvedic formulations, commercial and not. When gathering the required plants, 80% come from woods and wastelands and the other 20% are obtained from agricultural fields.

For centuries, Ayurvedic doctors utilised their own powers of selection and preparation to prepare treatments for their patients. However, this approach is used by a very small number of therapists today. The Ayurvedic medication production and marketing sector has grown and is now estimated to be over \$4 billion. Over 8,500 Ayurvedic medicine factories have licences to manufacture Ayurvedic medicines in India. Several problems concerning the quality of Ayurvedic medicine ingredients have now been brought to the forefront due to the expansion of the Ayurvedic industry.

Today, women and children harvest medicinal herbs from the forest; these individuals are not adequately trained in plant identification. Inappropriate or replaced medicinal herbs are quite commonly provided to manufacturing plants. This makes it very difficult to effectively screen these plants, which is a problem considering the poor quality control procedures most of these units lack. There is also variance in the local names, which adds to the confusion. Manual identification can be more challenging when plants come in dried state. The Ayurvedic medicine loses its effectiveness if used incorrectly. The possible side effects are unpredictable. To preserve the present growth of the sector, strong quality control procedures must be implemented on Ayurvedic medicines and raw materials utilised by the company.

A skilled botanist inspects all of the plant's attributes, such as leaves, flowers, seeds, roots, and stems, in order to identify the many kinds of plants. Aside from the leaf, everything else is a 3D object, increasing the number of analyses computers need to do. Leaves, being two-dimensional objects, are able to capture substantial plant information, making them suitable for taxonomic classification. The harvesting of leaves is rather simple, and photos can be taken with basic digital cameras, smartphones, or scanners. In contrast to seeds and blossoms, the product is available throughout the year. As the plant grows, it adopts several different colours, textures, and shapes, and these variations are inconsequential. The accuracy of plant recognition that depends on detecting certain descriptors and extracting feature vectors from it is dependent on how precisely those descriptors are identified. A classifier is then used to determine how similar the feature vectors of the training samples are to the feature vectors of the test sample.

Herbs are plants which are known to provide specific nutrients that are useful for maintaining good health. Medicinal plants have been found to contain qualities that help prevent, cure, or treat a wide range of diseases. There are 30,000 varieties of plants in Indonesia, and 7000 of them are called therapeutic herbs. The usage of medicinal plants has the added benefit of being a natural remedy compared to medications that include active components. Active components (phytochemicals) in medicinal plants can be identified by screening for molecules known as phytochemicals (phyto-chemicals).

Chemical medicines contain substances that are inorganic and pure, while the human body is complicated and organic. Chemical medications are not recommended for human ingestion, because long-term use can actually be detrimental to human health. There are, however, some chemical medications that are only symptomatic (temporary) and people with certain illnesses must take them for the rest of their lives. There is still a lot of knowledge about medical plants that is not equally distributed to the number of medicinal plants, which causes people to prefer medications with chemicals since they are regarded more practical and available. To make the people more aware of therapeutic plants, a system is needed that is able to highlight medicinal leaves. Using Neural Network[8], [9], [10], leaves can be recognised by a combination of colour, size, texture, and form.

The findings of previous studies were able to identify herbal medicinal plants based on leaf imagery utilising ANNs, Gray Level Co-occurrence Matrix, and K-Nearest Neighbor Algorithms. Medicinal plant leaves, including bay leaves, avocado leaves, cat's whiskers leaves, celery leaves, soursop leaves, african leaves, starfruit leaves, grass jelly leaves, and betel leaves, are the leaves to be identified. The leaves were selected due to the fact that hypertension is considered a significant medical condition that affects most individuals. [12].

2. Literature Survey

Using random leaf samples from 40 distinct species, Manojkumar P., Surya C., and others [1] compiled a collection of Ayurvedic front and rear side leaves. Weka is used for identifying medicinal plants based on machine learning methods. Features of leaves, such as colour and texture, are derived from photographs of leaves that have been processed to have both binary and colour features. The leaves can be identified using support vector machine (SVM) and multilayer perceptron (MLP) classifiers. CR distance geometric centroid-radii (HU) invariant moments textural characteristics Zernike moments MLP outperformed support vector machine by a wide margin (at 94.5 percent accuracy) (SVM).

Neural networks have been presented as a system for identifying of medicinal leaves. The five species of medicinal plant leaves are believed to be of high importance. Prewitt Edge Detection algorithm does the edge detection on the leaf edges. An Artificial Neural Network (ANN) classifier trains the data and compares it to others, which produces good accuracy.

Adams Begue et al. [3] proposed a method for automatic identification of medicinal plants (using leaf attributes) in a machine learning framework utilising machine learning techniques. Classifiers are used for classification purposes to discriminate between different types of data. roughly 90.1% of the accuracy is found in random forest classifiers, which is an especially accurate method in comparison to k-nearest neighbour, naive Bayes (NB), SVM, and neural networks. Shapes and textures of leaves are taken into consideration in order to identify and classify the medicinal plants in TCM [4] system. features that have been extracted are fed into the Support Vector Machine (SVM) classifier, which can accurately classify 93.3% of the images.

Gray-Level Co-occurrence Matrices (GLCM) and Color moment extract colour and texture features from flowers. Neural Network classifiers [5] feed extracted features into their structure. GLCM and Color moment are accurate to the tune of 40% and 65% respectively. There is a hybrid combination of accuracy that has a success rate of 95%. To give examples, authors first looked at three datasets, Oxford Flower 17, Oxford Flower 102, and Jena Flower 30, which were used for classifying flowers based on colour and form attributes. A variety of different approaches are applied to fusion, pooling, extraction, and flower detection. Other two data sets' classification accuracies were higher in Jena Flower 30, however Jena Flower 30 had the best classification accuracy at 94%.

Automatic recognition system for mango fruits [created by students in the field of Electronics and Telecommunication Engineering] using segmentation and edge and colour analysis techniques. One method utilised in picture segmentation is the method of K-means clustering and canny edge

detection. Edge-based algorithm outperforms and yields 85% of the total score while color-based method outperforms and yields 88% of the total score.

Throughout the aforementioned publication, the researchers highlighted and reviewed four classification techniques utilised in identifying medicinal plants: Support Vector Machine, Principal Component Analysis, and Probabilistic Neural Network. For identifying reasons, they determined the Aspect Ratio, Centroid, Area, Perimeter, and Roundness.

To construct a fruit recognition system, authors focused on various fruit characteristics, including form, colour, and texture. The features are trained by three separate classifiers, each one corresponding to a different type of classification algorithm. K-Nearest Neighbor (k-NN), Binary Classification Tree, and Support Vector Machine (SVM). With 100% accuracy, the SVM classifier yields the best result.

Ayurvedic plants are categorised in a creative way depending on their leaf structure. There were twenty-six different species, and a 208-leaf image dataset was constructed with them. Laplacian filtering is used to give an edge detection filter. After the morphological features of the leaf images are determined, the morphological features of all the training and testing leaf images are found to provide a list of candidate leaves. The proposed approach results in a level of accuracy of 93.7 percent.

Ayurvedic medicinal herbal plants can be identified using a smartphone application under Android platform. This plant identification system will find therapeutic plants by examining a set of leaves. In texture extraction, Gray-Level Co-occurrence Matrices (GLCM) are utilised, while in plant species image processing, processing approaches are applied. This device is completely free, it saves money and time, and doesn't require any assistance from specialists. Plant identification and querying information assists in botanical gardening, in medicine, and in the beauty sector.

Classifying medicinal plants using leaf photos is aided by the usage of Gray-level, Grey Tone Spatial Dependency Matrix (GTSDM) and Local Binary Pattern (LBP) characteristics. There are five plant species and 250 leaf photos for each species. Stochastic Gradient Descent (SGD), k-Nearest Neighbour (KNN), Support Vector Machines (SVM), Decision Trees (DT), Extra Trees (ET), and Random Forests (RF) are all used to categorise the retrieved features. These four statistics are generated from Gray-level values. GTSD and LBP give rise to entropy and mean, respectively. These computed properties are used to categorise the therapeutic plants in different ways. This system had an accuracy rating of 94.7 percent for SGD, DT, and k-NN classifiers. Feature extraction and preprocessing produce similar results for identification and categorization of medicinal plants, although feature extraction is faster.

Leaf photos were used to devise a system for identifying medicinal plants, as proposed by E. Plants, especially Hibiscus, Betel, Ocimum, Leucas, Vinca, Murraya, Centella, Ruta, and Mentha, are evaluated when doing a post-analysis on plants. Canny edge detection technique detects leaves that grow along the edges of features. The data is obtained by a histogram in which the colour information is retrieved. The calculation of the "area" is done using the algorithm proposed. As a result, the authors drew a conclusion based on the difference in the average values of three parameters between the database images and the test images. The value will be compared to other photos in the database if the database and test images are identical.

Color histogram (CH), Edge Histogram (EH), and Edge Direction Histogram are utilised to separate medicinal plants into the following three classes: herbs, shrubs, and trees (EDH). An image noise reduction filter, called a Gaussian filter, is used. In applying the K-means segmentation algorithm, medicinal plants are divided into background and foreground using the image. For the colour histogram, authors employed RGB (red, green, blue), HSV (hue, saturation, value) and Y₃ (color-space Y (value)), a colour space that includes all of the possible values and colours. Euclidean distance (ED) and sum of Absolute Difference (AD) are used to calculate image similarities. Mean Square Error (MSE) is employed to improve precision. They discovered that on average, the edge histogram, the edge direction histogram, and their combined features provided an accurate representation of the edge. Trees have a very good precision rate, with an average accuracy of 95%, 96%, and 99% correspondingly.

Color histogram, edge histogram, and edge direction histogram are retrieved to identify the medicinal plants as herbs, shrubs, and trees. Some plant classification techniques, such as support vector machines (SVM) and artificial neural networks (ANN), have been applied to plants. SVM is able to achieve a maximum and minimum classification accuracy of 94% and 70%, and ANN is able to achieve a maximum and minimum classification accuracy of 90% and 65%. Trees provide the most level of accuracy while Shrubs provide the least level of accuracy because there are insufficient photos for the stem component of the plant.

The Flower Identification System (FIS) is intended to provide both simple flower knowledge and detailed categorization information to a person with no prior botanical expertise. The most accurate way to identify a flower is by its colour. The flower image is categorised based on its colour histogram in this manner. Grabcut is used to separate the input image into a number of segments. With 80.67% accuracy, our classification system can classify floral photos.

Authors in Malaysia devised a system for identifying herbal plants using leaf texture. The organisation conducted a comparative analysis on the following texture features. HOG, LBP, and

Speeded-Up Robust Features (SURF). Authors looked at the Flavia data set and on the other hand, they examined a new dataset where additional herb plants had been included. Multiclass SVM classifier and the extracted texture feature are both filled in with the extracted texture feature. In relation to several texture features, HOG (99% accurate, 97% accurate), LBP (99% accurate, 97% accurate), and SURF (99% accurate, 97% accurate) (74 percent, 63 percent). Though BSA outperforms the other two texture features, Surf outperforms it by a large margin.

A system that looks for therapeutic plants automatically was developed by Gopal et al. [18] employing 10 different plant species. Skin colour and other skin characteristic attributes are retrieved. According to the algorithm, the accuracy is estimated to be 92% on 100 (trained) and 50 (tested) leaves.

Twenty leaf attributes (variables) allow medicinal leaf photos to be identified by form, colour, and texture properties. To extract the form, colour, and texture aspects, seven moment, colour, and GLCM are employed. To train 63 photos to have five different leaves, an ANN is employed. By stacking several individual features, the accuracy increases to 94.4 percent.

The leaf images are identified by texture properties by using the CLBP algorithm. To test the experiment, the data was first processed and experimented on across multiple datasets such as the Austrian Federal Forest (AFF) datasets, Foliage dataset, Flavia leaf dataset, Swedish leaf dataset, and Middle European woods (MEW) dataset. The SCLBP signature approach has an accuracy of 84.234 on all datasets with an average classification rate of 84.234. No classification rate is achieved with the MCLBP, which yields a large magnified image of a binary pattern. The component from CSMCLBP achieves an average classification rate of 89.346 when included in the mix.

In this experiment, a combination of Local Binary Patterns (LBP) texture characteristics and a Probabilistic Neural Network classifier [21] is used on tropical medicinal plants and domestic plants. Medical plants, leaves, and houseplants are included for identification purposes. To improve accuracy and efficiency, the fusion characteristics are categorised based on the classifier combination they're used with. For both medicinal plants and houseplants, the effectiveness of the off-the-top offusion operator is 70.66% and 84.44%. For medicinal plants, the classifier combination accuracy is 77 percent, while for domestic plants, it is 86 percent.

Leaf shape and texture extraction is used to identify medicinal plants. Canny Edge Detection has been used to remove noise and to determine the region on the image that is sharpest. After morphological process images are fed into the neural network, the neural network performs the morphological process. To find out if a Hibiscus leaf image works as well as a Hibiscus, Betel,

Castor, and Manathakali training image, we apply the trained images to Hibiscus leaf images. The proposed approach overall has an accuracy of 70.87 percent [22].

Shamrie et al. [23] in Malaysia provided a framework for the identification and classification of tropical medicinal plants based on 5 species with 65 leaves. In edge detection, the Prewitt Edge Detection Algorithm is employed. Classifiers used in the WEKA tool are plugged into extracted features, which are then fed into the WEKA tool. Some classifiers, such as Sequential Minimal Optimization (SMO), J48 decision tree, NB, and kNN (k=3) are regarded as practically good classifiers. To conduct this experiment, DECIML for Imbalanced Multiclass Learning was utilised. The DECIML classifier outperforms other WEKA classifiers in terms of accuracy (65% better).

A computer-automated technique created by Iyan Mulyana et al. [24] resolved problems with medicinal plant identification in Indonesia. Fractal dimension (pattern) and fractal coding (texture) are applied to medicinal plant leaf features. The classification of plant species is done using fuzzy C-Means clustering using a particular F-Measure clustering method (85.04% accurate) (79.94 percent).

3. Proposed Model

Digital cameras, smartphones, and the dataset for plants are used to take photos of plants. Machine-generated image samples are referred to as the machine input images. Before use of classification techniques, we analyse the input picture to eliminate noise, improve sections, and filter it. As at this step, only areas of interest in the image will be focussed, while extraneous data is deleted. Computational speed and machine efficiency are both boosted.

- Once you've obtained the original plant images, the next step is to extract the unique properties, and then use those to train a classification system to identify them. A variety of techniques are used to extract and classify information from images.
- The leaf, flower, or fruit researchers in plant identification often employed characteristics such as texture, shape, colour, and the character of the edge. In order to identify things, classifiers are employed.

The picture will often throw a shadow under the leaf, as the leaves aren't completely flat. The shadow will create uncertainty about the location of the borders because it contrasts with the background and muddies algorithms' ability to correctly identify shadow-based limits instead of real-world boundaries. Thus, picture segmentation should be disabled before image segmentation is performed. In the first step, the RGB image value was transformed to HSV. The shadow of the object

was then marked using the specified channel to help distinguish the shadow from the object. The next step is to determine the RGB value for photos as the HSV value modification alters the original colour.

After the preceding processes, leaf characteristics were defined as a set of parameters. Moral qualities, invariant characteristics, textural characteristics, and histograms with oriented gradients make up the features of this study. These design elements have been purposefully picked to meet the requirements for prominent picture leaves, and also to help distinguish various image leaf kinds based on their numerical values.

Two copies of the same data were generated from the same source. hence, the developed models were evaluated and compared using three established performance measures: accuracy, ROC, and AUC (AUC). To show the differing results between the predicted outcomes contrasted to those that are labelled, the confusion matrix has been created. Precision is being able to estimate percentages and values with no variance. A confusion matrix can be found by taking the formula shown and simplifying it.

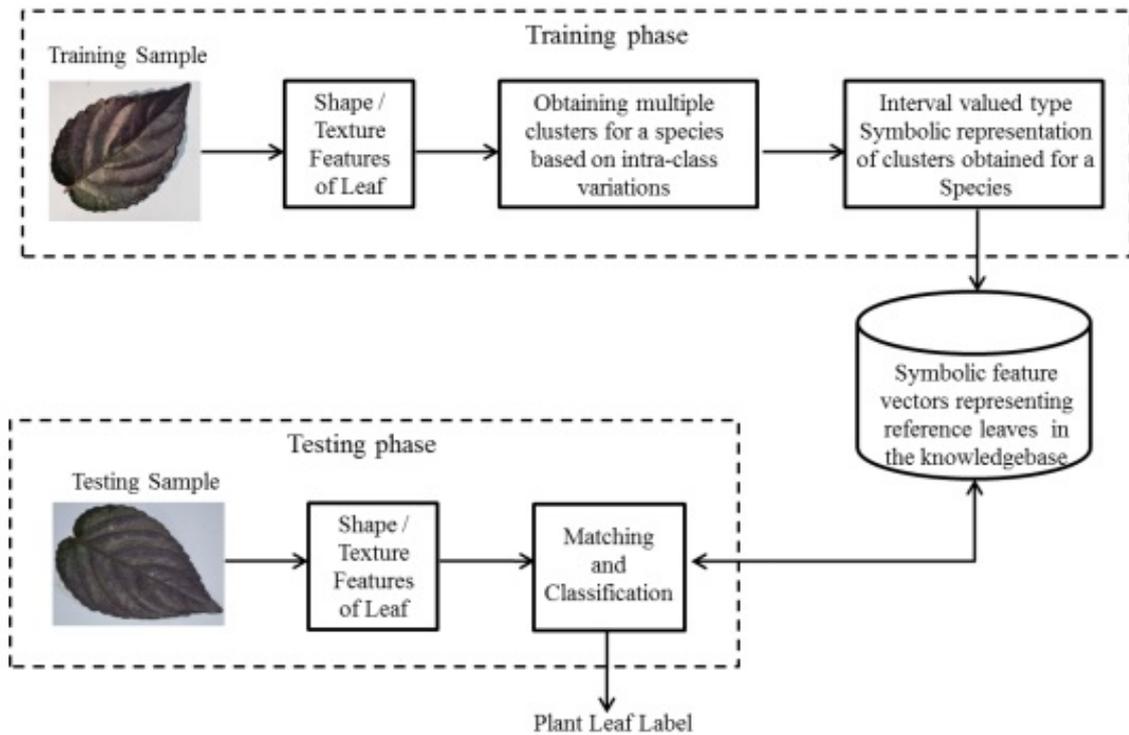
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}},$$

The TN result is a true negative, while the FP and FN results are false positives and false negatives, respectively.

The ROC value shows how well the binary classifier method is able to discriminate and label. A delicate and genuine story that tells of a falsehood toward happiness. Trapezoidal rule calculations of AUC begin with this plot. AUC is generally described as ranging from 0 to 1, with values higher than 0.9 suggesting outstanding discrimination, values between 0.8 and 0.9 are excellent, and values lower than 0.7 are suitable.

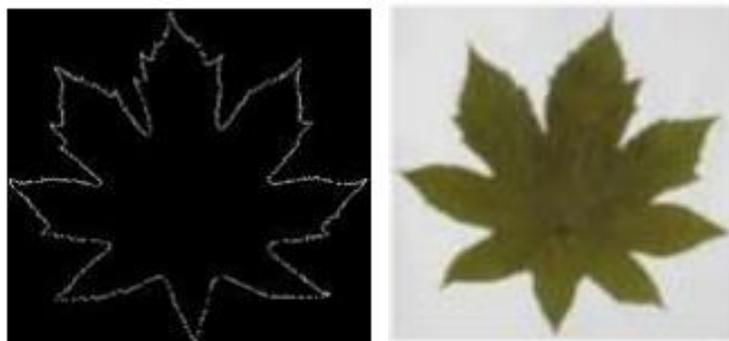
In addition to forms, textures, and veins, there are numerous other aspects associated with the leaf, including scents, tastes, and colours. By their form, texture, and scent, most individuals can identify leaves. In addition, a vein pattern is a property of the leaf. However, to recognize a vein pattern leaf, there must be a leaf that is located very far away from the same branch of a single plant. In addition to the above parameters, light and water supply influence the overall leaf structure because of environmental conditions. Ayurveda leaves are commonly found in ayurvedic areas, and the scholars who studied in those regions are well acquainted with their usage. Users are looking to recognize leaves in the actual world in this assignment. This is seen in Figure 1.

Fig. 1 - Proposed Machine Block Diagram



To identify the sheet, compare the leaf to the dataset's attributes. To begin, you need to build a learning basis of facts and ideas. This procedure contains two phases: start-up and production. Once the pre-processing is completed, the feature extraction will be done. Using a project outline is viewed as one of the most essential project specifications. Constant points can be represented by a curve, and the form can be defined as a continuous curve, using the same or similar hue or intensity. The detection of the leaf margin is shown in Figure 2.

Fig. 2 - Shape of Castor Oil Leaf



The residual characteristics are found by searching for the remaining shape after a removal.

Area: An area that represents the amount of pixels in a leaf region. In the regions of the image, the two-pronged shape is black and white. The area of the sheet is calculated to equal the number of white pixels.

A line that runs at the midpoint of the apex and base of the leaf is illustrated in the major axis.

The minor axis, a small leaf axis, and the second central standard axis of the leaf region are all perpendicular to each other.

Also, another type of picture moment is a weighted average of the image pixel intensities (moment).

Also known as the center of mass

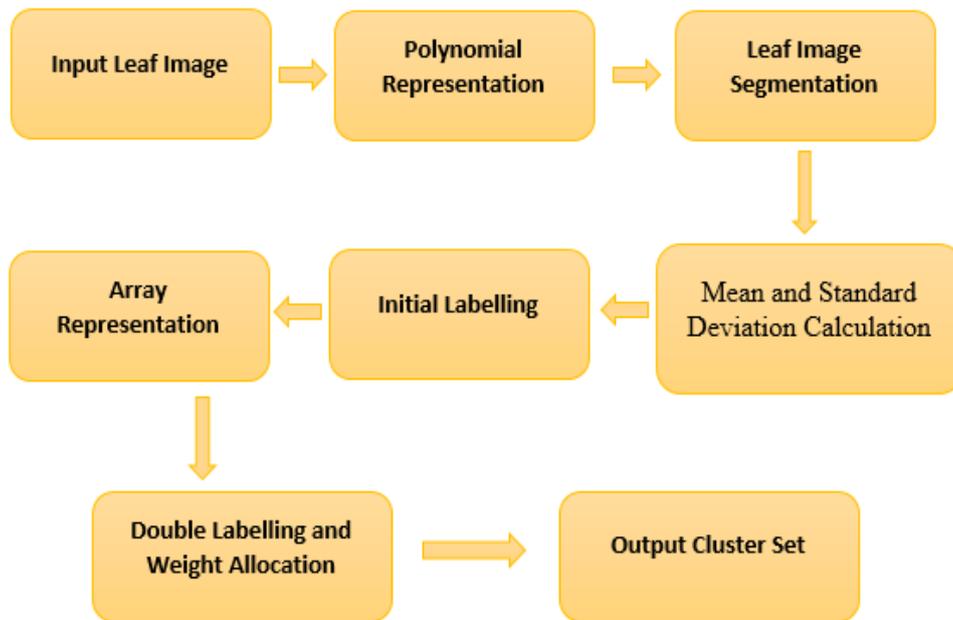
Perimeter: the leaf border width is known as the perimeter.

Convex Hull: A convex polygon with the area of all the leaves buried within it is known as a convex hull.

Digital image processing alters an image and carries out assorted duties, including quality enhancement, development, and processing. Many digital image processing processes, including picture acquisition, enhancement, and restoration, feature transformation, morphology, segmentation, definition, and object recognition, make up the foundation of digital image processing. The separation of images into meaningful sections based on criteria such as grey, spectrum, texture, and colour is the fundamental aspect of image processing. Many plant features are valuable for therapeutic purposes. The following strategies employ photographs as its input and output. Segmentation methods segment and characterize an image based on the attributes retrieved from that image.

Photographs are divided into different sections or sections that do not overlap. Data exclusion, topic identification, and features identification are needless segmentation phases. Separated images are typically defined by regions, boundaries, statistical information, and the number of distinct variables. There are many ways in which the pixels in an image might be organized: by colour, intensity, or texture. Segmentation techniques can be classified into two categories: block segmentation methods and layer segmentation methods. To produce reliable results, only models which consider block segmentation are used. Using these two qualities, similarity and discontinuity, the segmentation approach can be employed here. The partitioning of similarities of an image is done by breaking the image into sections along the lines where a partitioned image divides into border regions, as the intensity of pixels change suddenly. The model separates leaf images into different partitions in order to facilitate the proper evaluation of medicinal plant species. Figure 3 depicts the proposed segmentation model framework.

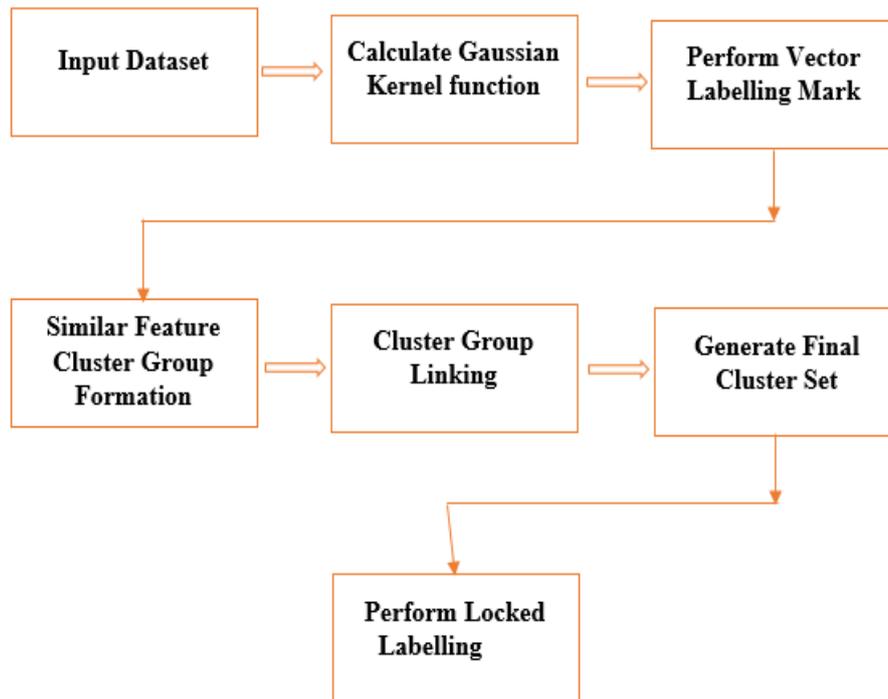
Fig. 3 - Proposed Segmentation Process



Plant leaves or root powder are the base materials for making medicinal goods. It will be necessary to have more experience in determining the medication through pharmacognosis if the dosage of the herbal drug to powder is reduced. Plants that are inaccurate can harm patients significantly. The standardization and quality control of medicinal products rely on the precise identification of the powder shape of medical plants. Medicinal plants are currently grouped into three main groups: biological assessment and physical evaluation based on chemical leaves, as well as a chemical analysis of the plant. Medicinal plants play a critical role in the manufacture of pharmaceuticals. Besides their appearance, leaf shape, colour, and texture are key features for determining whether a medicinal plant is present.

Deterministic species recognition criteria are present on both sides of the leaf, along with a wide variety in colour and texture. A database of medicinal plants leaves is built by scanning the front and back of regularly used medicinal plants and uploading the photos into a database. The leaves are divided into several groups based on the same set of features. When trying to determine intra-class changes, multiple samples of leaves of plants are analyzed by hierarchical clustering algorithms, which groups together leaves based on similarities. In hierarchical clustering technology, the coefficient of inconsistency is employed to construct natural clusters. A closer look at the plant species results in the discovery of clusters that can be used to separate within the class. The sum of the associated sample vectors is found for a single cluster representation. Cluster groups of entities connected by links are outlined in Figure 4.

Fig. 4 - Proposed Model Architecture

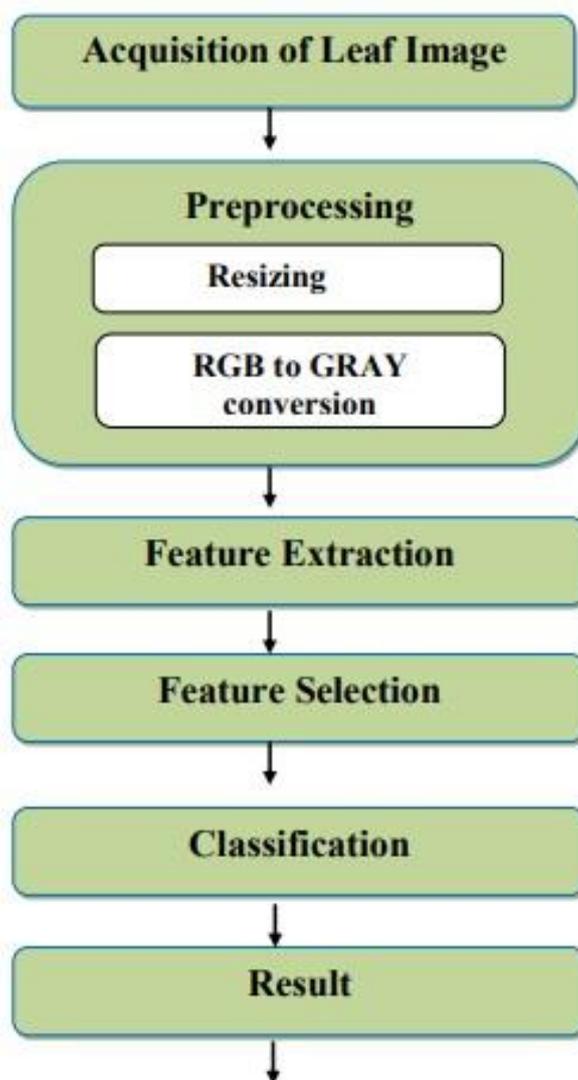


Discovering novel plant discoveries and computerizing the management of plant species are both challenging tasks in biology and agriculture. A plant must also be observed in order to help farmers spread garbage. The identification is a process that involves exposing all of the plants to each other on a spectrum of similar traits. While environmental protection, plant resources surveys, and learning applications are all vital, automatic classification systems place their reliance on a diverse array of programs including those which include environmental protection, plant resources surveys, and learning. The biggest difficulty is in deciding which plant group belongs to which plant. A number of experiments were done to allow this automated categorization project to proceed when the input image is supplied.

An identity mark is best manifested in the form of a pattern or shape. In the same way that any plant recognition algorithm can be used to process images, an advanced variation of the algorithm can be used on the image. Another identifying aspect is the texture. It's a vital part of the rehabilitation process. Color serves as an effective way of differentiating diverse species of plants, as their leaves change colour. However, when employed as a feature, chlorophyll may vary. Leaves are organ plants, which are necessary for plant survival. It is uncomplicated and straightforward to feed into a computer. A tremendous number of images of different kinds of leaf specimens are available for plant species identification in natural history institutions all around the world. Even though there are an almost infinite number of image databases on leaves, it is absolutely necessary to know the plant's

name in order to search for images. All the steps involved in identifying medicinal plants are clearly demonstrated in Figure 5.

Fig. 5 - Medical Plant Recognition Process



The proposed project aims to develop a system that automates leaf identification. A computer vision and machine learning-based automated method for locating medicinal plants was proposed. Features such as the leaf's longitudinal, distance, perimeter, size, number of vertices, colour, and hull surface have been retrieved from each leaf. Based on the different derived qualities, these attributes have specific values assigned to them. The data complexity has exponentially increased due to the increase in genomic data, which has caused the selection techniques to turn into a game-changer that is capable of minimizing data complexity dramatically and making interpretation and conversion of data into useful information significantly easier.

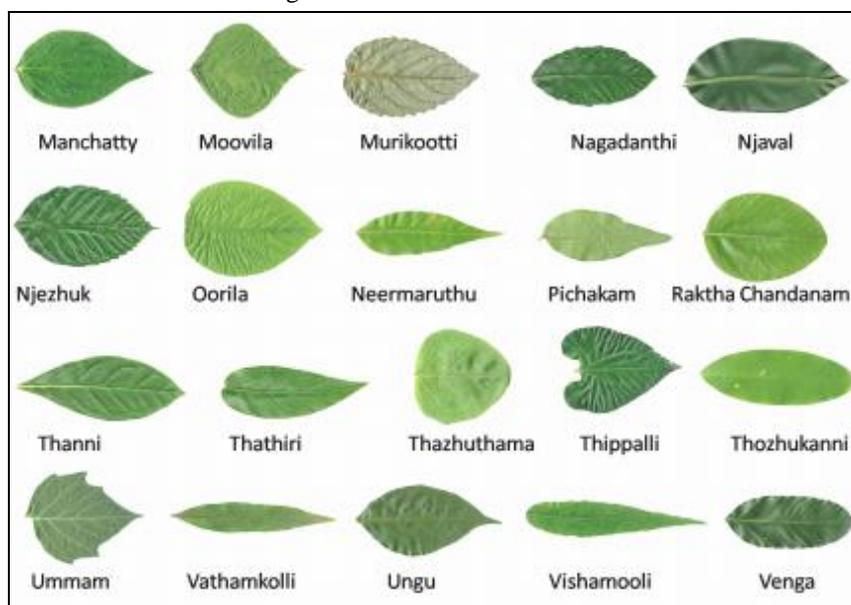
4. Results

ANACONDA SPYDER is built on the proposed image segmentation model and uses the leaf image dataset available on the link provided at the top of this page: <https://data.mendeley.com/datasets/nnytj2v3n5/1>. The result of various picture segmentation analysis was measured in all processing factors in order to increase consistency. By calculating a selection of images and their accompanying problems, in each step of the processing, the processing selects images that recognise each disadvantage, as well as ones that propose, simulate, and assess solutions for optimal and effective solutions. In the comparison between the suggested Double Labelling Image Segmentation Model (DLISM) and the Adaptive Dropout Depth Calculation (ADDC), the proposed model has greater accuracy and efficiency than the ADDC approach. Fig. 6 shows how to correctly identify the medical leaves, and Fig. 7 shows numerous examples of medical leaf structures.

Fig. 6: Medical Leaf Extraction

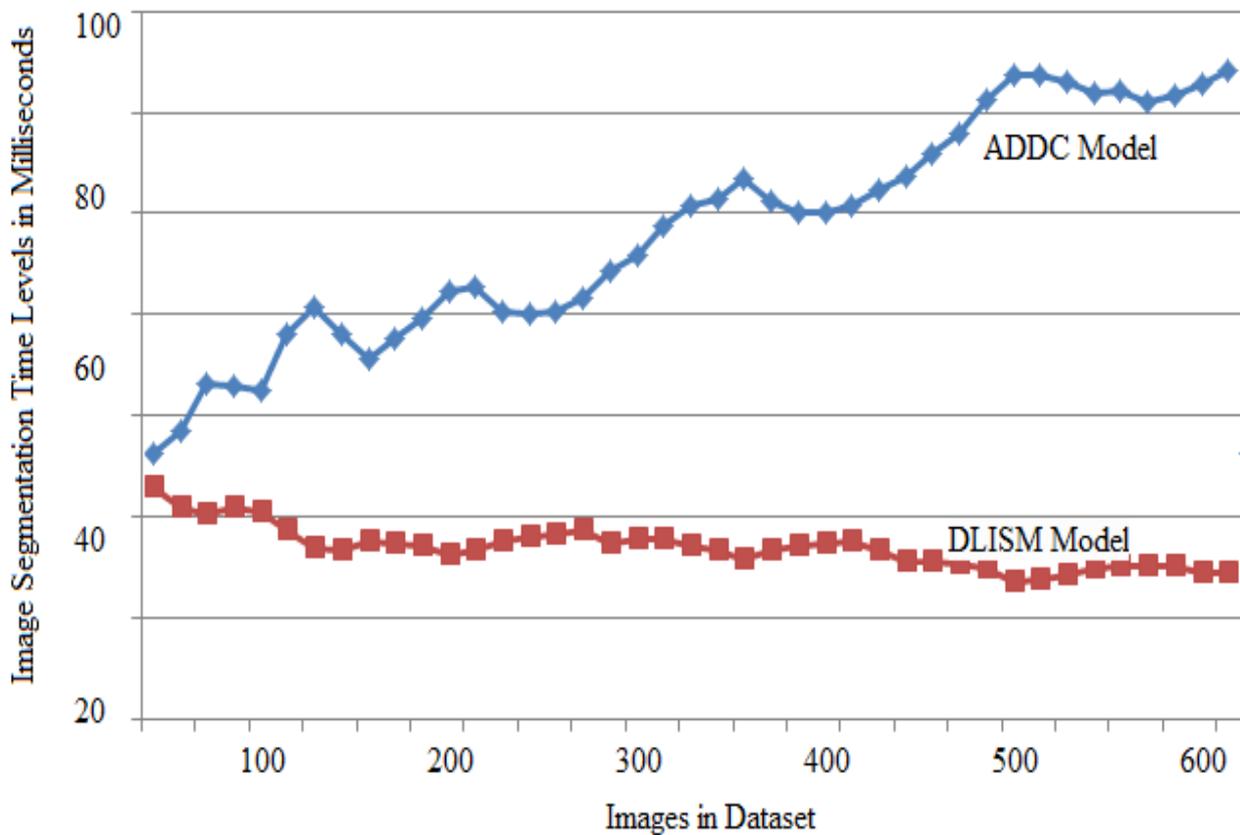


Fig. 7 - Medical Leaf Structures



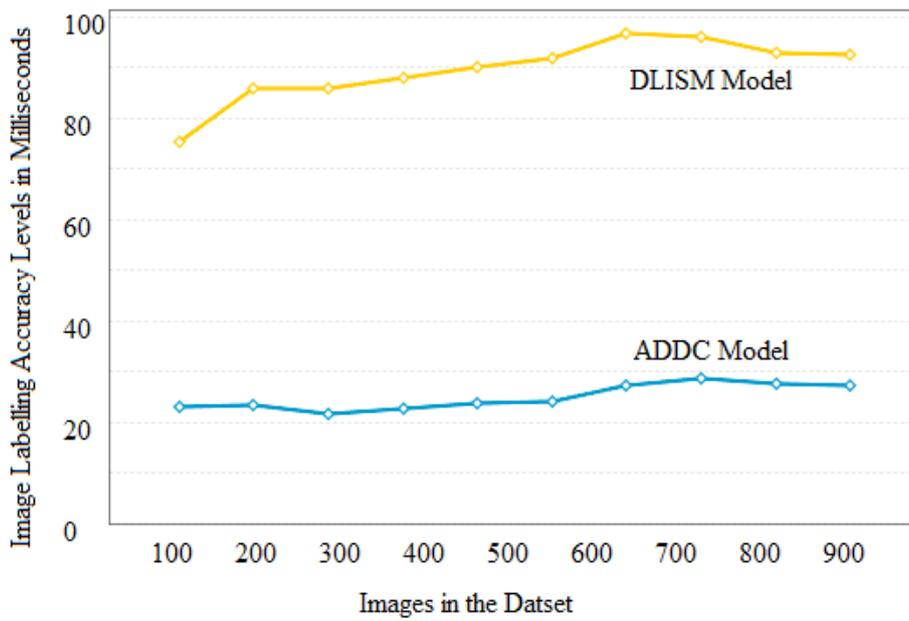
Subdividing photos into numerous pieces is used for additional investigation because it is focused on getting the sole detail or a fragment. Fractal imagery is a fractal process that generates fractal images based on specific aspects, such as colour, texture, pixel strength, etc. Thresholding, region-based approach, edge-based method, clustering method, and the like are often employed techniques. As seen in Figure 8, image segmentation involves the various temporal levels. The suggested approach spends less time on image segmentation than the traditional method.

Fig. 8 - Image Segmentation Time Levels



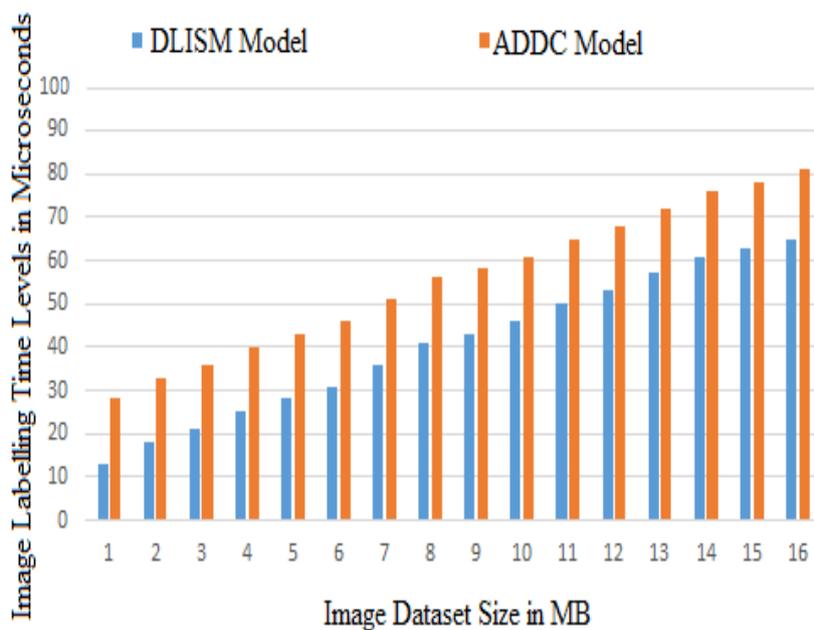
A crucial aspect of our image processing model is the segmentation of images. This is employed everywhere, because it lets our model perceive pixels inside the image. In the image, the images are separated into smaller sections. There's the issue of how much of the image is being supported. Segmentation must end after an image object is properly isolated. The framework only applies appropriate features. The figures in Figure 9 show the proposed methods and those already in use.

Fig. 9 - Image Labelling Accuracy Levels



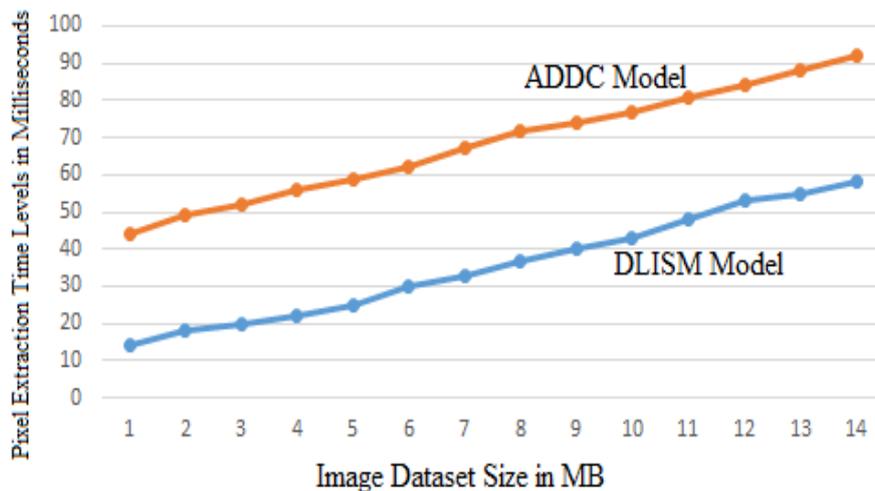
The figure shown in Figure 10 clearly depicts the time periods for labelling of the proposed and existing version of the image. The paradigm proposed involves pixel labelling as a fundamental component in image processing. It is easier to further process the division of the image into many chunks after the data processing has been completed. Object segmentation enhances image accuracy while reducing loss. In order to investigate these pixels for the identification of medical plants, segmented and extracted pixels are used.

Fig. 10 - Image Labelling Time Levels



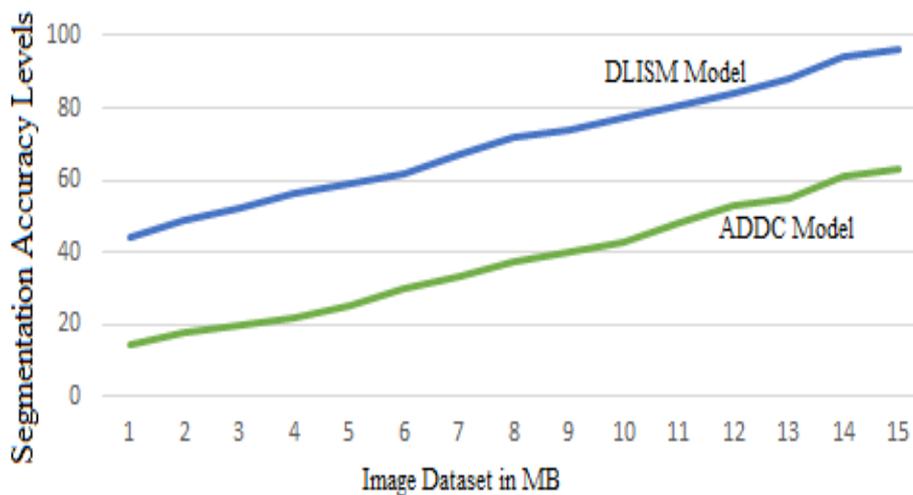
Dimensional reduction by dividing and lowering raw data into smaller groups that may be controlled begins with the extraction of pixels. Many variables serve as the key aspect in these enormous data sets. In order to accomplish this, various computer resources are necessary to process these variables, and then pixel extraction can be used to choose and combine variables into functions, resulting in better feature extraction in huge data sets through the selection and combination of variables. See Figure 11 for the estimates of the proposed and traditional pixel extraction times. On average, the suggested system pixel extraction times are lower than those of competing systems.

Fig. 11 - Pixel Extraction Time Levels

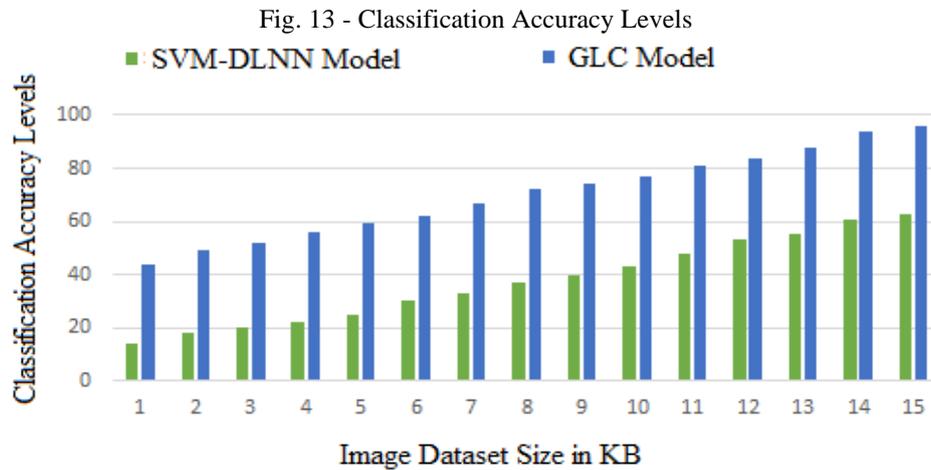


The image segmentation process is intricate and difficult, due to irregularity impacting various parameters, such as low contrast, light, and the presence of noise. Figure 13 shows the results of the segmentation degree of precision being compared to the present model.

Fig 13: Segmentation Accuracy Levels



When measuring accuracy, cost-sensitive classification systems are calibrated with a mistake that is designed to enhance the classifier. When you misclassify anything, you also pay a price. For error-based classification systems, failure in classification is similarly popular; in other words, failure in classification is widely accepted, even though it does not apply in all real-life applications. Classification accuracy levels are illustrated in Figure 13. The proposed model classification accuracy is significantly greater when compared to conventional methods.



The forecast of a given class is a matter of the class's attributes pertaining to data. It is important to remember that class measurements are used to determine information generalization. From Figure 14, it is apparent that suggested and actual classification systems go through several stages of development. Classification of the proposed model is projected to be done with less time levels than with the existing system.

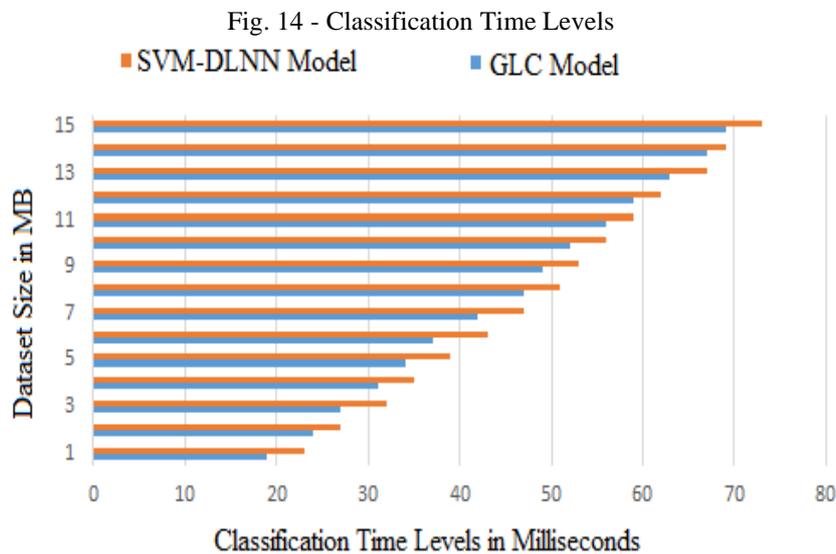
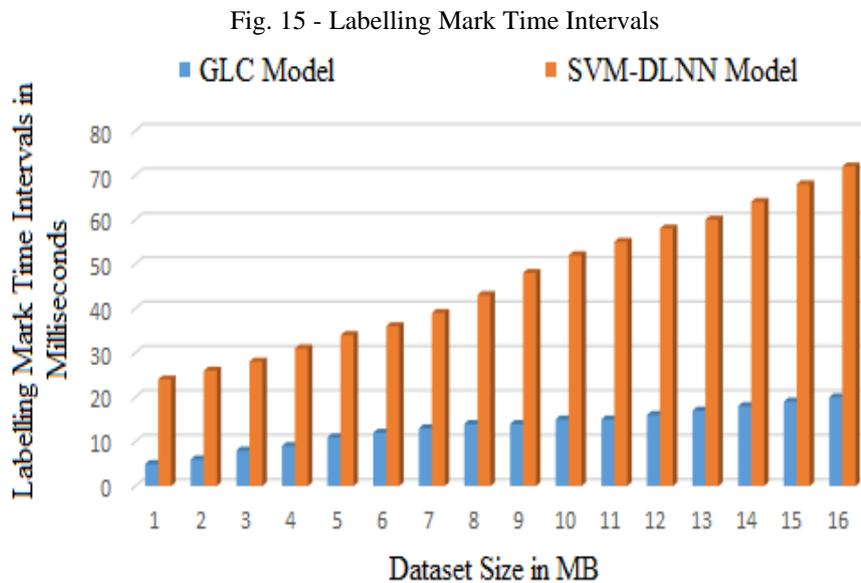
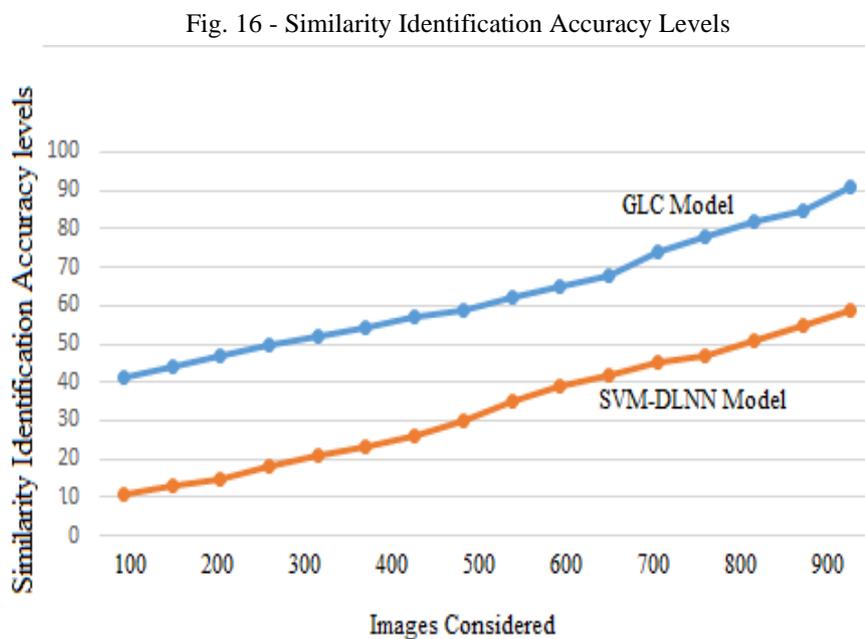


Figure 15 depicts the time intervals of classifying both current and proposed models. Using the new model, we are able to save more energy as compared to the previous models that enhance system efficiency.



In Figure 16, the similarity values of the characteristics are identified and the accuracy of the similarity detection is given. In this proposal, model accuracy levels are above what is already practiced.



The proposed model minimizes the cost and improves the performance levels by performing feature optimization. Figure 17 depicts the optimization time levels of the proposed and current models.

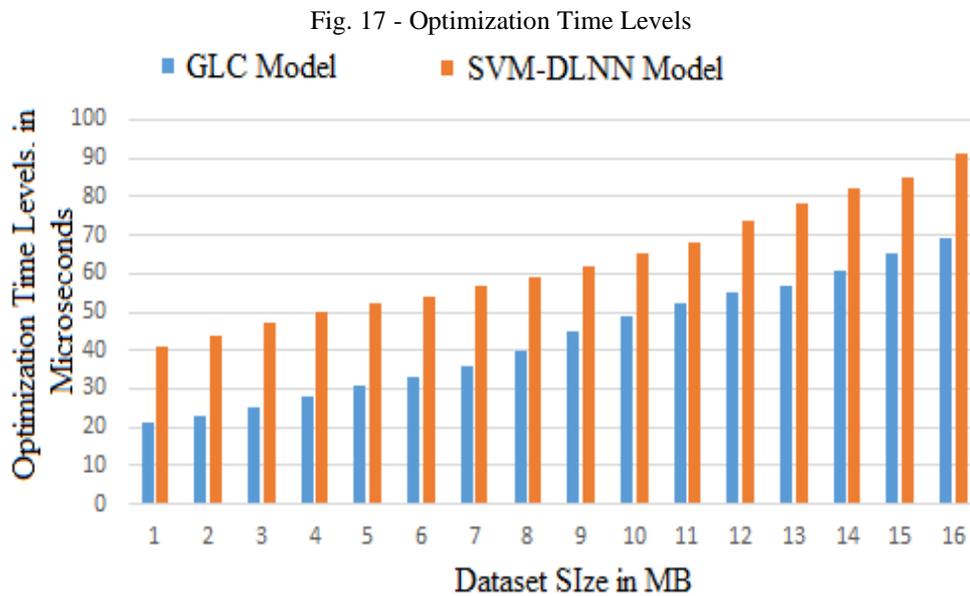
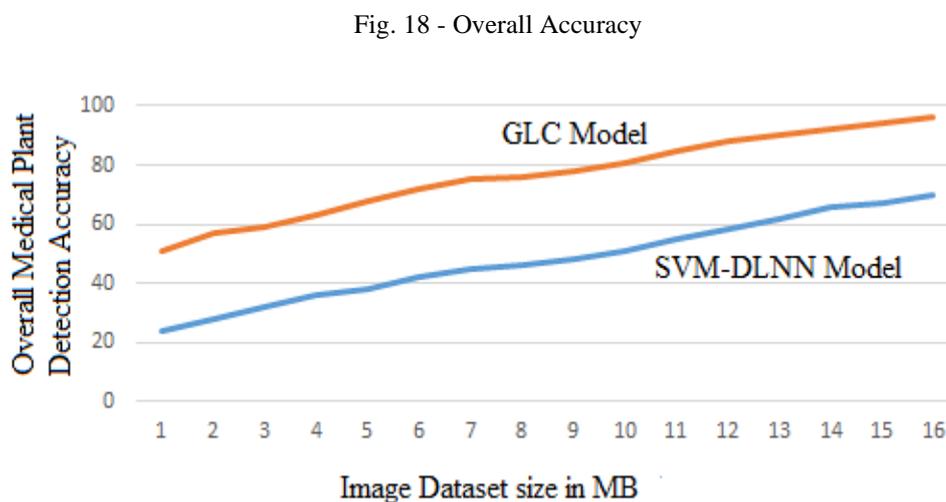


Fig. 18 shows the overall accuracy of the medical plant detection. The proposed model has a detection rate that is higher than the existing model. Compared to traditional methods, the proposed medical plant measuring accuracy is superior.



Overall, the medical plant detection accuracy shown in Figure 18. Table 1 displays the many characteristics of the proposed model compared to current approaches, including the efficiency levels.

Table 1 - Parameter Values

Parameters	LTrP	ELTrP	Hybrid method, ELTrP with LBP
Correct rate	0.8723	0.9333	0.9556
Error rate	0.1277	0.0667	0.0444
LastCorrectRate	0.8723	0.9333	0.9556
LastErrorRate	0.1277	0.0667	0.0444
InconclusiveRate	0.0600	0.1000	0.1000
Classified Rate	0.9400	0.9500	0.9549
Sensitivity	1	1	1
Specificity	0.9388	0.9480	0.9490
PositivePredictiveValue	0.2500	0.1667	0.1667
NegativePredictiveValue	1	1	1
Prevalence	0.0200	0.0200	0.0200

Training and accuracy levels of the proposed model are displayed in Table 2, where batch size of photos is evaluated. model performance values are better than proposed.

Table 2 - Accuracy Levels

Training Parameters			Outputs	
Mini-Batch Size	Initial Learn Rate	Epochs	Accuracy(%)	Training Time (min)
5	0.0001	10	98.99	7.31
10	0.0001	10	98.14	6.88
15	0.0001	10	98.82	6.70
25	0.0001	10	98.31	6.73
15	0.0001	5	97.64	3.41
15	0.0001	10	98.99	6.926
15	0.0001	15	98.65	10.20
15	0.0001	25	98.31	13.76

5. Conclusion

A detailed description of Ayurvedic history holds that each plant is therapeutically beneficial, meaning it is essential for people to know which part of the plant has medical benefit for which ailment. Medicinal plants, such as leaves, bark, nuts, berries, roots, and stem, are commonly found in the medical literature as the diagnoses for various diseases. Botanists and herbalists manually identify therapeutic plants based on the time it takes for these plants to do so. The study's goal is to use imaging techniques to save manual labour while enhancing productivity by automatically identifying therapeutic plants. Floral and Fruit and Seed classification of medicinal plants receive far less attention compared to classification of medicinal plants using leaves. Our goal is to boost plant identification and classifying by using flowers, fruits, and seeds in addition to leaves. This system can give individuals and farmers access to the most vital therapeutic plants by identifying and classifying plants automatically. In this model, botanists, end users, forest services, taxonomists, pharmaceutical corporations, and Ayurvedic practitioners will work together to identify and classify medicinal plants without the intervention of humans. The proposed model accuracy rates are high in comparison to traditional methods since the proposed model's detection rate and the time it takes to conduct the investigation are both quick.

References

- Manojkumar P., Surya C.M., and Varun P. Gopi, "Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples", *2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 978-1- 5386-1931-5
- K. Nithiyanandhan and Prof. T. Bhaskara Reddy, "Analysis of the Medicinal Leaves by using Image Processing Techniques and ANN", 8(5), ISSN No. 0976-5697, May-June 2017.
- Adams Bogue, Venitha Kowlessur, Fawzi Mahomoodally, Upasana Singh and Sameerchand, "Automatic Recognition of Medicinal Plants using Machine Learning Techniques", *International Journal of Advanced Computer Science and Applications*, 8(4), 2017.
- H.X. Kan, L. Jin, and F.L. Zhou. Classification of Medicinal Plant Leaf Image Based on Multi-Feature Extraction. *Pattern Recognition and Image Analysis*, 27(3), 2017, 581–587, 1054-6618. © Pleiades Publishing, Ltd.
- Riddhi H. Shaparia, Dr. Narendra M. Patel and Prof. Zankhana H. Shah," Flower Classification using Texture and Color Features", *International Conference on Research and Innovations in Science, Engineering & Technology*, Volume 2, 2017, Pages 113–118.
- Marco Seeland, Michael Rzanny, Nedal Alaqraa, Jana Wa ¨ldchen, Patrick Ma ¨der, "Plant species classification using flower images—A comparative study of local feature representations", *PLOS ONE*. DOI:10.1371/journal.pone.0170629 February 24, 2017.

Pradeep kumar Choudhary, Rahul Khandekar, Aakash Borkar, and Punit Chotaliya. Image processing algorithm for fruit identification. *International Research Journal of Engineering and Technology (IRJET)*, Vol 4 Issue 3, 2017.

D Venkataraman and Mangayarkarasi N, "Computer Vision Based Feature Extraction of Leaves for Identification of Medicinal Values of Plants", *IEEE International Conference on Computational Intelligence and Computing Research*, 978-1-5090-0612-0/16/\$31.00 ©2016 IEEE.

Ruaa Adeeb Abdulmunem Al-falluji, "Color, Shape and Texture based Fruit Recognition System", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 5(7), 2016.

Pushpa BR, Anand C and Mithun Nambiar P," Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images", *International Journal of Applied Engineering Research*, 11(7), 2016, 0973-4562.

Sana O M1, R. Jaya, "Ayurvedic Herb Detection Using Image Processing", *International Journal of Computer Science and Information Technology Research*, 3(4), 134139, Month: October - December 2015.

C.H. Arun, W.R. Sam Emmanuel, and D. Christopher Duraira," Texture Feature Extraction for Identification of Medicinal Plants and Comparison of Different Classifiers", *International Journal of Computer Applications*, 62(12), 2013, 0975 - 8887.

Sandeep Kumar E, "Leaf color, area and edge features based approach for identification of Indian medicinal plants", *Indian Journal of Computer Science and Engineering (IJCSE)*, 3(3) 2012, 0976-5166.

Basavaraj S. Anami, Suvarna S Nandyal, Govardhan. A, "Color and Edge Histograms Based Medicinal Plants' Image Retrieval", *I.J. Image, Graphics and Signal Processing*, 2012, 8, 24-35, 2074-9074.

Basavaraj S. Anami, Suvarna S. Nandyal, and A. Govardhan, "A Combined Color, Texture and Edge Features Based Approach for Identification and Classification of Indian Medicinal Plants", *International Journal of Computer Applications*, 6(12), September 2010.

Warisara Pardee, Prawaran Yusungnern, Peeraya Sripian, "Flower Identification System by Image Processing", *3rd International Conference on Creative Technology CRETECH 2015*, 1, August 2015.

Zaidah Ibrahim, Nurbaity Sabri, Nur Nabilah Abu Mangshor," Leaf Recognition using Texture Features for Herbal Plant Identification", *Indonesian Journal of Electrical Engineering and Computer Science*, 9(1), 2018, 152~156.

A Gopal, S. Prudhveeswar Reddy and V. Gayatri, "Classification of Selected Medicinal Plants Leaf Using Image Processing", 978-1-4673-2322-2.

R Janani and A Gopal, "Identification of selected medicinal plants leaves Using Image Features and ANN", *2013 International Conference on Advanced Electronic System (ICAES)*, 978-1-4799-1441-8.

Anilkumar. Muthevi and Dr. Ravi babu. Uppu, "Leaf Classification Using Completed Local Binary Pattern of Textures", *2017 IEEE 7th International Advance Computing Conference*, 978-1-5090-1560-3.

Iyos Kusmana, Yeni Herdiyeni, "Fusion of Local Binary Patterns Features for Tropical Medicinal Plants Identification". *2013 International Conference on Advanced Computer Science and Information Systems (ICACISIS)*. 353-357.

C. Ananthi, Azha. Periasamy and S. Muruganand, “Pattern Recognition of Medicinal Leaves Using Image Processing Techniques. *Journal of Nano Science and Nano Technology*, 2(2), Spring Edition, ISSN 2279 – 0381.

Mohd Shamrie Sainin, Taqiyah Khadijah Ghazali and Rayner Alfred, “Malaysian Medicinal Plant Leaf Shape Identification and Classification”, *Knowledge Management International Conference (KMICe)* 2014, 12 – 15 August 2014, Malaysia.

Iyan Mulyana, Yeni Herdiyeni and Sony Hartono Wijeya “Identification of medicinal plant based on fractal by using clustering fuzzy c-means”, *International Conference on Information Technology and Business Application Palembang*, Indonesia, 22-23 February 2013.

https://www.ayurveda.com/pdf/intro_ayurveda.pdf

B.S. Harish, “*Classification of plant leaves using Morphological features and Zernike moments*”, 978-1-4673- 6217-7, 2013 IEEE.

Rafael C. Gonzalez and Richard E. Woods (2014), *Digital Image Processing*, 3rd ed., Pearson publication, India.