Automatic Recognition of Medicinal Plants using Machine Learning Techniques

S. Vinothini[1], J. Agnes[2]

[1]Student, M.E, Dept. of Computer Science and Engineering, T.J.S Engineering College, Tamil Nadu, India
[2]Assistant Professor, Dept. of Computer Science and Engineering, T.J.S Engineering College, Tamil Nadu, India

---

Abstract - *A fully automated method for the recognition of medicinal plants using computer vision and machine learning techniques has been presented. Leaves from 24 different medicinal plant species were collected and photographed using a smart phone in a laboratory setting. A large number of features were extracted from each leaf such as its length, width, perimeter, and area, number of vertices, color, perimeter and area of hull. Several derived features were then computed from these attributes. The best results were obtained from a random forest classifier using a 10-fold cross-validation technique. With an accuracy of 90.1%, the random forest classifier performed better than other machine learning approaches such as the k-nearest neighbor, naïve Bayes, support vector machines and neural networks.*

Key Words: Medicinal Plants, attributes, features, classifier cross-validation, leaves

1. INTRODUCTION

Identifying unknown plants relies much on the inherent knowledge of an expert botanist. The most successful method to identify plants correctly and easily is a manual-based method based on morphological characteristics. Thus many of the processes involved in classifying these plant species are dependent on knowledge accumulation and skills of human beings. However, this process of manual recognition is often laborious and time-consuming. Hence many researchers have conducted studies to support the automatic classification of plants based on their physical characteristics. Systems developed so far use a varying number of steps to automate the process of automatic classification, though the processes are quite similar. Essentially, these steps involve preparing the leaves collected, undertaking some pre-processing to identify their specific attributes, classification of the leaves, populating the database, training for recognition and finally evaluating the results. Although leaves are most commonly used for plant identification, the stem, flowers, petals, seeds and even the whole plant can be used in an automated process. An automated plant identification system can be used by non-botanical experts to quickly identify plant species quite effortlessly.

2. EXISTING SYSTEM

This research proposed a new mobile application based on Android operating system for identifying Indonesian medicinal plant images based on texture and color features of digital leaf images. In the experiments we used 51 species of Indonesian medicinal plants and each species consists of 48 images, so the total images used in this research are 2,448 images. This research investigates the effectiveness of the fusion between the Fuzzy Local Binary Pattern (FLBP) and the Fuzzy Color Histogram (FCH) in order to identify medicinal plants. The FLBP method is used for extracting leaf image texture. The FCH method is used for extracting leaf image color. The fusion of FLBP and FCH is done by using the Product Decision Rules (PDR) method. This research used Probabilistic Neural Network (PNN) classifiers for classifying medicinal plant species. The experimental results show that the fusion between FLBP and FCH can improve the average accuracy of medicinal plants identification. The accuracy of identification using fusion of FLBP and FCH is 74.51%. This application is very important to help people identify and find information about Indonesian medicinal plants.

3. PROPOSED WORK

The proposed technique was tested on a dataset of 55 medicinal plants from Vietnam and a very high accuracy of 98.3% was obtained with a support vector machines (SVM) classifier. The size of each image was 256*256 pixels. Proposed an approach based on fractal dimension features based on leaf shape and vein patterns for the recognition and classification of plant leaves. Using a k-nearest neighbor classifier with 20 features, they were able to achieve a high recognition rate of 87.1%. Using a volumetric fractal dimension approach to generate a texture signature for a leaf and the GLCM (Gray level co occurrence matrix) algorithm.

4. METHODOLOGY

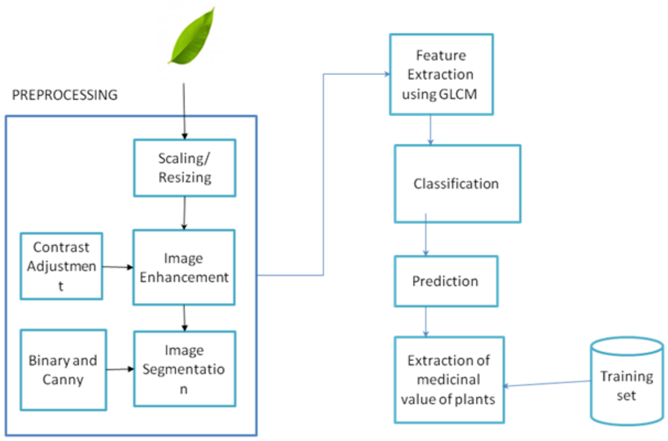

Fig -1: Proposed Methodology

4.1 Preprocessing

Preprocessor does Preprocessing of the input image. There are 130 leaf images in our trained set. Pre-processing is a common name for operations with images at the lowest level of abstraction both input and output are intensity images The aim of Pre-processing is to scale the image size and improve the contrast of image that suppresses unwanted distortions or enhances some image features important for further processing. The next step is to perform a thresholding operation which will convert the image into a binary image with only two values: black and white pixels. This is achieved using the Binary and canny method. This operation is important in order to clear the image from many small noisy pixels, which are the artifacts of the thresholding operation. Preprocessor initially does Scaling or Resizing and gives Query Leaf image. Later the Image enhancer adjusts the Contrast of the resized image. The Image Segmenter segments the image using Binary and Canny. Now the preprocessor gives a noiseless segmented image as input to the Feature Extractor.

4.2 Feature Extraction

The feature selection method is applied to take out unnecessary or irrelevant redundant properties from the set of extracted features that do not possess a substantial contribution to the accuracy of the classification model or may even reduce the model accuracy. GLCM method is used to picks up the relation between two pixels at a time, called the reference and the neighbour pixel. GLCM expounds the distance and angular spatial relationship over an image sub-region of specific size. GLCM is prepared from gray scale values. It is taken into account how often a pixel with gray level(gray scale intensity or gray tone) values come either horizontally, vertically and diagonally to leveled the pixels with the value j. GLCM directions are: Horizontal(0) Vertical(90) Diagonal a)bottom left to top right(-45) b)top left to bottom right (-135). Feature Extractor uses Gray-Level Co-Occurrence Matrix(GLCM) method to extract features like grayco properties such as contrast, correlation, energy, homogeneity, mean, standard deviation, variance etc and region properties like area, centroid, bounding box, convex hull, eccentricity etc. We extract nearly 25 features which produces more accuracy while identifying leaf. We train 130 leaf images in the dataset and store the identified feature values. The features such as area, perimeter, convex hull mean, standard deviation, variance, horizontal distance, vertical distance, bounding box etc are extracted using glcm. In total, 25 features are extracted to achieve higher level of accuracy.

4.3 Classification of Medicinal Leaves

Support vector machine isolates the input data linearly once it is directed to high dimensional space non-linearly and subsequently support vector machine is classified as non-linear classifier. This phenomenon of SVM enhances the classification performance. Basically SVM is a machine learning tool which has turned out as one of the capable problem solving method, problem related to binary classification and learning from data. SVM's main concept between the two transformed space classes is, using kernel function transforming given data into higher dimensional space and developing OSH (Optimal Separating Hyperplane).

This OSH is built by maximizing the edge between classes. A line is constructed in the changed space and the information vectors which are closest to this constructed line are known as support vectors. Function used for classifying the data into two unique classes is inspected by the SVM. Input vectors are linearly isolated which are put into high dimensional feature space by using non-linear transformation which is dependent on a regularization parameter. Inner product (x,y) is supplanted to construct a non-linear product by kernel function K (x,y) which can be expressed as below: Where f indicates the member of x. From given kernel set, the basis K(xi,x) where i=1,2,....N is selected using the first layer of the SVM and linear function in the space is created by second layer. SVM is independent of dimensionality of the input space. It has simple geometric construction and it generates spare solution. Classification is performed by support vectors n large number which is obtained by training set. We then compare the trained features with the features extracted by us along with the labels of the leaf using SVM Classifier. We use multi svm with these three parameters to compare and give the result based on the label. We predict the accurate matching and display the leaf with its name, medicinal values and the place it grows by comparing with the trained leaf dataset.

4.4 Dataset

The dataset used in our experiment is collected by our self. We surfed the name, medicinal value of the plant by self. We have taken 130 images as training set and each plant class contains the 10 leaf images in different degrees of rotation and different leaf images. After classification, we compared the label value, trained features, the features we extracted and the system displays the results of the name of the plant, its medicinal values and the place it grows.

4.5 Leaves Recognition

The various outputs are shown below,

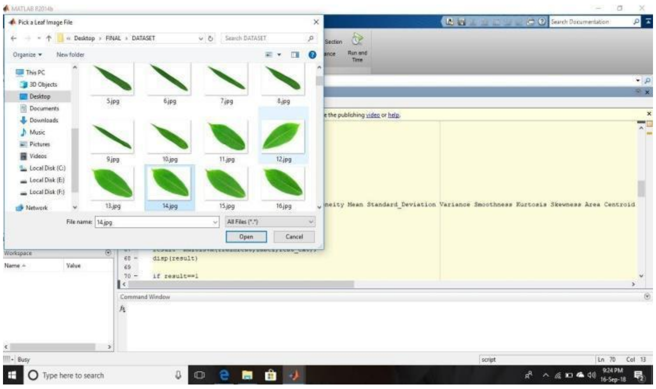

Fig -2: Leaves Dataset

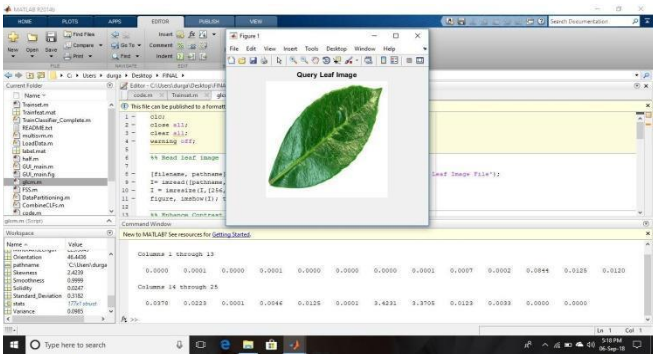

Fig -3: Resized Image

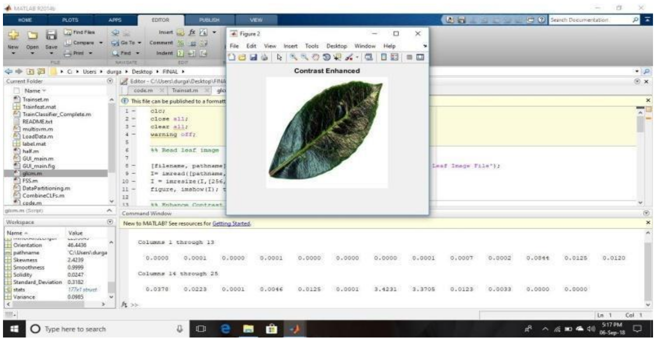

Fig -4: Contrast Enhanced Image

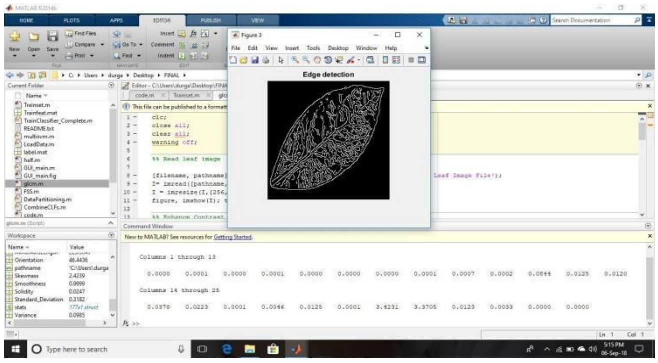

Fig -5: Canny Edge Detection

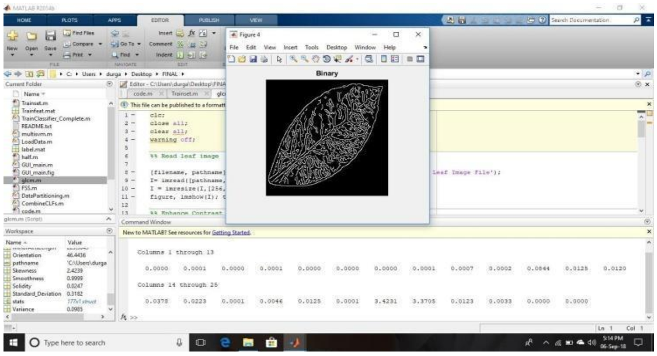

Fig -6: Binary Image

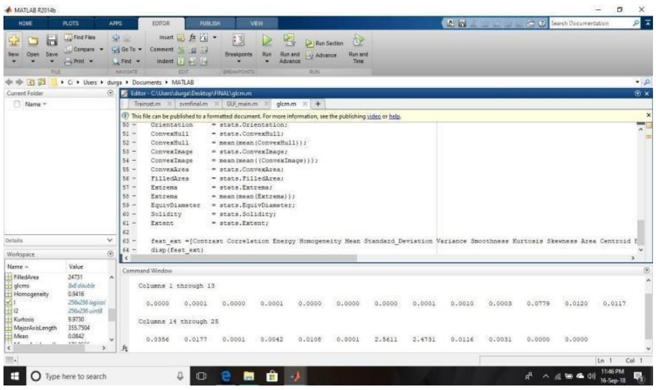

Fig -7: Feature Extraction Values

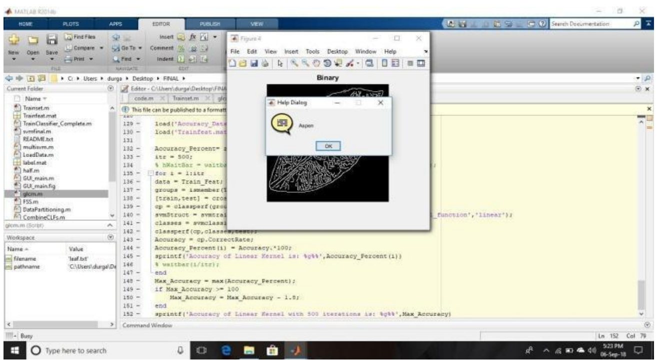

Fig -8: Classification Results

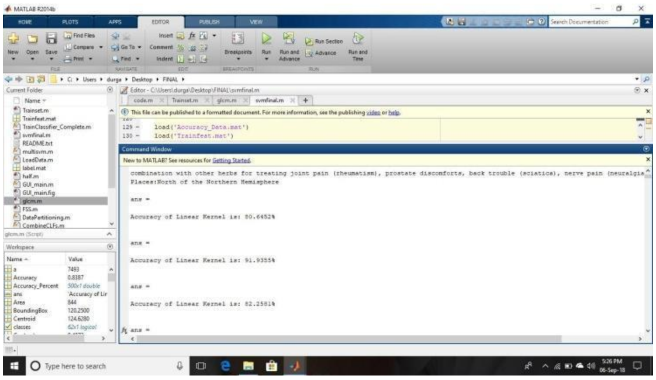

Fig -9: Medicinal value and the place it grows

5. CONCLUSION

A new dataset on medicinal plants of Mauritius has been made publicly available on the machine learning repository portal. In this paper, computer vision techniques have been used to extract several shape-based features from the leaves of medicinal plants. Machine learning algorithms were then used to classify the leaves from 13 different plant species into their appropriate categories. The accuracy of our proposed method is comparatively higher than Existing system. For future research, in an attempt to achieve even higher accuracies, probabilistic neural networks and deep learning neural networks would be investigated.

6. FUTURE ENHANCEMENTS

My future goal is to extend the current image processing desktop application to an android based smart-phone app, one way to effectively achieve is to offload the processing of algorithms on to a high performance server over the network and this would typically work as a client server image processing system. Using the Android based smart-phone; the input image to be identified is captured and sent to a server (eg. PHP Server) via HTTP. A script on the server invokes the server-side application (eg. MATLAB) to process the image which in turn sends the results to the smartphone via http. In an attempt to achieve even higher accuracies, probabilistic neural networks and deep learning neural networks would be investigated. Different datasets may be incorporated to validate the obtained results and other effective classifiers may be applied. Additionally, deep learning technique can be used. This work is performed on the regional leaves. It can also be extended to a larger number of leaf classes with improved efficiency. More extraction features can be used using increased number of leaf dataset in future work for leaf classification.

REFERENCES

[1] Gao, W. and Lin, W., 2019. Frontal Parietal Control Network Regulates the Anti-Correlated Default and Dorsal Attention Networks. Human Brain Mapping.

[2] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L., 2017. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. 7th IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 11-16.

[3] Zhang X., Liu Y., Lin H., Liu Y. (2018) Research on SVM Plant Leaf Identification Method Based on CSA. In: Che W. et al. (eds) Social Computing. ICYCSEE 2016. Communications in Computer and Information Science, Vol 624, Springer, Singapore.

[4] Hossain, J. and Amin, M.A., 2019. Leaf Shape Identification Based Plant Biometrics. 13th International Conference on Computer and Information Technology, Dhaka, Bangladesh, 458-463.

[5] Du, J.X., Wang, X.F. and Zhang, G.J., 2017. Leaf shape based plant species recognition. Applied Mathematics and Computation, 185, 883-893.

[6] Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for Plant Leaf Image Classification. 5th International Conference on Emerging Intelligent Computing Technology and Applications, Ulsan, South Korea, 627-634.

[7] Herdiyeni, Y. and Wahyuni, N.K.S., 2020. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. International Conference on Advanced Computer Science and Information Systems (ICACSIS), West Java, Indonesia, 301-306.

[8] Prasvita, D.S. and Herdiyeni, Y., 2019. MedLeaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image. International Journal of Advanced Science, Engineering and Information Technology, 3, 5–8.