International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **05** Special Issue **01** | **2020**

Comparison of Classification Algorithms for Identification of Ayurvedic Plant Species

P.P. Sadgir[1] and Dr. V.R. Ratnaparkhe[2]
[1,] Research Sholar, Department of Electronics Engineering, Government College of Engineering,Aurangabad
[2] Department of Electronics Engineering, Government College of Engineering, Osmanpura, Station Road, Aurangabad, Maharashtra 431005
Email: psadgir@yahoo.in

Abstract

Identification will facilitate for improving quality of medicines if plant leaf is scrutinized before extraction process takes place. This work proposes method for precise plant identification to facilitate the operator with limited domain knowledge. Medicinal plants which are used as analgesic and anti-inflammatory, typically *Neem, Datura, Aloe, Tulsi, Nirgundi, Parijat, Rui, Bel, Castor* are selected for the study. Different classification techniques are implemented in the paper for recognition of species. SVM classification method achieved accuracy of 89 % while tree classifier implemented improves accuracy with reduced operating time.

Keywords. Ayurvedic, Symetry test, Width Plot Analysis, KNN, SVM,Tree Classification.

1. INTRODUCTION

Recent advances in digital technology, coupled with rapidly increasing interest in the creation and dissemination of digitized specimen data for use in broad-scle research by botanists and other organismal scientists, have encouraged the development of a variety of new research opportunities in the botanical sciences. It is now increasingly possible to collect, use, re-use, and share data more easily and effectively. Automation in plant identification facilitates segregation of plants from different species also the intra species classification which is a very huge task. The identification of plants is based on patterns of shape, size, texture, color and phyllotaxy. The identification of plant species by the botanist is usually by considering the study of external traits of plant architecture, especially ontogeny and number of elements forming reproductive organs and dispersion entities. Though important, the reproductive organs are available only for very short period during the year. The reproduction through flowers and fruits occupies a relatively short part of plant life, while leaves are present for most of all of its existence. Plant leaf is widely used for classification owing to its availability and ease of collection in on-field study and database creation. Conventionally the plant identification and classification is done on visual characteristics of plant morphology.

Plant identification is an challenging and important topic for the research as different plant species produce leaves that are very diverse in shape and size, ranging from huge banana leaves to tiny tamarind leaves. Between accessing of species and even within a single plant leaf, e.g. pinus monophylla, characteristics can differ significantly. Many conditions such as water and nutrient availability, light day length and their interactions play crucial role in determining varied shape and size of leaves even within the same species. When leaves from different species are similar in shape, size and venation then conventional morphological features are not sufficient to recognize and determine the discrimination. So in the recent decades computer based analysis enables a more precise identification of plant leaves[1]. This paper presents work on plant classification based on plant leaves with macroscopic data of species useful as analgesic and anti-inflammatory treatment.

2. PROPOSED METHOD

The plants which are anti-analgesic and anti-inflammatory, according to Ayurveda, were selected for study. In processes of production of medicines from plant leaves, identification of species should be precise. The proposed method for identification consists of steps as image acquisition, preprocessing, feature extraction, decision tree classification stages.

2.1. Image Acquisition

Among all the parts of plants, leaves are easily available throughout the year; hence plants whose leaves are useful for medicinal purpose are selected for this project. Nine species of plants, Aegle Marmelos (Bel), Aloe Vera, Ricinus Communis (Castor), Ocimum Tenuiflorum (Tulas), Azadirachta Indica (Neem), Datura, Calotropis Gigantea (Rui), Vitex Negundo (Nirgundi) were considered for the study. Images of ten leaves of each type of plant were selected to create the database. Figure i shows the setup developed for database acquisition with customized illumination and rotating platform, with this rotating platform angular position can be changed with respect to X and Y axis. Figure ii shows outer view of setup. Leaf images from both dorsal as well as ventral side are considered for work. Thus database of each leaf of each plant was obtained with 15 different conditions of illumination, different angular positions of the leaf and from dorsal and ventral sides of leaf. In all 4200 images of each species are collected.

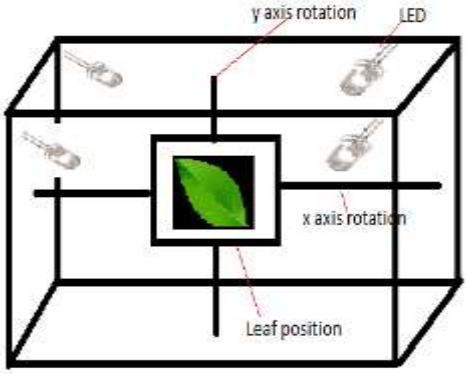


Figure i. Inner view of database creation setup

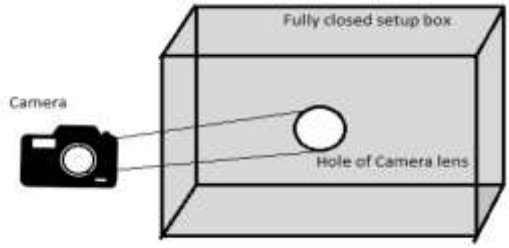


Figure ii. Outer view of the database reation setup

2.2. *Preprocessing*

Preprocessing helps in reducing the noise in image. Improved image quality leads to better accuracy in classification. Implemented preprocessing includes conversion of RGB image to gray scale image, local normalization and background noise cancellation[2,3].

2.2.a. Conversion of RGB image to gray scale image

Color features of leaves are not very significant as all leaves are green in selected species. Hence RGB image is converted to grey scale image before the local normalization stage. RGB to grayscale conversion is done using equation (i)[4,5,6].

$$val = 0.30 * R + 0.59 * G + 0.11 * B \qquad (i)$$

2.2.b. Normalization

Local normalization algorithm is used in this step. To estimate of local mean variance is local spatial smoothing is performed. Fast recursive Gaussian filters were implemented. The local normalization of $f(x,y)$ is computed as follows in equation (ii).

$$g(x,y) = \frac{f(x,y) - m_f(x,y)}{\sigma_f(x,y)} \qquad \text{(ii)}$$

where, $f(x,y)$ is original image, $m_f(x,y)$ is an estimation of mean, $\sigma_f(x,y)$ is an estimation of local variance, $g(x,y)$is the output image[7].

2.2.c. *Background noise cancellation*

While capturing the image, unnecessary background data is also captured. It may lead to false feature calculation. So background noise cancellation stage is incorporated in the work. The algorithm includes thresholding image, erosion operation, identification of all zeros and calculation of corner points of the image considering the major object and finally cropping the image for selected object [4,5,6].

2.3. Feature extraction

Classification of plants can be based on features such as color, texture and shape of the leaf. The proposed system uses shape and texture for identification.

2.3.a. *Ridge filter*

Texture feature is prominent in this form. A ridge filter is used to accentuate the ridge patterns of the leaf surface[8]. The image is binarized with average threshold and the total proportions of white pixels are used to discriminate between plants having single midrib, no midrib and multiple midrib as represented in Figure v.

2.3.b. Shape feature extraction

These are very important features in the leaf segregation. This stage consists of three types of shape measures such as symmetry test, geometrical & morphological features and width plot based parameters.

2.3.b.i. Symmetry test

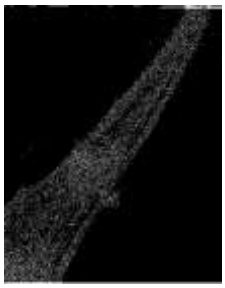
Images indicate that some leaves like *Neem* and *Datura* are not symmetrical along the midrib. This feature is used to classify the plants. The distance between the edges and the midrib is calculated according to equation iii. Equality of both side distances with respect to midrib indicates symmetry of the leaves as shown in figure vi.

$$ST = \sum_{k=i}^{n} (Side\ 1(k) - Side\ 2(k)) \quad \text{(iii)}$$

where k is point selected for test in image

Side 1 is distance between k^{th} point on midrib to right side edge
Side 2 is distance between k^{th} point on midrib to left side edge

(a)
Single midrib



(b)
No midrib

(c)
Multiple midrib

Figure. v. Midrib of leaves from database

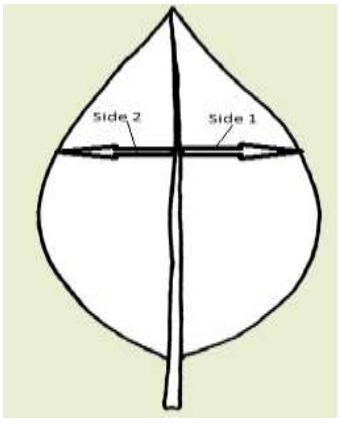


Figure vi. Image showing distances for symmetry test.

2.3.b.ii *Geometrical and morphological features*

The Geometrical and morphological features of the plant leaf is calculated and are concatenated to form the feature vector. Classification based on the features Major axis length, Minor axis length,Area, Convex area, Eccentricity, Perimeter, Solidity, Orientation, Extent, EquivDiameter is implemented.

2.3.b.iii Width Plot Analysis:

Complete width profile of leaf specifies the shape and hence the type of leaf. Width contour signal is considered as the digital signature of the leaf under consideration. Two level decomposition using Wavelet transform of this digital signature gives four feature values. Figure vii shows the width plot of the Parijat Leaf.

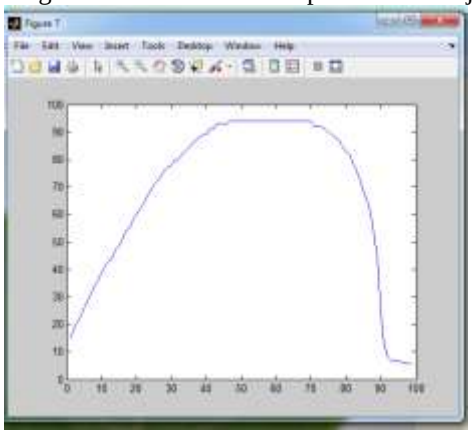


Figure vii. Width plot of Parijat Leaf

2.4. Classification

Different classification algorithms were implemented for classification of database. The feature vector derived from the above methods was analysied using statistical methods. Then the most prominant features which could be of higher importance were selected for better classification accuracy.

2.4.a. K- Nearest Neighbors

K Nearest Neighbors based classifications is an instance-based method for predicting class or value of a given query. It does not construct a general internal model. Classification is computed from a simple majority vote of the nearest neighbors of each pointc. New data point is assigned a class which has the most data points in the nearest neighbors of the point. Suited for classification where relationship between features and target classes is numerous, complex and difficult to understand. The method was implemented after the feature optimization. In the paper the distance formula used is Minkowski distance.

2.4.b. Support Vector Machine(SVM)

SVM was implemented in the paper as the data was complex and there was no clear distiction in the data. There are many hyperplanes that might classify the data.One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized.If such a hyperplane exists, it is known as the maximum-margin hyperplane. The best hyperplane finding is possible easily using SVM. SVM achived better accuracy than the KNN algorithm.

2.4.c. Tree Classifier

Tree classifier is implemented to separate gives easy and accurate way to differentiate the plants with reduced feature computation complexity and reduced time[14]. The stepwise classification starts with the ridge based classifier. The plants are categorized into three classes i.e. no ridge, multiple ridges, single ridge. Among database plants only Castor plant in multiple midrib and only aloe-vera with no ridge. In the remaining plant database now, all are single ridge plants. By symmetry test *Neem* and Datura are segregated as they are asymmetric and single ridge. So after symmetry test 2 classes will be formed symmetrical and asymmetrical.

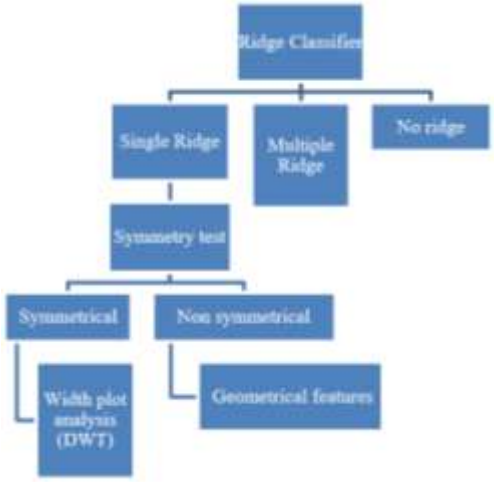


Figure viii. Tree Classification method flowchart

Within asymmetrical leaves category plants, further classification is done based on geometrical and morphological features. The sub classification of symmetrical leaves plants is proposed based on width plot analysis. Wavelet decomposition based features of width contour of leaves gives classification for five types of plants *viz. Tulas, Nirgudi, Parijat, Rui, Bel*. Figure viii. depicts the tree classification flow chart.

4. RESULT

In the addressed case study, the total database of 38000 images handled for implementation and validation of system. Among the database images of each species 70% of images i.e. 2940 images are used for training of system. While rest case study species images as well as random plant species images are used for validation of algorithm. The results of algorithm with 3 different Classification methods implemented as mentioned in table III.

Table III. Accuracy of different feature extraction algorithms.

Sr. no.	Features	Accuracy
1.	KNN	82%
2.	SVM	89%
3.	Tree Classifier	96%

In order to tune the accuracy, image normalization methods are implemented. When the tree classification is implemented the accuracy improves. The accuracy is 96% with higher speed.

5. CONCLUSION

Computer based identification of plants enables to precisely identify plants on large scale with less domain knowledge. The normalization of image leads to improve the plant images to enhance the features which will assist in better classification. Statistical analysis of the features extracted will lead to better feature optimization. Among the different classification methods implemented in the paper KNN doesn't work satisfactorily on the data. While the data is better classified by the SVM which gives the accuracy of 89% when tree classification is implemented accuracy is improved to 96% and it also provides increase in speed. Tree classification reduces load on classifiers and facilitates effective use of features.

6. REFERENCES

[1] Joao B. F., Odemir M. B., Davi R. R., Rosana M., Maria C. and Gabriel L. 2016, Identifying plant species using architectural features in leaf microscopy images, *Botany*. 94, No.1, 15-21.
[2] Hiew B.Y., Teoh A. B.J., Nago 2006, Automatic Digital Camera Based Fingerprint Image Pre-processing, *International Conference on Computer Graphics, Imaging and Visualisation*, 182-189.
[3] Madhava P. S. Seema V., 2019, Comparative Analysis of Segmentation techniques for Progressive Evaluation and Risk Identification of Diabetic Foot Ulcers, *4th MEC International Conference on Big Data and Smart City.* ii-iii.
[4] Petrou M., Pedro G. S. 1999 *Image Processing the fundamentals, Wiley Publication*, ISBN 0-471-99883-4.

[5] Gonzalez, Woods and Eddins ,2002, *Digital Image Processing Using MATLAB*, Prentice-Hall, Inc. ISBN 0-201-18075-8.
[6] Biva Shrestha 2010, Thesis on Classification Of Plants Using Images Of Their Leaves, Appalachian State University, Master of Science.9-20.
[7] Daniel S. 2018 Local Normalization: Filter to reduce the effect on a non-uniform illumination, *Biomedical Image Group, EPFL*, Switz, Retrieved from http://bigwww.epfl.ch/sage/soft/localnormalization/.
[8] Jyotismita C, Ranjan P. and Samar B. 2018, Plant leaf classification using multiple descriptors: A hierarchical approach, *Journal of King Saud University – Computer and Information Sciences*, 1-15
[9] Nursuriati J., Nuril A. C., Sharifalillah N., Khalil A., 2015 Automatic Plant Identification: Is Shape the Key Feature?, *IEEE International Symposium on Robotics and Intelligent Sensors, Procedia Computer Science* 436 – 442.
[10] Dengsheng Z., Guojun Lu. 2004 Review of shape representation and description Techniques, Pattern Recognition ,Pattern Recognition Society. Elsevier Ltd.1-19.
[11] Jyotismita Chaki, Ranjan Parekh,2011 Plant Leaf Recognition Using Shape Based Features And Neural Network Classifiers, *International Journal of Advanced Computer Science and Applications*, 2, 41-47
[12] James S. C., David C., Jonathan Y. Clark, Paolo R., Paul W., 2012, Plant Species Identification using Digital Morphometrics: A Review, *Expert Systems with Applications, Elsevier*, 7563-7577.
[13] Ji-Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang, 2007, Leaf shape based plant species recognition, *Applied Mathematics and Computation*, 185, 883–893
[14] David Dagan Feng, Wan-Chi Siu, Hong-Jiang Zhang (Eds.), 2003 Multimedia information Retrieval and Management Technological Fundamentals and Applications, *Springer-Verlag Berlin Heidelberg, Softcover reprint* ,1 st edition. 432-457