Identification of Indian Medicinal Herbs using Image Processing Techniques

Preshit Sharma[1], **Rashika Sharma**[2], **and Nitika Bagga**[3], **R.Prithiviraj**[4]

[1−3]Student,
Department of Electronics and Communication Engineering,
S.R.M Institute of Science and Technology Kattankulathur, Tamil Nadu, India

[4]Assistant Professor,
Department of Electronics and Communication Engineering,
S.R.M Institute of Science and Technology Kattankulathur, Tamil Nadu, India

ps6198@srmist.edu.in[1]
rr7165@srmist.edu.in[2]
nc9851@srmist.edu.in[3]
prithivr@srmist.edu.in[4]

1. Abstract

Identification of plants is a difficult task considering the lack of awareness about the local flora. Thus an automated identification system is a highly beneficial and superior option. Information about local plants is essential for plants and herbs-based medicine systems like Ayurveda. This paper focuses on the utilization of image processing techniques to extract morphological, texture-based, and color-based features from test images scanned using a camera and to compare these features using a support vector machine (SVM) for classification to detect leaf species with accuracy higher than 85 percentage.

Index Terms—Medicinal plants, Ayurveda, Automated, Image processing, Identification

2. Introduction

Ayurveda is Indian traditional system of medicine. It is related to preservation of health by maintaining a balance between mind, body, and spirit. It focuses on prevention of disease rather than the treatment. It is used to make the immune system stronger. It is an ancient medical practice that is mostly based on plants using leaves, roots, fruits and seeds of the plant. These medicinal herbs are used to cure a multitude of ailments. The identification of these leaves using naked eyes is tough and at high risk. Methodology to identify the plant species using its leaf image comprises of preprocessing of the query leaf image followed by segregation and differentiation based on various factors. The uses and properties of these herbs include curing of ailments, boosting immunity, homeopathic medicines, etc.

3. Problem Statement

Due to the lack of awareness about commonly found medicinal plants, it is difficult for a layman to identify the species of the plant. Computer vision models distinguish plants on either of the features of leaf i.e. shape, texture, color, and veins alone. Moreover, previous models demand huge datasets of the stored information regarding the uses and properties of plants with leaf images. Therefore, it is necessary to reduce the memory requirement and improve the accuracy of the model.

4. Aim/Objective

This paper aims to reduce the man work and make an automated application that uses leaf images to detect its species and results in the uses and properties with the botanical name of that Ayurvedic leaf. The uses will also include the contribution of that leaf in Ayurvedic practices. We also aim to increase the accuracy in identification using less storage space. The uses and properties are not stored but displayed on Google using the chrome control feature. The application uses data stored on the cloud storage which in turn shuns the use of hard disk or any external storage device.

5. Existing System

Plant identification has been proposed using image processing earlier using ANN, KNN, and SVM classifiers extracting features like shape, color, or veins. With a dataset of 1000- 1500 images and returning closest matches.

a) : Thibaut Beghin, James S Cope et al. [3]used incremental classification using only shape and texture based features.While incremental classification increases process time and is computationally intensive,dependence on just contour score and sobel score is only efficient in broader classifications thereby reducing accuracy for wider range of species.The method was only tested on limited dataset.

b) : Cem Kalyoncu, Onsen Toygar [2]are highly inclined ¨ towards geometrical features.Due to less inter species differences based on just geometrical shape ,model is less feasible for separating species with similar shape.

c) : Sulc M., Matas J. [10]relies on histograms for texture on border and interior of leaf and leaf orientation is hurdle for accurate classification.High accuracy was achieved for standard datasets with clear pictures.Pictures scanned using mobile camera cannot provide clear texture information.

d) : H.X.Kan ,L.Jin et al. [11]used 10 shape based features and 5 texture based features with support vector machine as classifier to classify 12 different species with accuracy greater than 90 percentage.Use of medicinal plants is TCM(Traditional Chinese Medicine) was considered.Lack of proper database for medicinal plants acts as big hurdle in testing reliability of model.

e) : D. Venkataraman and N. Mangayarkarasi [12]used HOG(Histogram of Oriented Gradients) ,texture based features and color moments for classification of medicinal plants using support vector machine as classifier.

From the previous studies it is evident that morphological or geometrical features ,texture based features and color moments are used to identify the plant using sample leaf images.

Local Name	Botanical Name
Neem	Azadirachta indica
Holy Basil	Ocimum tenuiflorum
Ashoka	Saraca asoca
Papri(Indian Elm)	Holoptelea integrifolia
Ayapan	Eupatorium ayyappana
Great Basil	Ocimum basilicum
Tej patta	Laurus nobilis
Elaichi	Elettaria cardamomum
Bhang	Cannabis
Giloi	Tinospora cordifolia
Bhurat(Sundakkai)	Solanum torvum
Paan	Piper betle
Indrajao(Dhudhi)	Wrightia Tinctoria Pala
Adrak(Ginger)	Zingiber officinale
Marorphali	Helicteres isora
Bamboo	Phyllostachys edulis
Horse Chestnut(Kanor)	Aesculus chinensis
Anhu Barberry	Berberis anhweiensis Ahrendt
Chinese redbud	Cercis chinensis
True indigo	Indigofera tinctoria L.
Nanmu	Phoebe nanmu (Oliv.) Gamble
Japanese maple	Acer Palmatum
Castor aralia	Kalopanax septemlobus (Thunb. ex A.Murr.) Koidz.
Cinnamon	Cinnamomum japonicum Sieb
Goldenrain tree	Koelreuteria paniculata Laxm.
Big fruited Holly	Ilex macrocarpa Oliv
Japanes cheeswood	Pittosporum tobira (Thunb.) Ait. f.
Wintersweet	Chimonanthus praecox L.
Camphor tree	Cinnamomum camphora (L.) J. Presl
Arrowwood	Viburnum awabuki K.Koch
Sweet osmanthus	Osmanthus fragrans Lour.
Deodar	Cedrus deodara (Roxb.) G. Don
Maidenhair	Ginkgo biloba L.
Crape myrtle	Lagerstroemia indica (L.) Pers.
Oleander	Nerium oleander L.
Yew plum pine	Podocarpus macrophyllus (Thunb.) Sweet
Flowering cherry	Prunus serrulata Lindl. var. lannesiana auct.
Glossy Privet	Ligustrum lucidum Ait. f.
Chinese toon	Tonna sinensis M. Roem.
Peach	Prunus persica (L.) Batsch.
Ford Woodlotus	Manglietia fordiana Oliv.
Trident Maple	Acer buergerianum Miq.
Barberry	Mahonia bealei (Fortune) Carr.
Southern magnolia	Magnolia grandiflora L.
Poplar	Populus ×canadensis Moench
Tulip tree	Liriodendron chinense (Hemsl.) Sarg.
Tangerine	Citrus reticulata Blanco

Fig. 1. List of leaf names

6. Proposed System

To eliminate the challenges faced in previous models, the proposed system architecture consists of different steps involved. The whole process is divided into two stages: Training and testing. The training stage is about storing the dataset and training the classifier on the same dataset.Dataset is a collection of feature vectors for the respective leaf image. Three main features of leaf i.e. Shape, texture, and color-based features are extracted and combined in a single feature vector. Features of multiple sample images of different leaf species are thus extracted and stored in their respective feature vectors. .In the testing stage same feature extraction is incorporated on query image followed by classification among the various classes or species of leaves. When the feature vector of the query image is fed to the classifier it returns the botanical name of the species. The botanical name of the species identified opens the gateway to access information regarding the species from the internet. Webbrowser module in python allows searching of user-defined strings on the internet. E-encyclopedias of medicinal plants and webpages dedicated to Ayurvedic plants are preferred sources for information regarding these plants such as uses, properties, and various local names and better alternative to storing of this information on local system. List of the leaves included in custom dataset used for training the model with common and botanical names is shown in Fig.1

7. Methodology

7.1 Algorithm

- Upload query image
- Extract features from leaf image
- Store Features in vector
- Feed vector to classifier
- Comparison with already stored feature vectors
- Output as botanical name of leaf species
- Uses and properties of species displayed

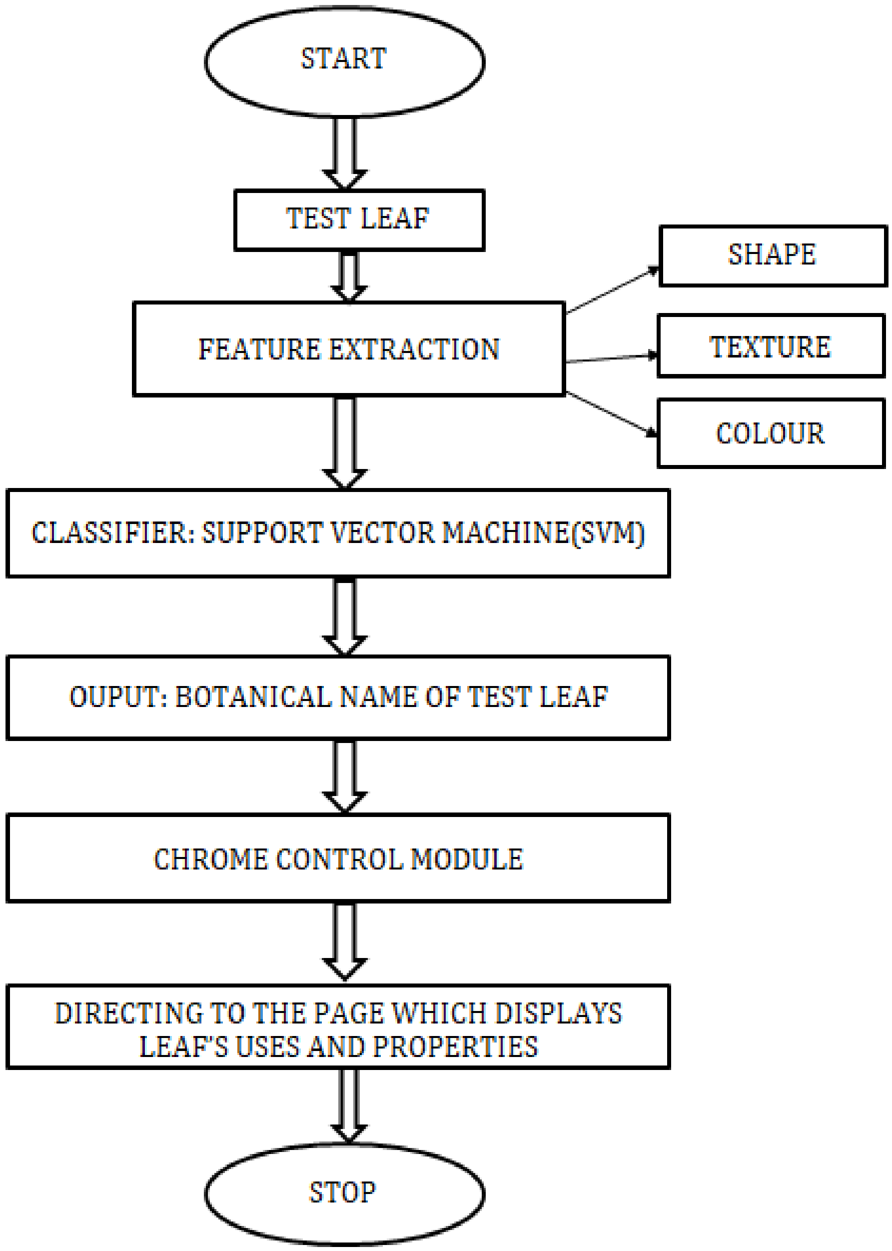

Fig. 2. Flowchart

7.2 Working Principle

7.2.1 Feature Extraction

a) **Shape**:

- Morphological features – Contouring[6][7] aids to find different contours in leaf image enclosing all closed shapes in image from this largest contour is the one enclosing the leaf as depicted in Fig.3(a,b and c). Opencv package of python provides with different functions for several morphological features - Image moments: Image moments aids in calculation center of mass and area of the object in image.Centroid is given by the relations:

$$C_x = \frac{M_{10}}{M_{00}} \tag{1}$$

$$C_y = \frac{M_{01}}{M_{00}} \tag{2}$$

Area is given by:

$$A = M['m00'] \tag{3}$$

here Mij is raw moment of image with pixel intensity I(x,y) and can be calculated by

$$M_{ij} = \sum_x \sum_y x^i y^j I(x,y) \tag{4}$$

,area can also be calculated by cv.contourArea() function.

- Contour perimeter: It can be found out using cv.arcLength() function
- Circularity: it can be found using area and arc length.

$$Circularity = \frac{perimeter^2}{area} \tag{5}$$

- Rectangularity and aspect ratio: Width and height of bounding rectangle is used to calculate rectangularity and aspect ratio

$$Aspectratio = \frac{width}{height} \tag{6}$$

$$Rectangularity = \frac{width + height}{area} \tag{7}$$

Fig. 3. Query Image

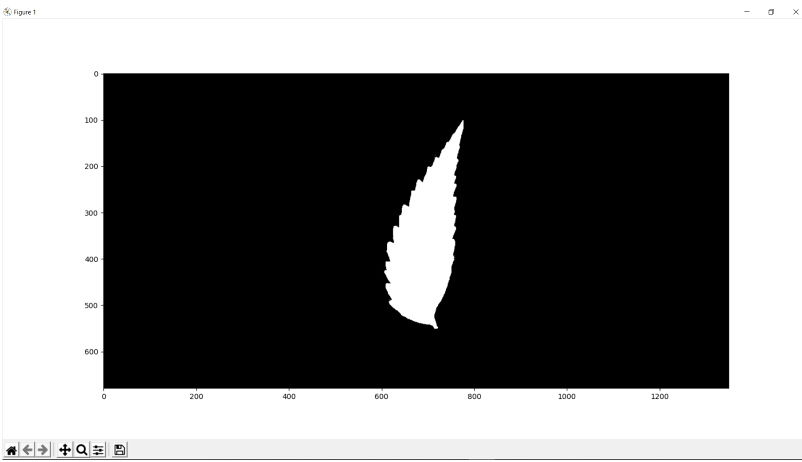

Fig. 4. Mask to extract region of interest

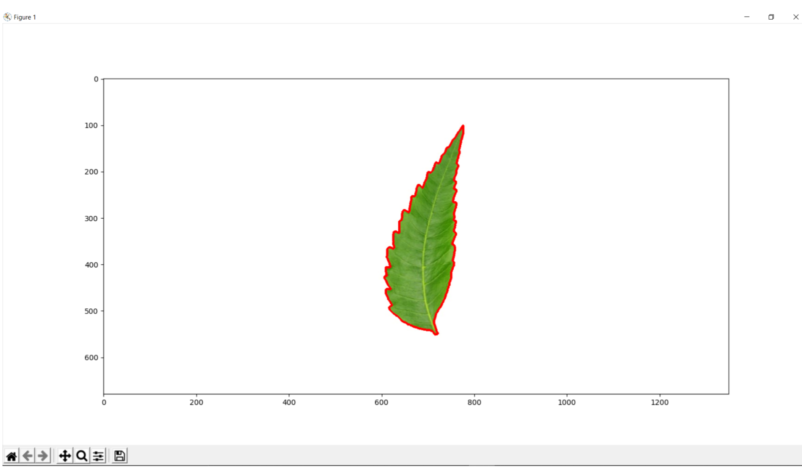

Fig. 5. Largest contour enclosing the leaf

b) **Texture**: To extract features from leaf image which can provide info of different vein structure and patterns on leaf surface ,image is transformed to grayscale . Haralick features are extracted from GLCM (Gray level Co-occurrence matrix) of grayscale image[5]. This matrix records how many times two gray-level pixels adjacent to each other appear in an image. This allows to identify particular vein pattern and texture on surface of different leaves and makes addition to features on the basis of which leaves are differentiated.[4] For a normalized symmetric GLCM matrix (G) of dimensions N X N where N is number of gray levels and G(i,j) is (i,j)th element of matrix ,various texture features are given by:

- Contrast is a measure of intensity or gray level variations between the reference pixel and its neighbor:

$$Contrast = \sum_i \sum_j (i-j)^2 G(i,j) \tag{8}$$

- Correlation feature shows the linear dependency of gray level values in the cooccurrence matrix:

$$Correlation = \sum_i \sum_j G(i,j) \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y} \tag{9}$$

  where μ_x, μ_y, σ_x and σ_y are the means and standard deviations and are expressed as:

$$\mu_x = \sum_i \sum_j iG(i,j) \tag{10}$$

$$\mu_y = \sum_i \sum_j jG(i,j) \tag{11}$$

$$\sigma_x = \sqrt{\sum_i \sum_j (i-\mu_x)^2 G(i,j)} \tag{12}$$

$$\sigma_y = \sqrt{\sum_i \sum_j (j-\mu_y)^2 G(i,j)} \tag{13}$$

- Inverse difference moments (IDM) or Homogeneity measures how close the distribution of elements in the GLCM is to the diagonal of GLCM:

$$IDM = \sum_i \sum_j \frac{1}{1+(i-j)^2} G(i,j) \tag{14}$$

- Entropy is the randomness or the degree of disorder present in the image. The value of entropy is the largest when all elements of the cooccurrence matrix are the same and small when elements are unequal:

$$Entropy = -\sum_i \sum_j G(i,j) lnG(i,j) \tag{15}$$

c) **Colour**: Colour based features[8] aids to differentiate between plants with different coloured leaves .Image is divided into three channels i.e. (red green and blue) Mean and standard deviation of theses channel values are stored as colour features of leaf image.

7.2.2 Classification

For classification of query image among the samples in database Support Vector Machine or SVM classifier is used . SVM classifies datapoints using N dimensional hyperplane where N is number of features. SVM natively is a binary classifier i.e. each datapoint can either lie on one side of the hyperplane .Support vectors are datapoints nearer to hyperplane which defines the margin. The generalized equation of a hyperplane is given by:

$$w^T x = 0 \tag{16}$$

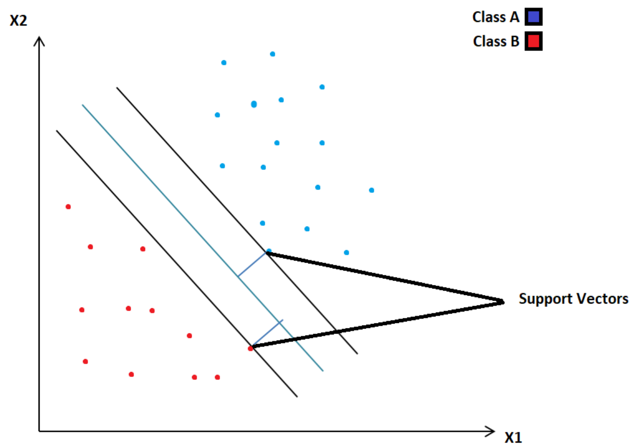

Fig. 6. Binary SVM Model

For multiclass classification[9] ,binary SVM is modified with the help of algorithms namely one vs all and one vs one which includes binary classifier per each class . In the One-to-Rest approach, hyperplane is used to separate between a class and all others at once while in One vs One approach it ignores all other classes while a hyperplane seperates two classes in current split as shown in Fig.5(a,b).

Normalization of feature vectors is done to match standard range for SVM from 0 to 1 or -1 to 1. After training the classifier on dataset , parameter tuning is used to enhance accuracy testing for different values of C (regularization parameter which deals with misclassification) and gamma and selecting the ideal kernel for the testcase.

Representation of user point of view of working, including custom GUI is shown in Fig.6.

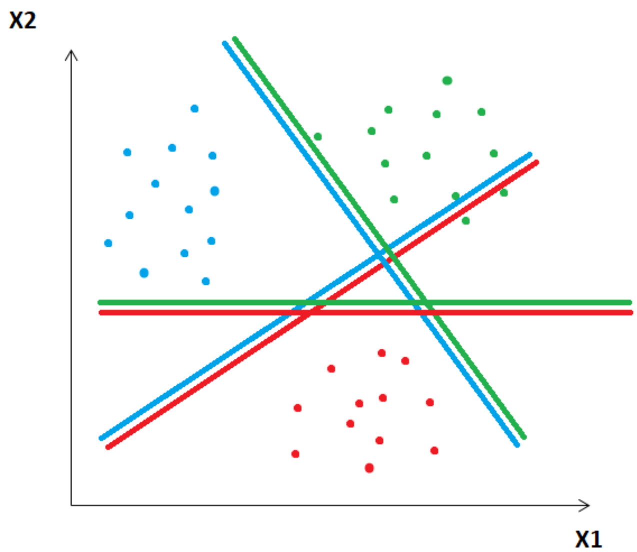

Fig. 7. One vs One Approach

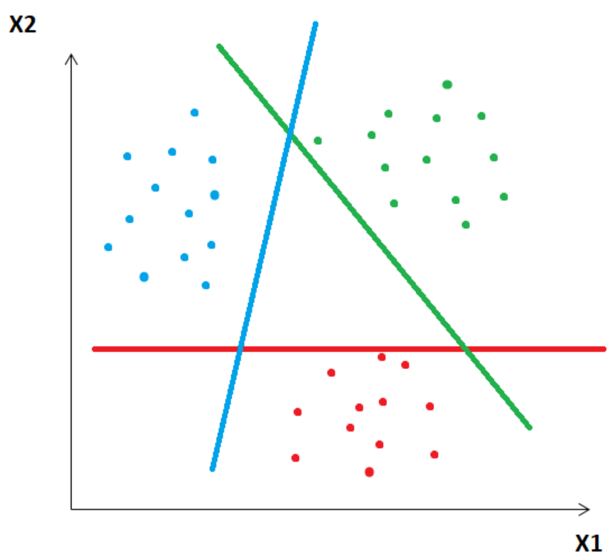

Fig. 8. One vs All Approach

7.2.3 Websearch

The webbrowser module provides interface to allow users to access web pages and display them. Browser application is set as environment for the code and controller.open(url) function can be used to open a new window in browser which displays the page corresponding to "url" string. url = 'http://www.google.com/' comment-Open URL in a new window, if a browser window is already open. webbrowser.open(url)

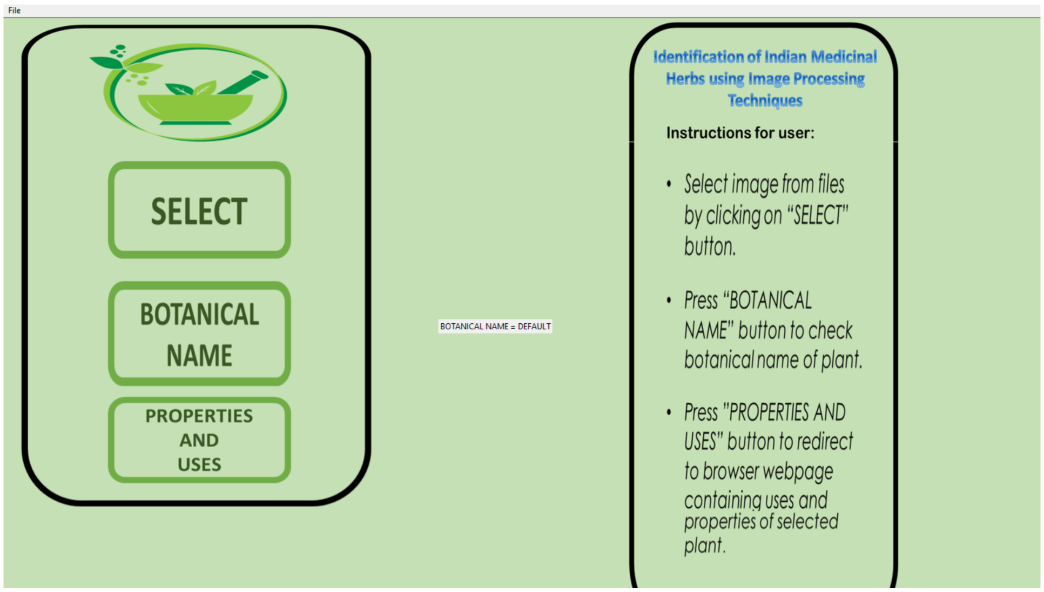

Fig. 9. User Interface

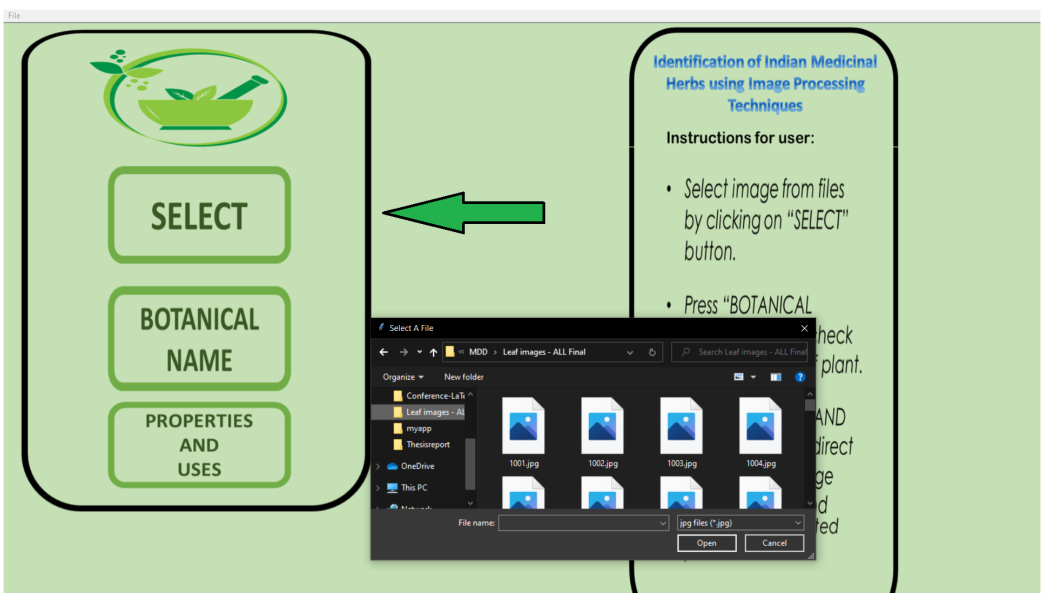

Fig. 10. Selecting query image

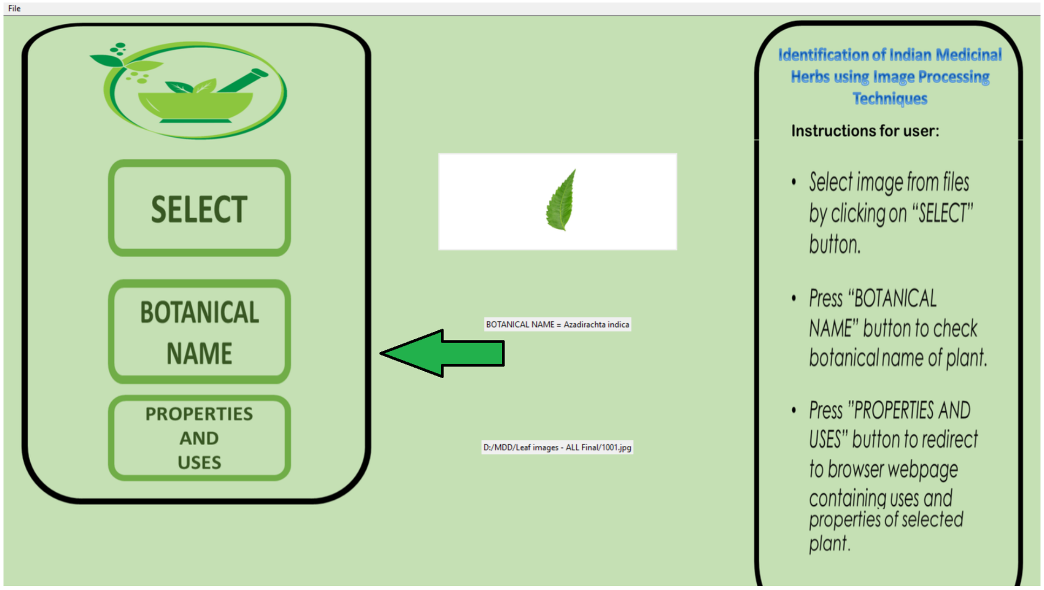

Fig. 11. Botanical Name displayed

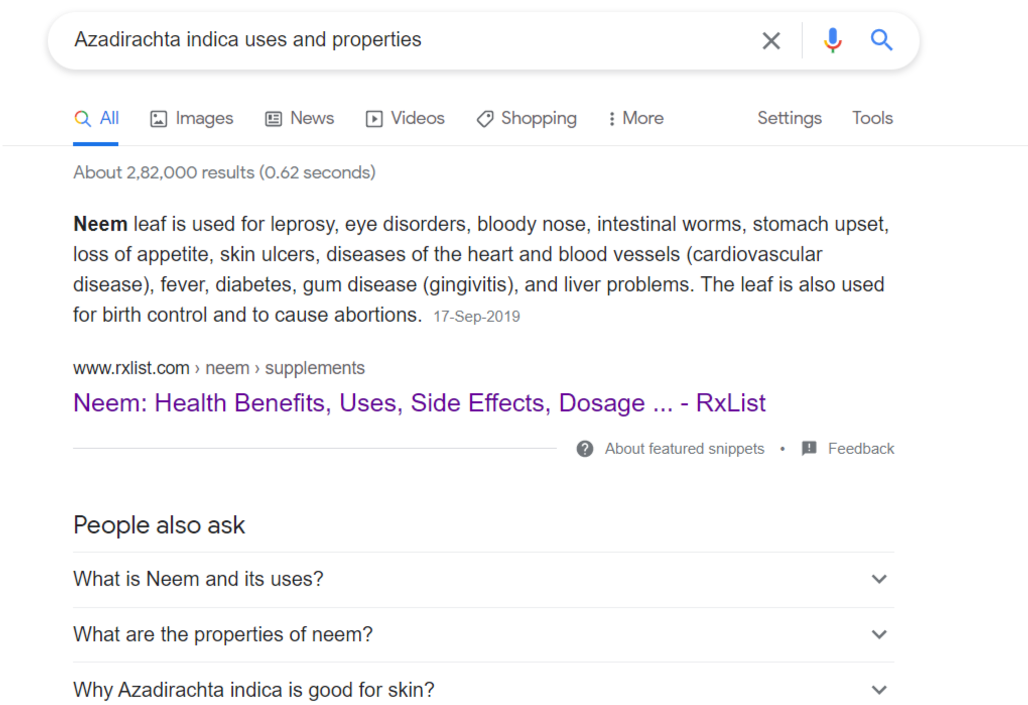

Fig. 12. Redirected to web page with uses and properties of species

8. Analysis and Results

The system contains a dataset of 470 sample images comprising 47 different species having 10 samples each from a custom made dataset comprising images from web and from Flavia leaf dataset. The dataset is divided into training and testing sets by a ratio of 0.33 i.e. 315 training samples and 155 testing samples. The classifier model is trained on the various morphological,texture-based and color gradient-based features to provide an accuracy of 86.5 percentage.

Features	Classifier	Accuracy	Dataset	Training	Testing	Species	Samples
Shape, texture, color	SVM	86.5%	470	315	155	47	10

```
..          ...          ...          ...   ...          ...        ...        ...
465  693327.0  4429.687838      1.346591  ...  36.301181  64.008354  28.171358
466  626601.0  4342.280765      1.378026  ...  44.432803  67.482810  30.603159
467  663465.5  4266.332500      1.430815  ...  46.771753  71.488590  27.488877
468  719563.5  4449.871604      1.397338  ...  41.220509  68.795816  27.017941
469  760375.0  4479.653348      1.410765  ...  37.616911  62.360694  28.105651

[470 rows x 15 columns]
best para =  {'C': 10, 'kernel': 'linear'}
score after tuning 0.8653846153846154
```

Fig. 13. Results

9. Conclusion and Future work

Ayurveda, being an ancient practice based on plants, plays a crucial role in the medicinal industry. Extracting and identifying these medicinal plants is even more important. This paper proposes multiple algorithms to identify the leaf species using image processing. In this paper, we have discussed Contouring, feature extraction of the leave samples, and finally, by the use of SVM classifier, system displays output as botanical name of the leaf. Furthermore, the incorporated GUI system with browser control module provides user with leaf uses and properties displayed on web page.This addition with access of information from internet reduces the on-system memory requirements and is a good upgrade to previously made models. Future scope of the study includes augmentation of new features in feature vector like venation patterns of leaf.In addition to this more number of samples per species can enhance the accuracy of the model.Moreover drafting of standard database of Ayurvedic plants can be supportive.Standalone dedicated hardware implementing the following algorithms can be constructed.Software application based on this methodology can be developed for portable devices.A cloud based storage for dataset will further reduce the on-system memory requirements.

10. Bibliography

[1] Dahigaonkar, Tejas D., and R. Kalyane. ”Identification of ayurvedic medicinal plants by image processing of leaf samples.” International Research Journal of Engineering and Technology (IRJET) 5, no. 05 (2018).

[2] Kalyoncu, Cem, and Önsen Toygar. ”Geometric leaf classification.” Computer Vision and Image Understanding 133 (2015): 102-109.

[3] Beghin, Thibaut, James S. Cope, Paolo Remagnino, and Sarah Barman. ”Shape and texture based plant leaf classification.” In International conference on advanced concepts for intelligent vision systems, pp. 345-353. Springer, Berlin, Heidelberg, 2010.

[4] Nourhan Zayed, Heba A. Elnemr, ”Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities”, International Journal of Biomedical Imaging, vol. 2015, Article ID 267807, 7 pages, 2015. https://doi.org/10.1155/2015/267807

[5] R. C. Gonzalez and R. E. Woods, “Digital Image Processing”, Prentice hall, 2008

[6] Mathew. M. and Gopi V. P., “Transform based bleeding detection technique for endoscopic images”, 2nd International Conference on Electronics and Communication Systems (ICECS), pp. 1730-1734, Feb. 2015

[7] Kumar, P. and Surya, C. and Gopi, Varun. (2017). Identification of ayurvedic medicinal plants by image processing of leaf samples. 231-238. 10.1109/ICRCICN.2017.8234512.

[8] van der Walt, Stéfan and Schönberger, Johannes and Nunez-Iglesias, Juan and Boulogne, François and Warner, Joshua and Yager, Neil and Gouillart, Emmanuelle and Yu, Tony and contributors, the. (2014). scikit-image: Image processing in Python. PeerJ. 2. 10.7717/peerj.453.

[9] Elhariri E., El-Bendary N., Fouad M.M.M., Platos J., Hassanien A.E., Hussein A.M.M. (2014) Multi-class SVM Based Classification Approach for Tomato Ripeness. In: Abraham A., Kromer P., Snasel V. (eds) Innovations in Bio-inspired Computing and Applications. Advances in Intelligent Systems and Computing, vol 237. Springer, Cham.

[10] Sulc M., Matas J. (2015) Texture-Based Leaf Identification. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8928. Springer, Cham.

[11] Kan, H. and Jin, L. and Zhou, F.. (2017). Classification of medicinal plant leaf image based on multi-feature extraction. Pattern Recognition and Image Analysis. 27. 581-587. 10.1134/S105466181703018X.

[12] D. Venkataraman and N. Mangayarkarasi, ”Support vector machine based classification of medicinal plants using leaf features,” 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 793-798, doi: 10.1109/ICACCI.2017.8125939.

[13] Maheswari, A. and Bharathi, N. and Neelamegam, Periasamy and Gayathridevi, T.. (2014). Classification and recognition of herbal leaf using SVM algorithm. Research Journal of Pharmaceutical, Biological and Chemical Sciences. 5. 415-423.

[14] R. Janani and A. Gopal, ”Identification of selected medicinal plant leaves using image features and ANN,” 2013 International Conference on Advanced Electronic Systems (ICAES), 2013, pp. 238-242, doi: 10.1109/ICAES.2013.6659400.

[15] Anh H. Vo, Hoa T. Dang, Bao T. Nguyen, and Van-Huy Pham, ”Vietnamese Herbal Plant Recognition Using Deep Convolutional Features,” International Journal of Machine Learning and Computing vol. 9, no. 3, pp. 363-367, 2019.