

MEDICINAL PLANT IDENTIFICATION

by Nitika Bagga Preshit Sharma

Submission date: 17-May-2021 05:03PM (UTC+0530)

Submission ID: 1587889836

File name: SP31-Report-converted.docx (1.43M)

Word count: 3074

Character count: 16847

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 project focuses on the utilization of image processing techniques to extract morphological, texture-based, and colour-based features from test images scanned using a camera and to compare these features using a support vector machine (Support Vector Machine (SVM)) for classification to detect leaf species with accuracy higher than 85 percentage.

CHAPTER 1

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.

1.0.1 Aim/Objective

Aim of this project is to recognize leaf using contour based edge detection for geometrical features and Gray level Co-occurrence matrices for texture features combined with color moments of RGB image. Then to study features using classifier i.e. SVM (Support Vector Machine) which will determine performance of the model. This model will provide botanical name of the leaf and display its uses and properties. This model aims to reduce the manual 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 leaf. The uses will also include the contribution of that leaf in Ayurvedic medicine compositions. This model aims to increase the accuracy in identification while having less storage requirements. The uses and properties are not stored but displayed on Google using the browser control feature. Database stored on the cloud storage can be used which in turn shuns the use of hard disk or any external storage device to store huge database.

CHAPTER 2

LITERATURE SURVEY

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

1. 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.
2. Cem Kalyoncu, Önsen 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.
3. 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.
4. 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.

5. D. Venkataraman and N. Mangayarkarasi [12] used HOG(Histogram of Oriented Gradients) ,texture based features and color moments ³ 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.

CHAPTER 3

SYSTEM DESIGN

3.1 Proposed System

- The proposed system architecture consists of different steps involved. The whole process ⁶ 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. These feature vectors are collection all the extracted features from sample image. The trained classifier model is then ready for testing on query images.
- 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. Web browser module in python allows searching of user-defined strings on the internet. E-encyclopedias of medicinal plants and web-pages 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.

3.2 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

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

Figure 3.1: Leaf names list

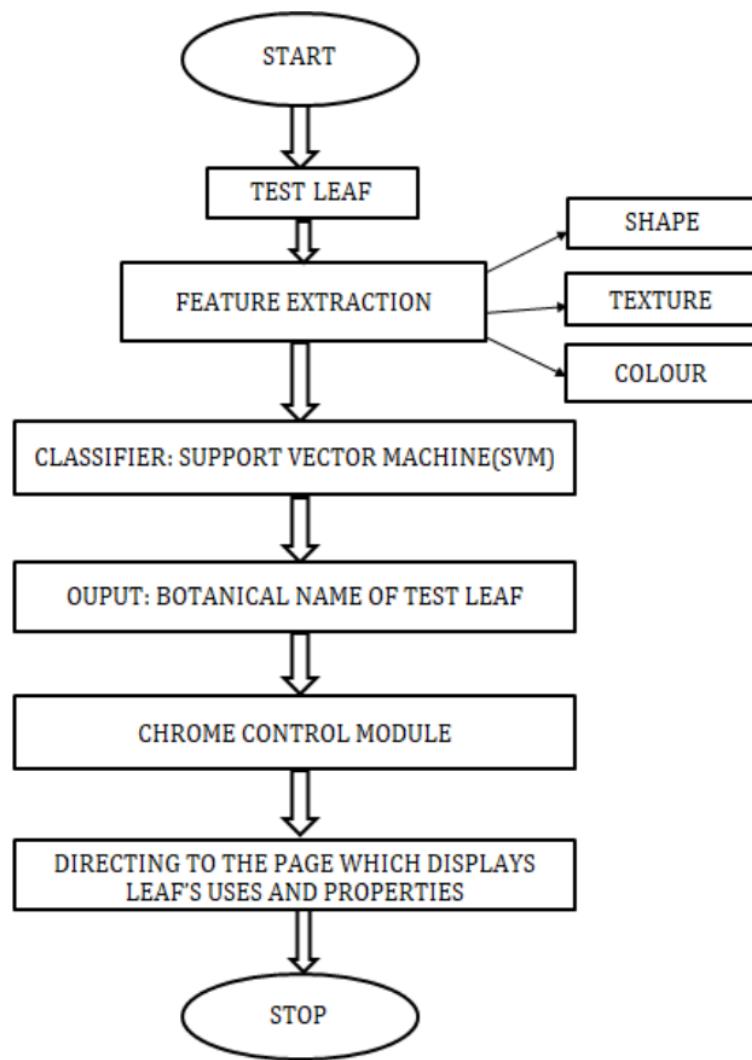


Figure 3.2: Flow chart

CHAPTER 4

METHODOLOGY

4.1 Working Principle

4.1.1 Feature Extraction

Shape



Figure 4.1: Query Image

- 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:

$$Q = \frac{M_{10}}{M_{00}} \quad (4.1)$$

$$C_y = \frac{M_{01}}{M_{00}} \quad (4.2)$$

Area is given by:

$$A = M [^m 00'] \quad (4.3)$$

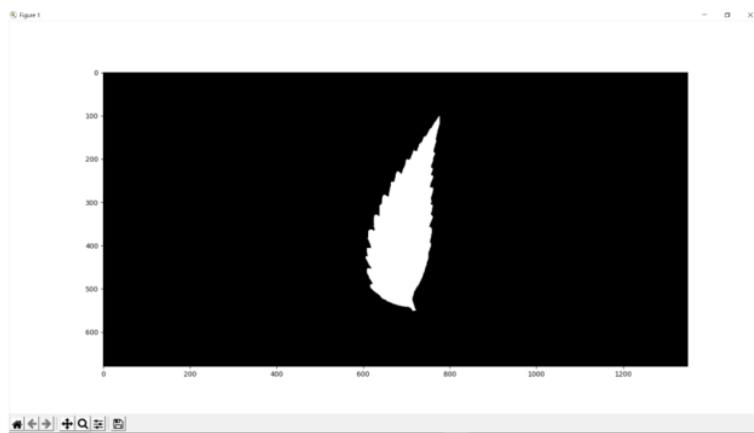


Figure 4.2: Mask to separate leaf from background

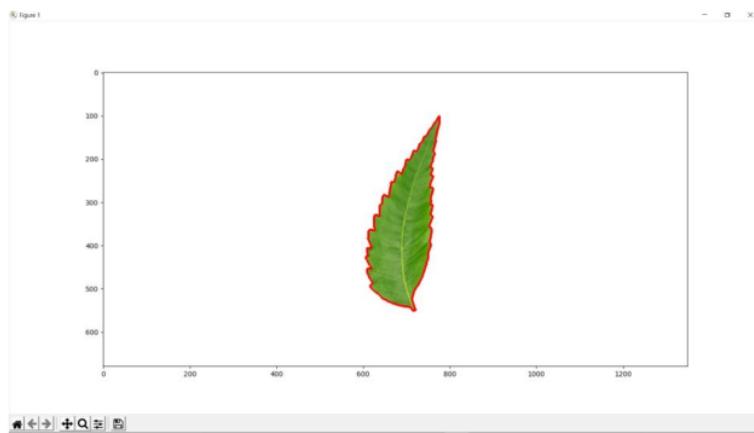


Figure 4.3: Largest contour enclosing the leaf

here M_{ij} is raw moment of image with pixel intensity $I(x,y)$ and can be calculated by

$$M_{ij} = \sum \sum x^i y^j I(x, y) \quad (4.4)$$

, area can also be calculated by `cv2.contourArea()` function.

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

$$\text{Circularity} = \frac{\text{perimeter}^2}{\text{area}} \quad (4.5)$$

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

to calculate rectangularity and aspect ratio

$$Aspectratio = \frac{width}{height} \quad (4.6)$$

$$Rectangularity = \frac{width+height}{area} \quad (4.7)$$

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 Gray Level Co-Occurrence Matrix (GLCM) (Gray level Co-occurrence matrix) of grayscale image. This matrix records how many times two gray-level pixels adjacent to each other appear in an image [4] .

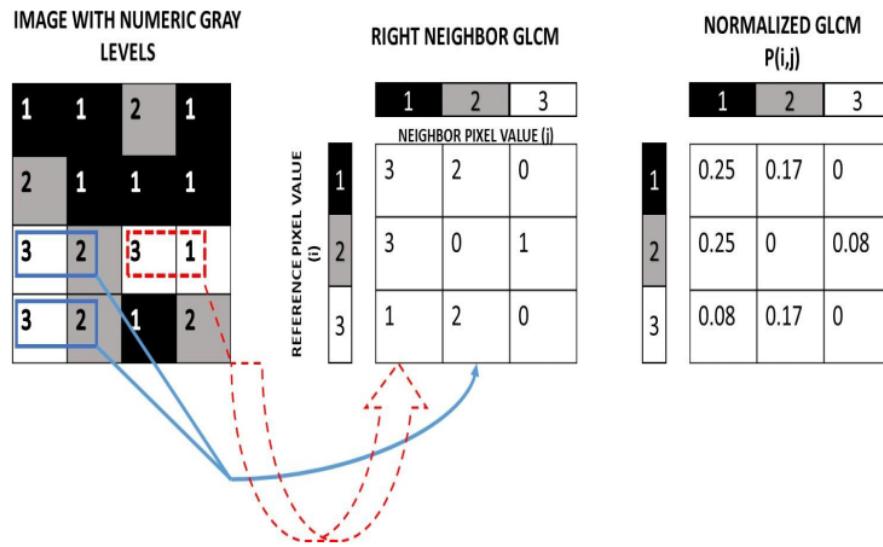


Figure 4.4: Gray Level Co-occurrence Matrix

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. For a ² normalized symmetric GLCM matrix (G) of dimensions ¹ $N \times N$ where N is number of gray levels and $G(i,j)$ is (i,j) th element of matrix ,various texture features are given by:

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

$$\text{Contrast} = \sum_{i} \sum_{j} (i - j)^2 G(i, j) \quad (4.8)$$

- 2 Correlation dictates the linear dependency of gray-level values in the co-occurrence matrix:

$$\text{Correlation} = \sum_{i} \sum_{j} G(i, j) \frac{(i - \mu_x)(j - \mu_y)}{\sigma_x \sigma_y} \quad (4.9)$$

4 where μ_x, μ_y, σ_x and σ_y are the mean values and standard deviation values and are generally expressed as:

$$\mu_x = \sum_{i} \sum_{j} i G(i, j) \quad (4.10)$$

$$\mu_y = \sum_{i} \sum_{j} j G(i, j) \quad (4.11)$$

$$\sigma_x = \sqrt{\sum_{i} \sum_{j} (i - \mu_x)^2 G(i, j)} \quad (4.12)$$

$$\sigma_y = \sqrt{\sum_{i} \sum_{j} (j - \mu_y)^2 G(i, j)} \quad (4.13)$$

- Inverse difference moments (Inverse Difference Moments (IDM)) or Homogeneity is a measure of closeness between the distribution of elements in the GLCM and the diagonal of GLCM:

$$\text{IDM} = \sum_{i} \sum_{j} \frac{1}{1 + (i - j)^2} G(i, j) \quad (4.14)$$

- Entropy is the degree of disorder present in the image. Entropy shows largest value when all elements of the co-occurrence matrix have similarity and lowest value when elements have inequalities :

$$\text{Entropy} = - \sum_{i} \sum_{j} G(i, j) \ln G(i, j) \quad (4.15)$$

Colour

Colour based features 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 these channel values are stored as colour features of leaf image [8].

4.1.2 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:

$$w^T x = 0 \quad (4.16)$$

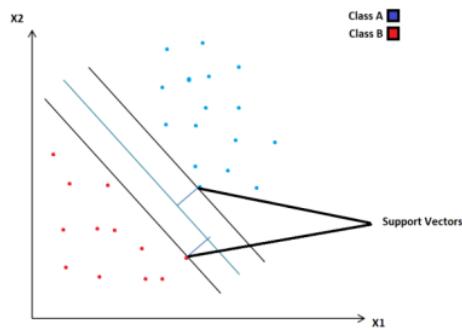


Figure 4.5: Binary SVM Model

For multiclass classification ,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 separates 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.

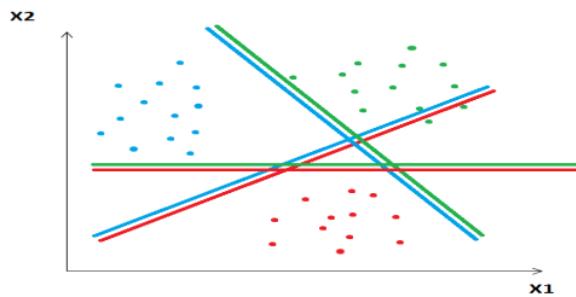


Figure 4.6: One vs One Approach

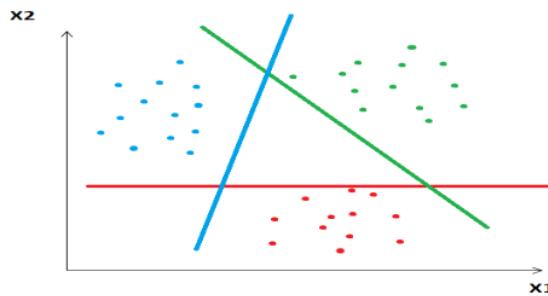


Figure 4.7: One vs All Approach

4.1.3 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(Uniform Resource Locator(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)

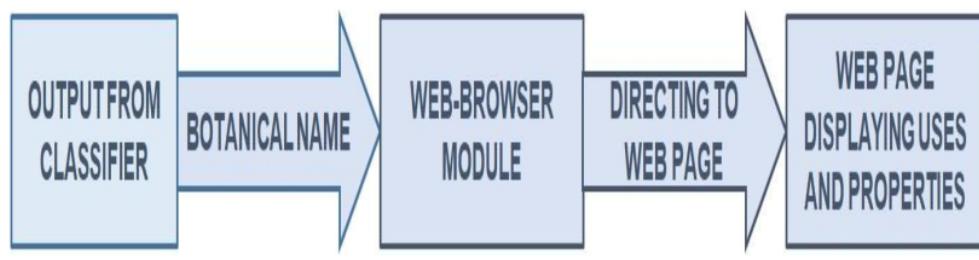


Figure 4.8: Browser Control

CHAPTER 6

RESULTS

6.1 Classifier Results

Classifier is trained on total of 470 sample images including 47 different plant species with 10 samples each. Dataset is divided in two sets : Training and Testing with 315 and 155 images respectively.

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 ... 41.228509 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
```

Figure 6.1: Classifier Results

6.2 User Interface

For User friendly outlook, a Graphical User Interface (GUI) (graphical user interface) is developed which allows user to select query image,display its botanical name and to find uses and properties of identified plant. Tkinter package for Python provides all required functions and utilities to develop a GUI. Custom buttons and labels are used to match the theme.

- Main page of application containing several buttons to interact with on left side and instruction manual on right Fig 6.2.
- Pressing SELECT button to open file browser for query image selection Fig 6.3.
- Pressing BOTANICAL NAME button to display botanical name of species Fig 6.4.
- Pressing PROPERTIES AND USES button to display uses and properties of identified species Fig 6.5.

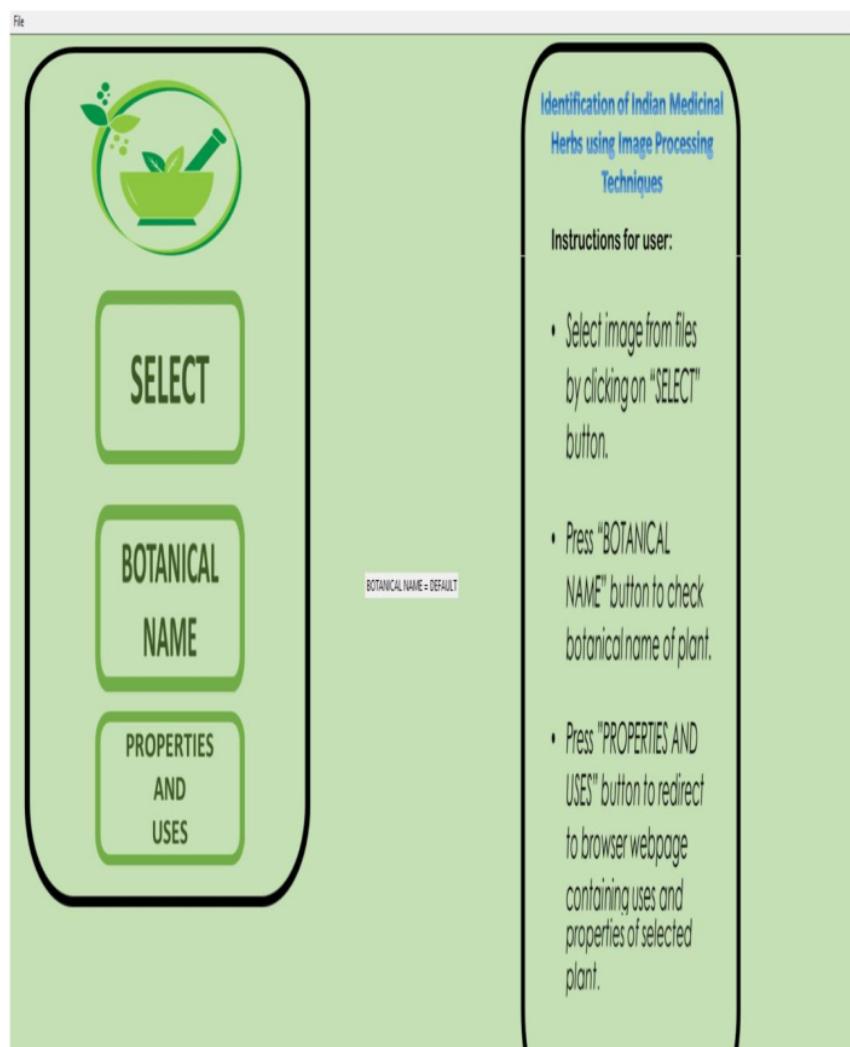


Figure 6.2: Main page of User Interface



Figure 6.3: SELECT button for query image selection

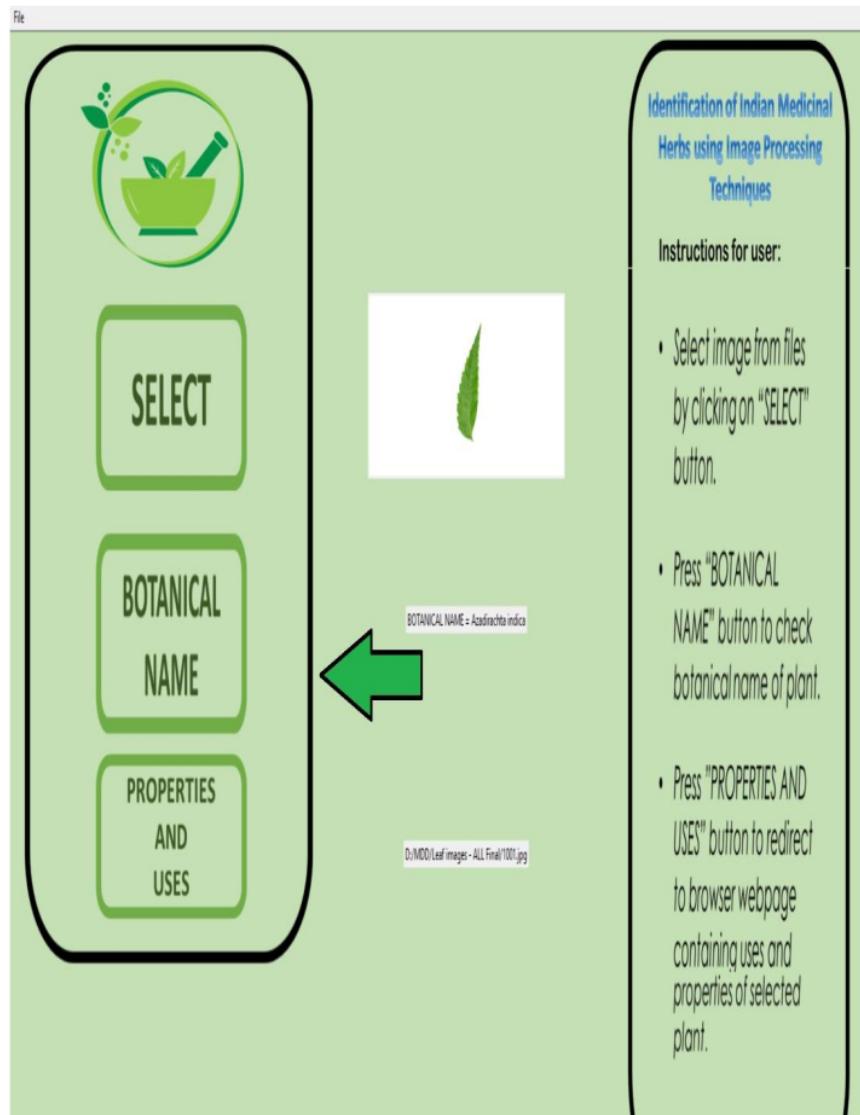


Figure 6.4: BOTANICAL NAME button to display botanical name

Azadirachta indica uses and properties

X |

All Images News Videos Shopping More Settings Tools

About 2,82,000 results (0.62 seconds)

Neem leaf is used for leprosy, eye disorders, bloody nose, intestinal worms, stomach upset, loss of appetite, skin ulcers, diseases of the heart and blood vessels (cardiovascular disease), fever, diabetes, gum disease (gingivitis), and liver problems. The leaf is also used for birth control and to cause abortions. 17-Sep-2019

www.rxlist.com › neem › supplements

[Neem: Health Benefits, Uses, Side Effects, Dosage ... - RxList](#)

About featured snippets •

People also ask

What is Neem and its uses? ▾

What are the properties of neem? ▾

Why Azadirachta indica is good for skin? ▾

Figure 6.5: Uses and properties of identified species on web page

CHAPTER 7

CONCLUSION

Study of medicines plays an important role in our lives. It is a never ending process and needs improvement and enhancement in techniques as and when needed. Ayurveda is an ancient practice which used to boost immunity and cure plethora of ailments. Ayurveda is the study of different medicinal plants and their uses. In this model different algorithms like contouring, feature extraction are performed on leave samples and finally by the use of SVM classifier, system displays the botanical name of the plant . This botanical name is fed to browser control feature which directs user to web page with uses and properties of the sample leaf . This incorporation reduces the on-system memory usage.

CHAPTER 8

FUTURE ENHANCEMENT

- Future scope of the study includes augmentation of new features in feature vector like venation patterns of leaf. Venation patterns will benefit the study in ways like scale-space analysis which could be a promising research direction to pursue.
- In addition to this, more number of samples per species can enhance the accuracy of the model. Having a huge number of species in a training set can lead to a better understanding of the model, thereby, giving us more precise results.
- Moreover, drafting of a standard database of Ayurvedic plants can be supportive in doing the analysis of the training model and would give more insightful results with better understanding of species.
- A standalone dedicated hardware implementing the algorithms used in our model can be constructed using digital circuitry like microprocessors, microcontrollers by feeding them with the codes.
- Software applications based on this methodology can be developed for portable devices. A simple example would be an app based on the methodologies used in this project.
- A cloud based storage for dataset will further reduce the on-system memory requirements. This will enhance our memory storage and will definitely speed up the process.

APPENDIX A

MACHINE LEARNING FOR CLASSIFICATION

A.1 Support vector machine 5

Support vector machine or SVM is a machine learning algorithm based on statistical learning theory by Vapnik et al[16] which is useful for training sets containing limited number of samples [15]. SVM is mainly used in classification and database technology. It is based on the principle of structural risk minimization which is different from conventional empirical risk minimization and thus allows learning on small number of samples. Basic concept of SVM consists of a decision boundary or Hyperplane which separates datapoints from different classes. This Hyperplane is a N dimensional plane where N is directly connected to number of features used for classification. 6

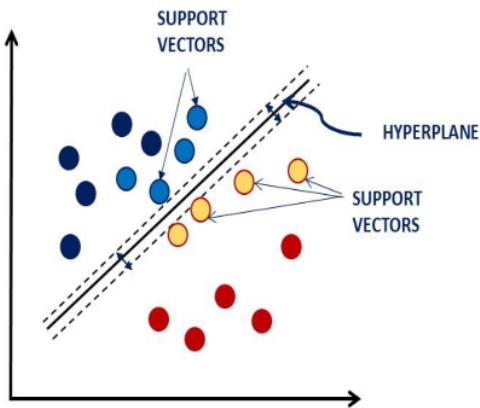


Figure A.1: Hyperplane separating two classes

APPENDIX B

B.1Contours and Edge detection

OpenCV package for Python programming language allows user to find and draw contours from binary image .The function "cv.findContours()" is based on the algorithm of Suzuki Be (1985) [17].Contours are usually a curve joining continuous points having same color or intensity near the boundaries.These contours encloses every single closed shape in a test image providing ability to separate required object from its background.This contour enclosing the object, then supplies with morphological or shape based features such as area, perimeter etc.

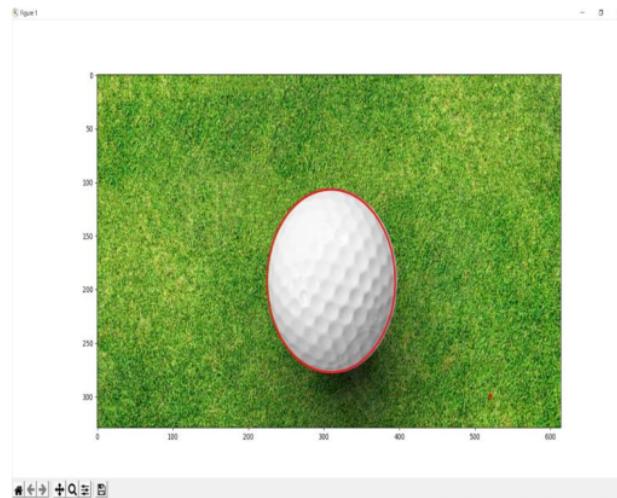


Figure B.1: Contour enclosing the ball on grass

REFERENCES

- [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/ICR-CICN.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, Emmanuel 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] Zhang Y. (2012) Support Vector Machine Classification Algorithm and Its Application. In: Liu C., Wang L., Yang A. (eds) Information Computing and Applications. ICICA 2012. Communications in Computer and Information Science, vol 308. Springer, Berlin, Heidelberg.
- [16] Vapnik, V.N.: Statistical learning theory. Springer, New York (1995)
- [17] Suzuki, S., and Be, K. (1985). Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing 30, 32–46. doi:10.1016/0734-189X(85)90016-7.

MEDICINAL PLANT IDENTIFICATION

ORIGINALITY REPORT



PRIMARY SOURCES

1	www.ijccr.com Internet Source	1 %
2	Nourhan Zayed, Heba A. Elnemr. "Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities", International Journal of Biomedical Imaging, 2015 Publication	1 %
3	www.science.gov Internet Source	1 %
4	"Intelligent Systems'2014", Springer Nature, 2015 Publication	1 %
5	bmcbioinformatics.biomedcentral.com Internet Source	1 %
6	www.ijedr.org Internet Source	1 %

Exclude quotes On

Exclude bibliography On

Exclude matches < 1%