Implementation and Evaluation of fuzzy Thresholded machine Learning Classifiers to Recognize Medicinal Plant Leaves

K. Ponmalar [#1], Dr. K. Krishnaveni [#2]

[#1] Phd scholar ,Manonmaniam Sundaranar University, Tirunelvelli, Tirunelvelli Dist.

[#2] Head, Department of Computer Science, Sri.S.R.N.M College, Sattur, Virudhunagar Dist.

[1] kpmalar28@gmail.com [2]kkveni_srnmc@yahoo.co.in

Abstract

In this paper, Fuzzy morphological thresholded machine leaning classifiers are proposed to recognize the class of the medicinal plant leaves and their medicinal usage. A data set containing 760 fresh medicinal plant leaf images of thirty different classes are captured, pre processed by iterative morphological filtering and fuzzy threshold value, morphological shape and texture features of the leaf image are extracted and stored as a feature data set. This feature set is used by the classifier to train the medicinal plant images of the training data base. From the data base ten images of each class is taken as the test images, features are calculated, mapped with the training feature set and finally classified into a particular class. Once the class of the medicinal leaf is identified and recognized, the medicinal usage of that particular plant will be displayed. In this research work, three machine learning classifiers Random Forest, K-Nearest Neighbour and Decision Tree are proposed, their results are analysed based on different performance metrics and proved that Random forest classifier outperforms compared to other classifiers even for high volume and degraded images. KNN gives best result when the data set is small. The Decision Tree classifier yields to misclassification when the images are having shadow effects.

Keywords: machine leaning classifier, medicinal plant, iterative morphological filtering, Random Forest, K-Nearest Neighbour, Decision Tree

1. Introduction

The key source of traditional medicines which provides the basic protection of human health is the Medicinal plants. The species of medicinal plants can be identified manually from the characteristics of the whole plant or individual part (leaf, flower, fruit, or bark) by using eyes, noses, hands, or other human organs based on either reference or experience. But this time consuming process heavily depends on people's level of knowledge and subjective experience. The computer based automatic image recognition is now widely used due to the development of image processing and pattern recognition technology. Since it is easy to collect, distinguish, and capture the characteristics of plant leaves, they are used as the primary basis for the identification of medicinal plants.

Feature extraction and classifier selection play a vital task in image recognition. Shyam Vijayrao et al used a texture investigation to categorize and classify the medicinal plants. The texture analysis generates a feature set to query the image from the database [2].Harish, B.S. et al used Morphological features and Zernike moments to classify plant leaves. The features extracted are not reliant on the scale, leaf growth and image translation and rotation [3].In this paper, shape and texture features of the medicinal plant leaf images are extracted, and the machine learning classifiers Random Forest, Decision Tree and K_Nearest Neighbour are proposed for classification. Finally the performance of the classifiers are compared and analysed based on different metrics. The paper is organized as follows: The proposed methodology is discussed in Section II. The experimental results and the performance evaluation is presented in Section III, and Section IV concludes the whole paper and gives related conclusions.

2.Proposed Methodology

In this research paper, Medicinal plant identification and recognition using machine learning classifiers is proposed. A real medicinal plant data set is created, trained; features are extracted and stored as a feature data set. For each test image taken, similar features are extracted, mapped with the training feature data set and classified using Decision tree, K-Nearest Neighbour and Random Forest Classifiers. The morphological Shape and texture features of leaf images are considered for the classification process. The proposed method contains five phases: Preparing Medicinal leaf Data set, Image Pre-processing, Feature Extraction, Image Classification and Identification and Recognition. The flow diagram of the proposed method is shown in figure 1.

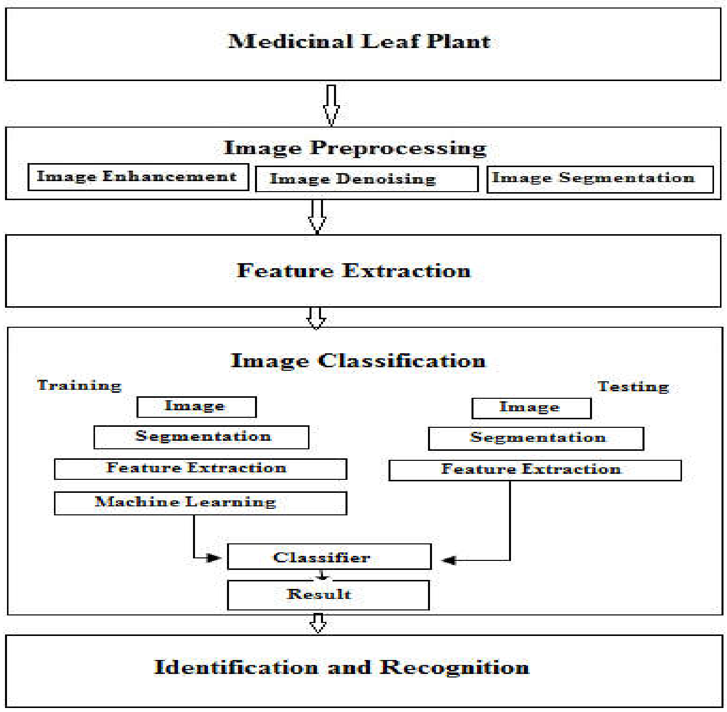

Figure 1. Flow Diagram of the Proposed Method

1.1. Medicinal Plant Leaf Data set

The plant leaves are more suitable for machine processing due to their two dimensional nature. The leaf images are captured through the Nikon camera P5 and stored in RGB format. The light intensity of the images is avoided by two methods:

- Keeping the leaf to be captured in a black box
- By placing leaf on a white background with light sources at 45 each side of the leaf to introducing better brightness and eliminate reflection.

A medicinal plant leaf data base with 816 leaf images of 30 different plant classifications of size 1024 x 600 stored in jpeg format is created. The taxonomy file containing the medicinal details of the specimen leaves is also created in Excel. Twenty leaf images of different classes and different leaf images of same class are shown in figure. 2 and figure 3.

Figure 2. Medicinal Plant Leaf Images of Different Classes.

Figure 3. Leaf images of same class.

2.1 Leaf Image Preprocessing

To eliminate the intensity variations or poor contrast Image Pre-processing is applied to get an enhanced image or to extract some useful information from it. The preprocessing techniques proposed in this research work are:

- Image Enhancement
- Image Denoising
- Binarization

2.1.1 Image Enhancement

The objective of image enhancement is transforming the RGB image to a suitable color space which is device independent in nature. Before transforming the input leaf image into suitable color transform, first the acquired leaf image will be resized. Then the image will be converted into a suitable color space so that the required information can be easily extracted more efficiently. The features may be extracted from grayscale images efficiently when compared to color images. So, the input RGB image is converted into a grey-level image using (1).

$$Y = (0.299 \times R) + (0.587 \times G) + (0.114 \times B) \tag{1}$$

where Y represents the gray value equivalent of the color pixel, R represents the Red component, G represents the Green component and B represents the Blue component of a pixel in the color image.

2.2.2. Image Denoising

An Iterative Morphological Filtering operation with the selected size and shape of the Structure Element (SE) is proposed to remove the shadows and undesired perturbations like impulse noises which may affect the computation of leaf Feature values. The impulse noises which are less than the size of the selected SE are removed effectively and this process is iteratively repeated until the spatial variation of the intensity becomes locally monotonic with respect to the selected SE. Any feature of size smaller than that may be treated as noise. The procedure for Iterative Morphological Filtering is given below [10].

1) Perform conventional morphological opening and closing on the input image using an SE of small size. The size of the SE is greater than that of noise particles.

2) Construct the output image by averaging the images resulting after opening and closing.

3) Compare the output image with the input image.

If they are identical, then halt. Otherwise consider the output image as the input image to the next iteration and go to step 1.

2.2.3 Image segmentation

Image segmentation is a process, where every pixel presents in an image is assigned a label so that same label of pixels shares the visual characteristics. Thresholding is a technique normally used to isolate the background or foreground of an image. The segmentation results depend on the selection of correct threshold value. There are various methods available but each has its own limitations and advantages in terms of applicability and suitability. Each threshold technique works best for certain set of images. In this work, fuzzy optimal Threshold value using Gaussian measure of fuzziness is calculated and applied [3].

Normal Image **Preprocessed Image**

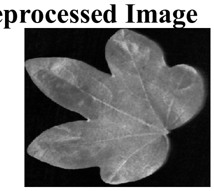

Figure 4. Leaf Image Preprocessing

2.3 Feature Extraction

The leaf images are classified based on the morphological shape and texture features.

2.3.1 Morphological Shape Features

A set of morphological features Smooth factor, Form factor, Narrow factor, Perimeter ratio of diameter and Vein features are computed for training feature data set.

1. Smooth factor: Image smoothness is measured by the effect of noise on the image area. It is a ratio between the area of the leaf image smoothed by 5 × 5 rectangular average filter and the one smoothed by 2 × 2 rectangular average filter.
2. Form factor: This feature is defined as 4πA/P 2, where A is the leaf area and P is the perimeter of the leaf margin.
3. Narrow factor: Narrow factor is defined as the ratio of the diameter D and physiological length LP (D/LP).
4. Perimeter ratio of diameter: Defined as P/D, where P is the leaf perimeter and D is the leaf diameter.
5. Vein features: Each leaf can be uniquely identified by its veins. The background information is subtracted and only the vein patterns are extracted by applying morphological opening operation.

The morphological shape features computed for the Sample image is given in Table1.

Table 1. Shape Features

S.NO	Feature	Value
1.	Smooth Factor	1.0028
2.	Form Factor	0.1587
3.	Narrow Factor	1.0222
4.	Vein Features Av1/A	0.1232
5.	Vein Features Av2/A	0.5339
6.	Vein Features Av3/A	0.6614
7.	Vein Features Av4/A	0.6975
8.	Vein Features V4/Av1	5.6595
9.	Perimeter Ratio Of Diameter	7.1182

2.3.2 Texture Features

A texture based Local Vector Pattern which possess high variance between inter class images such as robust against aging of images, illumination changes and other such factors has been proposed to extract the texture features of the leaf images [9]. The Local Vector Pattern of size 256 x 256 is computed for different orientations (0o,45o, 90o &135o) and stored as a training feature dataset with reduced dimension.

The feature set for all input images are computed and stored as s feature data set which can be given as the input to the classifier to train the data set.

2.4 Image Classification

Image classification is carried out in two phases.

2.4.1 Training Phase

Image classification is one of the most complex steps in image processing and pattern recognition. In the classification phase, the supervised machine learning classifiers Random Forest, K-nearest neighbour and Decision tree are proposed and their performances are evaluated [1].

I. Random Forest Classifier: Random Forest is a simple and flexible machine learning algorithm used for both classification and regression tasks. It creates a set of decision trees from randomly selected subset of training set and then aggregates the votes from different decision trees to decide the final class of the test object. The classification is performed in two stages namely Random forest creation, Random forest prediction

II. K-Nearest Neighbour (KNN): KNN classifier is instance based classifier that performs classification of unknown instances by relating unknown to known by using distance or similarities function. It is a kind of classifier which can train and test data at the same time. It takes K nearest point and then assigns a class of majorities to the unknown instance.

III. Decision Tree: Decision trees are powerful and popular tools for classification and prediction. They represent rules, which can be understood by humans and used in knowledge system such as database. Decision trees classify instances or examples by starting at the root of the tree and moving through it until a leaf node. Decision tree classifiers have the ability to convert the complex decision into understandable and easy decisions.

2.4.2. Testing Phase

From the training set, ten medicinal plant leaf images of each class are taken as query or test images, feature vector for each image is generated, tested against the training feature data set and classified by the proposed classifiers. The machine learning classification results are shown in Table 2.

3.Identification & Recognition Phase

From the output of the classifier, the class of the test image is identified, recognized and their medicinal details (name, usage) are displayed from the taxonomy file. The sample images, classification results of three proposed machine learning algorithms and their medicinal details are shown in figure 5.

Table2. RF,KNN and DT Classifier Output

Traditional Name of the plant	Scientific Name	LC	TR	TS	RF-CC	RF-MC	KNNCC	KNNMC	DT-CC	DT-MC
Kova	Psidium Guajava	1	26	10	9	1	8	2	9	1
Karuvapillai	Murraya Koenigii	2	28	10	9	1	9	1	10	0
Maruthani	LawsoniaInermis Linn	3	26	10	8	2	5	5	10	0
Kuppaimeni	Acalypha Indica	4	19	10	3	7	3	7	3	7
Vembu	Azadirachta Indica	5	31	10	9	1	8	2	7	3
Eucalyptus	Eucalyptus Citriodora	6	23	10	7	3	7	3	8	2
Paruthi	Gossypium Sps.	7	25	10	8	2	7	3	6	4
Pakarkai	*Momordica Charantia* L.	8	23	10	9	1	8	2	6	4
Thuthi	Abutilon Indicum	9	25	10	7	3	7	3	5	5
Omam	Carum Carvi	10	32	10	10	0	10	0	7	3
Vila	Limonia Acidissima	11	30	10	10	0	8	2	10	0
Vethalai	Piper Betle	12	22	10	7	3	5	5	4	6
Thoothuvalai	Solanum Trilobatum	13	34	10	9	1	8	2	8	2
Elumichi	Citrus Limon	14	21	10	10	0	10	0	9	1
Maa	Mangifera Indica	15	30	10	6	4	7	3	6	4
Adhatoda	Adhatoda Vasaca	16	41	10	3	7	3	7	1	9
Kalaa	Carissa Carandes	17	29	10	10	0	5	5	7	3
Thaanri	Terminalia Bellerica	18	56	10	9	1	9	1	9	1
Senbaruthi	Hibiscus Rosasinensis	19	23	10	8	2	7	3	6	4
Arali	Nerium Oleander	20	17	10	7	3	4	6	4	6
Nilavaembu	Andrographis Paniculata	21	24	10	10	0	9	1	10	0
Kadukkai	Terminalia Chebula	22	16	10	6	4	6	4	4	6
Etti	Strychnos Nux-Vomica	23	19	10	8	2	7	3	5	5
Malaivampu	Melia Azedarach	24	23	10	10	0	6	4	6	4
Semaiakathi	Cassia Alata L.	25	25	10	8	2	8	2	7	3
Mahilampoo	Mimusops Elengi	26	17	10	7	3	8	2	5	5
Vetpallai	Wrightia Tinctoria	27	32	10	0	10	0	10	2	8
Mathulai	Punica Granatum	28	31	10	10	0	7	3	7	3
Nellikai	Phyllanthus Emblica	29	33	10	9	1	9	1	9	1
Latchakattai	Pisonia Grandis	30	35	10	7	3	6	4	4	6

*LC Leaf Class, TR - Training Set, TS - Testing Set, CC - Correctly Classified using RF,KNN,DT, MC – Misclassified using RF,KNN,DT.

Input image	RF Classifier	KNN Classifier	Decision Tree Classifier
	Correctly classified as the name of this plant13_DSC_0262.JPG is : Thoothuvalai	Mis classification the name of this plant13_DSC_0262.JPG is:Nilavaembu	Mis classification the name of this plant13_DSC_0258.JPG is:Paruthi
	Correctly classified as the name of this plant9_DSC_0109.JPG is:Thuthi	Correctly classified as the name of this plant9_DSC_0109.JPG is:Thuthi	Mis classification the name of this plant9_DSC_0109.JPG is:Elumichi

Figure 5. Identification & Recognition of the Plant Leaves.

The Random Classifier outperforms when the image is clean as well as having shadow effect.
KNN misclassifies the both leaves. The decision tree classifier misclassifies the 'Thuthi' leaf into 'Elumichi' due to the shadow effect on the leaf.

4.Performance Evaluation

The classification results obtained from the machine learning classifiers for the test images are evaluated and compared based on the metrics Accuracy, Sensitivity, Specificity, Precision, Recall and Fscore and shown in fig 6. The scatter plot for the classification results is shown in fig 7.

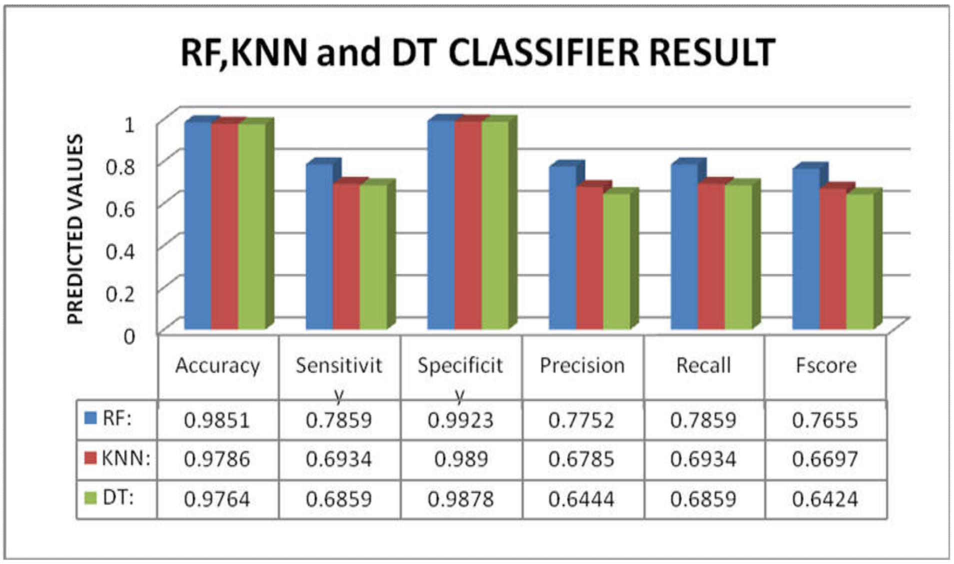

Figure 6. Performance Metrics of the Classifier Results.

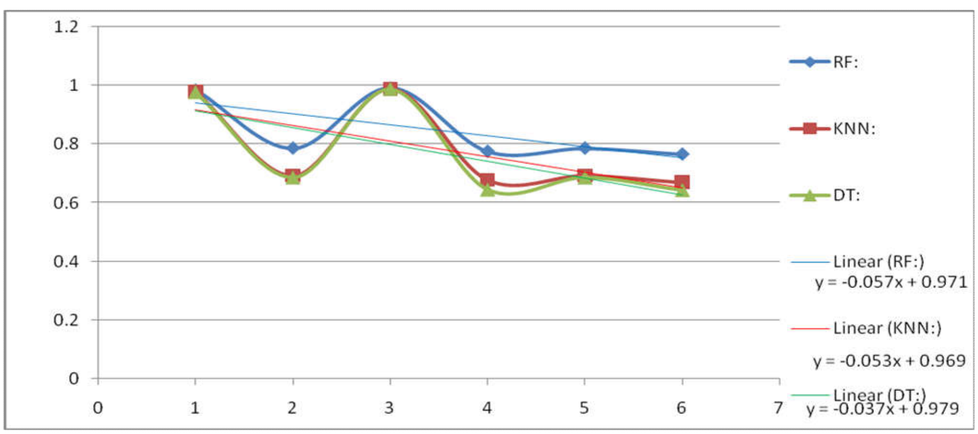

Figure 7. Scatter Plot for the Classification Results

5.Conclusion

Thirty different classes of 816 real medicinal plant leaf images are collected, Fuzzy thresholded morphological preprocessing is applied, Morphological shape features and the Local vector Patterns for each medicinal plant leaf are computed and stored as a training feature data set. From the training set, Ten leaf images of each class are taken as query or test images, feature vector for each image is generated, tested against the training data set and classified by the machine learning Decision tree, K-Nearest Neighbour and Random Forest Classifiers. From the experimental analysis it is found that the

Random Forest Classifier performed better by giving 99% of accuracy for a large data set and both KNN and Decision Tree classifiers yield better results when the input data set is small.

References

[1] K.Ponmalar, Dr.K.Krishnaveni, "Random Forest Classification of Medicinal Plant Leaves Using Shape and TextureFeatures", Journal of Applied Science and Computations, pp. 461-569, 2018 .

[2] Shyam Vijayrao Pundkar, Prof. M.M. Waghmare"Study of Various Techniques for Medicinal Plant Identification" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 11 3340 – 3343.

[3] Herdiyeni, Y. ; Wahyuni, N.K.S.,"Mobile application for Indonesian medicinal plants identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram"Advanced Computer Science and Information Systems (ICACSIS), 2012 International Conference,Page(s): 301 - 306,2012.

[4] Stephen gang wu, forrest sheng bao, eric you xu, yu-xuanwang, yi-fan chang and qiao-liangxiang (2007) "a leaf recognition algorithm for plant classification using probabilistic neural network", arxiv:0707.4289vl [cs.ai].

[5] Krishnan.N, Krishnaveni.K, et. al., "A Fuzzy Thresholded Multiscale Morphological Segmentation of Digital Images", International Journal of Signal Processing, Special Issue on "Soft Computing in Image Processing Techniques &Systems", vol. 3, no. 1, pp. 71-78, January, 2005.

[6] Krishnan.N, Krishnaveni.K, et. al., "Fuzzy Optimal Thresholded Multiscale Morphological Segmentation of Digital Images", Proceedings of IEEE Third International Conference on Wireless and Optical Communications Networks WOCN 2006, ISBN: 1-4244-0340-5, April 2006.

[7] Shyam Vijayrao Pundkar, Prof. M.M. Waghmare"Study of Various Techniques for Medicinal Plant Identification" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 11 3340 – 3343.

[8] Gopal, A. ; Prudhveeswar Reddy, S. ; Gayatri, V. "Classification of selected medicinal plants leaf using image processing" Machine Vision and Image Processing (MVIP), 2012 International Conference, Page(s): 5 – 8, 2012.

[9] Bhandarkar, P. ; Doshi, H. ; Kadakia, M. ; Ahmed, R. "Leaf identification using morphology and structural decomposition" Signal Processing and Integrated Networks (SPIN), International Conference on DOI: 10.1109/SPIN.2014.6776934 Page(s): 125 – 129 2014.

[10] Herdiyeni, Y. ; Wahyuni, N.K.S.,"Mobile application for Indonesian medicinal plants identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram"Advanced Computer Science and Information Systems (ICACSIS), 2012 International Conference,Page(s): 301 - 306,2012.

[11] A. Jenitta & R. Samson Ravindran. "Image Retrieval Based on Local Mesh Vector Co-occurrencePattern for Medical Diagnosis from MRI Brain Images" Springer Science+Business Media, LLC 2017Page(s): 41:157, 2017.

[12] Krishnan.N., Krishnaveni.K.*, et. al., "A Novel Multiscale Morphological Watershed Segmentation ", International Journal of Imaging Science and Engineering (IJISE), vol. 1, no.2, April 2007.