Advanced Leaf Recognition Algorithm using Artificial Neural Network

Voligi Monica Bharathi[1] D. Bhanu Prakash[2]
[1]M. Tech Student [2]Associate Professor
[1,2]CVR College of Engineering, India

Abstract— Plant identification is an important field of biological and medical sciences. Medicinal plants must be classified and recognized with high accuracy. Classification errors can lead to high costs and losses. Several methods have been proposed so far for plant classification based on leaf image. Generally, the methods don't include all the physical attributes of the leaf which entails important information. In this paper, various characteristics shape, color, texture and vein have been computed which resulted in more reliable results. The use of neural networks in science has seen a huge revolution in the last decades. Neural networks have been introduced in different scientific applications and proved their high efficiency with minimum costs. The identification of plant's category of input images has been accomplished along with the classification of various plants into their respective categories. In this case, combined classifier using majority voting technique has been proposed for recognizing leaf's category along with SVM, Naïve Bay's and Decision tree classifiers. The experimental results show that combined classifier outperforms the performance of SVM; Naïve Bay's and Decision tree classifiers.

Keywords: Plant Identification; Combined Classifier; Naïve Bayes; Decision Tree; SVM; Visual Features

I. INTRODUCTION

Plants are essential for humans and animals survival on the earth [1] as these are the sources of many natural resources. As there are plethora of plants present on the earth which look identical so it is difficult to differentiate them just from their physical appearance. Moreover, correct identification of plants are required in order to avoid any hazard due to various poisonous or dangerous plants. Also, correct identification of plants also helps botanist in their researches and for the ease of farmers. Thus, digital system for plant identification from digital image is the solution to this problem as the error rate of digital computer systems is very less as compared to the human visual inspection.

In digital systems, the input is a leaf image from which features are extracted and based on those, classifier performs classification. Plants can be distinguished on the basis of their parts such as fruits, flowers, leaves and so on. Flowers and fruits of the plants are seasonal and may dry up but leaves remain the entire year and hence leaves are the best way for their identification [2, 3,4]. In the previous work done, several methods have been proposed based on either shape, vein, color, texture or all physical characteristics, and various classifiers such as SVM, PNN, Naïve Bays have been used for classification purpose.

Here, a combined classifier has been implemented which shows better results by improving accuracy that gives more reliable system. In this paper, proposed approach is discussed in Sec II, Result analysis discussion in Sec III, Conclusion in Sec IV.

II. PROPOSED APPROACH

A. Image Preprocessing

The input colored digital images are resized to 256*256 for appropriate processing. Input image is in RGB color space. Firstly, Gray image is determined for vein feature extraction as shown in fig. 1(b). Further, R planes, G planes, B planes of the image are extracted for segmentation purpose. OR operation is performed on the extracted R, G, B planes of the image to obtain binary image which is a black image on white background as shown in fig. 1(c). Further, the resultant is complemented, the derived image contains noise thus morphological hole filling operation [5] is performed to remove noise as shown in fig. 1(d). Moreover, leaf's boundary image as shown in fig. 1(f) is obtained by retaining only those pixels that indicates boundary of the leaf where there is instant intensity variation. A dilated complemented image is subtracted from the original image to obtain leaf boundary. Also, from Red, Green and Blue planes as shown in fig. 1(g), 1(h) and 1(i) color information is derived. In addition to it, gray scale cropped image for texture features is extracted as shown in fig. 1(e).

B. Feature Extraction

In this paper, combination of visual low level features like shape, color and texture have been extracted for each image. There are various shape descriptors [6] extracted in this work. Starting from the binary image, 28 shape descriptors also known as basic and derived morphological features have been extracted and out of these 15 are boundary descriptors [7], 12 are regional descriptors [8] and an euler number. The boundary descriptors extracted are perimeter, diameter, radius, minor axis length, bounding box height and width, eccentricity, major axis length, convex perimeter, elongation, perimeter ratio of length and width, perimeter ratio to diameter, narrow factor, extent, equivdiameter. area, compactness, convex area, circularity ratio, rectangularity, solidity, convexity, smooth factor, minimum bounding box area, filled area, and orientation forms the regional descriptors. These shape descriptors have the ability of discriminating the various shapes thus helpful in discriminating different leaves species. Also, dispersion is calculated for the irregularity in the shape of the leaves [9].

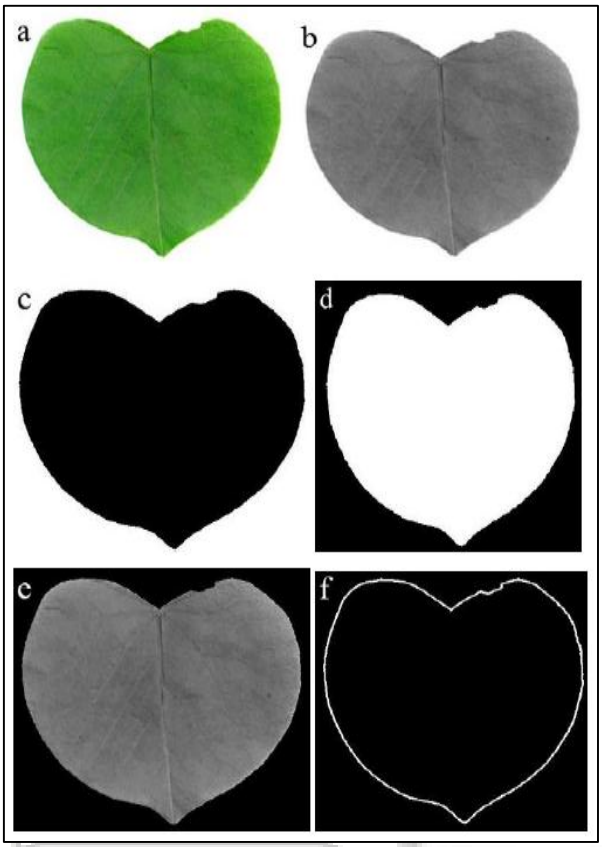

Fig. 1: (a) RGB scaled input image; (b) Gray level image; (c) Binary image; (d) Binary complement image; (e) Gray scale cropped image; (f) Boundary extracted;

$$D = \frac{\max\sqrt{((x(i) - x^2) + (y(i) - y^2))}}{\min\sqrt{((x(i) - x^2) + (y(i) - y^2))}} - -(1)$$

Vein features of the leaves are extracted by implementing morphological opening operation with flat disk shaped structuring element of radius 1,2,3,4 on grayscale image. Resultant images are subtracted from the grayscale image.

Thus on total we have five vein features v1 = a1/a, v2 = a2/a, v3 = a3/a, v4 = a4/a, v5 = a4/a1 where a1, a2, a3, a4 are the areas of the images obtained after subtraction operation and 'a' corresponds to the area of the input binary image.

Color moments [10, 11] that are mean, skewness, kurtosis, standard deviation, entropy, smoothness, and uniformity have been implemented for extracting color features. Since leaves are of one color only, thus color distribution in image follows probability distribution which makes use of color moments optimal. Color moments [12] are calculated for i-th color channel at j-th image pixel as P_{ij} for M total number of pixels in image. Therefore twenty one color moments have been calculated for the image, including seven for each color channel.

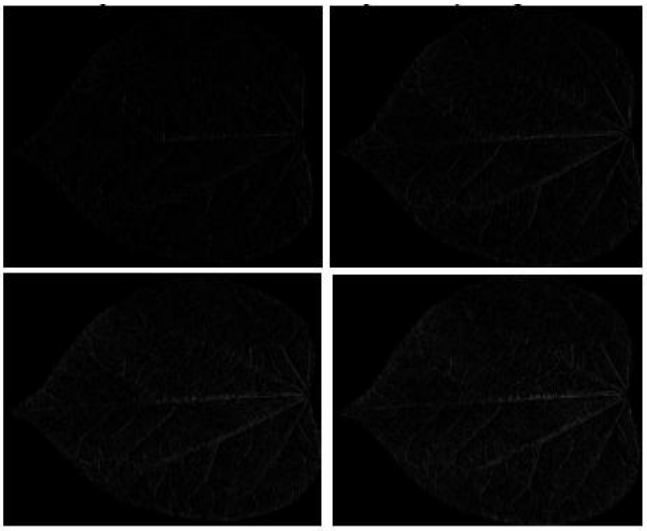

Fig. 2: Leaf vein images obtained with structuring disk 1, 2, 3, 4 respectively

Gray level co-occurrence matrix (GLCM) [13, 14] has been used for the extraction of texture features as it counts the occurrence of pixels with specific grey values at a given distance. GLCM twenty features given by Harlick [15] have been calculated at the distance 1 for angles 135, 90, 45 and 0 degrees. Thus on total eighty features have been determined at four inclinations. The features calculated are homogeneity, energy, variance, contrast, dissimilarity, correlation, autocorrelation, information measure of correlation1, cluster shade, sum entropy, difference variance, difference, inverse difference normalized, inverse difference moment normalized, maximum probability, information measure of correlation2, cluster prominence, entropy, sum average and sum variance.

C. Classification and Identification

Classifier performs the operation of categorization of input data into various classes based on spectral features. The classifiers can be differentiated on the basis of learning technique as supervised and unsupervised. In this paper, supervised classification is used which is a two phase process that employs training and testing. In the training phase, set of input data of known categories are constantly presented to the system. While in testing phase, a new data but of same category as the training data is presented to the classifier. In this paper, various supervised classification techniques such as SVM [16, 17], Naïve Bayes, Decision tree and combined classifiers have been implemented for plant identification based on leaf image. SVM (support vector machine) is non-probabilistic linear classifier. It is inherently a binary classifier, thus input data can be classified only in one of the two categories but it can be used as multiclass classifier by reducing multiclass problem into several binary class problem by using approaches one versus one, one versus all and error correcting output codes [18]. In this work, multiclass SVM using one versus all have been implemented. Naïve Bayes classifier is a probabilistic classifier i.e. based on Baye's theorem. It assigns labels to testing images based on feature vector. In decision tree classifier, different nodes are formed based on attributes which further branches out to various resultant outputs.

The resultants of the different classifiers can be combined using different approaches to increase the accuracy of the system. In this work, the resultants of the SVM, Naïve Bayes and Decision tree have been combined using majority voting technique to form combined classifier. In majority voting approach, the output of the system is the predicted class label with maximum votes.

D. Image Identification

In image identification, pattern of the image is recognized and the output displayed is the specific name of the detected pattern. In this work, various classification algorithms have been used to correctly identify the plants based on the leaf in the image. For instance, the fig.3 is given as an input to the system and name of the plant of the leaf in the image 'Chinese redbud' is displayed as the output.

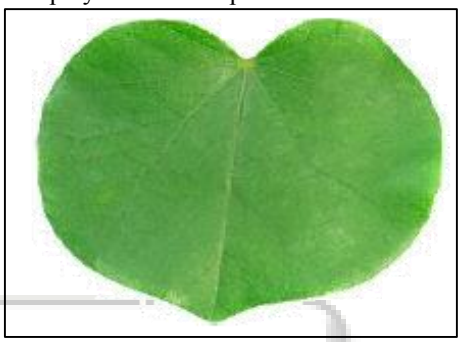

Fig. 3: Chinese redbud leaf

III. RESULT ANALYSIS

The performance of combined classifier is tested on Flavia dataset containing plants of 32 categories. Further, the performance of the implemented system using various feature extraction techniques and classifiers are evaluated on the basis of the parameters- accuracy, precision and recall.

The experiment is conducted with different number of input images in training and testing to calculate the variation in the accuracy of system. Also, images are tested with SVM, Naïve bayes and decisioin tree classifiers to compare their accuracy achieved with combined classifier. For example, from the table when number of input leaf images are 1536, Accuracy achieved by combined classifier is 93.11% while with SVM it is 91.8576, Naïve bayes accomplishes 88.2736% and Decision tree attains 85.6678%. It is ascertained from the various simulated results concluded in the table that accuracy is escalated by proposed combined classier as compare to already existing classifiers.

Moreover, in this work, number of feature extracted are 139 which are much more than already done researches based on feature extraction. As 33 shape features, 5 vein features, 21 color features and 80 texture features have been calculated for each image which leads to better precision of results. 139 which are much more than already done researches based on feature extraction. As 33 shape features, 5 vein features, 21 color features and 80 texture features have been calculated for each image which leads to better precision of results.

Classifier	Input Parameters		Evaluation Parameters		
	Number of input leaf images	Number of input classes	Accuracy	Precision	Recall
SVM Naïve Bayes Tree Combined Classifier	1536 Train images:1229 Test Images:307	32	91.8576 88.2736 85.6678 93.811	.8940 .8729 .8385 .9278	.8946 .8729 .8385 .9278
SVM Naïve Bayes Tree Combined Classifier	1280 Train images:1024 Test Images:256	32	93.8576 84.2736 82.6678 93.811	.9240 .8729 .8285 .9278	.9240 .8729 .8285 .9278
SVM Naïve Bayes Tree Combined Classifier	1170 Train images:936 Test Images:234	20	95.8576 91.2736 85.6678 94.811	.9140 .9129 .8185 .9078	.9140 .9129 .8185 .9078
SVM Naïve Bayes Tree Combined Classifier	1162 Train images:930 Test Images:232	20	90.8576 85.2736 86.6678 93.811	.9040 .8529 .8585 .9378	.9040 .8529 .8585 .9378
SVM Naïve Bayes Tree Combined Classifier	320 Train images:256 Test Images:64	32	78.8576 81.2736 65.6678 82.811	.7840 .8129 .8385 .9278	. .7840 .8129 .8385 .9278
SVM Naïve Bayes Tree Combined Classifier	301 Train images:241 Test Images:60	6	100 98.2736 98.6678 100	1 .8729 .8385 1	1 .8729 .8385 1

Table 1: System analysis for 139 features using different classifiers with different training and testing images

IV. CONCLUSION

The research paper demonstrates the enhanced performance of the technique used for the classification and identification of plant leaf images using combined classifier. The research methodology illustrates feature extraction like shape, color, texture and vein. Where in total 139 features have been extracted followed by classification and identification process using different classifiers like SVM, Decision tree

and combined classifier. The results obtained shows that for different number of input leaf images and classes taken during the classification process, the accuracy varies, where it has been observed that maximum accuracy have been achieved with combined classifier. The performance of the combined classifier have been compared with the SVM, Naïve Bay's and Decision tree as in the aforementioned results, where it has been proved that the performance accuracy is intensified as compared to existing methodologies.

V. FUTURE SCOPE

Furthermore, the advancement of the algorithm can be accomplished in order to identify category of leaf image from colored background as well as in multiple object image since in this work dataset is based on white background. Likewise, more efficient system with less computational time can be chased. Moreover, the work can be extended to identify plant's category based on attributes roots, barks etc.

REFERENCES

[1] Sapna Sharma and Dr. Chitvan Gupta, "A Review of Plant Recognition Methods and Algorithms", IJIRAE, ISSN: 2349-2163 Issue 6, Vol June 2015.
[2] Sachin D. Chothe, V.R.Ratnaparakhe "Plant Leaf Identification System", International Journal on innovative Research in Science, Engineering and Technology Vol.4, Special Issue 6, May 2016.
[3] Nisar Ahmed, Usman Ghani Khan and Shahzad Asif, "AN Automatic Leaf Based Plant Identification System", 5th International. Multidisciplinary Conference, 29-31 Oct.
[4] Andrade, M., Mayo, S. J., Kirkup, D., & Van Den Berg, C., "Comparative morphology of population of Monstera Adans. (Araceae) from natural forest fragments in northeast Brazil using elliptic Fourier analysis of leaf outlines", Kew Bulletin, 63, 193–211, 2008.
[5] R. Gonzales, R. E. Woods, "Digital Image Processing," 2nd Ed., New Jersey Prentice Hall, 2002.
[6] JX Du, XF Wang, GJ Zhang, "Leaf shape based plant species recognition", Applied Mathematics and Computation, Volume 185, Issue 2, Pages 883–893, 15 February 2007.
[7] Mingqiang Yang, Kidiyo Kpalma and Joseph Ronsin, "A Survey of Shape Feature Extraction Techniques", Pattern Recognition, IN-TECH, pp.43-90, 2008. <hal-00446037>.
[8] Dengsheng Zhang, Guojun Lu, "Evaluation of MPEG-7 shape descriptors against other shape descriptors", Multimedia Systems springer 9: 15–30, 2003.
[9] Nixon, M.S. & Aguado, A.S. Feature Extraction and Image Processing. Woburn: Newness, 2002.
[10] Noah Keen and Dr. Bob Fisher "Color Moments" https://pdfs.semanticscholar.org/bccc/f2b2142f62a217c035c1e8003b0363c43d2a.pdf.
[11] Martinez, W., & Martinez, A. 2002. Computational Statistics Handbook with MATLAB. Bocca Raton: CRC Press LLC.
[12] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto and Paulus Insap Santosa, "Leaf Classification Using Shape, Color, and Texture Features", International Journal of Computer Trends and Technology-July to Aug Issue 2011.
[13] Acharya, T., & Ray, A. 2005. Image Processingg Principles and Applications. New Jersey: John Wiley & Sons, Inc.
[14] Abdul Kadir, "A Model of Plant Identification System Using GLCM, Lacunarity and Shen Features", https://pdfs.semanticscholar.org/a70a/c3a6ffc33055a4ae9e40d68f8bb70096efe4.pdf.
[15] R. Haralick, "Statistical and Structural Approaches to Texture", Proc. IEEE, 67, 5, 786-804, 1979.
[16] Cortes C, Vapnik VN (1995) Support vector networks. Machine Learning 20:273–297.
[17] Pradeep Nijalingappa and V. J. Madhumathi, "Plant identification system using its leaf features", IEEE 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 29-31 Oct. 2015.
[18] Mohamed Aly, "Survey on Multiclass Classification Method", https://www.cs.utah.edu/~piyush/teaching/aly05multiclass.pdf. , November 2015.

