www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issues 6 June 2016, Page No. 17022-17025

Comparative study of various classification algorithms combined with K means algorithm for Leaf Identification

Nisha Pawar, Dr.K.Rajeswari
Department of Computer Engineering
Pimpri Chinchwad College of Engineering
Pune-44
pawarnisha222@gmail.com
Professor and Head, Department of Computer Engineering
Pimpri Chinchwad College of Engineering
Pune-44
raji.pcccoe@gmail.com

***Abstract*— Plants play a vital role in our daily life. Plants are a great source of medicine for many diseases. Due to their fewer chances of side effects on human body and also better compatibility with humans, using plant for treating diseases is considered to be safer. Other items like paper, bio-diesel are also obtained by using plant material. Hence identification of plants is a very important task helpful for various areas such as Agriculture, Ayurveda, Botanical research, Biological research, etc. Leaf based features can be used for appropriate results than other parts of the plant in the plant identification. Manual leaf identification is very difficult and time consuming task. To implement automatic leaf identification, classification techniques like Naïve Bayes classification, Neural Network, etc can be used. To get the leaf based features, image processing techniques are applied on the image of leaf. After leaf based feature extraction, a plant's leaf is classified based on the leaf features. The main objective of this paper is to present a survey of different classification methods for plant leaf identification. At the end, this paper concludes better classification method with more accuracy when compared to other classification methods.**

***Keywords*— Plant leaf identification, Classification, Leaf based features, Image processing.**

I. INTRODUCTION

Humans have been associated with the plant kingdom since for many years. Plants such as medicinal plants, crops or other useful plants are very important for the human's ecosystem. We can get the food, shelter and other useful things in our daily life. The plant also helps in reducing the pollution on the earth. The plant has plenty of use in foodstuffs, botany and many other industries [1]. In ayurvedic treatment, use of correct medicinal plant is very important. But manual identification of various medicinal plants is very critical. For this, automatic recognition of plants will be helpful. The automatic recognition method will save the time and will increase the speed of identification of plants. The automatic recognition of plants will identify and classify the plants using some available information of plants. Leaf has several advantages over flowers or fruits in identifying the plant such as its 2-dimensional nature and availability at all seasons worldwide [2]. Hence the leaf based features can be used to get appropriate results than other parts of the plant in the plant identification.

The goal of this paper is to focus on features extraction and classification on the leaf based features. In this paper, some useful leaf based features are used to get the appropriate results. This paper also gives the overview of comparative study of different classification techniques used in leaf identification.

The rest of the paper is organized into the following sections: Section 1 gives an introduction about the importance of plants in our life and use of the plant leaf identification and classification. Section 2 describes a brief literature review on the various leaf features based plant leaf identification methods. Section 3 describes various leaf based features used in leaf recognition. Section 4 describes proposed model for leaf based classification. Section 5 includes the comparative study of various popular classification methods used for leaf identification. Finally, Section 6 concludes this paper**.**

II. LITERATURE REVIEW

In the paper [10], Shilpa Ankalaki, Laxmidevi Noolvi, Dr. Jharna Majumdar proposed system in which Fuzzy C means clustering method for clustering the images and Naïve Bayesian classification to classify the leaf image. In this paper, the work consists of two phases first is feature extraction phase and another is classification phase used in image analysis and image processing. This paper used the leaf extraction method of the given leaf image by using some shape based features. The proposed methodology gives the 83.24% accuracy in leaf identification.

In the paper [9], Liwen Gao, Xiaohua Lin studied and implemented an automatic recognition system of medicinal plants. In this paper, the system used neural network classifier for leaf identification. The proposed system can be used to identify the different species of medicinal plants. This system

transformed manual identification system into semi-automatic leaf recognition system.

In the paper [7], E. Sandeep Kumar and Viswanath Talasila proposed to perform automatic recognition of medicinal plants and to analyze the statistical nature of the image features used for recognition. In experimental analysis, ten plant species belonging to the same families were used. In leaf feature extraction, the leaf features are Gaussian distributed, which served as a useful result in accurate classifier design by calculating the precise decision boundaries between the classes [7]. In this paper, the given work has more and better accuracy than their previous work.

In the paper [6], Aamod Chemburkar, Anand Sartape, Ajinkya Gawade, Prasad Somawanshi, Jayshree Ghorpade developed an automated tool which would compare trained tests, detect and classify the plant leaf species. After image processing, proposed system used artificial neural network. In this paper, edge detection and vein structure extraction was done in training phase. Prewitt edge detection filter was used for edge detection. In this proposed work, leaf identification was done based on edge detection and vein structure extraction for leaf identification.

III. Leaf Based Feature Extraction

Leaf based features can be extracted from vein patterns, color, textures and shape information. But many algorithms focus on shape based features. By using shape based features accuracy increases with increasing speed of image processing of a leaf image [3]. Some of the leaf based features are as follows:

A. Aspect ratio:

The aspect ratio is the ratio between the maximum length and the minimum length of the minimum bounding rectangle or the ratio between length and width of the minimum bounding box of a leaf image [10] as given in Eq.1

$$AspectRati \quad o = \frac{Length}{Width} \qquad (1)$$

B. Perimeter:

The total number of pixels on the leaf boundary [10].

C. Roundness:

Roundness is the measure of how closely the shape of leaf approaches that of a circle [10]. The difference between a leaf and a circle is calculated by using the Eq.2

$$Roundness \quad = \frac{4\Pi * Area}{Perimeter^{2}} \qquad (2)$$

D. Eccentricity:

Eccentricity is defined as the ratio of minor principal axes to major principal axes [10].

E. Rectangularity :

Rectangularity is the measure of how closely the shape of leaf approaches to rectangle or it can be defined as the similarity between leaf and rectangle [10].

IV. Proposed Architecture

Proposed leaf identification system has two phases, first is training phase and second is identification phase. Before starting any phase, image acquisition is done. Images of leaves can be captured by using camera having high resolution. To avoid unnecessary background, image can be captured by placing on white background.

In the next step, image can be preprocessed to make a leaf identification easier and to increase speed of identification. The input image of a leaf is converted to gray scale image. Then a binary image of gray scaled image is created. And at the last in the preprocessing stage, the boundary is extracted using boundary extraction algorithm such as Canny edge detection. The Figure 1 shows the steps of leaf image preprocessing.

After image preprocessing, leaf based features as given in Section III are calculated. Then, if it is a training phase, then add these feature values in database. Repeat this process for all plant's leaf images. After finishing of features insertion in the database, apply clustering algorithm such K means algorithm to create clusters of similar images. It will assign the leaf images to specific cluster based on similar feature based values.

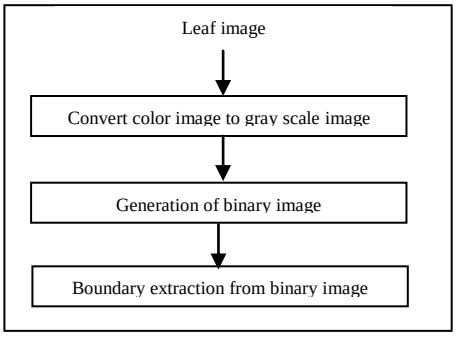

Figure 1. Preprocessing of input leaf image

DOI: 10.18535/ijecs/v5i6.40

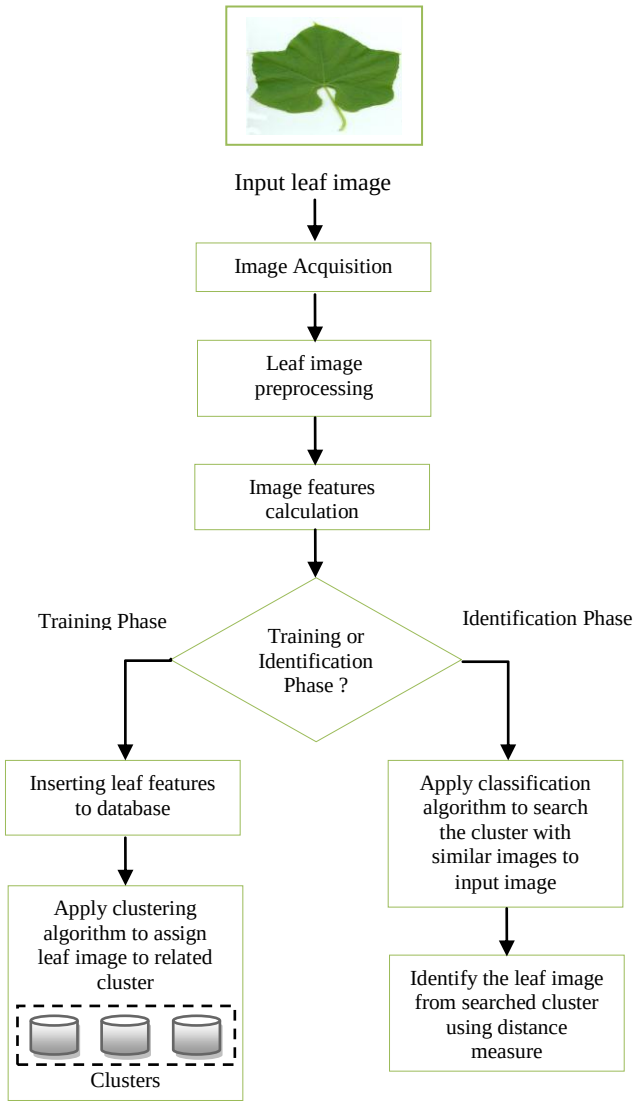

Figure 2. Flowchart of training and identification phase of Leaf identification.

In case of identification phase, after preprocessing, takes the features of input image, search the cluster using classification method and identify leaf image from given cluster using any distance measure. Figure 2 shows the flowchart of training phase and identification phase.

V. Comparative Study of Classification Methods With K Means Clustering

After extraction of leaf based features, clustering method can be used to make a group of plant leaves. So that leaves having similar features are placed in the same cluster. When new image comes, classification method is used to identify the similar cluster for that new image. In following table Table 1, comparative study of different classification methods with K means used for the leaf identification is given.

	Combination of K Means Clustering and Classification Methods for Leaf Identification		
	Naïve Bayesian Classification	***Decision Tree***	***Neural Network***
Precision	0.565	0.514	0.473
Recall	0.438	0.406	0.531
F-measure	0.466	0.405	0.498
Execution Time (in seconds)	0.01	0.02	0.09

Table 1. Comparative study of different classification methods with K means clustering

This comparative study of classification methods shows that execution time required for Naïve Bayesian classification is very less than the other two classification methods. Also the accuracy of the Naïve Bayesian classification method is better than decision tree classification and Neural network classification for the given leaf image database.

As the K means is used for clustering in learning phase, it will be useful to identify the leaf using classification method more accurately in leaf identification phase.

VI. Conclusion

This paper discusses comparative study of different classification methods used along with K means algorithm for clustering the leaf images for leaf identification.

This work uses a hybrid approach of clustering and classification algorithms. In this paper, K-means clustering has been used to cluster the leaf image dataset in the training phase. In this paper, the comparative analysis of different classification techniques concludes that Naïve Bayes classification has more accuracy than other two classification techniques for the given image dataset.

References

[1] N. Valliammal and Dr. S.N Geethalakshmi, "*Analysis of the Classification Techniques for Plant Identification through Leaf Recognition*" Ciit International Journal of Data Mining Knowledge Engineering, Vol 1, No. 5,August 2009, pp. 239-243.

[2] Vishakha Metre,Jayshree Ghorpade, "*An Overview of the Research on Texture Based Plant Leaf Classification*", IJCSN International Journal of Computer Science and Network, Vol 2, Issue 3, 2013 ISSN (Online): 2277-5420

[3] Mohd Shamrie Sainin, Taqiyah Khadijah Ghazali and Rayner Alfred, "*Malaysian Medicinal Plant Leaf Shape Identification and Classification*", Knowledge Management International Conference (KMICe) 2014, 12 – 15 August 2014, Malaysia.

[4] C.Ananthi, Azha.Periasamy, S.Muruganand, "*Pattern Recognition Of Medicinal Leaves Using Image Processing Techniques*", Vol 2, Issue 2, Spring Edition, DOI : February 2014, Pp 214-218, ISSN 2279 – 0381

[5] Bhupendra Fataniya, Prachi Manishkumar Patel, Tanish Zaveri, Sanjeev Acharya, "*Microscopic Image Analysis Method for*

Identification of Indian Herbal Plants", 978-1-4799-6052-1/14/$31.00 ©2014 IEEE

[6] Aamod Chemburkar, Anand Sartape, Ajinkya Gawade, Prasad Somawanshi, Jayshree Ghorpade, "*Automated Tool for Plant Leaf Classification Using Morphological Features*", International Journal Of Engineering And Computer Science ISSN:2319-7242, Volume 3 Issue 11 November, 2014 Page No. 9098-9102

[7] Sandeep Kumar and Viswanath Talasila "Leaf Features based approach for Automated Identification of Medicinal Plants", International Conference on Communication and Signal Processing, April 3-5, 2014, India

[8] Jeselle Petrize M. Sosa, Geoffrey A. Solano, Jasper A. Obico "*An Application Using Canny Edge Detection and Multilayer Perceptron for Recognizing Leaves of Tropical Plants*", Mathematics and Computers in Biology and Biomedical Informatics, ISBN: 978-960-474-333-9

[9] Liwen Gao, Xiaohua Lin, "A Study on the Automatic Recognition System of Medicinal Plants", 978-1-4577-1415-3/12/$26.00 ©2012 IEEE.

[10] Shilpa Ankalaki, Laxmidevi Noolvi, Dr. Jharna Majumdar, "leaf identification based on fuzzy c means and naïve bayesian classification", ISSN 0976 - 6480 (Print), ISSN 0976 - 6499 (Online), Volume 5, Issue 7, July (2014).