2nd National Level Students' Research Conference on
"Innovative Ideas and Invention with Sustainability in Computer Science and IT-2021"
In association with International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Leaf Classification Techniques for Medicinal Plants : A Detailed Survey

Gayatri Ambulkar, Prajakta Soman
School of Computer Science, MIT World Peace University, Pune, Maharashtra, India

ABSTRACT

Leaf classification is an important use case and has been explored by main researchers. The main features focused for leaf classification in previous work were shape and texture. In this paper we have discussed the comparative study of different technologies used for feature extraction and selection. We observed that less amount of work is done on Medicinal leaves data. As India is a rich country for being the habitat for a variety of medicinal plants it will be helpful if we can detect such leaves for better use which will be our part of future work after comparing and finding best method and model use for leaf classification.

Keywords – Feature extraction, canny edge detector, Gabor wavelet, PCA, PNN, leaf classification

I. INTRODUCTION

to Herbal medicinal plants are known to treat myriad of diseases and illness. India is home to many Ayurvedic herbs and plants. These plants are not just grown in off-forests but may be found all over in India. Leaves of plants like Tulsi, Ashwagandha, Peppermint and many more are known to be home remedies since ages. Current Covid situation has also taught us to boost our immunity by including medicinal plants in our daily dose. It is important for us to be able to recognize a medicinal plant in our vicinity so that we can use it for our health benefits. Recognition of medicinal leaf can also help prevent adulteration and better quality control while preparing Ayurvedic medicines. Leaf classification, thereby, becomes an important application of using AI in medicine. In this paper, we have done an extensive survey of AI techniques used for classification of leaf.

II. RELATED WORK

In this section, we present our survey done on work related to Leaf classification. There are different approaches, considering different features and their analysis on leaves sample. Collection of these leaves data in some work is done manually or using dataset available on internet.

In [1], the leaves data for Artificial Neural Network model was collected with help of digital cameras or scanners. Data collected was very huge with 1800 leaves of 32 types of plant species. The feature extraction of leaves was done without human intervention using PCA (Principal Component Analysis) and 12 features were extracted. Basic physical features were length, width, diameter, leaf area, leaf perimeter, vein feature and some digital feature. Images were convert from RGB to binary form. The author proposed the classification model

using PNN which is derived from RBF (Radial basis function) and ANN (Artificial Neural Network). PNN is user for classification and pattern recognition. PNN is easy and fast for training the model and gives the result with an accuracy greater than 90% [1].

There are many plant species with unique shape, texture and margin. The development of classification model using shape and texture analysis is done by Thibaut Beghin, James In this paper [5], the main task of feature extraction of leaves sample was done by canny edge detector [16]. This edge detector extract 17 features and 15 features were used out of it. Some of them were Perimeter, Area, Orientation, compactness, complexity, filled area, convex area, length, width and etc. Data use for classification model was flavia dataset [14] which consist of 1800 scans images of leaves that belong to 32 species from which 22 species were used for building classification model. After image processing and feature extraction with the help of canny edge detector author applied SVM (Support Vector Machine) algorithm classifier on flavia data And for Test data they have used around 10 samples for each species. Result of the overall accuracy of the system is 85% to 87% considering worst and best cases of the model [5, 8]. The development of classification model using shape and texture analysis is done by Thibaut Beghin, James S. Cope, Paolo Remagnino, and Sarah Barman in 2010 [2]. The data collected for this process was manually in the Royal Botanic Gardens, Kew, UK which contains 3 to 10 leaves from each of 18 different species. Here, Shape analysis is done with the help of contour signature of the leaves [2]. Contour signature is use to showing ridge and valley using contour outline which can help in recognizing shape of leaves. Distance calculated for this process is Jeffrey distance. Shape based analysis can classify more efficiently by first classifying lobed and unlobed leaves. Similarly, for recognizing the macro texture of leaves was done with the help of sober operator. Sober operator is use for computer vision within edge detection algorithm and creates emphasize the image edges. In this paper [2], classification using sober operator was 62.9% and the result of contour-based method was quite good up to 75%. The best result was given by considering both shape and texture analysis with overall rate of 81.5%.

In paper [3], feature extraction of leaf is done using gabor wavelets filter. Gabor wavelet filter is use for Shape and texture analysis. With shape and texture, colour and contour analysis is also taken into consideration for classifying plant leaf. LDA (Linear Discriminant Analysis) technique is use for feature reduction and pattern classification. Leaves data was trained on K-nearest neighbour algorithm and cross validation was performed for more accurate results. The collection of leaves sample were collected with the help of vivo camera. Correctness rate of the model was approx. 84%.

In paper [4], the author has worked on 1600 images of 100 different species with 16 images of each species and iris plant dataset was used to build the model. Features extracted of leaves were leave margin, height, width, texture and shape. Three features types were separately analyse: shape, texture, and margin which then combined using a probabilistic framework. The texture and margin features use histogram accumulation, while a centroid contour distance curve (CCDC) is used for the shape [4]. Previously published methods are used to generate separate posterior probability vectors for each feature, using data associated with the k-Nearest Neighbour classifier. This model is also trained on KNN classifier and also done cross validation for better accuracy. KNN classifier was trained on the 1600 images of 100 species which increases the accuracy and tested the model on fisher's iris dataset with 3 tree species. The PROP and WPROP density estimation methods were tested and both density estimators achieved a 96% mean accuracy of classification.

In this paper [7] by author T. Le, D. Tran and N. Pham, Leaf identification is done by kernel based approach. Kernel based approach was performed in two phases: First phase include the pre-processing of image and path by patch feature extraction of leaves. In Second phase training of images is done. Kernel descriptor is use to reduce the effect of size in terms of feature dependence without altering the effective feature and work on the area of leaf which is actually important for leaf identification. The work done by author was on two dataset: First was Flavia dataset consisting images of 32 plant species [14] and second dataset was Imageclef 2013 [15] which contains different organs (leaf, flower, stem) of 250 plant species in France. Result using KDES (Kernel descriptor) and SVM on Flavia dataset was having average accuracy of 97.5 % and on Imageclef 2013 dataset was less average accuracy with 58 % [7]. It concludes that the method use in this paper works well on dataset having no background and only leaf image and not if background and other leaf organs are present in the image.

In yet another paper by author E. Elhariri, N. El-Bendary and A. E. Hassanien [8], Task was divided into 3 phases: Pre-Processing, Feature Extraction and Classification Phase. Color, shape, texture, vein physical features were extracted and analyze. This paper uses the approach of Random forest and LDA (Linear Discriminant Analysis) classification algorithm and also perform the comparative analysis between the both algorithms. Random forest is a classification and regression algorithm which is give more accuracy on large dataset. The dataset consists of 340 images with 30 plant species which was downloaded from UCI-Machine learning repository and trained the data on LDA and RF algorithm with 10-folded cross validation. The result of work showed that LDA achieved classification accuracy of 92.65 % against the RF that achieved accuracy of 88.82 % with combination of shape, first order texture, Gray Level Co-occurrence Matrix (GLCM), HSV color moments, and vein features [8].

In paper [13], authors have divided the work in three phases: Pre-processing, Feature extraction and classification. In pre-processing, Images were converted into binary formats, morphological operation were performed and image were resize. They performed Feature extraction with the help of HOG (Histogram of Oriented Gradients) and SURF (Speeded up Robust Features) [13]. HOG is a descriptor which mainly focuses on the structure or the shape of an object whereas SURF focuses on features related to texture and vein. Strongest features were normalized and combined for building classification model. Image data was then trained and tested using KNN algorithm. Precision rate of the Performed work was 0.96 [13].

We have also tabulated the comparative study of all the above works in Table I

III. FEATURE EXTRACTION TECHNIQUES

1. Shape, edge and margin Analysis:

The most effective way to analyse and extract all these 3 features is contour approach. Contour is a tool which joins all the coordinates including boundary and perform shape analysis and edge detection. Shape analysis consist features as size, perimeter, circularity and compactness. Contour tool is also use for object detection and recognition.

2. Texture and colour analysis :

Texture is the spatial and visual quality of an image based on the co-occurrence matrix.

Macro-Texture analysis and feature extraction can be done using multiple approach as GLCM (Gray Level Co-occurrence Matrix) which is used to estimate the joint probability distribution for the grayscale values in an image and SURF (Speeded up Robust Features) is a descriptor that provide a unique and robust description of an image feature, e.g. describing the

intensity distribution of the pixels within the neighborhood of the point of interest.

3. Vein analysis :

Vein analysis is done by emphasizing the edges within the margin which mainly done using sobel operator, CNN (Convolutional Neural Network) and many other approaches.

IV. PROPOSED DIAGRAM

The proposed system architecture is given in figure I. In our system, the process will be divided into 4 tasks as image annotation, image pre-processing, feature extraction and Training and testing the model. The dataset use will be of medicinal leaves which consist of 30 plant species containing around 1500 images [17]. Then we will be labelling each leaf in the dataset using labelling or any other online annotation tool. Images will be then resized, converted to gray scale, enhance image and then perform feature extraction using different method as Canny Edge detector or by performing PCA and applying neural network to select important features regarding texture, shape, size, margin, etc. After the feature extraction process neural network can applied on data to train, as neural network works efficiently on large dataset and increases the performance of model. Performance of the model will be then check by testing the model on test data.

V. CONCLUSION AND FUTURE WORK

In this survey we address the problem of insufficient Image collection of each classes to train the data which results in incorrect classification of plant species [2, 3, 4, and 11]. Work done on building classification model using Machine learning algorithms is quite traditional approach as it reduces the performance of test data and less efficient to train huge data. Major features for leaf recognition which are considered in many papers are shape, texture, vein and margin. Large number of leaves data for each plant species should be used for training to increase the performance of classification model. The effective classification technique use for leaf detection and classification can do using ANN (Artificial Neural Network) as it works well on huge amount of data. Neural Network also automated the feature extraction, effective feature selection from image, classification and segmentation task. The best method from all the paper was proposed by author Wu [1], In which Neural network was use for training huge data where effective feature selection was done using Principal Component Analysis with the highest average accuracy.

From an experimental point of view, there are many medicinal plants which are useful in many ways such as drug formulation, production of herbal products, and medicines to cure many common ailments and diseases. Our contribution will be on mmedicinal leaves data with 30 different medical plant species each consist of 50 images which will help to identify useful medical plant species [17]. We will mainly focus on the features as: Shape, Texture, Margin, and Vein. The data then can be trained using neural network technique for effective result

VI. REFERENCES

[1]. Wu, Stephen & Bao, Forrest & Xu, Eric & Wang, Yu-Xuan & Chang, Yi-Fan & Xiang, Qiao-Liang. (2007). A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. Centre, Kingston University, London, UK {t.beghin,j.cope,p.remagnino,s.barman}@kingston.ac.uk.

[2]. Thibaut Beghin, James S Cope, Paolo Remagnino, and Sarah Barman Digital Imaging Research Centre, Kingston University, London,

UK
{t.beghin,j.cope,p.remagnino,s.barman}@kingston.ac.uk

[3]. Casanova, D., de Mesquita Sá Junior, J. J., & Bruno, O. M. (2009). Plant leaf identification using Gabor wavelets. International Journal of Imaging Systems and Technology, 19(3), 236–243. doi:10.1002/ima.20201

[4]. Mallah, Charles & Cope, James & Orwell, James. (2013). Plant Leaf Classification using Probabilistic Integration of Shape, Texture and Margin Features. Pattern Recognit. Appl. 3842. 10.2316/P.2013.798-098.

[5]. A. Salman, A. Semwal, U. Bhatt and V. M. Thakkar, "Leaf classification and identification using Canny Edge Detector and SVM classifier," 2017 International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICISC.2017.8068597.

[6]. C. A. Priya, T. Balasaravanan and A. S. Thanamani, "An efficient leaf recognition algorithm for plant classification using support vector machine," International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), Salem, India, 2012, pp. 428-432, doi: 10.1109/ICPRIME.2012.6208384.

[7]. T. Le, D. Tran and N. Pham, "Kernel descriptor based plant leaf identification," 2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, France, 2014, pp. 1-5, doi: 10.1109/IPTA.2014.7001990.

[8]. E. Elhariri, N. El-Bendary and A. E. Hassanien, "Plant classification system based on leaf features," 2014 9th International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt, 2014, pp. 271-276, doi: 10.1109/ICCES.2014.7030971.

[9]. Leafsnap: A Computer Vision System for Automatic Plant Species Identification Neeraj Kumar1, Peter N. Belhumeur2, Arijit Biswas3, David W. Jacobs3, W. John Kress4, Ida Lopez4, and Jo˜ao V. B. Soares3 1 University of Washington, Seattle WA 2 Columbia University, New York NY 3 University of Maryland, College Park MD 4 National Museum of Natural History, Smithsonian Institution, Washington DC

[10]. Ayurvedic leaf recognition for Plant Classification Sethulekshmi A V#1, Sreekumar K#2 # Department Of Computer Science, College Of Engineering Poonjar, Cochin University of Science and Technology Poonjar, Kottayam District, Kerala, India.

[11]. Z. Zulkifli, P. Saad and I. A. Mohtar, "Plant leaf identification using moment invariants & General Regression Neural Network," 2011 11th International Conference on Hybrid Intelligent Systems (HIS), Melacca, Malaysia, 2011, pp. 430-435, doi: 10.1109/HIS.2011.6122144.

[12]. Hall, David, McCool, Christopher, Dayoub, Feras, Suenderhauf, Niko, & Upcroft, Ben (2015) Evaluation of features for leaf classification in challenging conditions. In Das, S, Sarkar, S, Parvin, B, & Porikli, F (Eds.) Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision (WACV 2015). Institute of Electrical and Electronics Engineers (IEEE), United States of America, pp. 797-804.

[13]. A. Sabu, K. Sreekumar and R. R. Nair, "Recognition of ayurvedic medicinal plants from leaves: A computer vision approach," 2017 Fourth International Conference on Image Information Processing (ICIIP), Shimla, India, 2017, pp. 1-5, doi: 10.1109/ICIIP.2017.8313782.

[14]. Flavia Dataset : http://flavia.sourceforge.net/

[15]. ImageCLEF 2013 https://www.imageclef.org/2013

[16]. Canny edge detector in python- Computer Vision https://towardsdatascience.com/canny-edge-detection-step-by-step-in-python-computer-vision-b49c3a2d8123

[17]. Medicinal Leaf Dataset 1 https://data.mendeley.com/datasets/nnytj2v3n5/

TABLE I: Comparative Study

Paper	Dataset	No. of Images	No. of Classes	Technique for feature extraction	Features	Classification Approach	Accuracy
[1]	Manually collection using digital camera	1800	32	PCA(Principal component analysis)	Vein, leaf area, size, Leaf perimeter	PNN(Probabilistic Neural Network) In MATLAB	>90%
[2]	Manually collection in Royal Botanic Gardens,UK	110	18	Contour Signature(Shape and margin analysis), Sober operator(macro texture analysis)	Shape, margin, texture, colour	Increment Classification Approach	81.5%
[3]	Manually collection of data	-	20	Gabor Wavelet	Shape and Texture	LDA (Linear Discriminant Analysis) and K-Nearest Neighbour Algorithm	84%
[4]	Manual collection and fisher's iris dataset	1600	100	Contour(shape analysis)	Shape, Texture, Margin, length, width	K-Nearest Neighbour Classifier	96%
[5]	Flavia dataset	1800	32 (worked on 22)	Canny edge detector	Margin, Shape, size, Texture, Area	SVM (Support Vector Machine)	85% to 87%
[8]	UCI – Machine Learning Repository	380	30	Statistical method	Colour, Shape, Area, Texture, Margin, vein	LDA (Linear Discriminant Analysis) and Random forest	LDA : 92.65% Random Forest : 88.82

[13]	Ayurvedic leaf datasets (Western ghat India)	200	20	HOG (Histogram of Oriented Gradients) and SURF (Speeded Up Robust Features)	Shape, Texture, Vein	K-Nearest Neighbour	Average precision rate : 0.96

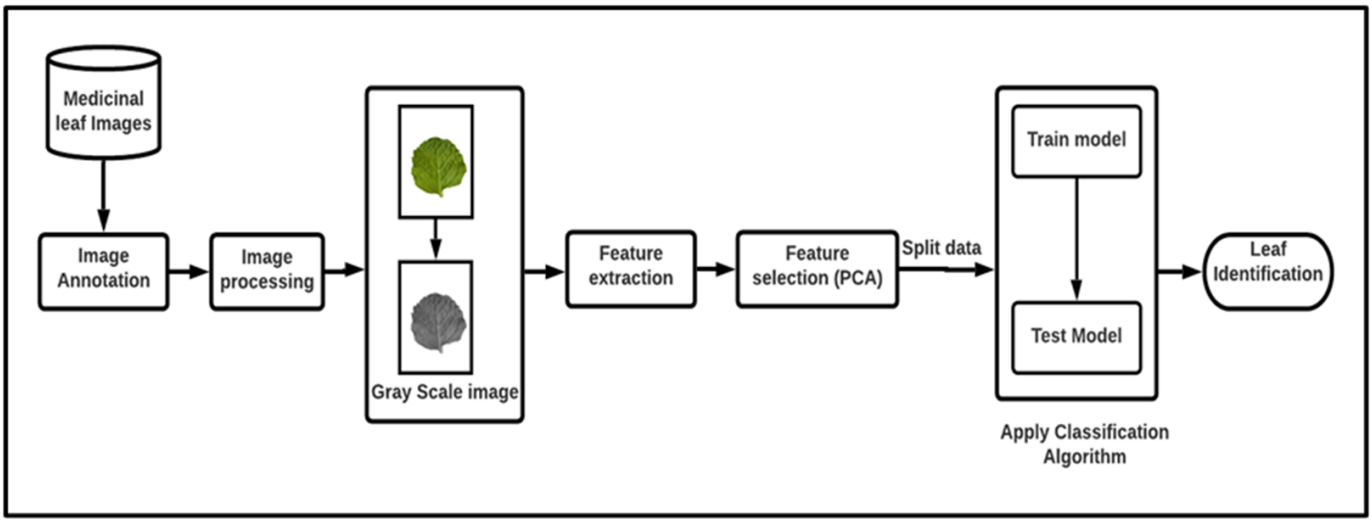

Fig I: Proposed Architecture for Leaf Classification