

# Machine Learning for Plant Species Classification using Leaf Vein Morphometric

Malarvizhi K<sup>1</sup>

Department of Information Technology  
Coimbatore Institute of Technology  
Coimbatore, India.

Sowmithra M<sup>2</sup>

Department of Information Technology  
Coimbatore Institute of Technology  
Coimbatore, India.

Gokula Priya D<sup>3</sup>

Department of Information Technology  
Coimbatore Institute of Technology  
Coimbatore, India.

Kabila B<sup>4</sup>

Department of Information Technology  
Coimbatore Institute of Technology  
Coimbatore, India.

**Abstract** - The number of plant species are extremely huge, with about 391,000 vascular plant species all over the world. Hence, it is impossible and not practical for a botanist or an expert, to be able to identify and classify all the species. In addition, some plant species may have high similarity between each other, taking a long time to differentiate them. Hence, there is a need to develop an automated or computerized system to identify and classify plants. With the advancement of science and technology, machine learning has been widely employed for classification and recognition tasks in many domains especially in the biological fields. Automated plant species identification system could help botanists and lay man in identifying plant species rapidly. Machine learning is robust for feature extraction as it is superior in providing deeper information of images. Plant species identification is a research field area, which includes pre-processing, segmentation, feature extraction and classification. In this field, there exists many algorithm to analyse the image. The paper aims at understanding pre-processing, extracting the leaf features using Contours and classifying using various machine-learning techniques, namely, Support Vector Machine (SVM), k-Nearest-Neighbour (k-NN), and Random Forest (RF). It presents the leaf of various plant species from which the vein characteristics extracted, presented to detect and classify various kinds of plant species and other artificial intelligent techniques used to perform pattern recognition. The features extracted using contours found to fit well with the RF classifier which is little greater than SVM and k-NN classifier respectively.

**Keywords:** Leaf vein; segmentation; feature extraction; classifier algorithms.

## I. INTRODUCTION

Plant species classification is one of the interesting fields in which machine learning techniques are applied, to differentiate between the species. Automation in plant species classification is done using extracting features of the plants. Leaf shape is the most commonly used feature, that is used to develop such automated plant classification systems. Other than shape, leaf features like colour, texture and veins can also provide additional information that may help in automation process. With the advancement of science and technology, machine learning has been widely employed for classification and recognition tasks in many domains especially in the biological fields. Machine learning

techniques, such as, the Support Vector Machines, k-Nearest Neighbour, and others are artificial intelligent techniques are employed in pattern recognition. In this work, Contours is applied to extract the features from images of selected tree species. The dataset used for this project is the Flavia dataset. The data pre-processing technique used here is data transformation. The numerical dataset has obtained as a transformation of image dataset. The extracted features that are in the csv file fed into few classification approaches for learning and training purposes. Classifiers that had employed in this work are, Support Vector Machine (SVM), Random Forest (RF) and k-Nearest Neighbour (k-NN). A conventional method Sobel edge detection was replaced with contours as the latter is memory efficient by removing redundant points.

## II. RELATED WORK

The number of plant species are extremely huge, with about 391,000 vascular plant species all over the world [1]. Currently, machine learning, a subfield of artificial intelligence (AI), is a popular and widely used technique, that has been applied in various domains including biology, medical, computer vision, speech recognition and others [2-5]. Deep learning is a modern AI approach, which contributes a robust framework towards supervised learning [6]. It is able to map an input vector rapidly and efficiently to an output vector even in a large dataset [7]. Image enhancement is a process that is used to emphasise the features of an image [8]. Texture is one of the important features of the plant identification system, which can be used to characterize the leaves based on the surface structure of the leaves. It is a non-consistent spatial distribution pattern of different image intensities [9,10], which concentrates mainly on each single pixel of an image.

Cope et al. [11] introduced an evolved vein classifier based on genetic algorithms (GA) and Ant Colony algorithms to extract the vein structure. Anami et al. [12] had proposed a combination of colour and texture features based plant identification system. Larese et al. [13] constructed an automated leaf identification approach for legumes based on the vein architecture only Simple measurements were

applied on the vein morphology and then identified by a Random Forests approach. Kadir et al. [14] proposed another method on the Foliage dataset and the Flavia dataset. Lee et al. [15, 17] proposed a CNN technique to identify 44 plant species acquired from the Royal Botanic Gardens of Kew, England. The extracted features were then classified with a Multilayer Perceptron (MLP) and a SVM. Two different datasets were used, namely, whole image (D1) and leaf patches (D2). Both datasets achieved an accuracy of more than 97%. Furthermore, researchers had combined both local and global features together in [18] and achieved more than 91% accuracy. Furthermore, Sladojevic et al. [19] employed CNN for plant diseases recognition. Another study based on leaf vein morphological patterns and using deep learning technique for plant identification was proposed in Grinblat et al. [20].

### III. PROPOSED WORK

The classification of plant species is considered vitally important in preserving biosphere and helps botanists to identify huge plant varieties without more human effort. The data pre-processing technique known as, Data transformation is used to transform the plant image data set into numerical data set with necessary features as attributes using contours method. Further, Classification is done using three machine learning algorithms namely, Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbour (k-NN) and the accuracy of these algorithms are compared and the algorithm that gives best accuracy is used for further testing of real time images. Fig.1 shows the work flow of the proposed work

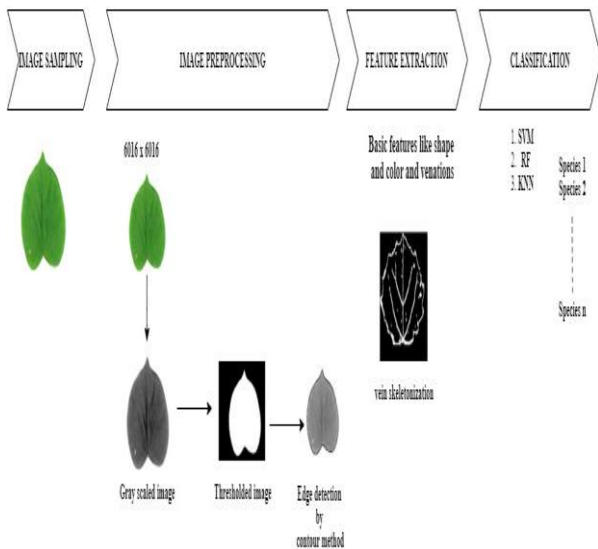


Fig. 1: Work flow diagram of proposed Plant species classification using leaf veins.

### IV. METHODOLOGY

The proposed work has four main steps as show in Fig. 2 which are collecting dataset, image pre-processing, feature extraction and classification. First, the leaf samples were collected and images were acquired from Flavia dataset. The leaf images were then pre-processed and fed into the feature

extraction step to retrieve the important information from the leaves using contours approach. Lastly, the extracted features were trained and classified by using various machine learning methods.

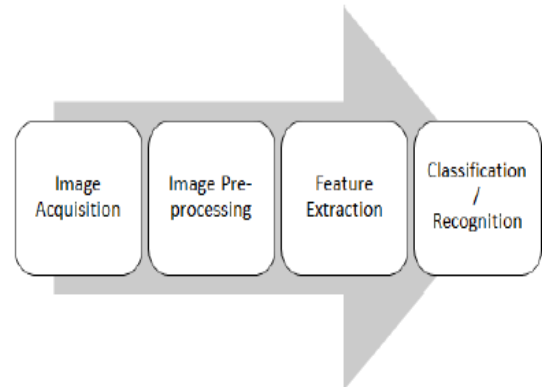


Fig. 2: General Methodology for Automated Plant classification

#### A. Collecting Data-set:

The data set used for this work is Flavia dataset which contains about 1907 images of 32 different species. Fig 3. Shows some images from the Flavia dataset.



Fig. 3: images of 32 different species present in Flavia dataset. Here we show one image per species.

#### B. Data pre-processing:

Data pre-processing is essential in machine learning as data we use may be subjected to noises, which may affect or suppress the quality of the data. Noises occur as pixel values, which do not represent the true intensities of an image during the image acquisition. It is a necessary step to remove the image noises in order to highlight or enhance the important features of an image. The leaf images were reconstructed into square dimension (m x m), which was required as the inputs of data pre-processing. The original images with 6016 x 4016 resolution were resized into 1600 x 1200 resolution in order to maintain the ratio of the leaf shape. The images were then retained in the RGB format. This background-removed image is used for further processing. Fig 4 shows the pre-processed image and segmented images. Data is transformed from image format

(.jpg) to numerical data that is well organized and improves data quality and protects applications from potential landmines such as null values, unexpected duplicates, incorrect indexing and incompatible format.

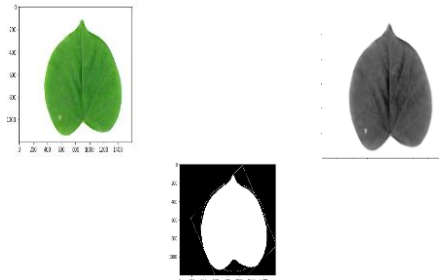


Fig. 4: Pre-processed image formats

**C. Feature Extraction:**

Features play an important role in differentiating the different hand signs. However, the selection of feature requires the proper understanding and interpretation of extracted feature values. First, all images were converted from RGB images into grey-scale images. Then, Contour was used to segment out the region of interest (ROI) from the images. After segmentation, the images were then post-processed and skeletonized to ensure a clean image of the leaf obtained. The vein features were extracted from the segmented images by measuring the vein morphological features. The leaf area was computed for the shape, color and vein texture, end points, entropy, aspect ratio, areoles, inverse-difference-moments, rectangularity and so on.

**D. Classification:**

Classification, as the last step for an automated plant recognition system, is an intelligent algorithm in training data to recognize the specific features of each individual plant species and categorizing a new sample as the correct species. The favoured machine learning methods for plant identification are Artificial Neural Network (ANN), Support Vector Machines (SVM) and k-Nearest Neighbour (kNN). Three classification methods that was used in this work were SVM, kNN, and Random Forest (RF).

Support Vector Machines (SVM), a supervised machine learning approach, is conceded as one of the powerful classification methods due to its high capability in dealing with high dimensional space and data points that are not linearly separated. Applying linear SVM on feature-mapped data can execute speedy with low storage and improve the classification performance. Linear SVM with ‘One versus all’ scheme was employed in this research since it involves a multi-classes dataset.

A k-NN is a classification approach, which classifies a sample according to the majority vote of its neighbours. The number of neighbours in this research is fixed as one with the city block distance metric.

Random Forest is an adaptable, simple to utilize machine learning calculation that produces, even without hyper-parameter tuning, an awesome outcome more often than not. It is additionally a standout amongst the most utilized calculations, since its effortlessness and the way that

it can be utilized for both grouping and relapse errands. In this context, it can be inferred the appropriateness and usability of random forest calculation and a few other vital things about it. The flow of execution is shown in Fig.5.

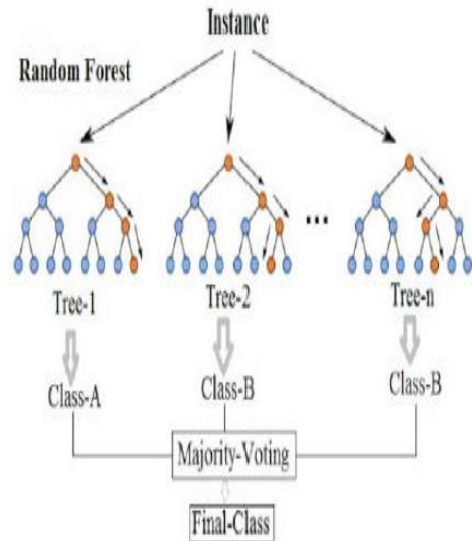


Fig.5: Random Forest Classification

**V. RESULTS AND DISCUSSION**

From the context of this work here we have discussed the performance of the different models, after the data pre-processing, for performance evaluation the data was partitioned into training and testing set in the ratio of 70:30 and fitted to three different classifiers and then we extended our work with testing with some real time images of the plant species. The performance of each classifier is discussed in TABLE I.

s.no	classifiers	Accuracy
1	Random Forest	90
2	Support Vector Machine	88
3	k-Nearest Neighbours	84

TABLE I: Classification accuracy of different classifiers.

It is observed that Random forest classifier achieved the best performance of 90% for accuracy. However, the other classifiers obtained accuracy of not less than 85%. Hence, in this research we can claim that RF is highly suitable with contours for feature extraction model. In contrast, SVM and k-NN were less suitable with the contours. On the other hand, conventional feature extraction method extracts each type of feature separately and manually consuming a lot of time. For example, if the shape features are considered in this work, different sets of processes will be required for segmentation followed by shape feature extraction. Thus, machine learning approach is more practical and appropriate than conventional methods for developing an automated plant species classification system.

## VI. CONCLUSION

We proposed a shape based approach for leaf vein morphometric with several steps including sampling, image pre-processing, edge detection, feature extraction, classification, comparison etc. The strength of this approach includes its simplicity, time management, accuracy, ease of implementation, and it does not required any significant amount of training or post processing, it provide us with the higher recognition rate with minimum computation time. The weakness of this method is that we define certain parameters and threshold values experimentally since it does not follow any systematic approach for vein recognition, and maximum parameters taken in this approach are based on assumption made after testing number of images.

## REFERENCES

- [1] Lottes, Philipp, et al. "UAV-based crop and weed classification for smart farming." *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018.
- [2] Ghazi, Mostafa Mehdipour, Berrin Yanikoglu, and Erchan Aptoula. "Plant identification using deep neural networks via optimization of transfer learning parameters." *Neurocomputing* 235 (2018): 228-235.
- [3] Azlah, Muhammad Azfar Firdaus, et al. "Review on techniques for plant leaf classification and recognition." *Computers* 8.4 (2019): 77.
- [4] Zhang, Shanwen, Harry Wang, and Wenzhun Huang. "Two-stage plant species recognition by local mean clustering and Weighted sparse representation classification." *Cluster computing* 20.2 (2017): 1517-1525.
- [5] Elsalamony, Hany A. "Bank direct marketing analysis of data mining techniques." *International Journal of Computer Applications* 85.7 (2014): 12-22.
- [6] Kolivand, Hoshang, et al. "A new leaf venation detection technique for plant species classification." *Arabian Journal for Science and Engineering* 44.4 (2019): 3315-3327.
- [7] Wäldchen, Jana, and Patrick Mäder. "Plant species identification using computer vision techniques: A systematic literature review." *Archives of Computational Methods in Engineering* 25.2 (2018): 507-543.
- [8] Masemola, Cecilia, Moses Azong Cho, and Abel Ramoelo. "Assessing the effect of seasonality on leaf and canopy spectra for the discrimination of an alien tree species, *Acacia meurnsii*, from Co-occurring native species using parametric and nonparametric classifiers." *IEEE Transactions on Geoscience and Remote Sensing* 57.8 (2019): 5853-5867.
- [9] Ferentinos, Konstantinos P. "Deep learning models for plant disease detection and diagnosis." *Computers and Electronics in Agriculture* 145 (2018): 311-318.
- [10] Kumar, Munish, et al. "Plant species recognition using morphological features and adaptive boosting methodology." *IEEE Access* 7 (2019): 163912-163918.
- [11] Chen, Junde, et al. "Using deep transfer learning for image-based plant disease identification." *Computers and Electronics in Agriculture* 173 (2020): 105393.
- [12] Lee, Sue Han, Chee Seng Chan, and Paolo Remagnino. "Multi-organ plant classification based on convolutional and recurrent neural networks." *IEEE Transactions on Image Processing* 27.9 (2018): 4287-4301.
- [13] Santhosh, S., et al. "Classification of Leaf Images for Species Identification." *2019 International Conference on Wireless Communications Signal Processing and Networking (Wisp NET)*. IEEE, 2019.
- [14] Goyal, Neha, and Nitin Kumar. "Plant species identification using leaf image retrieval: a study." *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2018.
- [15] Fan, Jianping, et al. "Hierarchical learning of tree classifiers for large-scale plant species identification." *IEEE Transactions on Image Processing* 24.11 (2018): 4172-4184.
- [16] Sahay, Aparajita, and Min Chen. "Leaf analysis for plant recognition." *2017 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2017.
- [17] Singh, Vijai, and Ak K. Misra. "Detection of plant leaf diseases using image segmentation and soft computing techniques." *Information processing in Agriculture* 4.1 (2017): 41-49.
- [18] Purohit, Suchit, et al. "Automatic plant species recognition technique using machine learning approaches." *2015 International Conference on Computing and Network Communications (CoCoNet)*. IEEE, 2015.
- [19] Lee, Sue Han, et al. "How deep learning extracts and learns leaf features for plant classification." *Pattern Recognition* 71 (2018): 1-13.
- [20] Saleem, G., et al. "Automated analysis of visual leaf shape features for plant classification." *Computers and Electronics in Agriculture* 157 (2019): 270-280.