

Identification of Plant Species using Deep Learning

Gargi Chandrababu

Department of CSE
College Of Engineering Kidangoor

Ojus Thomas Lee

Department of CSE
College of Engineering Kidangoor

Rekha K S

Department of CSE
College of Engineering Kidangoor

Abstract— In addition to maintaining the earth's ecosystem plants provide us with oxygen, food, medicine, and fuel. The accurate identification of plant species is a very challenging task because plant species identification requires specialized knowledge and in-depth training related to botany. Even for botanists themselves species identification is often a difficult task. Therefore, there is an urgent need to develop an automatic plant leaf recognition system. Many researches focuses on plant leaf based identification, since it's easier to access as compared to other parts of a plant. This paper provides a survey on the methods and classifications used to identify the various plants in the recent years. Moreover, this survey includes a comparative study of those methods according to the accuracy achieved by the classifiers. This review will be helpful to beginners in research field to understand and analyse the methods as a guideline.

Keywords— *Image processing; Machine Learning; Classifiers; Deep learning;*

I INTRODUCTION

Plants are really important for the planet and for all living things. Plants provide breathable climate, fuel, medicines, etc. There is a huge flora around the world. Therefore, it is not possible or practical for a botanist or expert to identify all the plant species. Some species are similar, so it takes a long time to distinguish between them. In addition, many plants face the problem of extinction. Therefore, an automatic plant identification system needs to be developed. Such a system is useful not only for general use, but also helpful for experienced botanists and plant ecologists. Leaf identification is an important part of identifying plant species. The leaf forms of the plants are different and each plant has relatively stable characteristics. These characteristics are the starting point for recognizing plants. Many researchers have made efforts to identify plants using the leaves of the plant. They chose it because the leaves are a two-dimensional planar system that is easy to process and accessible compared to other parts of the plant. Image-based methods are considered a good approach to plant identification. The plant recognition process is mainly divided into the following steps which is illustrated in Fig. 1.

- **Image Acquisition:** The purpose of this step is to obtain the image of plant species to perform preprocessing and classification operations.
- **Preprocessing:** The preprocessing process receives an image as input and generates a modified image as output, suitable for feature extraction step. Preprocessing typically includes operations like noise removal, image enhancement, and

segmentation. It improves the detectability of the plant species.

- **Feature Extraction:** Feature extraction is a key step in plant species recognition. It refers to extracting the image information to a manageable groups. Feature extraction can be divided into global and local features. The global features refers to the representation of features which describes the overall information of the image, such as shape feature, texture feature and color features. Local features refers to the partial feature of the image within a specific region.
- **Classification:** In the classification step, all extracted features are concatenated into a feature vector, which is then being classified. The state-of-the-art methods uses classification algorithms, such as k-Nearest Neighbour (k-NN), Decision Tree (DT), Support Vector Machine (SVM) and Artificial Neural Networks (ANN).



Fig. 1. Generic steps of an image-based plant classification process

II LITERATURE SURVEY

Recently, several state-of-the-art methods uses deep learning based methods for plant species identification. Deep learning based methods uses deep learning algorithms like convolutional neural network (CNN) to extract features and classify leaf images.

B. K. Varghese, A. Augustine, J. M. Babu, D. Sunny, and E. S. Cherian [5] proposed an android application called INFOPLANT to identify plants using the CNN (Convolutional Neural Network). Transfer learned MobileNet model is used in this system. The model is trained with customized dataset which is created from video of plant species. The dataset is then converted and stored as .tflite file. The application predicts the input plant image with the tflite model. After prediction the application will check all the labels and find the label with maximum probability which gives the plant name as the output. The obtained output is then connected to firebase. The output will be the details like biological name, common name, location, nutrient requirement needed by the plant and the medicinal value of the plant. The proposed model achieved a prediction result with accuracy of 99% and validation accuracy of

95%. The flowchart shown in Fig. 2 represents the working flow of the proposed method.

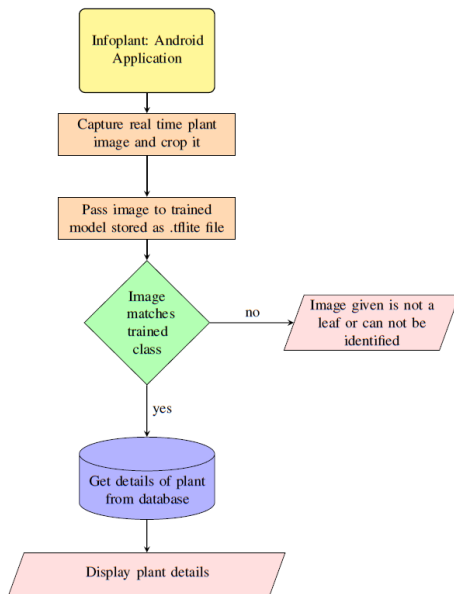


Fig. 2. Flow Diagram - INFOPLANT: Plant Recognition using Convolutional Neural Networks

S. A. Riaz, S. Naz, and I. Razzak [4] presented a plant identification method using multipath multi deep convolutional network. The proposed architecture illustrated in Fig. 3 which consists of multiple CNN blocks, max-pooling layers, flatten layer and soft-max layer for the classification of input plant images. Each block consists of three convolution, one batch normalization, one max-pooling layer and one dense layer. The features extracted from one block is concatenated with the features from second block. The soft-max layer classifies the plant species. This study uses two datasets LeafSnap and MalayaKew and achieved accuracy of 99.38% and 99.22% respectively.

N. Manasa, P. Herur, P. Shetty, N. Prarthana, and P. Venkatrao [3] proposed a plant identification method using watershed algorithm and convolutional neural network. The input image undergoes various pre-processing stages when the leaf is surrounded by multiple leaves watershed algorithm is used to separate each leaf. The proposed method consists of two phases training and testing phase as shown in Fig. 4. In training phase the images are pre-processed, extracted features from the pre-processed image and classified. In the testing phase the images not used for training is processed and tested by feeding them into neural network. In this model a pre-trained convolutional neural network is used to solve classification problem. Stochastic Gradient Descent algorithm with momentum is used to train the network and is successfully trained with a validation

accuracy of 100%. Confusion matrix is used to evaluate the performance of the network.

Different species of plants have different sizes. Moreover, even the same flora has different sizes due to different growth conditions. Based on this J. Hu, Z. Chen, M. Yang, R. Zhang, and Y. Cui [2] proposed a multi-scale fusion convolutional neural network (MSF-CNN) for plant leaf recognition at multiple scales. In the first step the input image is down-sampled into multiple low resolution images and then these images with different scales are fed into the MSF-CNN architecture to learn the features in different depths. At this stage the feature fusion between two different scales is realized by a concatenation operation, which concatenates feature maps learned on different scale images from a channel view. The last layer of MSF-CNN aggregates all different information to obtain the final feature for predicting the plant species of input image. The proposed method is evaluated on two datasets, MalayaKew and LeafSnap. The architecture of the proposed system is illustrated in Fig. 5.

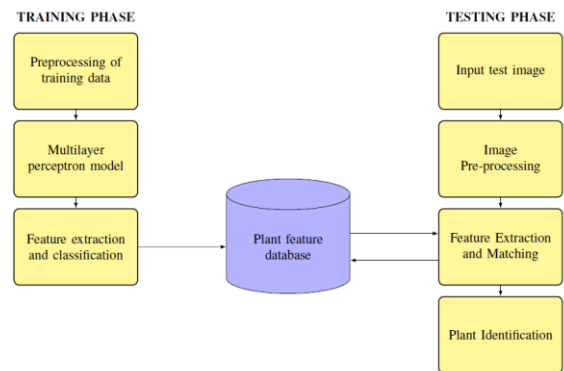


Fig. 4. Block Diagram- Plant Recognition using Watershed and Convolutional Neural Network

wei Tan, J, S.W. Chang, S. Abdul-Kareem, H. J. Yap, and K.T. Yong [6] proposed a CNN based method named D-Leaf. The plant leaf images are pre-processed and extracted features by using three CNN models namely pre-trained AlexNet, fine-tuned AlexNet, and D-Leaf. Five machine learning classifiers are used for classification such as SVM, ANN, k-NN, NB and CNN. In the preprocessing step the acquired images are converted to Tagged Image File Format (TIFF) and the noises are removed using Adobe Photoshop. Image is reconstructed into square dimension and then is resized into 250x250 resolution. Sobel edge detection method is used segment out the region of interest (ROI). For this all images are converted from RGB to grayscale, after segmentation the images were post-processed and skeletonized to obtain clean vein architecture.

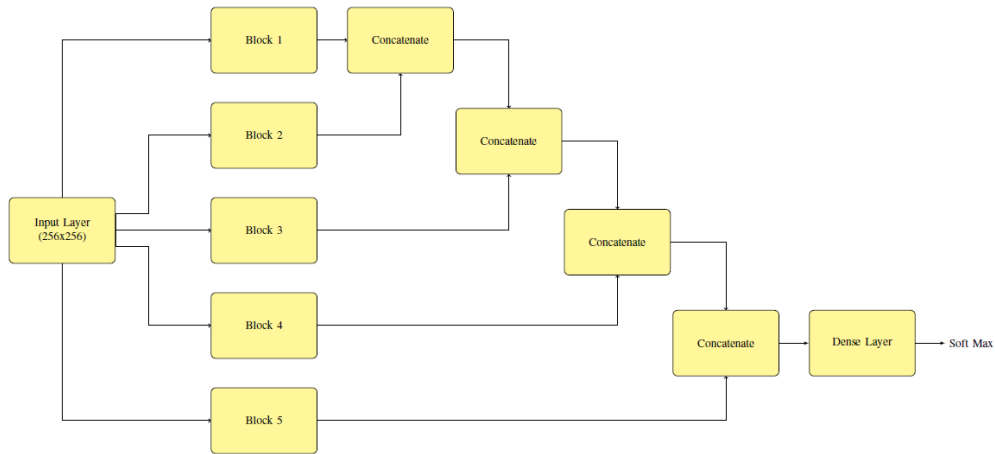


Fig. 3. Proposed Framework of Multi-path Multi-CNN Fusion Network for Plant Species Identification

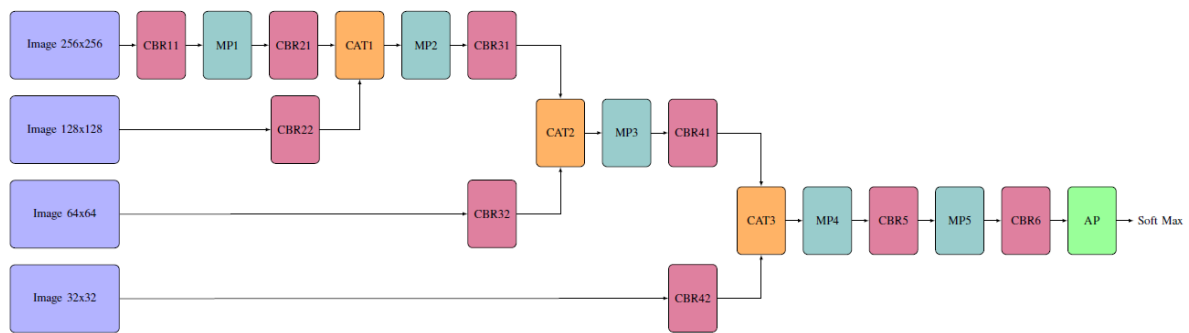


Fig. 5. Block diagram of Multi-Scale Fusion Convolutional Neural Network for Plant Leaf Recognition

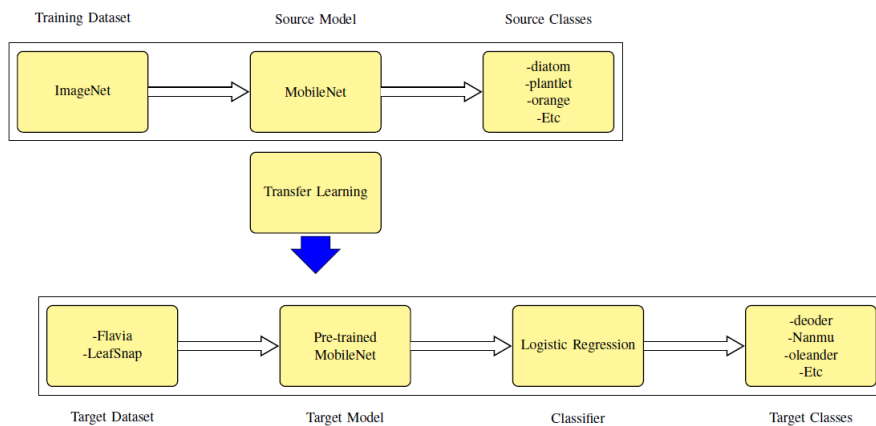


Fig. 7. Block diagram - Leaf Classification for Plant Recognition with Deep Transfer Learning

After pre-processing step feature extraction is performed using CNN models as mentioned earlier and also feature is extracted using a conventional morphological method which extracted vein features from the segmented image by measuring the vein morphological features. The evaluation shows CNN method is better than the conventional morphological method. The proposed D-Leaf for feature extraction and ANN as classifier gives 94.88% testing accuracy. Fig. 6 shows the block diagram of proposed method.

Beikmohammadi, and Faez. [1] proposed a method that uses transfer learning to identify plant leaf for classification. This method uses the MobileNet model as a feature extractor as shown in Fig. 7. With the help of image label and the feature

vector logistic regression classifier was trained to predict the target class. The presented method works directly with RGB images which eliminate the use of pre-processing stage and hand-crafted feature extraction. The proposed method is evaluated on flavia dataset and LeafSnap dataset and achieved accuracy of 99.6% and 90.54% respectively.

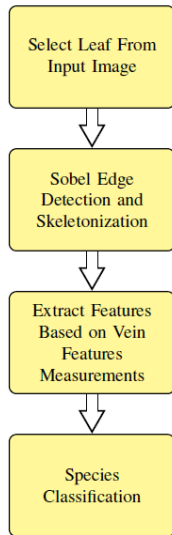


Fig. 6. Block diagram - Deep Learning for Plant Species Classification

III COMPARATIVE STUDY

TABLE I illustrate the comparison of various state-of-the-art methods based on dataset used, metrics and classifier.

TABLE 1. COMPARISON OF VARIOUS LEAF IDENTIFICATION SYSTEM BASED ON DEEP LEARNING

| Authors | Dataset | Metrics | Classifier |
|-----------------------------|---|-------------------------|-------------------------|
| Varghese, B.K et al. (2020) | Own datasets | Accuracy | MobileNet |
| Riaz, S.A et al. (2020) | Leafsnap, MalayaKew | Classification accuracy | MPF-CNN |
| Manasa, N et al. (2019) | Downloaded from Kaggle, GitHub, PlantVillage | Confusion matrix | AlexNet |
| Hu J et al. (2018) | MalayaKew (MK) Leaf Dataset and LeafSnap Plant Leaf Dataset | CMC curve, accuracy | MSF-CNN |
| wei Tan, J et al. (2018) | MalayaKew, Flavia and Swedish Leaf Dataset | Classification accuracy | SVM, ANN, k-NN, NB, CNN |
| Beikmohammadi et al. (2018) | Flavia, LeafSnap | Accuracy, CMC | Logistic Regression |

TABLE 2. DATA SETS AVAILABLE FOR PLANT IDENTIFICATION

| Datasets | Features | Location | Data Source |
|-----------------------------|---|-------------------------------------|---|
| Flavia | Flavia dataset contains 1907 leaf images from 32 different species. The number of leaf images for each species is between 50 and 77. | Yangtze Delta, China | https://sourceforge.net/projects/flavia/ |
| MalayaKew (MK) Leaf Dataset | It consists of scan-like images of leaves from 44 species classes. This dataset is very challenging as leaves from different species classes have very similar appearance. | Royal Botanic Gardens, Kew, England | http://cs-chan.com/downloads_MKLeaf_dataset.html |
| Swedish Leaf Dataset | Swedish leaf dataset consists of 15 tree classes with 75 samples per species and a total of 1,125 leaf images. | Swedan | http://www.cvl.isy.liu.se/en/research/datasets/swedish-leaf |
| LeafSnap | LeafSnap Plant Leaf Dataset consists of 7719 field images taken by mobile devices and 23147 lab images captured using a high-quality camera. This dataset currently covers 184 tree species from the north-eastern United States. | Northeastern United States | https://www.kaggle.com/xhlulu/leafsnap-dataset |

REFERENCES

[1] A. Beikmohammadi and K. Faez. Leaf classification for plant recognition with deep transfer learning. In 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pages 21–26. IEEE, 2018.

[2] J. Hu, Z. Chen, M. Yang, R. Zhang, and Y. Cui. A multiscale fusion convolutional neural network for plant leaf recognition. IEEE Signal Processing Letters, 25(6):853–857, 2018.

Most of the study uses publicly available datasets. For evaluating the performance of the system classification accuracy metrics is used.

IV DATASETS AND FEATURES

Data sets are used to train and test the model. Collecting and preparing the dataset is one of the most crucial task while creating a machine learning or AI project. There are different types of data sets publicly available for plant identification. Flavia, Swedish leaf dataset, LeafSnap and MalayaKew are some dataset available publicly the details are listed in TABLE II.

V CONCLUSION

Identification of plant species is a crucial task due to the wide varieties of plant species hence there is a need of automatic plant identification system. This paper focuses on various state-of-the-art methods for plant species identification based on deep learning. Based on the review it is found that size of database should be large to allow for better training. Which will increase the accuracy of developed system.

[3] N. Manasa, P. Herur, P. Shetty, N. Prarthana, and P. Venkatrao. Plant recognition using watershed and convolutional neural network. In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), pages 969–973. IEEE, 2019.

[4] S. A. Riaz, S. Naz, and I. Razzak. Multipath deep shallow convolutional networks for large scale plant species identification in wild image. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2020.

- [5] B. K. Varghese, A. Augustine, J. M. Babu, D. Sunny, and E. S. Cherian. Infoplant: Plant recognition using convolutional neural networks. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), pages 800–807. IEEE, 2020.
- [6] J. wei Tan, S.W. Chang, S. Abdul-Kareem, H. J. Yap, and K.T. Yong. Deep learning for plant species classification using leaf vein morphometric. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(1):82–90, 2018.