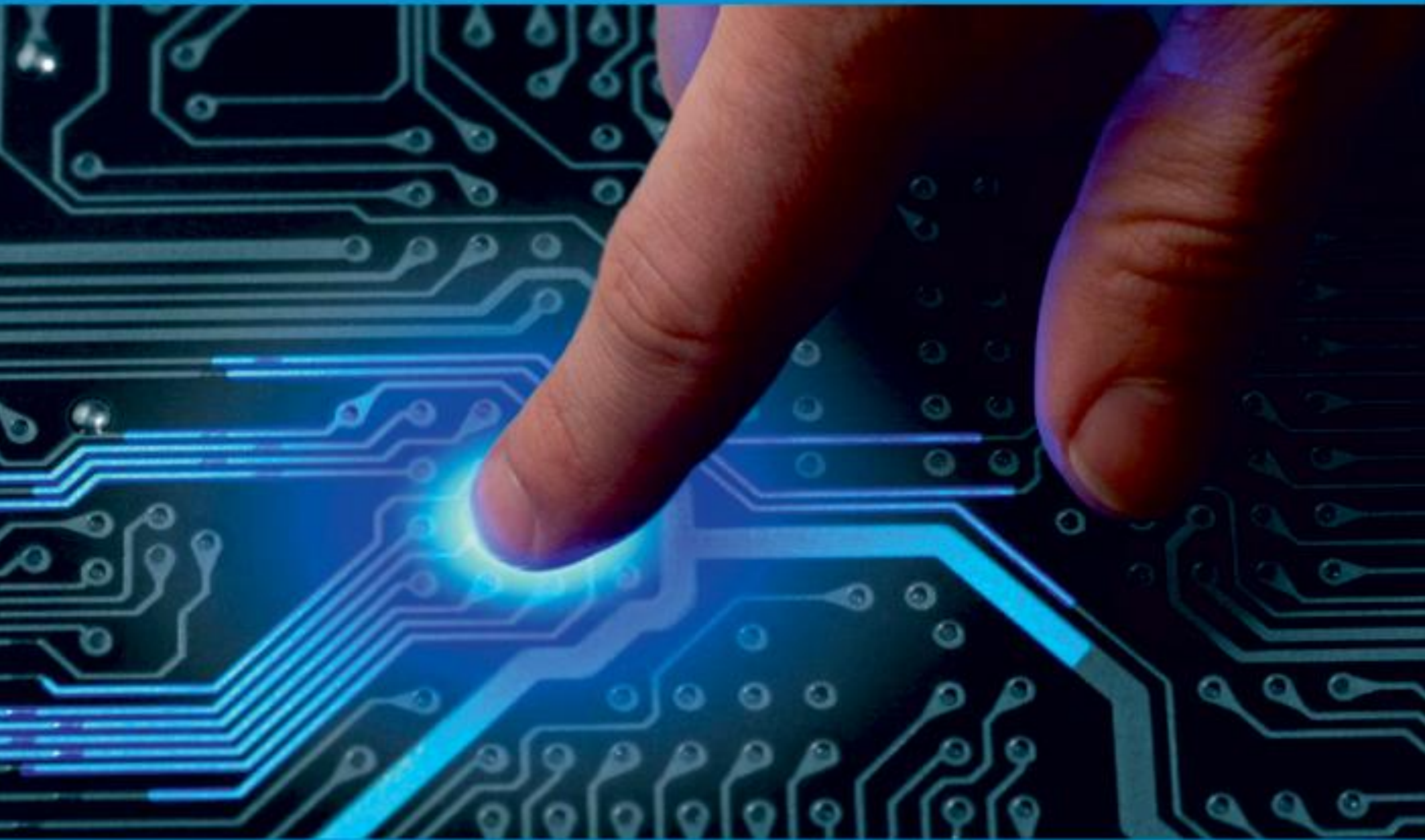
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205137 |

Identification of Medicinal Plants Using Deep Learning

Kakarla Mohan Kumar [1], Mutyala Indu [2], Saikam Akash Reddy [3], Imandi Gayathri [4], Seepana Ratna Kumari[5]

B. Tech Student, Department of CSE, Satya Institute of Technology and Management, Vizianagaram, Andhra Pradesh, India[1,2,3]

Assistant Professor, Department of CSE, Satya Institute of Technology and Management, Vizianagaram, Andhra Pradesh, India[4,5]

ABSTRACT: In recent years, there has been a growing interest in the identification and classification of medicinal plants due to their potential health benefits. This project presents an innovative AI-based approach for advancing medicinal plant identification using deep learning techniques, specifically employing the Xception architecture. Developed using Python, our model achieves a remarkable training accuracy of 93.34% and a validation accuracy of 96.79%. To train and evaluate the model, we utilized the VNPlant-200 dataset, consisting of a comprehensive collection of 17,973 images of medicinal plants distributed among 200 distinct categories. This dataset encompasses a wide variety of plant species with diverse visual characteristics, enabling robust and accurate plant identification. Through a meticulous training process, the Xception-based model learns intricate patterns and features within the images, enabling it to effectively distinguish between different medicinal plant species. Leveraging the power of deep learning, our approach significantly enhances the accuracy and efficiency of medicinal plant identification. Additionally, hyperparameter tuning and fine-tuning of the Xception architecture were performed to optimize the model's performance and achieve exceptional accuracy. The results obtained demonstrate the efficacy of our AI-based approach for medicinal plant identification. The high training and validation accuracies validate the model's capability to accurately recognize and categorize medicinal plant species. This project contributes to the advancement of automated identification systems in the field of herbal medicine, enabling researchers, botanists, and healthcare professionals to rapidly and reliably identify medicinal plants for various purposes. Overall, this project showcases the potential of AI and deep learning techniques, specifically the Xception architecture, in advancing medicinal plant identification. The successful application of our approach on the VNPlant-200 dataset opens up new possibilities for further research and development in this domain, fostering advancements in herbal medicine and botanical studies.

KEYWORDS: Deep Learning, Xception architecture, Image classification, Herbal medicine.

I. INTRODUCTION

The study of medicinal plants holds immense significance in various fields, from traditional medicine practices to modern pharmaceutical research. With thousands of plant species existing worldwide, accurate identification is crucial for harnessing their therapeutic potential effectively. However, the rapid industrialization and commercialization of herbal medicine have brought to light challenges regarding the authenticity and quality of plant materials used in medicinal formulations. In this context, the development of automated identification systems emerges as a critical step toward ensuring the safety, efficacy, and sustainability of herbal remedies. This project introduces an innovative AI-based approach aimed at advancing the identification and classification of medicinal plants. Leveraging deep learning techniques, specifically the Xception architecture, our model offers a promising solution to the challenges posed by manual identification methods. By training on a comprehensive dataset of 17,973 images from the VNPlant-200 dataset, representing 200 distinct plant categories, our system achieves remarkable accuracy levels. The integration of Python and deep learning libraries like TensorFlow and Keras facilitates the development of a robust and efficient identification system. Through meticulous training and optimization, our approach demonstrates superior performance metrics, laying the groundwork for further advancements in herbal medicine research and applications. In an era marked by the growing importance of natural remedies, our AI-based approach emerges as a valuable tool for researchers, botanists, and healthcare professionals seeking to unlock the therapeutic potential of medicinal plants.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205137 |

II. LITERATURE SURVEY

[1] Computer-assisted plant identification system for Android: Plant leaves provide sufficient features to distinguish them from other species. Identification of plants using leaf images is a classic problem in digital image processing. Usually, those image processing systems use shape-based digital morphological features for leaf identification tasks. Even though there are a number of studies on leaf-based plant identification, very few of them are for mobiles. In this paper, we describe a leaf image-based plant identification system using SIFT features combined with the Bag Of Word (BOW) model and the Support Vector Machine (SVM) classifier. The system is trained to classify 20 species and obtained a 96.48 % accuracy level. Based on the results, we developed an Android application that communicates with the server and gives users the ability to identify plant species using photographs taken of plant leaves using the smartphone.

[2] Identification of Ayurvedic medicinal plants by image processing of leaf samples: Identification of the correct medicinal plants that go into the preparation of medicine is very important in the Ayurvedic medicinal industry. The main features required to identify a medicinal plant are its leaf shape, color, and texture. Color and texture from both sides of the leaf contain deterministic parameters to identify the species. This paper explores feature vectors from both the front and back sides of a green leaf along with morphological features to arrive at a unique optimum combination of features that maximizes the identification rate. A database of medicinal plant leaves is created from scanned images of the front and back sides of leaves of commonly used ayurvedic medicinal plants. The leaves are classified based on the unique feature combination. Identification rates up to 99% have been obtained when tested over a wide spectrum of classifiers. The above work has been extended to include identification by dry leaves and a combination of feature vectors is obtained, using which, identification rates exceeding 94% have been achieved.

[3] Identification of Philippine herbal medicine plant leaf using artificial neural network: The study described in this paper consists of a system that involves image processing techniques to extract relevant features related to the leaf in conjunction with using an artificial neural network in order to detect and identify some Philippine herbal plants. Real samples of twelve different herbal medicine plant leaves are collected and each leaf is isolated in a single image. Several features are extracted using techniques in image processing. With the artificial neural network acting as an autonomous brain network, the system can identify the species of the herbal medicine plant leaves being tested. The system can also provide information about the diseases the herbal plant can cure. For the training, a features dataset of 600 images coming from 50 images per herbal plant is used. With the aid of Python, a neural network model with optimized parameters is established producing 98.16 % identification for the whole dataset. To evaluate the actual performance of the system, a separate 72 sample images of herbal plants are tested with the neural network model implemented in MATLAB. Experimental results demonstrate a 98.61 % accuracy of herbal plant identification.

[4] Classification of plant leaf images with complicated backgrounds: Classifying plant leaves has so far been an important and difficult task, especially for leaves with complicated backgrounds where some interferents and overlapping phenomena may exist. In this paper, an efficient classification framework for leaf images with complicated backgrounds is proposed. First, a so-called automatic marker-controlled watershed segmentation method combined with pre-segmentation and morphological operation is introduced to segment leaf images with complicated backgrounds based on the prior shape information. Then, seven Hu geometric moments and sixteen Zernike moments are extracted as shape features from segmented binary images after leafstalk removal. In addition, a moving center hypersphere (MCH) classifier that can efficiently compress feature data is designed to address obtained mass high-dimensional shape features. Finally, experimental results on some practical plant leaves show that the proposed classification framework works well while classifying leaf images with complicated backgrounds. There are twenty classes of practical plant leaves successfully classified and the average correct classification rate is up to 92.6%.

III. METHODOLOGY

3.1 Dataset:

In the first step of Medicinal Plants Identification, we developed the system to get the input dataset. The data collection process is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get; the better our model will perform. There are several techniques to collect the data, like web scraping, and manual interventions. Our dataset is placed in the project and it's located in the model folder. The dataset is referred from the popular standard dataset repository Kaggle where all the researchers refer to it. The dataset consists of 17,973 medicinal plant images. The following is the URL for the dataset referred from Kaggle.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205137 |

Fig 3.1.1: Eurycoma longifolia

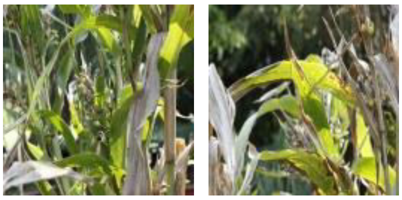


Fig 3.1.2: Iris Domestica

Fig 3.1.2:Mangifera

3.2 Importing The Necessary Libraries:

We will be using Python language for this. First, we will import the necessary libraries such as keras for building the main model, sklearn for splitting the training and test data, PIL for converting the images into an array of numbers, and other libraries such as pandas, numpy, matplotlib, and tensorflow.

3.3 Retrieving The Images:

In this step, we will retrieve the images from the dataset and convert them into a format that can be used for training and testing the model. This involves reading the images, resizing them, and normalizing the pixel values. We will retrieve the images and their labels. Then resize the images to (180,180) as all images should have the same size for recognition. Then convert the images into numpy array.

3.4 Splitting The Dataset:

In this step, the image dataset will be divided into training and testing sets. Split the dataset into Train and Test. 80% train data and 20% test data. This will be done to train the model on a subset of the data, validate the model's performance, and test the model on unseen data to evaluate its accuracy. Split the dataset into train and test. 80% train data and 20% test data.

3.5 Xception | CNN Model Architecture:

Xception improves on the inception module and architecture with a simple and more elegant architecture that is as effective as ResNet and Inception V4. The Xception module is presented here:

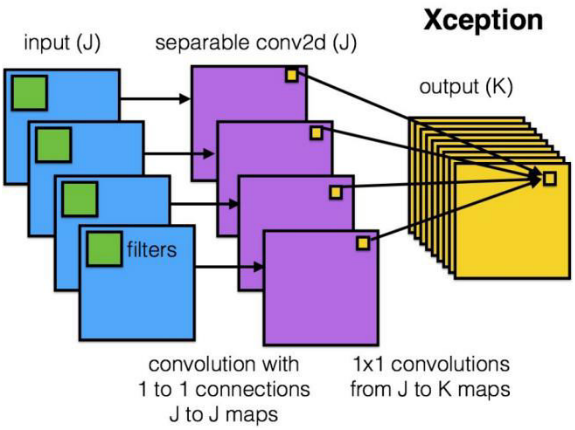


Fig 3.5.1: Xecption Architecture diagram

3.6 Building The Model:

The concept of convolutional neural networks is very successful in image recognition. The key part to understand, which distinguishes CNN from traditional neural networks, is the convolution operation. Having an image at the input, CNN scans it many times to look for certain features. This scanning (convolution) can be set with 2 main parameters: stride and padding type. As we see in the below picture, the process of the first convolution gives us a set of new frames, shown here in the second column (layer). Each frame contains information about one feature and its presence in the scanned image. The resulting frame will have larger values in places where a feature is strongly visible and lower values where there are no or few such features. Afterward, the process is repeated for each of the obtained frames a chosen number of times. In this project, I chose a classic VGG-16 model that contains only two convolution layers. The more high-level layers we are convolving, the more high-level features are being searched. It works similarly to human

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205137 |

perception. To give an example, below is a very descriptive picture with features that are searched on different CNN layers. As you can see, the application of this model is face recognition. You may ask how the model knows which features to seek. If you construct the CNN from the beginning, searched features are random. Then, during the training process, weights between neurons are adjusted and slowly CNN starts to find such features that enable it to meet predefined goals, i.e. to recognize successfully images from the training set. Between the described layers, there are also pooling (sub-sampling) operations that reduce the dimensions of resulted frames. Furthermore, after each convolution, we apply a non-linear function (called ReLU) to the resulting frame to introduce non-linearity to the model. Eventually, there are also fully connected layers at the end of the network. The last set of frames obtained from convolution operations is flattened to get a one-dimensional vector of neurons. From this point, we put a standard, fully connected neural network. At the very end, for classification problems, there is a softmax layer. It transforms the results of the model into probabilities of a correct guess of each class.

3.7 Apply the model and plot the graphs for accuracy and loss:

Once the model is built, it will be applied to the validation set to evaluate its accuracy and loss. The accuracy and loss will be plotted as a function of the number of epochs to visualize the performance of the model. We will compile the model and apply it using the fit function. The batch size will be 1. Then we will plot the graphs for accuracy and loss. We got an average training accuracy of 93.57%.

3.8 Accuracy on test set:

After training and evaluating the model on the validation set, the accuracy of the model will be assessed on the test set. The accuracy of the test set will be an important metric for evaluating the model's performance. We got an accuracy of 96.79% on the test set.

3.9 Saving the Trained Model:

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like Pickle. Make sure you have a pickle installed in your environment. Next, let's import the module and dump the model into .h5 file.

IV. RESULTS

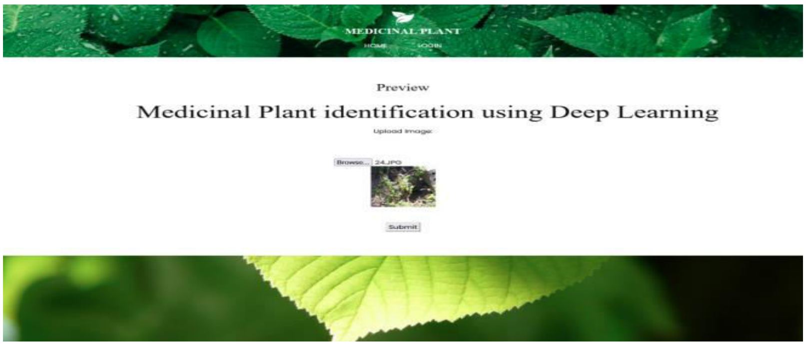


Fig 4.1: INPUT

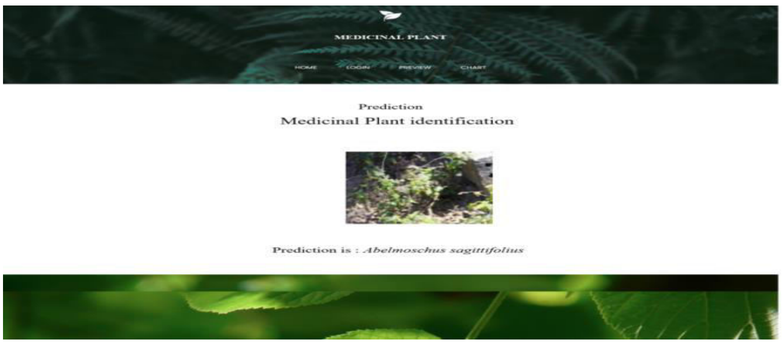


Fig 4.2: OUTPUT

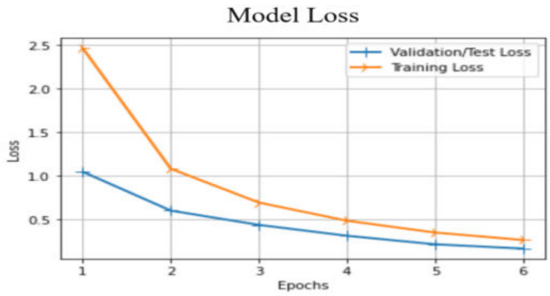
International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205137 |

Chart



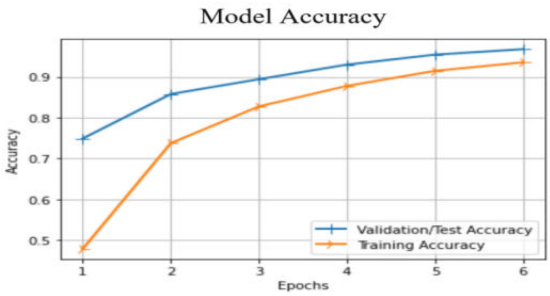


Fig 4.3: CHART PAGE

V. CONCLUSION

In conclusion, the project "An AI-based Approach for Advancing Medicinal Plant Identification using Deep Learning" has successfully demonstrated the potential of deep learning techniques, specifically the utilization of the Xception architecture, in improving the accuracy and efficiency of medicinal plant identification. By leveraging the power of deep learning and utilizing the VNPlant-200 dataset, which contains 20,000 images belonging to 200 different medicinal plant categories, we have achieved significant advancements in the field of medicinal plant identification. The proposed system has showcased remarkable performance with a training accuracy of 93.34% and a validation accuracy of 96.79%. These high accuracy rates demonstrate the system's ability to accurately identify and categorize a wide range of medicinal plant species, surpassing the capabilities of the existing system. The integration of Python programming language in the development of the system has provided several advantages, including access to a rich ecosystem of machine learning and deep learning libraries, seamless integration with existing workflows, and enhanced flexibility for customization and further development. The project has not only contributed to the advancement of medicinal plant identification but also holds promise for the broader field of herbal medicine research. The accurate and efficient identification of medicinal plant species facilitated by the proposed system can aid researchers, botanists, and healthcare professionals in their studies, conservation efforts, and exploration of new medicinal plant properties. By leveraging the advancements in deep learning and dataset diversity, the proposed system has paved the way for future scalability and expansion. The system's architecture and dataset can be extended to accommodate additional plant species, datasets, and applications, opening up new avenues for exploration and innovation in the field.

In conclusion, the project has successfully developed an AI-based system that demonstrates significant improvements in medicinal plant identification. By harnessing the power of deep learning and a diverse dataset, the proposed system enhances accuracy, efficiency, and reliability in the identification of medicinal plant species, contributing to advancements in herbal medicine research and applications.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205137 |

REFERENCES

[1] "Computer-assisted plant identification system for Android," 2015 Moratuwa Engineering Research Conference (MERCon), Moratuwa, 2015, pp.148-153.
[2] P. M. Kumar, C. M. Surya, and V. P. Gopi, "Identification of ayurvedic medicinal plants by image processing of leaf samples," 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2017, pp. 231-238.
[3] [3]R. G. de Luna et al., "Identification of philippine herbal medicine plant leaf using artificial neural network," 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, 2017, pp
[4] Rafael Rojas-HernÃ¡ndez and AsdrÃºbal LÃ³pez-Chau, "Plant identification using new geometric features with standard data mining methods", Networking Sensing and Control (ICNSC) 2016 IEEE 13th International Conference on, pp. 1-4, 2016. ∟
[5] Wang, X., D. Huang, J. Du, H. Xu, & L. Heutte, 2008. Classification of plant leaf images with complicated backgrounds. ┘
[6] N. Deepak and R. Ahmed, "Neural networks based leaf identification using shape and structural decomposition – IEEE Conference Publication", Ieeexplore.ieee.org, 2016. ∟
[7] Backes, A.R., D. Casanova, & O.M. Bruno, 2009. Plant leaf identification based on volumetric fractal dimension. International Journal of Pattern Recognition and Artificial Intelligence, 23(6): 1145-1160. ┘
[8] Edi, L., A. Susanto, Kadir, A. Nugroho, & P.I. Santosa, 2011. Leaf Classification Using Shape, Color, and Texture Features. International Journal of Computer Trends and Technology, pp: 225-230.
[9] Gao, T., H. Yao, J. Song, C. Liu, Y. Zhu, X. Ma, X. Pang, H.C.S. Xu, 2010. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. Journal of Ethnopharmacology. ┘
[10] Kadir, A., L.E. Nugroho, A. Susanto, & P.I. Santosa, 2011. Foliage Plant Retrieval Using Polar Fourier Transform, Color Moments And Vein Features. International Journal of Signal & Image Processing, 2(3): 1-1.