JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IDENTIFICATION OF MEDICINAL PLANTS WITH IMAGE PROCESSING USING MACHINE LEARNING

[1]Tara V K, [2]Yashashree R, [3]Shoba H N, [4]Sonam Vishwakarma

[1]Assistant Professor, [2]Student, [3]Student, [4]Student
[1]Computer Science Engineering,
[1]City Engineering College, Bangalore, India

Abstract: Creating an automated categorization system for medicinal plants is a time-consuming and challenging task. India is a nation with a varied variety of plant species, each with its unique set of therapeutic qualities. Because it is hard for humans to recall the names of all plant species and their applications, previous knowledge is essential for manual identification and categorization. The preservation of these therapeutic plants is crucial as it will help a broad variety of sectors, including medicine, botanic research, and plant taxonomy studies, among others. Existing technologies cannot imitate the range of therapeutic plant species present in India. The suggested technique facilitates in the classification of medicinal plants by exploiting textural aspects that are crucial in leaf recognition and identification. The three key phases of the proposed technique are picture enhancement, feature extraction, and classification. The photographs of the leaves are shot using cellphones and then processed using digital image processing algorithms to extract the features that may be compared between them. Finally, the CNN classifier is employed to develop an automated classifier.

***Index Terms* – Leaf Identification, Feature extraction, Image processing, CNN.**

I. INTRODUCTION

Ayurveda is an ancient medicinal system that is practiced in India and has its origins in the Vedic times, approximately 5000 years ago. Ayurveda is considered to be the oldest healing science. In the Sanskrit language, Ayurveda means "The Science of Life". In spite of being suppressed during years of foreign occupation, Ayurveda is being widely used in its native land and throughout the world. The main ingredients of ayurvedic medicines are plant leaves, roots, bark, fruits, seeds, etc. It is said that about 8000 plants of Indian origin are known to possess medicinal attributes. The fundamentals on which this Ayurvedic system is built are always true for all ages and can be easily adapted from generation to generation. This, in turn, makes Ayurveda one of the very few medicinal systems that were developed in ancient times and are still implemented to date. Ancient Tibetan and Traditional Chinese Medicine and Early Greek medicine have accepted many concepts that were described in the ancient Ayurvedic literature dating back to several thousands of years. This widespread and extensive use in the field of medicine makes Ayurveda the 'mother of healing'. These Ayurvedic plants are used for preparing medicines on a commercial basis. This has resulted in the production and marketing of Ayurvedic medicines to become a thriving industry with its annual turnover exceeding Rs 4000 crores. Because of this, the number of licensed Ayurvedic medicine manufacturers in India now exceeds 8500. Because of this increased commercialization of the Ayurvedic sector, several issues regarding the raw material quality used for their preparation need to be focused. These plants are usually collected by tribal masses that are not professionally trained in the work of identifying the correct plants. Even the manufacturing units, at times, receive improper or substituted medicinal plants. Most of these manufacturing units do not have proper quality control mechanisms that can screen these plants. In addition to this, confusion due to the different local names of these plants makes the matters worse. Because of the usage of improper raw materials, Ayurvedic medicine becomes inefficient. It may result in certain unpredictable side effects too. As a result, an intelligent system needs to be developed that can properly identify the ayurvedic plant based on the leaf samples. This will, in turn, improve the quality of the medicine and will also maintain its credibility.

II. SYSTEM DESIGN

Design overview explains the architecture that would be used for developing a software product. It is an overview of an entire system, identifying the main components that would be developed for the product and their interfaces.

2.1 System Architecture

A Deep-CNN is type of a DNN consists of multiple hidden layers such as convolutional layer, RELU layer, Pooling layer and fully connected a normalized layer. CNN shares. weights in the convolutional layer reducing the memory footprint and increases the performance of the network. The important features of CNN lie with the 3D volumes of neurons, local connectivity and shared weights. A feature map is produced by convolution layer through the convolution of different sub-regions of the input image with a learned kernel. Then, a non-linear activation function is applied through RELU layer to improve the convergence properties when the error is low. In pooling layer, a region of the image/feature map is chosen and the pixel with the maximum value among them or average values is chosen as the representative.

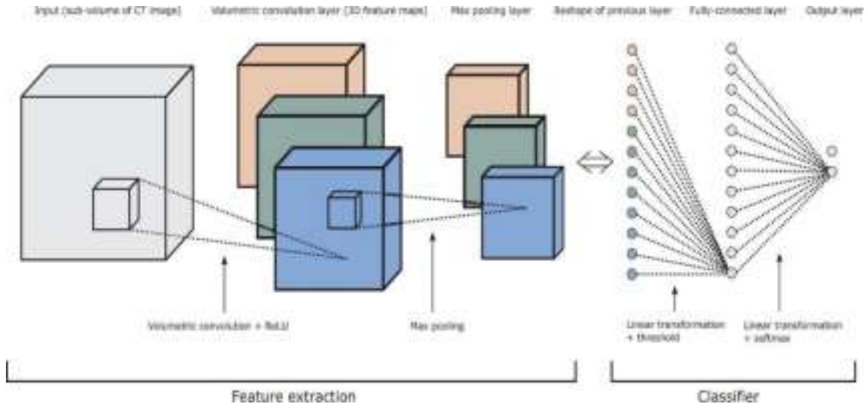

Figure 1: Deep-Convolutional Neural Network Architecture

This results a large reduction in the sample size. Sometimes, traditional Fully- Connected (FC) layer will be used in conjunction with the convolutional layers towards the output stage. In CNN architecture, usually convolution layer and pool layer are used in some combination. The pooling layer usually carries out two types of operations viz. max pooling and means pooling. In mean pooling, the average neighborhood is calculated within the feature points and in max pooling it is calculated within a maximum of feature points. Mean pooling reduces the error caused by the neighborhood size limitation and retains background information. Max pooling reduces the convolution layer parameter estimated error caused by the mean deviation and hence retains more texture information.

3.2 Convolutional Neural Networks

A CNN is type of a DNN consists of multiple hidden layers such as convolutional layer, RELU layer, Pooling layer and fully connected a normalized layer. CNN shares weights in the convolutional layer reducing the memory footprint and increases the performance of the network. The important features of CNN lie with the 3D volumes of neurons, local connectivity and shared weights. A feature map is produced by convolution layer through convolution of different sub regions of the input image with a learned kernel. Then, anon- linear activation function is applied through ReLu layer to improve the convergence properties when the error is low. In pooling layer, a region of the image/feature map is chosen and the pixel with maximum value among them or average values is chosen as the representative pixel so that a 2x2 or 3x3 grid will be reduced to a single scalar value. This results a large reduction in the sample size. Sometimes, traditional Fully-Connected (FC) layer will be used in conjunction with the convolutional layers towards the output stage. A CNN is composed of several kinds of layers:

- Convolutional layer: creates a feature map to predict the class probabilities for each feature by applying a filter that scans the whole image, few pixels at a time.
- Pooling layer (down-sampling): scales down the amount of information the convolutional layer generated for each feature and maintains the most essential information (the process of the convolutional and pooling layers usually repeats several times).
- Fully connected input layer: flattens the outputs generated by previous layers to turn them into a single vector that can be used as an input for the next layer.
- Fully connected layer: Applies weights over the input generated by the feature analysis to predict an accurate label.

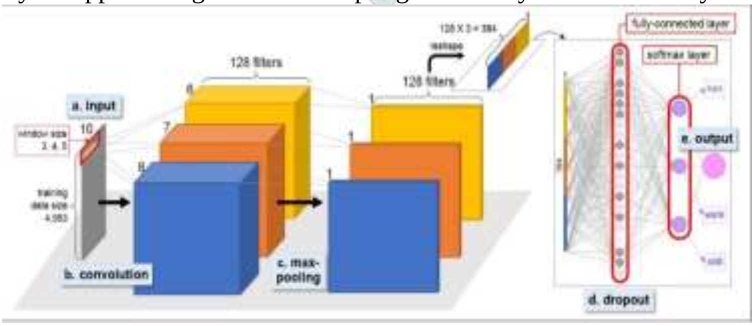

Figure 2: Convolutional Neural Network

3.3 Data Flow Diagrams

A dataflow outline is a tool for referring to knowledge progression from one module to the next module as shown in Fig 4.3 This graph gives the data of each module's info and yield. The map has no power flow and there are no circles at the same time.

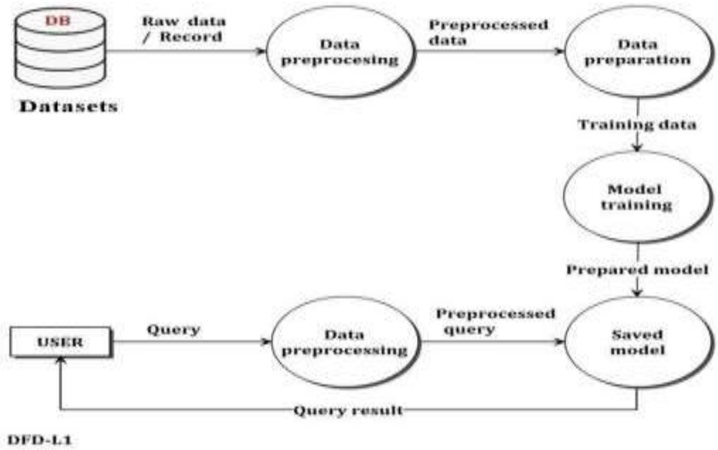

Figure 3: Data Flow Diagram

3.4 Use Case Diagrams

Use case diagram is the boundary, which defines the system of interest in relation to the world around it. The actors, usually individuals involved with the system defined according to their roles. The use cases, which are the specific roles played by the actors within and around the system.

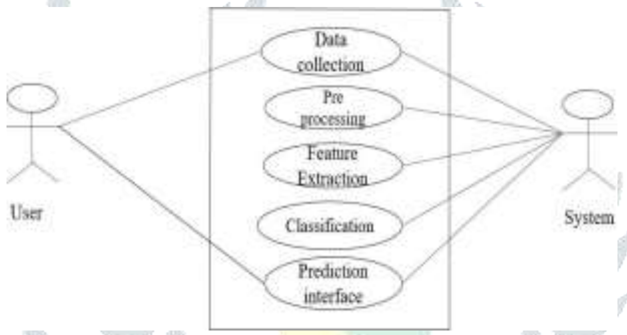

Figure 4: Use Case Diagram

3.5 Class Diagram

Class diagrams are the main building block in object-oriented modelling. They are used to show the different objects in a system, their attributes, their operations and the relationships among them as shown in the Figure.

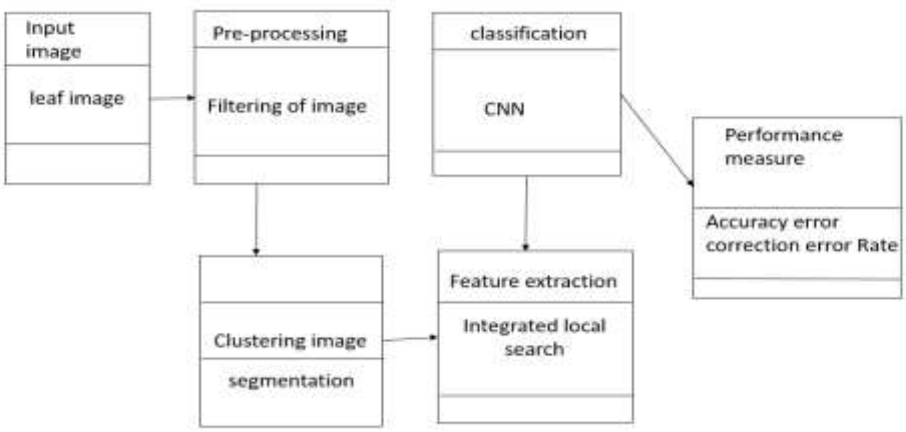

Figure 5: Class Diagram

III. METHODOLOGY

The methodology for identifying medicinal plants using image processing and machine learning involves several steps. First, a dataset of images of various medicinal plants needs to be collected and labeled with corresponding plant species. Next, pre-processing techniques such as image resizing, normalization, and noise reduction may be applied to enhance image quality and reduce variability. Then, a convolutional neural network (CNN) model can be trained on the pre-processed images to learn features and patterns indicative of different plant species. Finally, the trained model can be used to classify and identify medicinal plants in new images with high accuracy. The System design mainly consists of Image Collection, Image Preprocessing, Image Segmentation, Feature Extraction, Training, Classification.

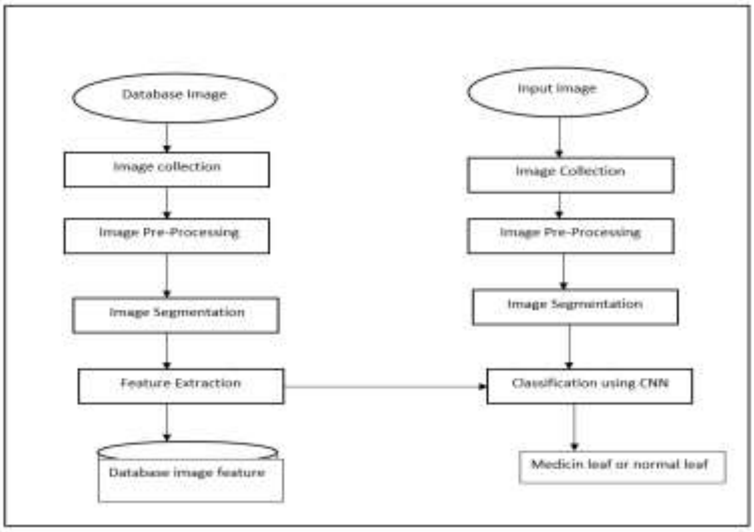

Figure 6: Data flow Diagram of Training and Testing Phase

4.1 CNN Algorithm Explanation

The invention of the CNN in 1994 by Yann LeCun is what propelled the field of Artificial Intelligence and Deep learning to its former glory. The first neural network named LeNet5 had a very less validation accuracy of 42% since then we have come a long way in this field. Nowadays almost all giant technology firms rely on CNN for more efficient performance. The idea to detect diseases in mulberry leaf incorporates the use of CNN before we dive into the "functionality and working of CNN" concept, we must have a basic idea on how the human brain recognizes an object in spite of its varying attributes from one another. Our brain has a complex layer of neurons ,each layer holds some information about the object and all the features of the object are extracted by the neurons and stored in our memory, next time when we see the same object the brain matches the stored features to recognize the object, but one can easily mistake it as a simple "IF-THEN" function, yes it is to some extent but it has an extra feature that gives it an edge over other algorithms that is Self-Learning, although it cannot match a human brain but still it can give it a tough competition . Image is processed using the Basic CNN to detect the diseases in leaves. The data training in our CNN model has to satisfy following constraints: 1) No missing values in dataset. 2) The dataset must distinctly be divided into training and testing sets, either the training or the testing set shouldn't contain any irrelevant data out of our model domain in case of an image dataset all the images must be of the same size, one uneven distribution of image size in our dataset can decrease the efficiency of our neural network. 3) The images should be converted into black and white format before feeding it into the convolution layer because reading images in RGB would involve a 3-D NumPy matrix which will reduce the execution time of our model by a considerable amount. Any kind of corrupted or blurred images should also be trimmed from the database before feeding it into the neural network. This is all about data pre-processing rules, let us dive right into the working of the convolution neural network.

A. Convolution layer

This layer involves scanning the whole image for patterns and formulating it in the form of a 3x3 matrix. This convolved feature matrix of the image is known as Kernel. Each value in the kernel is known as a weight vector.

Figure 7: Convolution layer

B. Pooling Layer

After the convolution comes to the pooling here the image matrix is broken down into the sets of 4 rectangular segments which are non-overlapping. There are two types of pooling, Max pooling and average pooling. Max pooling gives the maximum value in the relative matrix region which is taken. Average pooling gives the average value in relative matrix region. The main advantage of the pooling layer is that it increases computer performance and decreases over- fitting chances.

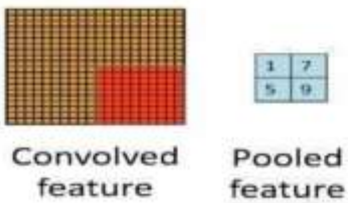

Figure 8: Polling Layer

C. Activation layer

Implementation It the part of the Convolutional Neural Networks where the values are Normalized that is, they are fitted in a certain range. The used convolutional function is RELU which allows only the positive values and then rejects the negative values. It is the function of low computational cost.

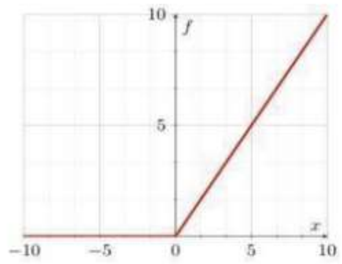

Figure 9: Activation Layer

D. Fully Connected layer

Neural networks are a set of dependent non-linear functions. Each individual function consists of a neuron (or a perceptron). In fully connected layers, the neuron applies a linear transformation to the input vector through a weights matrix. A non-linear transformation is then applied to the product through a non-linear activation function *f*.

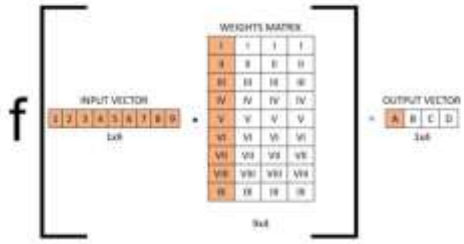

Figure 10: Fully Connected layer

4.2 Procedure

A. Dataset

The Dataset used for this work contains 5 classes of medicinal plant where 20 images in each class. 80% of the images is used as train data and 20% of the images is used as test data. Figure 1 below shows the sample images in the dataset.

Figure 11: Sample images in the dataset

B. Proposed Work

The overview of the architecture is shown in the Figure 2. The input image is in RGB color space. The leaf in the input image is segmented using ExG-ExR method. The texture and color features can be extracted from the segmented leaf. The Weka is used to classify the medicinal plant species.

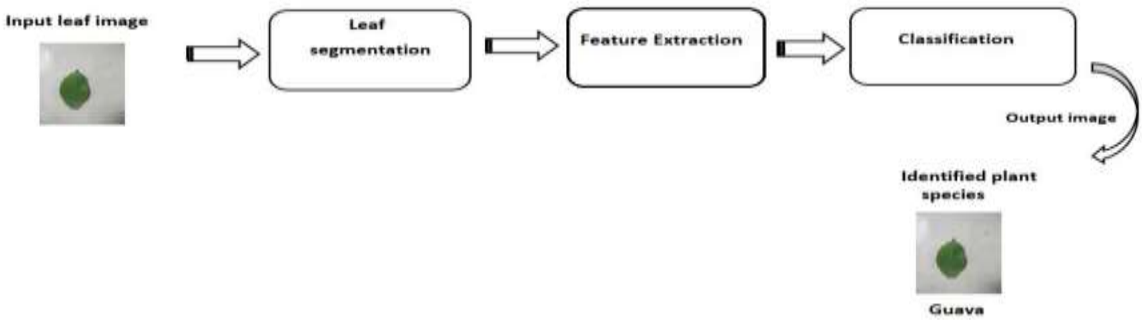

Figure 12: Overview of Architecture for the classification of plant species.

IV. RESULTS AND DISCUSSION

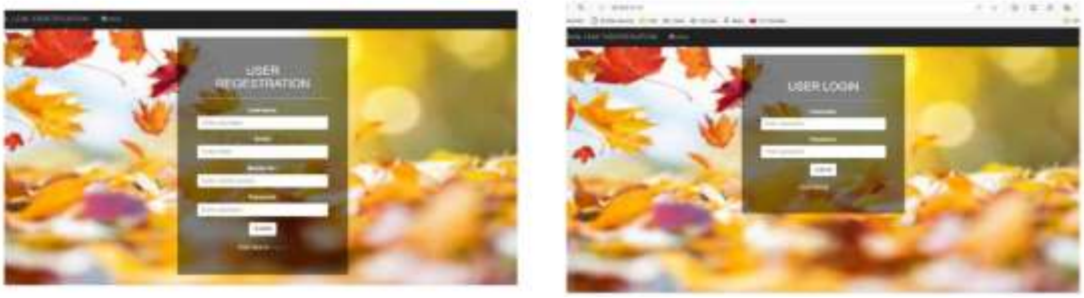

Figure 13: User Registration and login

Figure 14: Input image and Capture image

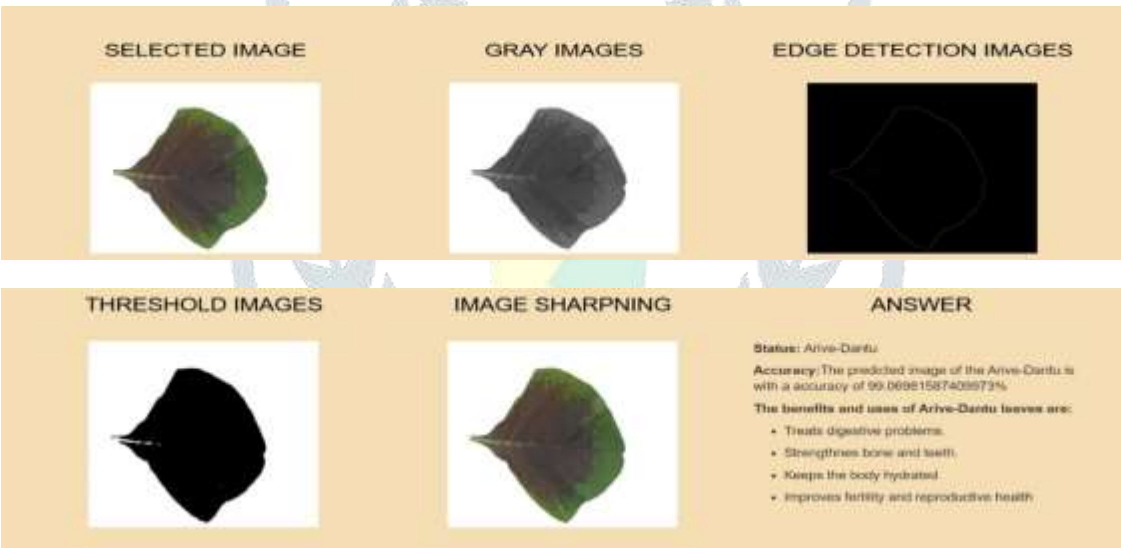

Figure 15: Image detection

V. ACKNOWLEDGMENT

We would like to acknowledge the invaluable support of our research team and collaborators for their expertise and contributions. Special thanks to [list names if applicable] for their guidance and assistance throughout this project. Additionally, we extend our gratitude to the funding agencies [mention if any] for their financial support.

REFERENCES

[1] Manojkumar P., Surya C. M., and Varun 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), pp 978- 1 5386-1931-5.

[2] Mr. K. Nithiyanandhan and Prof. T. Bhaskara Reddy, "Analysis of the Medicinal Leaves by using Image Processing Techniques and ANN", Vol 8, No. 5, ISSN No. 0976-5697, MayJune 2017.

[3] Adams Begue, VenithaKowlessur, FawziMahomoodally, Upasana Singh and Sameerchand, "Automatic Recognition of Medicinal Plants using Machine Learning Techniques", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 4, 2017.

[4] H. X. Kan, L. Jin, and F. L. Zhou," Classification of Medicinal Plant Leaf Image Based on Multi-Feature Extraction", Pattern Recognition and Image Analysis, Vol. 27, No. 3, 2017, pp. 581–587, 1054-6618. © Pleiades Publishing, Ltd.

[5] Riddhi H. Shaparia, Dr. Narendra M. Patel and Prof. Zankhana H. Shah," Flower Classification using Texture and Color Features", International Conference on Research and Innovations in Science.