WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analysis of Medicinal Herbs Using Machine Learning

V. Kokila[1], P. Ganeshamoorthi[2], P. Muthukrishnan[3], K. Saranarayanan[4], N. Shameer Hussian[5]

[1]Assistant Professor, Department of Electronics and Communication Engineering, Krishnasamy College of Engineering and Technology, Cuddalore, Tamilnadu, India.
[2,3,4,5] Students, Department of Electronics and Communication Engineering, Krishnasamy College of Engineering and Technology, Cuddalore, Tamilnadu, India

ABSTRACT—

Automatic plant image recognition is prominent option for closing the botanical taxonomic gap and it has attracted much interest from botany and the computing community. As machine learning technology progresses, more sophisticated models for automatic plant identification have been developed. Medicinal plants are gaining popularity in the pharmaceutical sector because they have fewer side effects and are less expensive than contemporary pharmaceuticals. Many researchers have expressed a strong interest in the study of automatic medicinal plant recognition as a result of these facts. There are several avenues for progress in developing a strong classifiers such as CNNs and active contour that can reliably categorize medicinal plants in real- time. The effectiveness and reliability of various machine learning techniques to plant classification using leaf pictures that have been utilized in recent years are discussed in this research. For some machine learning classifiers, the review includes the image processing methods used to detect leaves and extract significant leaf attributes. The performance of these deep learning classifiers when identifying leaf images based on typical plant properties such as shape, vein, texture and a combination of numerous aspects is categorized and the result provide information on the medicinal characteristics of the plants. This approach overcomes the drawbacks of SVM based classification and effectively increases the prediction accuracy of CNNs.The proposed system has a higher accuracy rate, according to the findings of the experiments.

Keywords— Taxonomy, CNN, Active contour, Machine learning, SVM .

Introduction

The Rich Biodiversity of India, an economically developing country in all fields ensures the maintenance of natural resources by boosting ecosystem productivity. It takes time and effort to develop an automated identification methods for classifying medicinal plants. There are many different kinds of plant species in India, each with its own special set of medicinal properties. The names and awareness of all plant species and their uses are difficult for humans to remember, thus prior information is crucial for manual identification and categorization. It is crucial to preserve these medicinal plants because they will benefit a wide range of fields including medicine, botanic research and plant taxonomy studies. However, increasing awareness of traditional medicinal herbs through automation provides the valuable insights into the use of medicinal herbs for therapeutic measures rather than allopathy treatments. The WHO revealed that eighty percent of people follow traditional medical practices as their primary one. The essential automated assisting algorithm for recognizing and identifying leaves suggested for recognition, feature extraction, classification and Identification of herbs that contains merits and therapeutic procedures involve the use of respective herbs to handle diseases and the geographical abundance of the species.

Literature review

Medicinal herbs identification and recognition methods with help of AI techniques such as Machine learning and Computer vision to make accurate diagnosis. Here discussing the related works as following

[1] demonstrated the deep learning Neural networks using LifeCLEF 2015 plant task datasets, with the help of ALEXNET, GOOGLENET and VGGNET pertained learning. The validation accuracies of 76.87 and 78.44 percent respectively. It has a drawback to take long time for computation.

[2] has used for the fine-grained plant classification problem which evidenced by external variable such as the data of flowering and the position of flowers utilized for the mode and it depends the image based on environmental condition. It has an advantage to detect species from the flower images and suffered with environmental condition of image fed as input.

[3] has investigated the models characteristics based on qualitative and quantitative way to determine height, colour and leaf shapes. Then it requires high level of expertise for classification for further studies deal with large volume of plants.

[4] used for utilizing the classification algorithms to discover the unique characteristics of a picture to improve quality of images by raising the resolution. It has a drawback for the prediction, disease segmentation and insect detection using CNN architechture.

[5] has revealed that the deep learning approaches a larger amount of data for diagnosis of plant. Generative Advertised Networks (GANs) can be used to generate synthetic data when there is enough instances in the data training and traditional procedures which doesn't improve the results sufficiently.
[6] has utilized RESNET 50 model for classification. It's additional layers enhance the feature extraction by consuming least amount of resources. CNN used to classify the plant defects and suffered for predict the true labels specification for plant leaves detection.
[7] has pre-trained CNN model such as VGG-16,VGG-19, Inception V3 and Xception on image net data used for feature extraction and this target model helps to classify the 40 different Indian herbs.

Proposed work

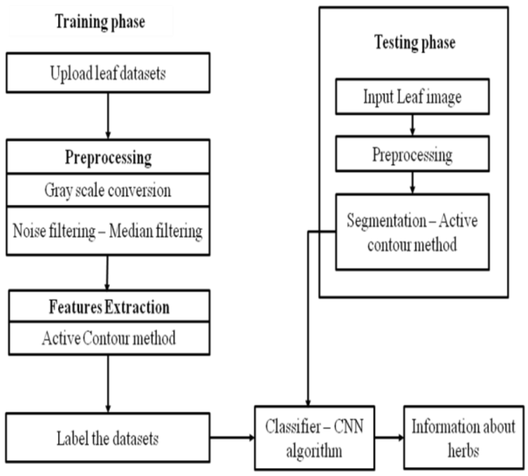

Fig. 1. Systematic Block diagram of Proposed work

The proposed system focused an automated plant identification system to assist users who lack specialized knowledge and in-depth training in botany and obtaining information about some herbal plants by photographing the plants and feeding the images into an automated plant recognition system. Botanists have requested that computer vision-assisted plant identification technologies created to help them recognize and identify unknown herbal plants more quickly. The systems main duties are image identification and retrieval, which have attained a lot of attention from computer vision experts. Identification of leaf species has a wide range of socio-economic implications. Pattern recognition is being used to identify plants and there is a lot of research going on this particular area. This project uses the.NET framework to discuss CNN-based methodologies for identifying Indian leaf species from a white backdrop. To segment the leaf parts in the suggested system, we can use the guided active contour method. For effective leaf classification, different CNN models are applied to features such as traditional shape, texture, colour and venation as well as additional minor features such as edge pattern uniformity, leaf tip, margin and other statistical features.

Training and Testing phases

Data Collection

This is the phase to upload various datasets as input about the medicinal herbs with its characteristics such as leaf dimensions which helps to train the model for the classification and labelled. We collected and processed the data and further examined and crosschecked the dataset against several available online source. It is ensured that the dataset consists of a comprehensive and diversified range of dataset to train the model and futher divided the datasets let into training and testing.

Pre-Processing

It is the step to create a raw data to the mode. The median filter is a non-linear digital filtering techniques, frequently used to suppress noise from an image or signal.
Data Cleaning: As a primary step in getting a herb's datasets ready for Machine learning model training is data cleaning. In this process, redundant datasets were eliminated, colour balance was adjusted, brightness firm and contrasting nature were adjusted, images were cropped, annotations were reviewed and corrected, data augmentation was used and the dataset was split. By ensuring the enhanced quality of the data used for training, we can develop more accurate and reliable systems for detecting and classifying the characteristics of medicinal herbs.
Data Balancing: Balanced datasets are needed to build reliable models. The algorithm is trained on a representative dataset where each class has equal samples. Before training the learning model, we balanced the dataset. Class imbalance was addressed by

undersampling and oversampling. Oversampling randomly repeats samples from minority classes to improve dataset representation. Undersampling scrutinizes majority class dominance by eradicating samples. Balanced datasets improve detection and classification by accurately reflecting the characteristics of herbs.

Data Augmentation & Resize: Data augmentation uses available data to expand the training dataset. This method is excellent for tiny or biased .The larger and more number of diversified dataset increased our plant classification system's accuracy and robustness. In addition to that of data augmentation, image resizing was performed to standardize image sizes. All images were scaled to 512 × 512 pixels. The model performed better when we increased the image size from 256 by 256 pixels to 512 by 512 pixels.

Annotation & Labelling: We utilised the Labelling tool to annotate, creating a bounding box to the affected part and label it, to assign the corresponding class category of herbs.

Methodology

This project is carried over by two machine learning classifiers as following

Convultional Neural Networks

Artificial Neural Networks (ANN) are capable of learning and may thus be trained to recognize patterns, develop solutions, predict future occurrences, and classify data. The use of ANN for traffic-related activities is well established. The way its separate computing parts are coupled, as well as the intensities of these connections or weights, determine how neural networks learn and behave. These weights can be automatically modified by training the network according to a learning rule until it completes the task appropriately. These well-known parameters assist the ANN in making predictions. The fundamental components of a training dataset are input data and their response values. The greatest strategy to increase predictive power and the ability to generalize across several new datasets is to employ larger training datasets. The back propagation algorithm can be used to classify. Back propagation is a typical way of training artificial neural networks to reduce the goal function to the smallest possible value. It is a generalization of the delta rule and is a supervised learning method. It necessitates a training set that contains a dataset of the intended result for a variety of inputs. It's perfect for feed-forward networks. The phrase is an acronym for "error propagation".

Steps in CNN algorithms:

Step 1: Randomly initialize the weights and biases.
Step 2: feed the training sample.

Step 3: Propagate the inputs forward; compute the net input and output of each unit in the hidden and output layers.
Step 4: back propagate the error to the hidden layer.
Step 5: update weights and biases to reflect the propagated errors. Training and learning functions are mathematical procedures used to automatically adjust the network's weights and biases.

Step 6: terminating condition.

some characteristic or computed feature, such as colour, intensity, or texture, each pixel in a region is comparable.. In image processing, segmentation is the most crucial step. A fence divides a complete image into many portions, making it more meaningful and easier to process. The complete image will be covered by these many portions that have been linked together.

Active Contour Method

The active shape model is a discrete variant of this approach in two dimensions, using the point distribution model to limit the shape range to an explicit domain learned from a training set. Because the method requires prior knowledge of the intended contour shape, snakes do not address the complete challenge of identifying contours in photos. Instead, they rely on additional mechanisms including human involvement, interaction with a higher-level picture comprehension process, or information from image data that is close in time or location. Contour models are used in computer vision to describe the edges of forms in an image. Snakes are especially well-suited to problems which the approximate geometry of the border is known. Snakes can adjust to variations and noise in stereo matching and motion tracking since they are a malleable model. In addition, by ignoring lacking boundary information, the approach can detect illusory shapes in an image. Because a polygonal model is unavoidably an approximation of the true contour of the leaf, basing its evolution on the edge information included in the gradient is irrelevant. To make the model suit the leaf in the image, we just need the colour information. It is, however, impossible to develop an a priori model for determining the hue of a leaf that is accurate regardless of the leaf, season, or illumination. The specific colour model must then be estimated for each new leaf we want to segment, which can only be done with an approximate idea of where the leaf lies in the image, implying limits on its position. We'll assume that the leaf is roughly centered and vertically oriented in the following, so that a template initialization in the centre of the image comprises virtually entirely leaf pixels. A leaf in an image captured on purpose by a user is usually large enough to verify that the initialization is proper. The internal elastic energy term E internal and the outward edge-based energy term E external define a simple elastic snake. The internal energy term's aim is to manage the snake's deformations, while the exterior energy term's function is to regulate the contour's fit onto the image. The external energy is frequentlymix of forces caused by the image itself E image and constraint forces imposed by the user E con.

The energy function of the snake is the sum of its external energy and internal energy as follows:

$$E_{snake}^{*} = \int_{0}^{1} E_{snake}\left(v(s)\right)ds = \int_{0}^{1}(E_{internal}\ (v(s) + E_{image}\left(v(s)\right) + E_{con}\ (v(s)))ds$$

The pseudo code as follows:

```
function v = snakes (I, v)
% INPUT: N by M image I, a contour v of n control points
% OUTPUT: converged contour v of n control points
E_image = generateImageEnergy (I);
while not converged
F_cont = weight.alpha * contourDerivative(v, 2);
F_curv = weight.beta * contourDerivative(v, 4);
F_image = interp2 (E_image, v(:,2), v(:,1));
F_image_norm = weight.k * F_image ./ norm (F_image);
F_con = inputForces();
F_internal = F_cont + weight.external * F_curv;
F_external = weight.external * (F_image + F_con);
    v = updateSnake(v, F_internal, F_external);
checkConvergence ();
end
end
```

We decided to reuse the dissimilarity map from the previous phase instead of having an external energy element based on colour consistency or distance to a mean, because we already had an efficient measure of how pixel should fit in the leaf in terms of colour. Despite the fact that the final contour must be exact in order to capture points and teeth on the leaf margin, some minor smoothing is required to keep the contour from becoming too noisy. Finally, the balloon energy is present to counterbalance and stabilize the other energies by exerting a steady push on the contour's outside.

Result & Analysis

This framework used the features extraction and classification techniques. Then can evaluate the performance using accuracy metrics. The accuracy metric is evaluated as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

The proposed algorithm provide improved accuracy rate than the machine learning algorithms.

TABLE 1. Accuracy Table of different algorithms

Algorithm	**Accuracy (%)**
Naives bayes	20
SVM	50
CNN	90

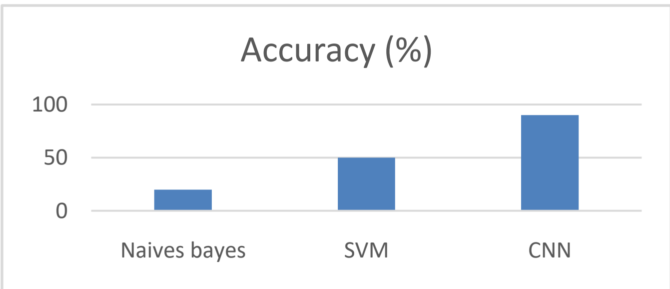

Fig 2: Performance report

From the performance chart, CNN provide high level accuracy than the existing machine learning algorithms.

Conclusion & Future work

Traditional medical systems are still frequently using for a variety of reasons. By increasing emphasis on the usage of medicinal herbs as a source of medicines for a wide variety of human ailments has resulted from population growth, insufficient drug supply, side effects of several synthetic drugs and the development of immune to resist against for infectious diseases. In this effort, CNN-based techniques for detecting Indian leaf species were proposed. Pre-training and edge detection were used in the trials. Softmax and sigmoid layers are using in CNN experiments. The results show that binary CNN with sigmoid can detect leaf species with accurate edge detection and pre-training. The project development provides the cost effective and ease method for classifying and characteristics of the medicinal herbs. We worked toward this successful methodology of automated plant species classification because of the medicinal properties of the plants and the strong demand for the plants. We can modify the framework in the future to implement various deep learning techniques to improve plant categorization accuracy. There are also many classifications for different portions of herb plants such as the root, flower and other parts.

References

[1] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, ‘‘Plant identification using deep neural networks via optimization of transfer learning parameters,’’ Neuro-computing, vol. 235, pp. 228–235, Apr. 2017.
[2] J. W. Lee and Y. C. Yoon, ‘‘Fine-grained plant identification using wide and deep learning model 1,’’ in Proc. Int. Conf. Platform Technol. Service (PlatCon), Jan. 2019, pp. 1–5.
[3] A. Kaya, A. S. Keceli, C. Catal, H. Y. Yalic, H. Temucin, and B. Tekinerdogan, ‘‘Analysis of transfer learning for deep neural network based plant classification models,’’ Comput. Electron. Agricult., vol. 158, pp. 20–29, Mar. 2019.
[4] V. Bodhwani, D. P. Acharjya, and U. Bodhwani, ‘‘Deep residual networks for plant identification,’’ Proc. Comput. Sci., vol. 152, pp. 186–194, Jan. 2019
[5] M. Arsenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, ‘‘Solving current limitations of deep learning based approaches for plant disease detection,’’ Symmetry, vol. 11, no. 7, p. 939, Jul. 2019.
[6] I. Z. Mukti and D. Biswas, ‘‘Transfer learning based plant diseases detection using ResNet50,’’ in Proc. 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT), Dec. 2019, pp. 1–6
[7] H. S. Abdullahi, R. E. Sheriff, and F. Mahieddine, ‘‘Convolution neural network in precision agriculture for plant image recognition and classification,’’ in Proc. 7th Int. Conf. Innov. Comput. Technol. (INTECH), Aug. 2017.
[8] C.-C. J. Kuo, ‘‘Understanding convolutional neural networks with a mathematical model,’’ J. Vis. Communication. Image Representation., vol. 41, pp. 406–413, Nov. 2016.
[9] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, ‘‘Deep learning for plant identification using vein morphological patterns,’’ Computer. Electron. Agriculture., vol. 127, pp. 418–424, Sep. 2016