ESP Journal of Engineering & Technology Advancements (ESP-JETA)
ISSN : 2583-2646
Special Issue

AUTOMATED VISION BASED SYSTEM FOR MEDICINAL PLANTS

[1]Bowmitha parveen S
Department of Agriculture
MCE-Mangayarkarasi college of Engineering
paravai-Madurai
bowmis593@gmail.com

[2]Malathi P
Department of Agriculture
MCE-Mangayarkarasi college of Engineering
paravai-Madurai
malathip803@gmail.com

[3]Nandhini M
Department of Agriculture
MCE-Mangayarkarasi college of Engineering
jothimaya729@gmail.com

[4]Mr V Muthukarthick
Assistant professor
MCE-Mangayarkarasi college of Engineering
muthukarthick96@gmail.com

Abstract— **India has the treasure of medicinal plants due to wide variability in Agro climatic condition spread all over the country. India has 16 Agro-climatic zones having 45000 different plant species in which 15000 species are medicinal that include 7000 plants used in Aurveda, 700 in Unani, 600 in Sidha medicine, 450 in Homeopathy and 30 in Modern Medicine. There are many different plant species in India, each with its own special set of medicinal properties. The names 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 essential to preserve these medicinal plants because doing so will benefit a wide range of fields, including medicine, botanic research, and plant taxonomy studies, among others. The variety of medicinal plant species that are present in India cannot be replicated by current technologies. This method makes it easier to classify medicinal plants by utilizing textural characteristics that are essential for recognizing and identifying leaves. There are three main steps consists of feature extraction, classification, and picture improvement. In this study identification of medicinal plant leaves was carried out using the Convolutional Neural Network method. Using training data of 3,265 images of Medicinal Plants images are collected. There are many medicinal species available on earth, we focus on Ten species including: Aloevera, Curry Leaf, Moringa Olifera, Henna, Hibiscus, Tulsi (Holy Basil), Lemon, Mint, Neem, Betel. It is carried out a computer dataset and then implemented in Web-based software to recognize the types and benefits of medicinal plant leaves identified. The final result of this project gives the 97% of accuracy, in future we upload more species and it will give added value to the Medicinal Field to recognized most suitable supplementary for Allopathy. Automatic Identification of Medicinal plants is essential and soon replace the manual recognition of enormous plant species by resulting in many benefits such as the reduced time and cost with acceptable classification Accuracy.**

KEY WORDS: Medicinal plants, Identification, Image processing CNN (Convolution Neural Networks), Feature Extraction, Deep Learning, Classification, Segmentation.

I. INTRODUCTION

There are almost 374,000 plant species around the world [1]. Among all the plants, some used in medicine, which provides many drugs from the ancient time to the present. Within the context of India, there are about 7500 enlisted medicinal plants [2]. Among them, a lot of traditional and modern medicine exists which can be derived from these plants. Considering this huge number, the medicinal plant classification is a fairly difficult task and lengthy process even for experienced botanists. Because it relies much on the inherited knowledge of an expert botanist. Also, the plants are hard to recognize because of their almost similar shape and color. So, it is important to study and classify plants correctly for further use. As manually identifying these plants requires a lot of time, an automation process is needed to automate this plant classification system so that the people with no botanical knowledge can identify plant easily. Our objective of this research is to make the process of plant classification easy for a human. In this regards, we propose an approach for classification of medicinal plants. This method offers an alternative solution to the traditional way of identifying medicinal plant images by the botanist and reduces time, the manpower and associated costs for the whole process. In this paper, we introduce an image dataset of 10 classes of medicinal plants which were taken in different conditions. Also, a three-layer convolutional neural network is applied to classify the plants. In CNN, there are feature maps which capture the result of the filters to an input image. That feature map is then passed to a fully connected layer to perform the classification. For performing the classification with convolutional neural network, it requires a huge amount of training data. Data augmentation is a technique used for enlarging the data. It was first introduced in 2001 [3]. This technique increases the multiplicity of data for training model without collecting new data. It is more effective for reducing overfitting and improving generalization [4]. Today, the applications of computer vision are trying to make it easier to identify plants for humans. Plants can be classified using its bark, roots, seeds, flower and leaves [5], [6]. According to Anant at el. [7] the plant leaves contain essential feature and also it is two dimensional. So it is easy to study the shape of leaves than the flower or morph of plants because of their three-dimensional structure. Previously, the plants were identified using many features of leaves including leaf margin, veins, etc. with standard classifiers such as support vector machine, KNN and random forest. But there are limitations in these approaches. Firstly, the images must be spotless without having any natural background that makes it difficult to classify in real-world situations. Secondly, the hand-crafted features are easily affected by the noise and the change of conditions such as light, rotation, etc. Thirdly, all these techniques were applied to a limited number of images with a limited number of features. These techniques can't extract the high-level feature on itself. To extract high-level features or key features, CNN is far better than other methods [8]. There were many approaches which used CNN for plant classification. In most cases, they applied their method on a dataset which was clicked on a lab environment or in a plain background. For which we add a novelty to our practice by using a dataset which contains both single and multiple leaf images also in the natural environment. The rest of the paper is organized as follows: Section II discusses the related works which have been done previously. Section III introduces the dataset and the data augmentation techniques used in this paper. In Section IV, the proposed model architecture is analyzed. Section V mentions the experimental result and discusses this result. Finally, Section VI concludes the paper while mentioning the future direction of our work.

II. LITERATURE REVIEW

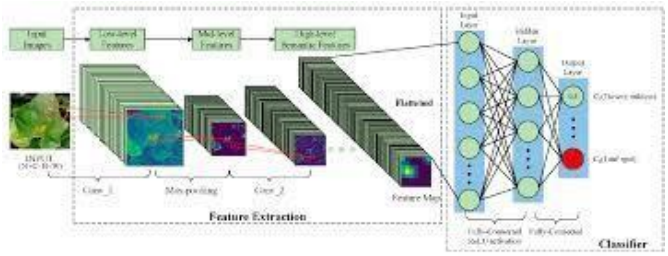
In the past decades, there were various works on plant classification, some of were based on low-level feature extraction such as color, texture and leaf vein extraction, etc. [9], [10], [11]. Also in 2013, R.Janani et al. [12] used a small dataset of 63 medicinal leaf images of 6 classes and implemented an ANN classifier. They used a minimum of eight input features and got 94.4% correct identification accuracy of leaves. In 2014, Mohd Shamrie Sainin et al. [13] proposed a framework for identifying the tropical medicinal plants in Malaysia, which was based on the extracted pattern from the leaf. They used an ensemble classifier and got almost 65% accuracy from five species of plants. In 2017, a new dataset on medicinal plants of Mauritius was introduced. They applied different algorithms on this dataset. They found the best accuracy from random forest classifier, which was 90.1% [14]. Also in 2017, Adil Salman et al. [15] studied different features related to leaf shape, size. They selected convex area, filled area and perimeter e.t.c for identifying the leaves of twenty-two class. They achieved an accuracy of 85% to 87%. In 2016, Codizar [16] proposed a desktop application LeaVes which aims to classify a plant species by leaf images alone. They used the handcrafted image featuring and different machine learning algorithm: centroid-radii model, moment-invariant model, canny edge detection, morphological operations, image difference, artificial neural networks as classifiers to achieve 95% accuracy. These approaches classified medicinal leaf images based on a limited number of hand-crafted features. Later, some researchers used deep learning techniques to automate the feature extraction procedure. In this regard, important work was done by Chaoyun Zhang et al. [17]. They achieved 94.6% accuracy using the combination of CNN and data augmentation together for plant identification. They used the Flavia dataset for classifying the plants. In 2016, Hulya Yalcin et al. [18] applied a pre-trained convolutional neural network to classify sixteen kinds of agricultural plants. They also evaluate performance with SVM classifier, and their experiment indicates that their CNN model gives a higher performance which is 97.47%. S.H Lee et al. [19] published a paper in which the use of deep learning to harvest discriminatory features from leaf images by learning and applied them as classifiers for plant identification was investigated. Their result showed that learning the features using CNN can provide better feature representation of leaf images as compared to using handdesigned features. M.M Ghazia et al. [20] and Aydin et al. [21] applied transfer learning to identify plant species using deep convolutional neural network on different dataset and got the accuracy of about 78.44%. Sue Han Lee et al. [22] developed a convolutional neural network architecture for the LifeCLEF dataset challenge, which was the competition of identifying 1000 species of images of plants. They got about 68.9% accuracy after applying augmentation techniques. Recently in 2019, a new dataset on medicinal plants images in the natural scene has been introduced. The dataset consists of a total of 10279 images of 10 plant species. The authors got 93.6% accuracy by using the VGG16 model to extract the features from the images and then applied the LightGBM classification method.

III. PROPOSED SYSTEM DATA COLLECTION

A medicinal plant dataset is introduced which contains 3265 images of 10 different medicinal plants with one default class. The following plants are selected for classification: ‘Aloevera’, ‘Curry Leaf’, ‘Moringa Olifera’, ‘Henna’,’ Hibiscus’, ‘Tulsi (Holy Basil)’,’ Lemon’,’ Mint’, ‘Neem’, ‘Betel’. As convolutional neural network requires a large number of training images, the augmentation technique can bring massive effect in this small amount of images. So we applied augmentation techniques such as flipping and different rotation angles. After that, the number of images becomes 7565.

NETWORK ARCHITECTURE

In this work, we build a small convolutional neural network from scratch to classify the exact classes. Technically in a convolutional neural network, every input images are passed through a series of convolutional layers with different kernels. Then the convolution operation has been done by element-wise matrix-multiplication and sums the result which turns into a feature map. Here we applied 128 different types of weight filter to produce the feature map. As we used different filters, it will make different types of feature maps. So we stack all the feature maps and pass these through an activation function for summing this up and also for having non-linearity effects. It describes the basic feature extraction procedure of CNN architecture.



The proposed CNN architecture as illustrated and it is consist of several layers, in which convolutional, Gaussian noise, Batch normalization, Max Pooling and Dropout are placed sequentially. This pattern was repeated three times. Then, a large dense layer was placed at the output end of the network in an attempt to better translate the large number of feature maps to class values. The input image was 128 x 128 with 3 channels and which was then fed with a CNN of 128 different kernels each having 3x3 filter size and also an activation function (ReLU) was added for non-linearity effect which is defined as below:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x > 0 \end{cases}$$

As the size of the dataset is comparatively small, it may represent a harder mapping problem to learn for the neural network. So to make this smoother, one way is to add noise to inputs during training. For this reason, we added Gaussian noise to make the inputs different every time and network will less able to memorize during training. We also added batch normalization which standardizes each layer input for every mini-batch. The batch normalization also stabilizes the learning process.

After that, a pooling layer with filters of size 2x2 was applied to reduce the total amount of parameters in the network. The pooling layer uses the MAX operation on every depth slices of the input and resizes. Lastly, a dropout rate is added. Hence, the network becomes less sensitive to the specific weights of neurons. This dropout technique results in a better- generalized network and is less likely to overfit during training

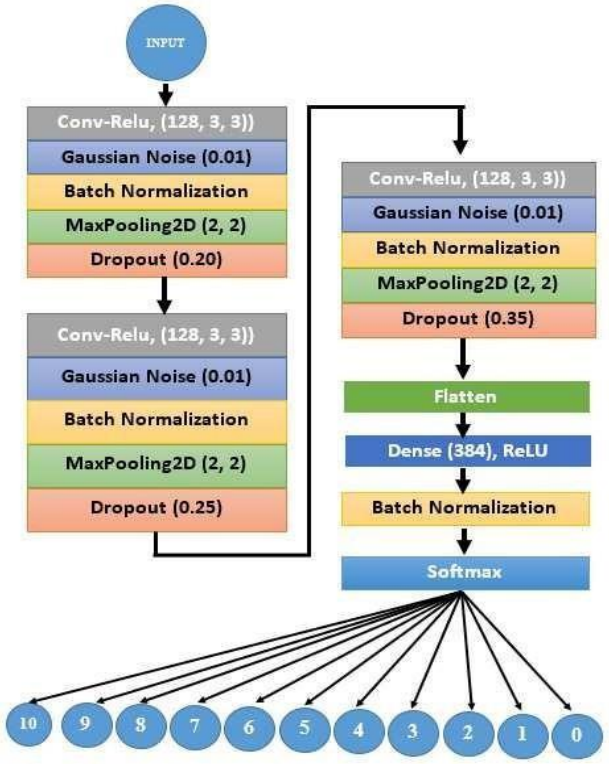


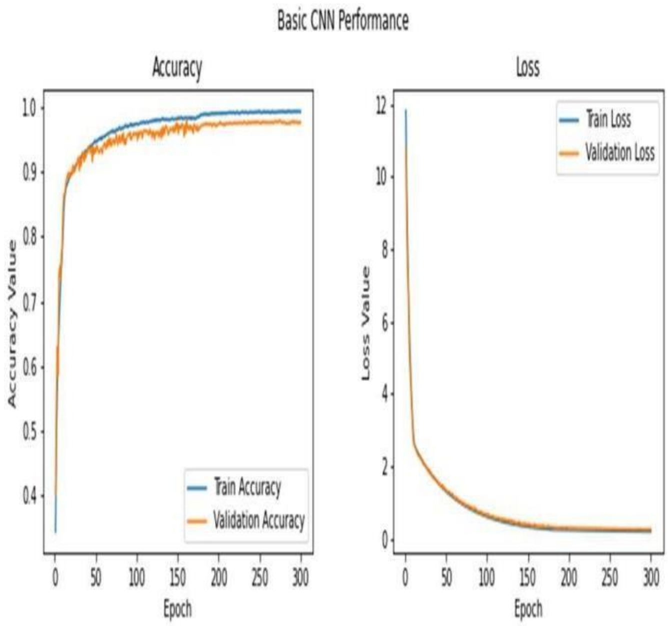
Figure: Network Structure of Proposed Model. Starting with an input layer, followed by a CNN layer, Gaussian noise and batch normalization layer. Then a MaxPooling layer to reduce the total parameter number and finally a dropout layer to disable a certain amount of neurons. This pattern is repeated three times. Lastly, a fully-connected layer is added before the Softmax output function.

This pattern is repeated three times with gradually increasing dropout rate. In the end, a fully-connected layer is added and a softmax output function is used to turn the outputs into probability-like values and allow one class of the 11 to be selected as the model's output prediction. We added an extra input and output class called 'none' class to classify that the input images belong to that 10 plant classes or not.

EXPERIMENTAL RESULT

In the experiment, the images were divided into two parts: the training and the test image dataset. To use the dataset, we first need to process the images because machines can't take the images as they are. So the data were preprocessed to become suitable for upcoming training steps. In the preprocessing techniques, first we resized the images into 128*128 dimension to establish a fixed size. Then the images are converted into numpy array. After that, each image was normalized by dividing each value of data by 255 which is maximum

observation and scaled the data into [0, 1].
After preprocessing the data, they were sent to the convolutional neural network for feature extraction. As said, we used double augmentation technique at preprocessing, in the training period the other augmentation method such as width shift, height shift, zoom range etc. was applied. The mentioned model was trained using multiple learning rates through the epochs, firstly 0.01 for 10 epoch, then 0.001 for 165 epoch and lastly 0.0001 for 125 epoch. It was trained with 300 epochs and a small batch size of 64. The entire training phase finished in approximately 7.5 hours, where each epoch took almost 87 seconds. In the training period, the weights were optimized through backpropagation. Backpropagation is a process where the weights of the layers were correctly tuned with respect to the optimization function while carrying out forward and backward passes.



it has been seen that the validation accuracy is above 95%. In the rightmost loss curve, in the initial training period the validation loss was almost 11, and in the end the loss was about 0.3.
The training was performed on a Google Colab environment featuring a GPU service, 25 GB RAM. The learning model was developed in Keras, a Python library for deep learning, using Tensorflow backend and trained over GPU.

CONCLUSION

In this work, we introduced an attention architecture for the task of extracting features for plant classification and identified medicinal plants from leaf images. We collected a dataset of Bangladeshi medicinal plants, preprocessed them, and classified the plant species using a deep learning technique. In our paper, we described the methodology of the architecture. We analyzed the results based on both training and testing sets, and our result is quite impressive, which correctly identifies 71.3 % of leaves. Besides, the results of the proposed architecture are promising and comparable to other existing methodology through in our dataset, we have both single and compound leaf images.

In the future, we will focus on exploring the CNN model for better performance in both single and compound images as well as we will extend our dataset. Finally, we aim to build a mobile application based on this model by which people can be able to identify medicinal plants more easily.

References

[1] Christenhusz, Maarten J.M, Byng James W. “The number of known plants species in the world and its annual increase”, Phytotaxa, [S.l.], 2016.

[2] Ghani, A. “Medicinal plants of Bangladesh: Chemical constituents and uses.” (1998).

[3] Dyk D A V, Meng X. “The Art of Data Augmentation,” Journal of Computational & Graphical Statistics 10(1) pp. 1-50, 2001.

[4] Pawara, Pornntiwa, Emmanuel Okafor, Lambert Schomaker, and Marco Wiering. "Data augmentation for plant classification." In International Conference on Advanced Concepts for Intelligent Vision Systems, pp. 615-626. Springer, Cham, 2017.

[5] Ganschow, Lene, Tom Thiele, Niklas Deckers, and Ralf Reulke. "Classification of Tree Species on the Basis of Tree Bark Texture." International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives 42, no. W13 (2019).

[6] Seeland, Marco, Michael Rzanny, Nedal Alaqraa, Jana Wäldchen, and Patrick Mäder. "Plant species classification using flower images—A comparative study of local feature representations." PloS one 12, no. 2 (2017): e0170629.

[7] Anant Bhardwaj, Manpreet Kaur, and Anupam Kumar. “Recognition of plants by Leaf Image using Moment Invariant and Texture Analysis”, International Journal of Innovation and Applied Studies, 2013.

[8] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, pp. 2319-2323, 2018.

[9] Basavaraj S. Anami, Suvarna S. Nandyal , A. Govardhan, India. “A Combined Color, Texture and Edge Features Based Approach for Identification and Classification of Indian Medicinal Plants”. International Journal of Computer Applications (0975 – 8887) Volume 6– No.12, September 2010.

[10] Herdiyeni, Y. and Wahyuni, N.K.S., “Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram”. International Conference on Advanced Computer Science and Information Systems (ICACSIS), West Java, Indonesia, 301-306. , 2012.

[11] Gopal, A., Prudhveeswar Reddy, S., & Gayatri, V. “Classification of selected medicinal plants leaf using image processing”. 2012 International

Conference on Machine Vision and Image Processing.

[12] R.Janani, A.Gopal. "Identification of Selected Medicinal Plant Leaves Using Image Features and ANN". International Conference on Advanced Electronic Systems (ICAES), 2013.
Mohd Shamrie Sainin , Taqiyah Khadijah Ghazali and Rayner Alfred. "Malaysian Medicinal Plant Leaf Shape Identification and Classification". Knowledge Management International Conference (KMICe) , 12 – 15 August 2014, Malaysia.

[13] Adams Begue, Venitha Kowlessur, Upasana Singh, Fawzi Mahomoodally and Sameerchand Pudaruth, "Automatic Recognition of Medicinallants using Machine Learning Techniques" International Journal of Advanced Computer Science and Applications(ijacsa), 8(4),2017. http://dx.doi.org/10.14569/IJACSA.2017.080424

[14] Adil Salman, Ashish Semwal, Upendra Bhatt. "Leaf Classification and Identification using Canny Edge Detector and SVM Classifier". International Conference on Inventive Systems and Control (ICISC- 2017)

[15] Azeil Louisse Codizar, Geoffrey Solano. "Plant Leaf Recognition by Venation and Shape Using Artificial Neural Networks".

[16] Zhang, Chaoyun, Pan Zhou, Chenghua Li, and Lijun Liu. "A convolutional neural network for leaves recognition using data augmentation." In 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 2143-2150. IEEE, 2015.

[17] Hulya Yalcin, Salar Razavi. "Plant Classification using Convolutional Neural Networks".

[18] S.H Lee , C.S Chan, S.J Mayo, P. Remagnino,2017. "How deep learning extracts and learns leaf features for plant classification".

[19] M.M Ghazia, B Yanikoglu, E Aptoula, 2017. "Plant identification using deep neural networks via optimization of transfer learning parameters".

[20] Kaya, Aydin, Ali Seydi Keceli, Cagatay Catal, Hamdi Yalin Yalic, Huseyin Temucin, and Bedir Tekinerdogan. "Analysis of transfer learning for deep neural network based plant classification models." Computers and electronics in agriculture 158 (2019): 20-29.

[21] Sue Han Lee, Yang Loong Chang, Chee Seng Chan, and Paolo Remagnino(2016). "Plant Identification System based on a Convolutional Neural Network for the LifeClef 2016 Plant Classification Task".

[22] Anh H. Vo, Hoa T. Dang, Bao T. Nguyen, and Van-Huy Pham. "Vietnamese Herbal Plant Recognition Using Deep Convolutional Features", International Journal of Machine Learning and Computing, Vol. 9, No. 3, June 2018.