



Medicinal Plant Recognition based on CNN and Machine Learning

¹Bhanuprakash Dudi, ²Dr.V.Rajesh

¹Research Scholar, Dept.of ECE, K L Deemed to be university, Vaddeswaram, Guntur, Andhra Pradesh

¹Associate Professor, Dept.of ECE, CVR college of Engineering, Ibrahimpatanam, RangaReddy, Telangana

²Professor, Dept.of ECE, K L Deemed to be university, Vaddeswaram, Guntur, Andhra Pradesh

pbhanududi@gmail.com, rajesh444@kluniversity.in

ABSTRACT

In the recent days automated plant species recognition systems are developed to help the ordinary people in identification of the different species. But the automatic analysis of plant species by the computer is difficult as compared to the human interpretation. The research has been carried out in this field for the better recognition of plant species. Still these approaches lack with exact classification of the plant species. The problem is due to the inappropriate classification algorithm. Especially when we consider the recognition of medicinal plant species, the accuracy will be the primary criteria. The proposed system in this research adopts the deep learning method to obtain the high accuracy in classification and recognition process using computer vision techniques. This system uses the Convolutional Neural Network (CNN) and the machine learning algorithms for deep learning of medicinal plant images. This research work has been carried out on the leaf dataset of flavia from sourceforge website. This data is fed as the training dataset for the CNN and machine learning based proposed system. An accuracy of 98% has been achieved in the recognition of the medicinal plant species. All the performance metrics like precision, recall, F1-score and support are calculated. Also the achieved training and validation accuracies are nearly equal.

Key words: Medicinal plant, image recognition, classification, CNN and Machine Learning.

1. INTRODUCTION

An image represents the most valuable information than the natural description in many applications like plant recognition, face recognition, etc. Unlike the human beings, it is very difficult process to extract the features by the computer/ system. To obtain the good accuracy the computer/system has to be trained in a well manner with the help of training data-set. The more the training data-set gives the more number of feature vectors in the extraction process. Also it gives the good amount of accuracy in recognition process. The accuracy in recognition is the most important parameter to identify the similar kind of objects as well as to distinguish the different type of objects. In the applications like face recognition, this parameter gives the permission to only authorized users whereas in the applications like medicinal plant recognition system, it identifies the absolutely needed medicinal plant to save the life of a patient.

Usually ordinary people are assigned with the job of collecting the plants from the forests. Sometimes they couldn't recognize the rare and important plants because of human error. These rare kinds of plants are very important to save the life of a patient. Also, sometimes these people could pick-up incorrect species which may be harmful plant. In such cases, it is necessary to use automatic plant recognition system. This system helps an ordinary people or any lay man to recognize the different plant species. These kinds of systems are also very helpful for the trekking people if they are interested to collect the plant species while trekking the mountains.

2. LITERATURE SURVEY

In the paper [1], a CNN based plant classification system has been proposed to classify the different types of plant species from the image database collected from smart agro-stations. In this method CNN architecture is used for feature extraction of different plant images. This approach is tested on the TARBIL database and obtained an accuracy of 97.47% on 16 different plant species. In this method the CNN based classification shows the more accuracy than the SVM based classification.

In the paper [2] by shah, sougatta singh et al., it has been proposed that a leaf classification system using the dual-path deep CNN. The dual-path CNN will do the following major operations

- i. Both the *shape* and *texture* characteristics are studied
- ii. Optimizes the obtained features for the classification.

Using this method, it has been obtained a good accuracy of about 99.28 % (flavia data-set) which is better than other uni-path CNN methods.

In the paper [3] by Sue Han, Seng, et al., it has been proposed that a plant identification by using the deep learning CNN algorithm. In this method the CNN algorithm is used to learn unsupervised feature representations. These experiments are carried out on 44 different plant types, which are collected at the Royal Botanic Garden. The experimental results show the consistency and superiority than the other hand-crafted feature extraction methods.

Wang-Su and Sang Yong proposed a plant leaf recognition using CNN model [4] and created two models by adjusting

the network depth using Google Net. Also the performance of each model is evaluated according to the discoloration/damage effects are applied on the leaves. Accuracy was obtained greater than 94% even with the 30% damaged leaves.

In the paper [5] by Habibollah Agh has developed a new CNN architecture which is applied for leaf classification. This CNN architecture consists of 5 layers. ReLU/ELU activation function is applied after each convolution layer. MaxPooling function is applied for each pooling layer. The fully connected layers are defined as convolutional layers with the filter size of 1x1. The final layer consists of 32 or 15 output units corresponding to 32 and 15 leaf species in Flavia and Swedish leaf datasets. After all layers, a SoftMax loss is placed. But this CNN model achieves an accuracy of 97.24%.

A novel method for plant leaf classification was proposed by Jiachun Liu, S. Yang and et al. [6]. This method uses the CNN algorithm for feature extraction and classification. 10 layers are introduced in this CNN architecture. In this method, an augment for leaf was applied to enlarge the database. So that it can improve the classification performance. The visualization method was used for the analysis of the factors affecting the accuracy. The experimentation of this CNN method was carried out on the Flavia dataset and obtained an accuracy of 87.92%.

3. SYSTEM IMPLEMENTATION

In this method we have used the flavia database for the training and testing of the system. The implemented system consists of 2 main modules as shown in the figure 1.

- i. CNN based Feature Extraction
- ii. Machine Learning based classification

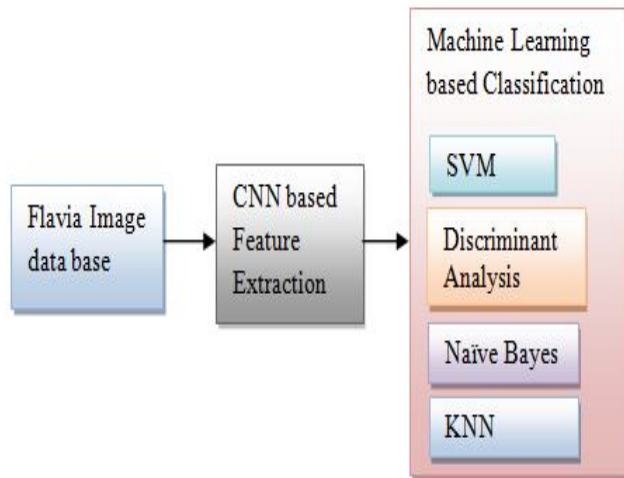


Figure 1: The implemented system

i. CNN based Feature Extraction

In this implementation, Convolutional Neural Networks (CNN) based feature extraction method is adopted to obtain the more accurate features. These features are like diameter, length, width, area and perimeter of the leaves.

Flavia data set is applied as the input training images for the CNN. The performance of CNN based approach is far better than the hand-crafted features of the previous approaches. The CNN architecture is shown in the figure 2.

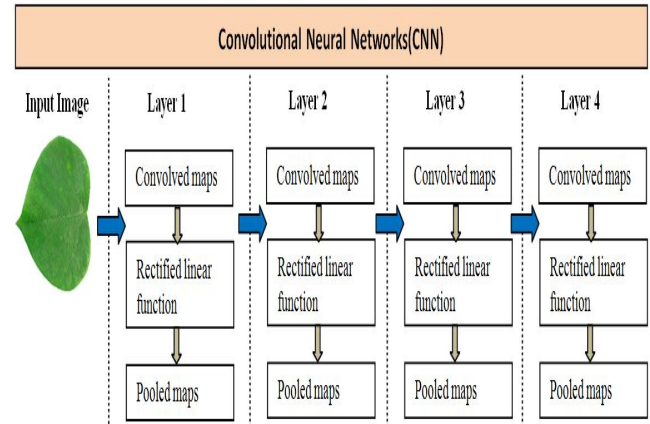


Figure 2: CNN Architecture

This CNN algorithm was implemented on falvia dataset used as the training images with 1800 leaves to extract the features from 32 different species. Here we have used 4 layer CNN model. The CNN architecture is shown in figure 3.1 the network layers are given below.

i) Input Layer

The input data set vector (p) is fed to this layer. Its dimensions are 5x1.

ii) Radial Basis Layer

In this layer, the vector distance (n) is calculated between the vector p and the weight vector ω. It is implemented as the dot product.

$$n_i = \| \omega_i - p_i \| \cdot * p_i \tag{1}$$

$$\text{radial basis } (n_i) = \exp (-n_i^2) \tag{2}$$

Where i is the row number of the vector.

iii) Some characteristics of Radial Basis Layer

$$n_i=1, \text{ if } p_i \text{ is identical with } \omega_i$$

In this case, its output weights in the competitive layer will pass their values to the competitive function.

iv) Competitive Layer

$$\text{The output vector } d = n_i * M_i \tag{3}$$

The competitive function C is calculated from d. C equals to 1 at the largest element of d and 0 elsewhere. It is used as the index to look for the scientific name of this plant.

ii. Machine Learning based classification

Classification is the final step in the automatic plant recognition process. This classification process can be implemented using the various machine learning algorithms like Artificial Neural Network (ANN), Support Vector Machines (SVM), K-Nearest Neighbour (KNN) and Naive Bayes (NB).

SVM [7] classification is characterized by an isolating hyper plane. The concept of SVM classification can be understand easily from the following training data set which is described into two different classes,

$$D = \{(x^1, y^1), (x^2, y^2) .. (x^k, y^k), x_i \in R_m, y_i \in [+1, -1]\} \quad (4)$$

$$\langle w, x \rangle + b = 0 \quad (5)$$

The set of vectors is said to be optimally separated by the hyper plane (Eq. (5)) if the separation error is '0' and the distance function is maximal.

Now a day's ANN [8] is the most widely used research method in the machine learning process. This concept of ANN has the similarity with the human brain neurons system. Performance of ANN depends on the neuron count and hidden layers. In this implementation, feed forward neural network with a single hidden layer was used. In training phase, the system uses the default Scaled Conjugate Gradient function. The minimum gradient is used for stopping criteria.

Also, KNN method can be used for the plant species recognition. KNN algorithm [9-10] is based on distance measure metrics in feature space. KNN classifier at first determines *k* nearest neighbors and then determines labels for the sample based on neighbor weight.

Bayesian classifier is a statistical approach. It performs the classification based on the concept of probability. Accuracy of Bayesian classifier method is less due to its conditional independence of Bayesian probability.

4. SIMULATION RESULTS

The experiment is carried out on the 10 different plant species [16] as shown in table 1. The table 2 shows the classification report (Precision, Recall, F1-Score & Support) [17] obtained using the proposed algorithm. The following are the metrics that are used in the classification process.

- i. **Precision** is the ratio of correctly predicted positive observations to the total predicted positive observations.
 $Precision (P) = TP / (TP+FP)$
 TP → True Positives
 FP → False Positives
- ii. **Recall** is the fraction of positives that are correctly identified.
 $Recall(R) = TP / (TP+FN)$
 FN → False Negatives
- iii. **F1-Score** is created by finding the harmonic mean of precision and recall.
 $F1-Score = 2 * (R * P) / (R + P)$
- iv. **Support** is the number of occurrences in each class.

Table 1: Different plant species used for training and testing data set.

Plant No.	Scientific Name	Common Name(s)	Number of images
1	Phyllostachys edulis (Carr.) Houz.	pubescent bamboo	1-59
2	Aesculus chinensis	Chinese horse chestnut	60-122
3	Berberis anhweiensis Ahrendt	Anhui Barberry	123-194
4	Cercis chinensis	Chinese redbud	194-267
5	Indigofera tinctoriaL.	true indigo	268-323
6	Acer Palmatum	Japanese maple	324-385
7	Phoebe nanmu (Oliv.) Gamble	Nanmu	366-437
8	Kalopanax septemlobus (Thunb. ex A.Murr.) Koidz.	castor aralia	438-496
9	Cinnamomum japonicum Sieb.	Chinese cinnamon	497-551
10	Koelreuteria paniculata Laxm.	goldenrain tree	552-616

Table 2: Classification report obtained using the proposed algorithm

Class/ plant	Precision	Recall	F1-score	Support
class 0 (plant1)	0.85	1.00	0.92	11
class 1 (plant2)	0.60	0.53	0.56	17
class 2 (plant3)	0.77	0.83	0.80	12
class 3 (plant4)	0.73	1.00	0.85	11
class 4 (plant5)	0.88	1.00	0.94	15
class 5 (plant6)	1.00	0.73	0.85	15
class 6 (plant7)	0.88	1.00	0.93	7
class 7 (plant8)	0.70	0.70	0.70	10
class 8 (plant9)	0.64	0.50	0.56	14
class 9 (plant10)	0.82	0.75	0.78	12

The performance of the classification model is shown in the table 3 as the confusion matrix (P represents plant species).

Table 3: Confusion matrix

	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10
P-1	11	0	0	0	0	0	0	0	0	0
P-2	1	9	3	1	1	0	0	1	1	0
P-3	0	0	10	1	0	0	0	0	0	1
P-4	0	0	0	11	0	0	0	0	0	0
P-5	0	0	0	0	15	0	0	0	0	0
P-6	0	1	0	0	1	11	0	0	2	0
P-7	0	0	0	0	0	0	7	0	0	0
P-8	1	1	0	0	0	0	1	7	0	0
P-9	0	4	0	1	0	0	0	1	7	1
P-10	0	0	0	1	0	0	0	1	1	9

The obtained results of training accuracy and validation accuracy values are shown in the graphical form in figure 3.

In this figure we can also observe that the validation accuracy is achieved same as the training accuracy for the less number of epochs (up to 12). In case of more number of epochs these two accuracy values are approximately comparable.

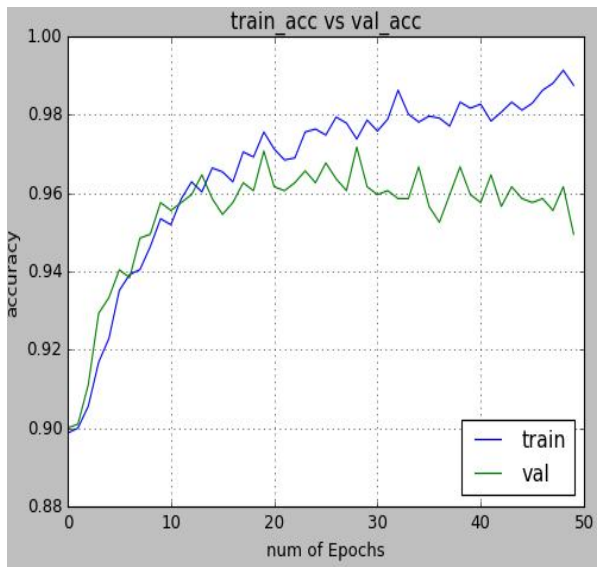


Figure 3: Training accuracy vs. validation accuracy graph

5. CONCLUSION

In this paper, we used CNN and machine learning methods for the feature extraction and classification of medicinal plant species. The different machine learning algorithms are ANN, SVM, KNN and Naive Bayes (NB). We have considered the Flavia dataset in this method. This data sets used for both and training and testing purpose. It has been achieved an accuracy of 98%. All the performance metrics like precision, recall, F1-score and support are calculated. Also the achieved training and validation accuracies are nearly equal. Images used in the training purpose are small and gray scale images. As a future work it is possible to implement the colour image classification for medicinal plant recognition.

REFERENCES

[1] H. Yalcin and Razavi, “Plant classification using convolutional neural networks,” *IEEE, Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Tianjin, 2016, p.p. 01 - 05, doi: 10.1109/Agro-Geoinformatics.2016.7577698.

[2] M. P. Shah, Singha, et. al., “Leaf classification using marginalized shape context and shape+texture dual-path deep convolutional neural network,” *IEEE, International Conference on Image Processing*, Beijing, 2017, p.p. 0860 - 0864. doi: 10.1109/ICIP.2017.8296403

[3] S. H. Lee, Chan, et. al., “Deep-plant: Plant identification with convolutional neural networks,” *IEEE International Conference on Image Processing*, Quebec City, QC, 2015, p.p. 0452 - 0456, doi: 10.1109/ICIP.2015.7350839

[4] Wang-Su Jeon and S. Y. Rhee, ” Plant Leaf Recognition Using a Convolution Neural Network, ” *International Journal of Fuzzy Logic and Intelligent Systems*, Vol. 017, No. 01, March 2017, p.p. 026 - 034, <http://dx.doi.org/10.5391/IJFIS.2017.17.1.26>

[5] Habibollah Agh “A Convolutional Neural Network With A New Architecture Applied On Leaf Classification,” *The IIOAB Journal*, Vol. 07, Suppl 05, ISSN 0326 - 0331.

[6] J. Li, Jia at al., “Unsupervised Representation Learning of Image-Based Plant Disease with Deep Convolutional Generative Adversarial Networks,” *IEEE, 37th Chinese Control Conference*, Wuhan, 2018, pp. 09159 - 09163, doi: 10.23919/ChiCC.2018.8482813.

[7] Y. Es-saady, Massi, et al., " Automatic recognition of plant leaves diseases based on serial combination of two SVM classifiers, " *IEEE, International Conference on Electrical & Information Technologies*, Tangiers, 2016, p.p. 0561 - 0566, doi: 10.1109/EITech.2016.7519661

[8]H. Zhang, Yanne and Liang, "Plant Species Classification Using Leaf Shape and Texture," *International Conference on Industrial Control & Electronics Engineering*, Xiann, 2012, p. p. 2025 - 2028.

[9] Y. Dandawate, et al., “An automated approach for classification of plant diseases towards development of futuristic Decision Support System in Indian perspective,” *International Conference on Advances in Computing, Communications & Informatics*, p.p. 0794 - 0799.

[10] A. Rastogi, R. Arora and S. Sharma, " Leaf disease detection and grading using computer vision technology & fuzzy logic, " *IEEE, 2nd International Conference on Signal Processing & Integrated Networks*, Noida, 2015, p.p. 0500-0505, doi: 10.1109/SPIN.2015.7095350

[11] M. Suresha, K. Shreekanth et. al., “Recognition of diseases in paddy leaves using kNN classifier,” *IEEE, 2nd International Conference for Convergence in Technology*, Mumbai, 2017, p.p. 0663 - 0666, doi: 10.1109 / I2CT.2017.8226213

[12] N. Krithika and Selvarani, "An individual grape leaf disease identification using leaf skeletons and KNN classification,” *International Conference on Innovations in Information, Embedded & Communication Systems*, 2017,

[13] A. Sabu & K. Sreekumar, “ Literature review of image features and classifiers used in leaf based plant recognition through image analysis approach, ” *IEEE, International Conference on Inventive Communication &*

Computational Technologies, Coimbatore, 2017, p.p. 0145 – 0149, doi: 10.1109/ICICCT.2017.7975176

[14] F. R. Padoa, et al., "Using Naïve Bayesian method for plant leaf classification based on shape and texture features," *International Conference on Humanoid, Nanotechnology, Information Technology, Communication & Control, Environment and Management*, p. p. 01 - 05.

[15] X. Jing, "A Bayesian Network Based Intelligent Plant Classification System," *4th International Symposium on Information Science and Engineering*, Shanghai, 2012, p. p. 0263 - 0265.

[16]...Flavia medicinal plant species dataset, website: <http://flavia.sourceforge.net/>, accessed on 01/01/2018

[17] N. Goyal, Kapil and Kumar, "Plant Species Identification using Leaf Image Retrieval: A Study," *International Conference on Computing, Power & Communication Technologies*, 2018, p. p. 0405 - 0411.

[18] Lavanya Maddiseti,Ranjan K.senapati "Training Neural Network as Approximate 4:2 Compressor applying Machine Learning Algorithm for Accuracy comparison," *International Journal of Advanced Trends in Computer Science and Engineering*, Vol(8),No(2),2019.