JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 498-502
https://publishoa.com
ISSN: 1309-3452

A Novel CNN-KNN based Hybrid Method for Plant Classification

Mr. P. Siva Prasad[1], Dr. A. Senthilrajan[2]

1. Research Scholar, Dept. of Computer Science, Alagappa University,
Karikudi 630 003, Tamil Nadu, India
E-mail id: prasadsiva_17@yahoo.com

2. Professor and Director, Dept. of Computational Logistics, Alagappa University,
Karikudi 630 003, Tamil Nadu, India
E-mail id: agni_senthil@yahoo.com

Abstract: Plant classification is an interesting problem in Computer Vision. Several researchers are completed to classification of plant by leaf of plant and flower of plant. After several research efforts, it has been confirmed that leaf of plant is the best and consistent source for classification of plant. However it is interesting to classify a plant through structure of leaf. Hence, it is mandatory to normalize the leaves of plant into the same size to acquire better performance. In this paper, we have proposed a hybrid method namely CNN-KNN based hybrid method. This method used on two datasets. These datasets are LeafSnap and Flavia. CNN + KNN achieved to get maximum accuracy of 98.4% and 96.5%, respectively, in LeafSnap and Flavia datasets.

Keywords: *CNN, KNN, LeafSnap dataset , Flavia dataset*

Introduction

The different kinds of plants exist in the earth. Plants play an important key role in the world by monitoring the environment and improving the climate. In the traditional plant classification systems, we have to know plant knowledge to detect plants by leaf. Due to manual plant classification, we face many problems. Due to all these difficulties, we have to develop an automatic plant identification system that can have better performance. In the automatic plant classification system, we need to have the simple information to classify the plants and no need to know plant taxonomy. Leaf of plant provides the medium for classification of plant. Even diseases of plant can be identified by leaves. The main key point of this paper is to get a best predictive model for classifying the plant. In this paper, we have addressed and proposed a hybrid method to get a better accuracy in result. Image Classification is an important and best application of Computer Vision. Computer vision methods has classified into two types. These are Machine Learning methods and Deep Learning methods. Machine Learning method is K-Nearest Neighbors(KNN). Deep Learning (DL) method is Convolutional Neural Networks (CNN).In this paper, we have addressed and proposed a hybrid method namely CNN-KNN based hybrid method for plant classification. In CNN-KNN based hybrid method, we have integrated CNN with KNN. This method used on two datasets. These datasets are LeafSnap and Flavia.

Related Work

Many research works finished on classification of plant. Different Researchers have classified the plant by getting the features from the leaf dataset. Some Research papers published on classification of plant in good Journals and best conferences. Several research scholars have worked on Artificial Neural Networks and Machine Learning approaches for classification of plant.

Convolutional Neural Network (CNN)

In Deep Learning, Convolutional Neural Network (CNN) can extract the leaf features automatically. Convolutional Neural Network (CNN) can handle two tasks called ‘convolution’ and pooling’. These tasks can produce features from leaf and then using these features to classify the plant.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 498-502
https://publishoa.com
ISSN: 1309-3452

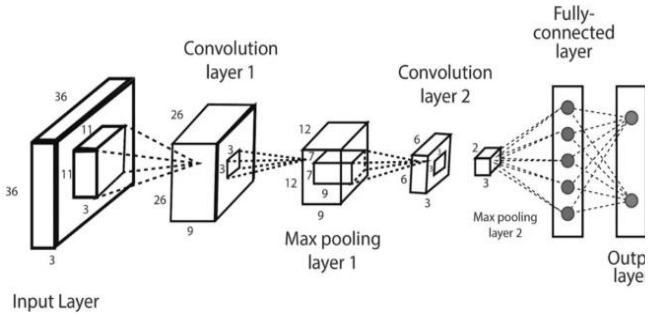


Fig. 1 Architecture of Convolutional Neural Network (CNN)

Proposed Work

In this paper, we proposed CNN-KNN Hybrid method. This method can predict the plant from the LeafSnap leaf dataset and Flavia leaf dataset. Initially, this proposed method is trained through these leaf datasets for predicting the plant class and then system extracts the features from train data. The user will feed the system with a leaf image that is to be tested for plant class prediction. The features extracted from test image will be stored. Later this, the system will compare the test data and train data and predict the plant using KNN algorithm.

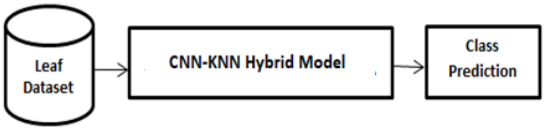


Fig. 2 Proposed System Architecture

ProposedAlgorithm

Step 1: *upload leaf image.*

Step 2: *The proposed system is trained by leaf images stored in train and tune folder using CNN algorithm.*

Step 3: *CNN is used to extract features and classify the image in different categories. The extracted features are written in file (name TRAIN with extension. arff) for further use.*

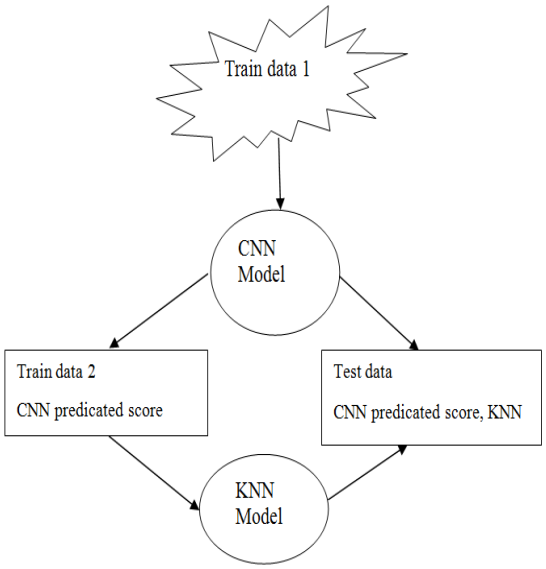
Step 4: *The image uploaded by the user is considered as test image. This test image undergoes feature extraction process using CNN algorithm.*

Step 5: *These extracted features are written in the file (name TEST with extension .arff)*

Step 6: *The TRAIN file and TEST file are compared for prediction using KNN algorithm where the K-nearest neighbor is being calculated.*

Step 7: *This gets the prediction of test image class.*

Proposed Flow Diagram



Experimental Environment

In our research study, Python has been used. In our research study, pandas, matplotlib, Open CV, etc. predefined libraries has been used for data analyzing and numerical plotting. Pandas is used for structuring of the data. NumPy is used for basic numerical calculations. Open CV is used for the computer vision task in plant classification.

Data Pre-processing

In this paper, we have transformed leaf image photos to a JPEG or JPG format before examination. Therefore, we have collected leaf image photos (JPEG). We have converted these image photos into JPEG format by a Python script. The noise in the leaf image photo has been reduced through scaling and converting the image using Python scripts. Both CNN leaf image reconstruction and leaf measurement has been used as data pre-processing approches.

Datasets

Flavia Leaf Dataset:

It is one leaf datasets for plant image classification. In this leaf datasets, 32 Species are covered. Leaf Images are named by a 4-digit numeric digit and having an extension of ".jpg". Plant leaf images of the Flavia dataset is given in the following figure

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 498-502
https://publishoa.com
ISSN: 1309-3452

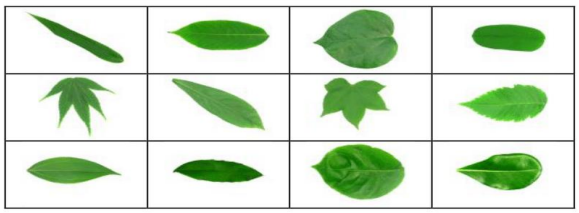


Fig 3: Flavia Leaf Dataset:

Leafsnap Dataset:

It is one of leaf datasets for plant image classification. The Leafsnap Dataset has leaf images .These leaf images are collected from Lab leaf images and Field leaf images. Lab leaf images are captured in bright lighting conditions and also pure white background. These Lab leaf images are having shape. Lab leaf image has 23147 leaf images. These are high-quality resolution. Field dataset has 7716 leaf images. these leaf images are having noise. Leaf Images of the snap dataset is given in the following figure.

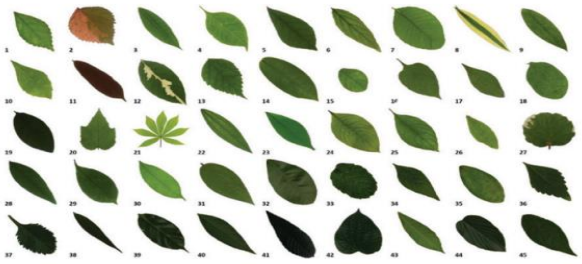


Fig 4: Leafsnap Dataset

Training

In this paper, we have addressed and proposed a hybrid method namely CNN-KNN based Hybrid Method. This method used on two datasets. These datasets are LeafSnap and Flavia. LeafSnap and Flavia datasets and structure of datasets, no.of species, total leaf images and total images (after augmentation) has been shown in below table. Images in the given LeafSnap and Flavia dataset are improved. In the total data, 60% data is being used for training and 40% data is being used for testing.

Dataset Name	Number of Species	No. of images	Total No. of images
LeafSnap dataset	15	50	750
Flavia leaf dataset	33	50	1650

Table 1: Training Dataset Structure

Result Analysis

In this paper, the outcome of CNN-KNN based hybrid method predicts the class of plant. Initially the CNN-KNN based hybrid method is trained by leaf dataset. Next training the system by these leaf dataset, and then the test image (input leaf image) is fed and then the system gives the prediction. CNN + KNN achieved to get maximum accuracy of 98.4% and 96.5%, respectively, in LeafSnap and Flavia datasets.

Dataset	Feature selector	Classifier	Accuracy
LeafSnap	kNN	kNN	78.2
Flavia leaf	kNN	kNN	77.8
LeafSnap	CNN	kNN	98.4
Flavia leaf	CNN	kNN	96.6

Table 2: Performance of dataset with different classifier

In this, through metrics, the CNN-KNN based Hybrid method is calculated

$$Sensitivity = \frac{TP}{TP + FN} * 100$$

$$Specificity = \frac{TN}{TN + FP} * 100$$

$$Error\ Rate = \frac{FP + FN}{TP + TN + FP + FN}$$

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

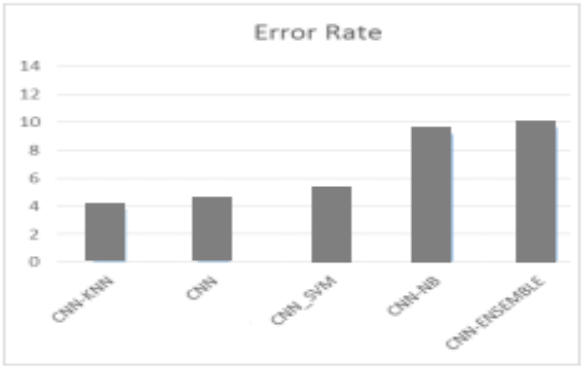
In the following table, the proposed CNN-KNN hybrid method has better performance. Due to least error rate, the CNN-KNN is in top place.

S.NO	MODEL	Sensitivity	Specificity	Error Rate	Accuracy
1	CNN	94.1	97.5	4.1	95.83
2	CNN_SVM	94.1	95.8	5.0	95.00
3	CNN-KNN	95.8	96.7	3.7	96.25

Table 3: Performance of methods

The following figures 5 and 6 shows error rate between various models and accuracy between various models. The CNN-KNN hybrid model is showing well results.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 498-502
https://publishoa.com
ISSN: 1309-3452



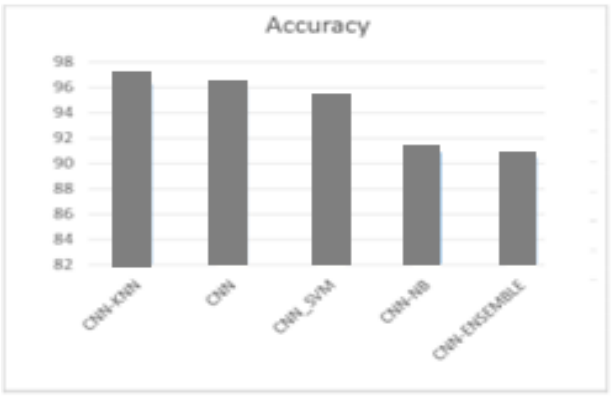
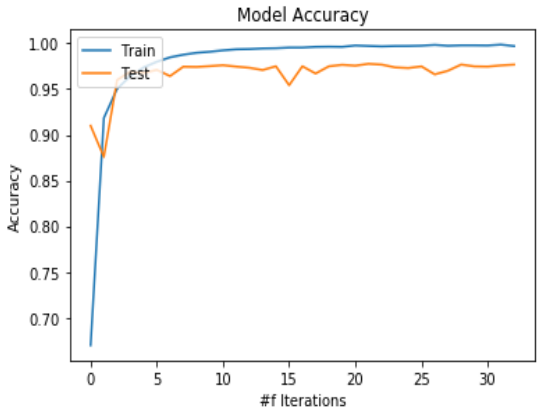


Fig. 5 Error rate between various models



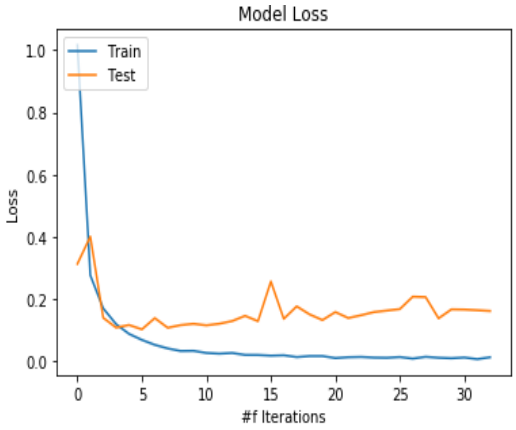


Fig. 6 Accuracy between various models

Proposed model accuracy and loss

Conclusion

In this paper, we have proposed a novel CNN-KNN based hybrid method for enhancing the performance in terms of accuracy in plant classification. We have used CNN for feature selection and also we have used KNN instead of SoftMax for the classification of plant. In this paper, we have increased the amount of training data for improving the performance. The CNN-KNN based hybrid method has been trained and tested on a dataset (13,600 images). On the Flavia dataset, this method for plant classification has achieved 99.4% accuracy. On the LeafSnap dataset, this method for plant classification has achieved 97.4% accuracy. The result outcomes deeply and totally depend on the size of the leaf dataset. When compared with other ML method namely SVM-KNN, the CNN-KNN based hybrid method has given better performance. In this paper, we have also addressed the reliability of the proposed method. We have gathered additional leaf images from various sources to increase the leaf dataset and enhance performance for upcoming research studies. In addition, we can enhance this method to analyse and also diagnosis plant leaves. In future, we will propose and address to conduct better study of the learning process without using labelled leaf images.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 498-502
https://publishoa.com
ISSN: 1309-3452

References

[1] S. S. Chouhan, U. P. Singh, A. Kaul and S. Jain, "A data repository of leaf images: practice towards plant conservation with plant pathology," in Proc. 4th Int. Conf. on Information Systems and Computer Networks (ISCON), Mathura, India: IEEE Access, pp. 700–707, 2019.

[2] A. A. Bharate and M. S. Shirdhonkar, "A review on plant disease detection using image processing," in Proc. Int. Conf. Intelligence Sustain System, Tripura, India: IEEE Access, pp. 103–9, 2017.

[3] Z. Liu, "Hybrid deep learning for plant leaves classification," in Proc. Intelligent Computing Theories and Methodologies, Lecture Notes in Computer Science, Springer, vol. 9226, pp. 115–132, 2015.

[4] A. Kaya, A. S. Keceli, C. Catal, H. Y. Yalic, H. Temucin et al., "Analysis of transfer learning for deep neural network based plant classification models," Computers and Electronics in Agriculture, vol. 158, pp. 20–29, 2019.

[5] S. H. Lee, C. S. Chan and P. Remagnino, "Multi-organ plant classification based on convolutional and recurrent neural networks," IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4287–4301, 2018.

[6] B. Wang and D. Wang, "Plant leaves classification: A Few-shot learning method based on siamese network,"IEEE Access, vol. 7, pp. 151754–151763, 2019.

[7] F. Marzougui, "A deep CNN method for plant disease detection," in Proc. Int. Arab Conf. on Information Technology, Giza, Egypt: IEEE Access, pp. 1–6, 2020.

[8] J. W. Tan, S. W. Chang, "Deep learning for plant species classification using leaf vein" Transactions on Computational Biology and Bioinformatics, vol. 17, 82–90, 2020.