International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 11, November-2019

Identification of Enhanced Green Natural Environmental Plant Species Using Deep Learning

Jakkala Priyanka[1], M.Jaiganesh[2], R.K. Selvakumar[3]

PG Scholar[1], Associate Professor[2], Professor[3]
[1, 2, 3]Department of Computer Science and Engineering,
CVR College of Engineering College, Hyderabad,Telangana, India.

Abstract— Many individuals have been trying to resolve the issue of recognizing weeds for several years. They used a broad variety of methods to identify weeds from ordinary crops with a common goal, but no system has been created that has made a business breakthrough. Botanists and those who study plants however, are able to identify the type of plants at a glance by using the characteristics of leaves but we are interested in identification of plants by using the digital cameras, mobile devices, like techniques of image processing and the pattern recognition. Features are shape; colour and texture have to be studies for the plant identification. This work proposes a good computer vision deep learning technique of convolution neural network consists of four layers, convolution layer, max pooling, dropout, average pooling. The optimum results were accomplished with much less computational effort and it shows the efficiency of the algorithm. We implemented the proposed model by using the Kaggle tool. We discussed the result and the accuracy of our work. We can come out with the 86% accuracy for the plant identification using plant image technique. This paper addresses the technique that is used for the identifying plant using plant leaves. This paper discusses the model for plant identification.

Keywords— Plant species, Deep learning, Pre-processing, Image classification, Identification

I. Introduction

Plant Identification is the determination of the identity of an unknown plant in comparison with previously collected specimen. The process of recognition connects the specimen with a botanical name. Once this connection is established, related details like name and other properties of the plant can be easily obtained. Plant Classification is the placing of known plants into groups of categories to show some relationship. They use features that can be used to group plants into a known hierarchy [1].

In the last few decades, the development of image and video processing algorithms and approaches has received significant attention from the scientific community. This development is a consequence of the increased hardware and sensing capabilities and the increased quality of the digitally recorded materials. The increase of quality was in the recorded materials and the increased computing capabilities have allowed the development of more complex and more powerful approaches.

They used a broad variety of methods to identify weeds from ordinary crops with the common goal, but no system has been created that has made a business breakthrough [2]. There are different reasons why these systems cannot fix the issue. One is the set of information Ji-Xiang Du, Xiao Feng Wang & Guo-Jun Zhang, Department of Automation, University of Science and Technology of China, publishes a study paper on comparable issues identifying crop species by leaf form [11].

Deep learning is a feature of artificial intelligence that intimates the functioning of the human brain in information processing and generating patterns for making decision. [14] Deep learning is a subset of artificial intelligence (AI) machine learning that has networks capable of learning from unstructured or unlabeled information without supervision, and also known as profound neural learning or profound neural network [5]. Deep learning has evolved hand in hand with the digital age, resulting in an explosion of data in all from all regions of the world.

This huge quantity of information is easily available and can be shared by fintech applications such as cloud computing. However, the data, which is usually unstructured, is so vast that human understanding and extracting appropriate information could take centuries. Companies are realizing the incredible potential that can result from unravelling this data huge wealth of and are increasingly adjusting for automated assistance to AI systems [10]. The main task of image classification is acceptance of the input image and the following definition of its class.
The example images of leaf variation depicted in Figure 1 seeding from data set.

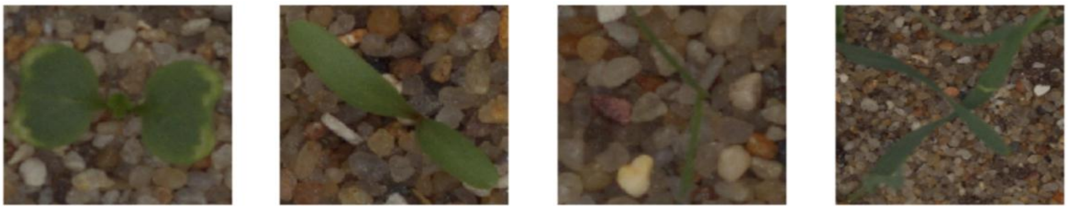

Figure 1 Example of leaf variation

The Computer Vision and Bios stems Signal Processing Group, Department of Engineering, Aarhus University gathered the information and made it accessible to the public free of charge to promote and encourage the creation of species recognition methods for the agricultural industry [12, 13]. The paper organized as follows, Section 3 delivers the system model and explanation of the architecture, Section 4 brings the implementation and results with their discussions.

Finally, it concludes with future enhancement. Major contribution of the paper

1. A novel plant identification scheme is proposed by applying CNN techniques.
2. In this work, CNN based classifier is developed for the seedlings class based on images.
3. The work focuses on identifying the raw image data set to classify plant species and their images.

II. **RELATED WORK**

The development of digital image processing, machine vision and pattern recognition, are the famous and numerous techniques for plant classification. Most of the investigations are performed using leaves. To contribute to these techniques, this paper proposes to develop a classification system using both shape-based and texture based analysis [9]. For each image we calculate the histogram gradient using sobel operator. This histogram provides a description of the relative directions of the main veins. The difference between the gradient histograms is again calculated using the Jeffrey-divergence distance measure. The Sobel method only achieved a rate of 66.1%, it can be seen that though some species are classified more accurately using the contour method, others do much better using the Sobel method [9].

In this paper, the geometrical feature and the texture feature are all extracted and used for plant species identification. It is operated in two phases: Training, Identification [4]. SOM is unsupervised network that exploits the natural structure of input feature space without using any a priori class information. SOM (Self organizing maps) neural network has been widely utilized in the field of pattern recognition.

The architecture of SOM neural network is designed to be a two-layer-based network which includes an input, a competition layer, and every neuron leaf [7]. Previous researches mainly focus on leaf identification by utilizing the shape features of the while ignoring the leaf venation that actually contains important texture information [3]. Plant leaf image passes through the feature extraction to obtain a set of feature values, which are given as input to the trained SOM artificial neural network, the trained SOM neural network will give decision of the query plant species [4].

Identifying plants using flowers and fruits are very time consuming task. Leaves also play an important role in plant identification. Plant leaves have two-dimensional nature and thus they are most suitable for machine processing [8]. Before applying the classification methods leaf image classification can performed. It consists of image acquisition and pre-processing.

K-Nearest Neighbour: Sample should be classified similarly to its surrounding samples. Before applying the classification methods leaf image classification can performed. It consists of image acquisition and pre-processing step. Image acquisition includes plucking leaf from plant and then, the digital color image of the leaf is taken with a digital camera .It could be predicted by considering the classification of its k nearest neighbour samples. Given an unknown sample and a training set, all the distances between the unknown sample and all the samples in the training set can be computed. [6]

Before applying the classification, methods leaf image classification is e performed. It consists of image acquisition and pre-processing procedures [15]. Image acquisition includes plucking leaf from plant and then, the digital color image of the leaf is taken with a digital camera. In cooperation with the University of Southern Denmark, the Aarhus University Signal Processing Group recently published a dataset comprising pictures of about 960 distinctive crops belonging to 12 species at several development phases. As shown in Figure 2, the CNN system composition model gets image input with RGB format.

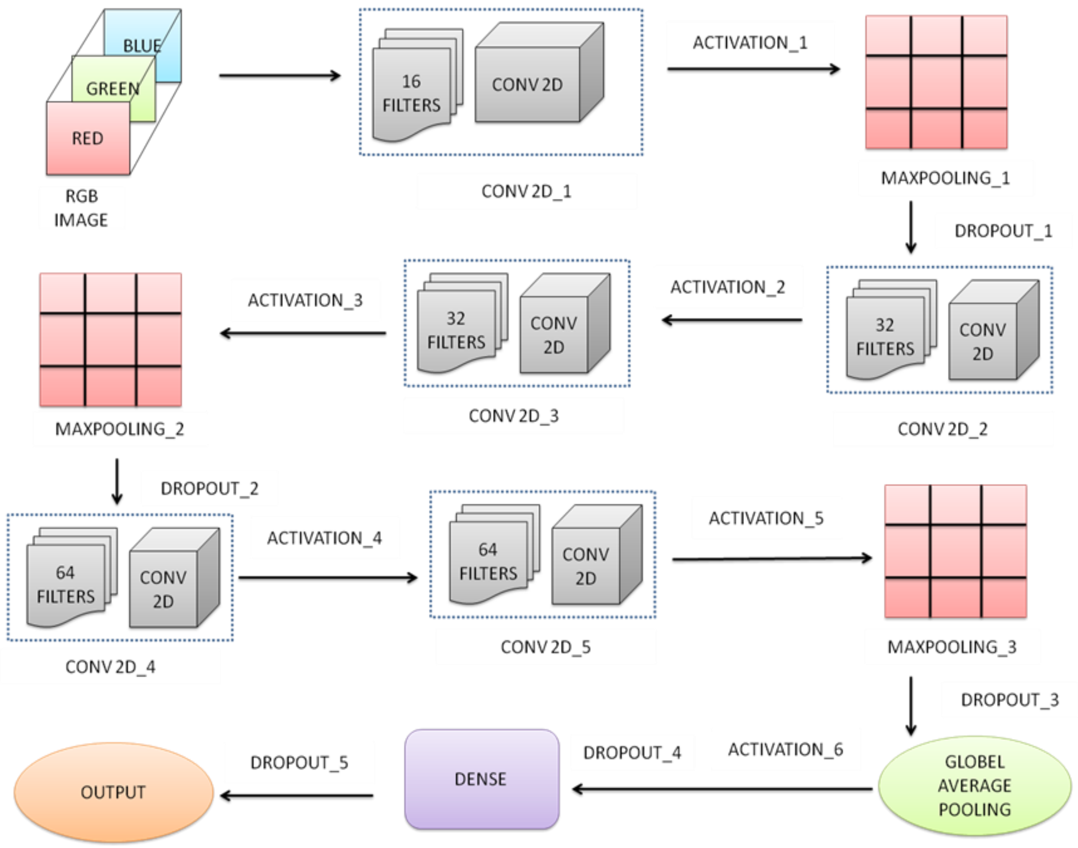

Figure 2 Plant Identification system composition model

The input image followed to perform the CNN operations called Input /Output layering, pooling and dropout operations shown in Figure 2,

Input Layer: From Input image, Conv2d Layer that accepts a (1*256*256*3) shaped image as an input. This layer has 16 filters and a 5*5 kernel size.

Convolution2D: convolution operation is performed by sliding the filter over input at every location by performing the element wise matrix multiplication and sum the result. The sum goes to the feature map.

$$\text{Output width} = \frac{w - F_W + 2P}{S_W} \quad 3\text{-}1$$

$$\text{Output height} = \frac{H - F_H + 2P}{S_H} \quad 3\text{-}2$$

In the above formula

W= Width of input image,

F_W = filter width,

S_W=Horizontal stride,

P=Padding

H=Height of input image,

F_H = Filter height,S_H=Vertical stride.

Stride mean number of pixels with which we slide our filter.

Maxpool2D: Max-pooling layer is used to decrease the dimensions (only width and height) of the image by the pooling size parameter. For, example the input was in size 256*256*3 and if the pool size is 2 then, the output size would be 128*128*3. It is applied in all Max-pool layers, the pool size is 2.So, after this layer, the width and height of the image are halved

Dropout: This layer is used as a regularization method to prevent over fitting. It drops the nodes at random when calculating forward and backward propagations. The dropout probability can be given as a parameter. In this architecture, we set probability of 0.2 to dropout nodes after several layer.

Output Layer: This layer is a dense layer that outputs class probabilities for 12 classes. This layer has ‘soft max’ activation function

which we are working .we conduct a sequence of convolution and pooling operations followed by several layers that they are fully attached. If we classify multiclass, the output is SoftMax [8]. The primary benefit of CNN is that it detects the significant characteristics automatically without any human supervision.

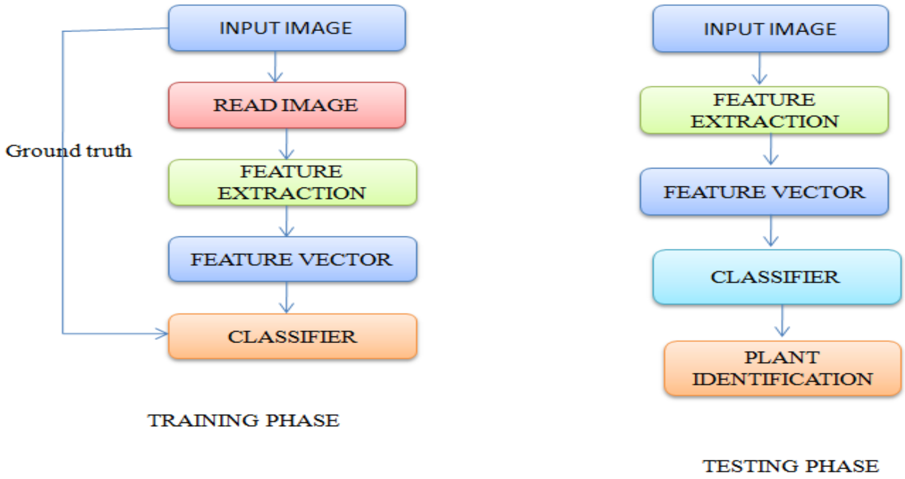

Figure 3 Plant identification CNN data flow diagram

Figure 3 represents the data flow of Plant identification model. Here first inserted images are inserted into the image database, where it read the image from database which extracts the feature from the feature vector. A successful plant identification system requires a set of plant features that best describe the plant image and which can provide maximum discrimination between different plants.

The classifier used to preprocess the image this step consists of two main tasks namely enhancement and segmentation. Enhancement techniques also include operations that can improve leaf image properties which help to increase the overall performance of the identification system. Enhancement techniques also include operations that can improve leaf image properties which help to increase the overall performance of the identification system. In segmentation, this subdivides the input leaf image into various parts of meaningful entities.

Segmentation techniques use the various features extracted like gray or color features or texture features to separate the various regions of the input image. The quality of a feature vector is related to its ability to discriminate examples from different classes. The system performs training and testing the image, that predicts the image by using CNN model.

III. Implementation and Results

Kaggle is an online community of data scientists and machine learners owned by Google LLC. Kaggle enables users to discover and publish data sets, explore and build models in a web-based data science setting, work with other data scientists and machine learning technicians, and join contests to address data science problems [13].
Kaggle allows information researchers and other developers to participate in competitions for machine learning write and share code and host information sets. Experimental setup of Kaggle includes create your own account in the portal following by data set uploading. In dataset uploading, we preferred plant seedling classifications then create a new kernel and upload note book. Note book is compressed file of our data set. We perform commit operation to initiate our classification. We add new kernel along with note book entity. We set as Graphical Processing Unit (GPU) and select programming platform as python. This Table1 shows the validation accuracy, validation loss, training accuracy and training loss

TABLE I

No. of epochs	Training Loss	Training accuracy	Validation Loss	Validation accuracy
1	2.37	0.17	2.42	0.23
2	2.09	0.26	2.06	0.25
3	1.99	0.28	1.99	0.27
4	1.92	0.31	1.66	0.39
5	1.81	0.35	1.43	0.51
6	1.63	0.42	1.63	0.42
7	1.50	0.49	1.33	0.53
8	1.39	0.53	1.23	0.57
9	1.45	0.54	1.11	0.59
10	1.21	0.58	1.41	0.50
11	1.18	0.60	1.36	0.54
12	1.11	0.63	0.72	0.76

13	0.98	0.66	1.04	0.62
14	0.91	0.68	0.68	0.76
15	0.88	0.70	0.62	0.77
16	0.76	0.74	0.68	0.74
17	0.78	0.74	0.63	0.79

Table 1: Loss and Accuracy table

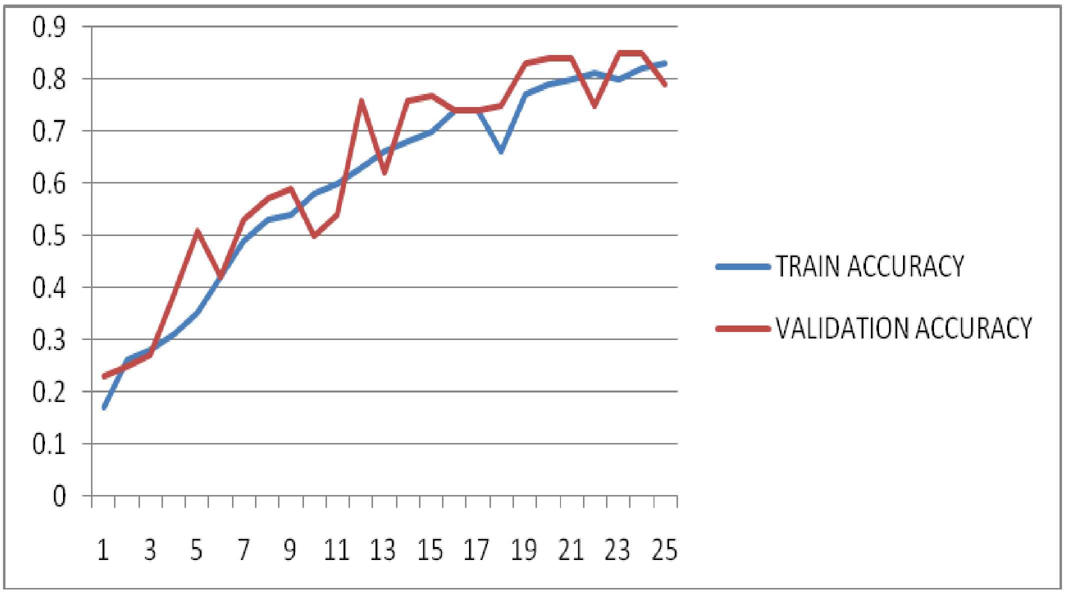

Figure 4 Training accuracy and validation Accuracy graph

Above figure 4 shows training accuracy, validation accuracy, which plotted in Table1.To train the model, the training set is used, while the validation set is only used to assess the output of the model. Let us consider that our model accuracy 82% on training set the validation accuracy is 86% then our model accuracy expect is 86%.

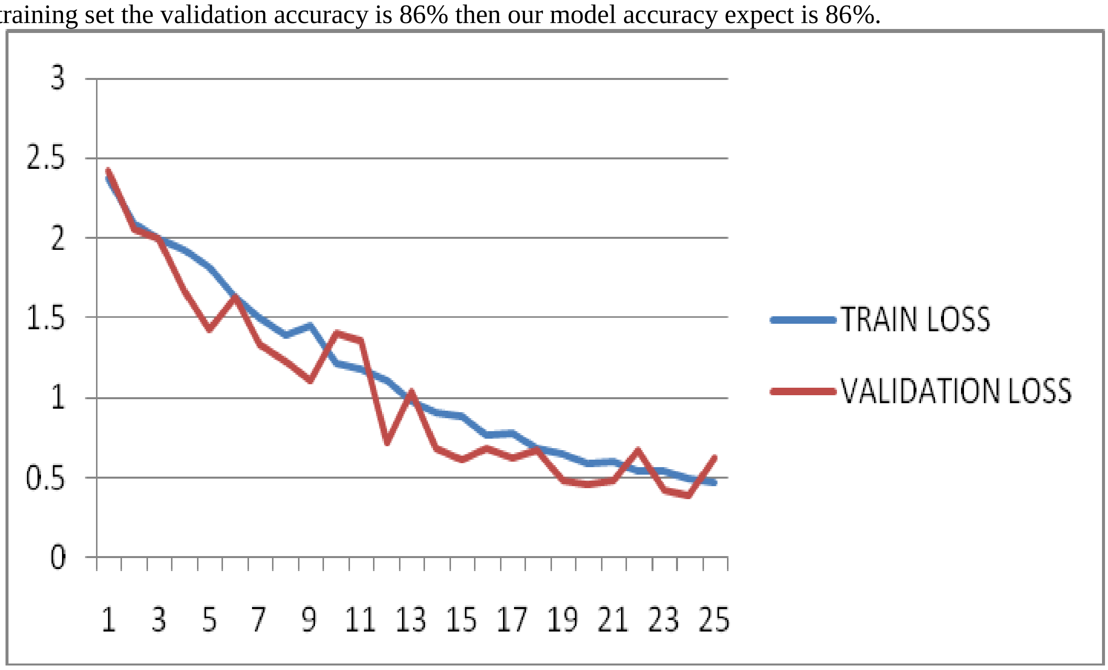

Figure 5 Training loss and validation loss graph

Above figure 5 shows the training loss, validation loss, which plotted in Table1.To train the model, the training set is used, while the validation set is only used to assess the output of the model. The training loss is the error that is calculated on the training set. The validation loss is calculated on fresh information for a model trained only on the training set. Training loss must be less than validation then the model is over fitting. Figure 6 shows the list of species used into the proposed model and their names.

Overfitting is statistical with more parameters and trains the data well. In over fitting process we perform many functions like add more data, data is augmented, batch normalization, training and testing, L1and L2 regularization, and dropout function.

S.NO	IMAGE	PLANT NAME
1		Common chickweed
2		Loose-silky bent
3		Charlock
4		Shepherds purse

Figure 6 Plant species identification

IV. CONCLUSIONS

In this paper, we proposed a plant species identification using convolution neural network model. The process of CNN is done through convolution, maximum pooling and global average pooling. As the further improvement to our model, we can build a model by over or under sampling the data because of the irregular count in images and build a more generalized model. We can also improve the accuracy by doing some preprocessing techniques such as image segmentation, masking etc. Finally, the model we built can be used for real-time identification of plants by the farmers if a mobile application is built on this model.

REFERENCES

[1] Jana Waldchen,Michael Rzanny, et.al.," Automated plant species identification Trends and future directions", PLOS Computational Biology (2018):1-19

[2] Jana Wäldchen, Patrick Mäder.: "Plant Species Identifcation Using Computer Vision Techniques: A Systematic Literature Review", Arch Computat Methods Eng (2018) 25:507–543

[3] Ji-Xiang Du, Xiao-Feng Wang, "Shape Matching and Recognition Base on Genetic Algorithm and Application to Plant Species Identification". Springer-Verlag Berlin Heidelberg, ICIC 2005, Part I, LNCS 3644, 282 – 290,

[4] Lei Zhang , Jun Kong, Xiaoyun Zeng, Jiayue Ren.: "Plant Species Identification Based on Neural Network", Fourth International Conference on Natural Computation (2018) :90-94

[5] Mostafa Mehdipour Ghazia,Berrin Yanikoglua, Erchan Aptoulab.: "Plant identification using deep neural networks via optimization of transfer learning parameters", Neurocomputing 235 (2017) :228–235

[6] Saguna Kailas Ingle, Vijayalakshmi Kandroli.:"plant species identification using leaf image and classification methods", International Journal Of Advance Research In Science And Engineering, Vol. No.4, Issue No.02, February (2015): 276-279

[7] Shulin Dave, Student Member, IEEE, Ken Runtz.: "Image Processing Methods for Identifying Species of Plants", IEEE WESCANEX '95 PROCEEDINGS ,403-409

[8] Surleen Kaur1, Prabhpreet Kaur.: "Plant Species Identification based on Plant Leaf Using Computer Vision and Machine Learning Techniques",Journal of Multimedia Information System VOL. 6, NO. 2, June 2019 (pp. 49-60): ISSN 2383-7632

[9] Thibaut Beghin, James S. Cope, Paolo Remagnino, and Sarah Barman.: "Shape and Texture Based Plant Leaf Classification", sci-hub.tw/10.1007/978-3-642-17691-3_32 (2010): 345-353.

[10] Vinit Bodhwania , D. P. Acharjyaa , Umesh Bodhwani.: "Deep Residual Networks for Plant Identification" , Procedia Computer Science 152 (2019) 186–194

[11] X.-F. Wang, J.-X. Du, and G.-J. Zhang.:"Recognition of Leaf Images Based on Shape Features Using a Hypersphere Classifier", Springer-Verlag Berlin Heidelberg 2005, pp. 87–96

[12] Yu Sun, Yuan Liu, Guan Wang, and Haiyan Zhang.: "Deep Learning for Plant Identification in Natural Environment", Computational Intelligence and Neuroscience Volume 2017, Article ID 7361042, 1-6.

[13] https://en.wikipedia.org/wiki/Kaggle

[14] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[15] https://www.investopedia.com/terms/d/deep-learning.asp