Optimized Sequential model for Plant Recognition in Keras

Shilpi Aggarwal[1], Madhulika Bhatia[2], Rosy Madaan[3], Hari Mohan Pandey[4]

[1] Faculty of Engineering and Technology, ManavRachna International Institute of Research and Studies,Faridabad, (Haryana), India

[2]Associate Professor, Department of Computer Science Engineering, Amity University, Noida, (Uttar Pradesh), India

[3]Associate Professor, Faculty of Engineering and Technology, ManavRachna International Institute of Research and Studies,Faridabad,(Haryana), India

[4]Department of Computer Science, Edge Hill University, Lancashire, United Kingdom

shilpi.16.g@gmail.com

Abstract. There are huge varieties of floras in the world. Lots of varieties of species are beneficial to the human's life. Plant recognition is a very important task to segregate the huge amount of floras which belong to various categories. Various researchers have applied different approaches to recognize the plant family. Deep learning is a subset of machine learning. This is one of the accepted technologies that automatically extracts features, processes them and yields the best results. Keras is a widely used deep learning framework which is employed in this work. Five different plant species are chosen as samples among the Indian species, namely OcimumTenuiflorum, Sansevieriatrifasciata, Chlorophytumcomosum, Azadirachtaindica, Aloe Vera. These samples are rich in oxygen. From these samples the features like shape, color, texture, corners are extracted. One hot encoding is also applied onto the target values to optimize the results of recognition. The extracted features are fed into the sequential keras model which recognizes the plant species. The accuracy of the training set is 100 percent and the testing set is 96.7percent.Confusion matrix is drawn to show the correctly classified and misclassified samples.

Keywords:Plant Recognition, Keras, Oxygen, Plants.

1. Introduction

Lots of machine learning [1] and deep learning algorithms [2-5] were proposed by the researchers to classify and recognize the several plant species present on the earth. As there are total 391,000 species known to biotic department of the world. The approaches like classification and recognition can be done on plants to separatethem into categories. Shilpi A et al. (2020) proposed the classification of oxygen rich plants based on morphological properties with the help of Random forest classifier [6].

Plant recognition can be done by considering the plant samples as images. There are various image processing applications in different areas like remote sensing, medical field [7-13] pattern recognition, astronomy and many more.

Guillermo L et al. (2016) proposed a deep learning model for plant recognition using morphological patterns. He analyzes the model by simple visualizing techniques [14]. Pacifico, L. et al. (2019) applied different classification algorithms to recognize the plant species based on texture, color and shape features [15].The observed gaps in the work done by the researchers is that the dataset belongs to normal plant species. As far as the plant recognition is concerned, it must be beneficial to the mankind. That's why in this work the dataset is self-collected plant images which are having most significant importance to the human's life. This dataset is chosen due to the extremely rise in pollution. All the plants are selected due to their antioxidant properties, absorbent of harmful gases (carbon monoxide, xylene, formaldehyde) present in the atmosphere. The variance on plants due to pollution can be shown by SSIM Technique[16].

Contribution of this work is as follows:

- The features (shape, color,texture,corners) are extracted from the 150 sample images(OcimumTenuiflorum, Sansevieriatrifasciata, Chlorophytumcomosum,Azadirachtaindicaand Aloe Vera) by using image processing techniques in MATLAB 2019a.
- Then these features are inserted into the keras sequential model to recognize the plant species belongs to 5 different categories.
- These five different species are very precious to the environment due to their positive properties.

The workflow of this paper is divided into several sections. First section consists of introduction and the background study. Second section comprises of materials and methods required for the work implementation. Third section contains the results and analysis of the model. At last, conclusion winds up the paper.

2. Materials and Methods

2.1. Dataset

The image dataset is taken from the oxygen rich plants which are collected from the Indian species as shown in Table 1.The details of the samples are shown in Table 2. Due to easy availability of these sample species and various beneficial remarks of these samples attracted to put into research. Holy Basil is having antioxidant effects. SansevieriaTrifasciata suck up the carbon monoxide from the atmosphere.

Table 1 Names of the Samples

Sample No.	Scientific Names	Common Names
1	OcimumTenuiflorum	Holy Basil
2	SansevieriaTrifasciata	Snake
3	ChlorophytumComosum	Spider
4	AzadirachtaIndica	Neem
5	Aloe Vera	Indian Aloe

Table 2 Information about no. of samples and capturing device.

Total numbers of samples	ImageAcquisition Device
150 RGB images	Digital camera having 12.1 Megapixel resolutions of 4000 x

3000 pixels.

ChlorophytumComosum also decontaminates the air and absorbs carbon monoxide and formaldehyde and xylene from the air. AzadirachtaIndica being a medicinal plant, used in heal many problems like liver problems, loss of appetite, gingivitis etc. Aloe Vera is having antioxidant and antibacterial properties. It protects from ultra violet rays.

2.2 *FeatureExtraction*

The features like shape color, texture, corners are extracted from the image samples as shown in Figure 1. These features are drawn with help of image processing techniques by using MATLAB 2019a. First, consider the texture feature. This belongs to the fluctuation in the intensities of gray level. The method called gray level co-occurrence matrix (GLCM) is used. 16 GLCM is created. Statistics like Contrast, homogeneity, correlation, energy gets calculated. Then the average value is taken into consideration. Secondly, the shape dimensions namely area, perimeter, minor axis length major axis length are calculated using the algorithm[17].Thirdly, Maximum of Red, green and Blue Color values are calculated from the RGB image samples. Fourthly, corners and valid corners are calculated by using Harris- Stevens algorithm [18].

2.3 One Hot Encoding

Pre-processing can be done by encoding. Generally, encoding schemes are used in case of categorical data. After encoding, the data converts in the form of 0's and 1's. In this work, one hot encoding scheme is applied on the target values. These values are the five distinct classes of sample species. As the encoding procedure comes to an end, the species belong to first category takes the form of [1 0 0 0 0], second category's form is [0 1 0 0 0] and so on. Here the position of 1's shows the class of species.

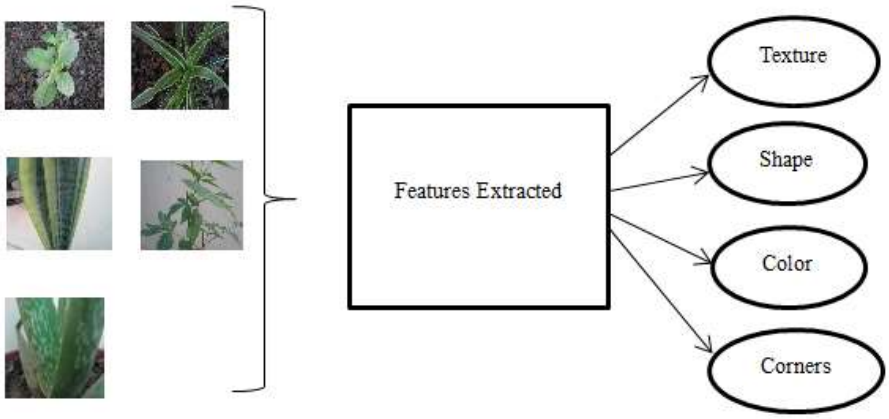

Figure 1 Feature extraction of input samples images belongs to five different species.

2.4 Proposed Keras Sequential Model

Keras is predominantly used framework of deep learning .Basically, various layers combined to form a model.A Sequential Model is implemented in keras which recognizes the plant belong to different five categories. The flow of the model is shown in Figure 2. The features are fed into the fully connected layer that is the dense layer. There are two dense layers and after that that sigmoid layer is there.

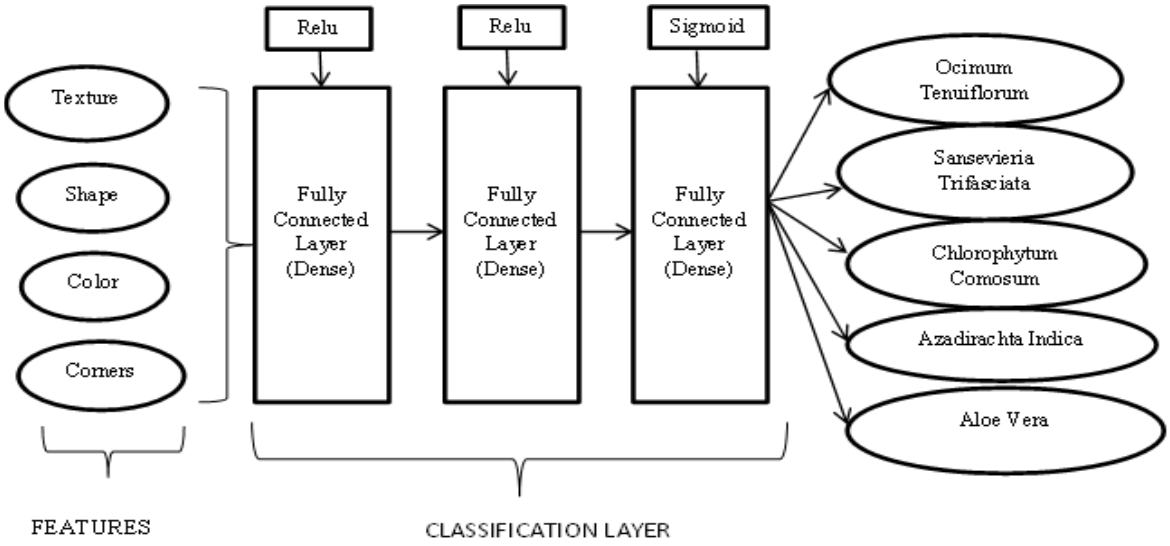

Figure 2 Layers included in Sequential Keras Model

There is sequence of steps that has been followed to achieve the aim are shown in Figure 3. As explained in the figure 3, starting from importing the required libraries like Keras, pandas, numpy, sklearn, matplotlib then read the features data collected from the input image samples. Then apply one hot encoding on the target values. After dividing the dataset into independent and dependent attributes, split the attributes into training data and test data. Further scale the features and apply the sequential model in keras. After that add two dense layers and one sigmoid layer. Then optimize the model by using Stochastic Gradient Decent. Then fit the model with epochs value to 500. The training and testing accuracy is evaluated as 100%.

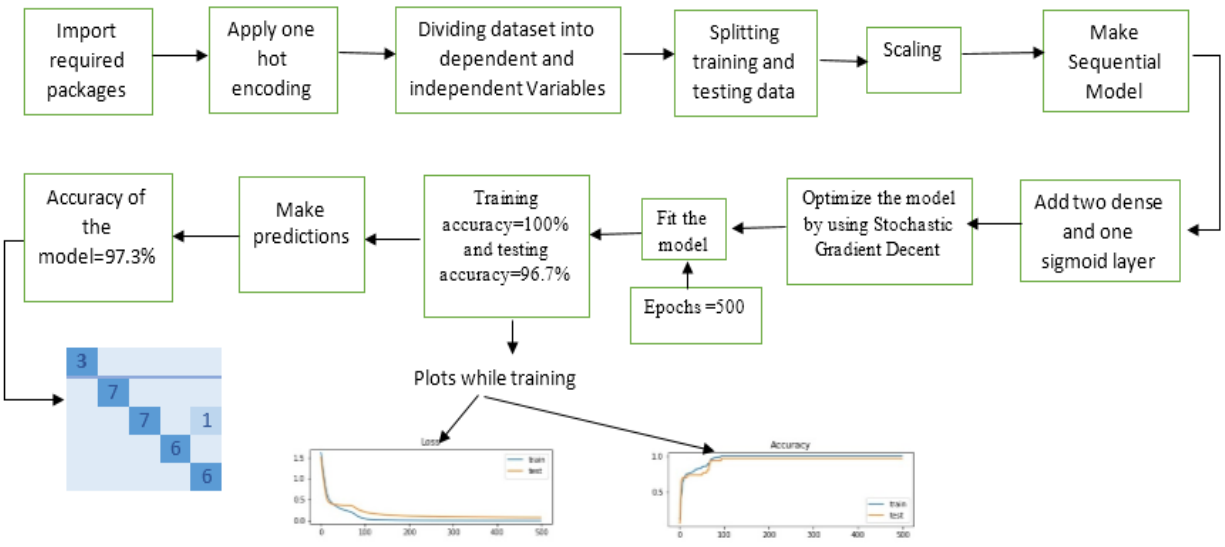

Figure 3 Working flow of the Model.

and 96.7%.After making predictions on the model,the average accuracy is found to be 97.3%.

3. Results and Analysis

Deep learning's mostly used framework is keras. Keras is a high level API to train and build models with full of flexibility and high performance. Guillermo L et al. (2016) classify the plants on the basis of morphological properties. He used simple visualization techniques for analyzation of model [14]. Pacifico, L. et al. (2019) applied different classification approaches to classify the plants based on the shape, color, and texture [15].

In this research the features like texture (Contrast, homogeneity, correlation, and energy), color, corners, valid corners and shape (area, perimeter, minor axis length major axis length) are evaluated. Further, the features size of (150, 14) is fed into the Keras sequential model. Three dense layers are added. First dense layer is added with 12 output dimension and 14 as an input dimension with *Relu* as an activation function and *he_uniform*as thekernel initializer. Second dense layer is added with 14 output dimension *Relu* as an activation function and *he_uniform*as thekernel initializer. In the third layer, 5 value is for output dimensions and *sigmoid* as an activation function. The model summary is shown in Figure 4. After defining the model, optimized with Stochastic Gradient Decent with learning rate of 0.01 and momentum of 0.9.Then compile the model with attributes loss as *categorical_crossentropy,* optimizer and metrics chosen is accuracy. Now fit the model with training and testing data and epochs value set to 500.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 12)                180
_________________________________________________________________
dense_2 (Dense)              (None, 14)                182
_________________________________________________________________
dense_3 (Dense)              (None, 5)                 75
=================================================================
Total params: 437
Trainable params: 437
Non-trainable params: 0
_________________________________________________________________
```

Figure 4 Model Summary

Training accuracy and testing accuracy is obtained as 100% and 96.7 % respectively. The plots of loss and accuracy while training is shown in Figure 5 and 6.Confusion matrix of the result is shown as in Figure 7.The average accuracy of the model is 97.2% while the 500 epochs as shown in Figure 8.

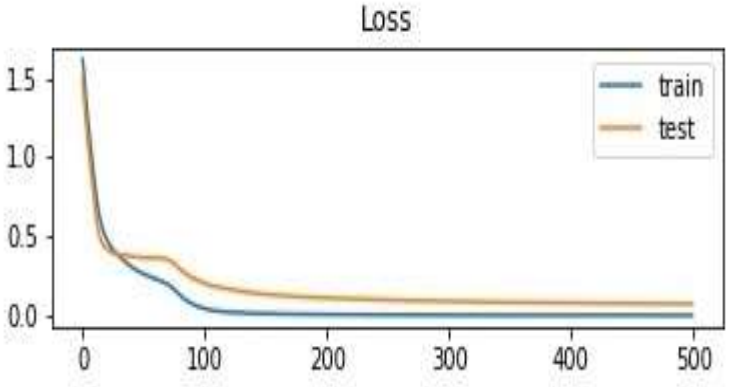

Figure 5 Loss while training

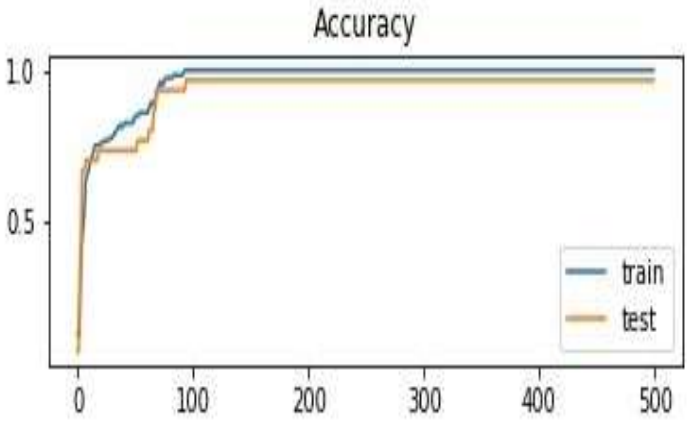

Figure 6 accuracy while training

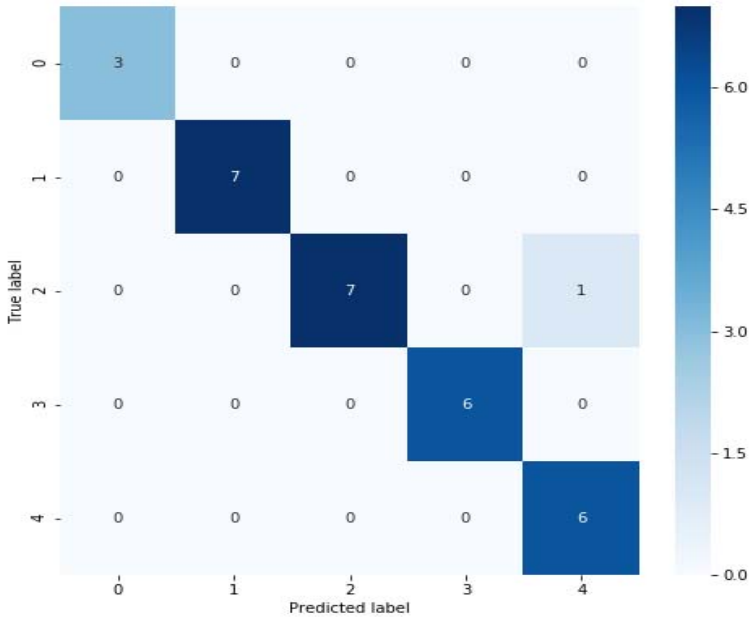

Figure 7 Confusion matrix.

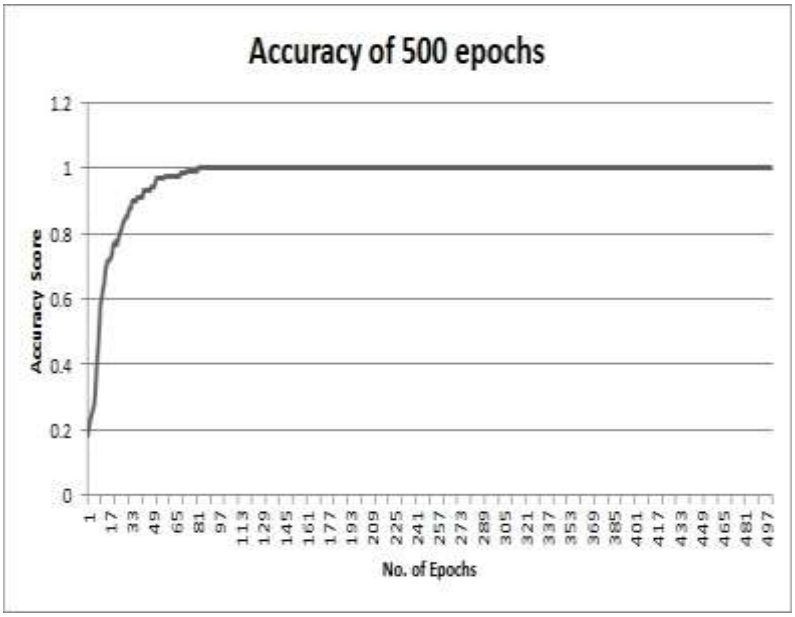

Figure 8 Plot of accuracies while 500 epochs.

4. Conclusion

The main purpose of plant recognition should be for the well-being of the ecosystem. So in this work we have tested the dataset which is rich in oxygen and absorbs harmful gases from the atmosphere. We have applies a sequential model in keras and tested over 500 epochs and achieved accuracy of 100% for the training dataset and 96.7% for the testing dataset.

References

[1] Aggarwal, S., & Bhatia, M. 2019. Anatomy of Leaf Classification Techniques. *Int. Conf. on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* .510-13. IEEE.

[2] Bhatia, M., & Mittal, M. 2017. Big data & deep data: minding the challenges. *Deep Learning for image processing Applications*, 31, p.177.

[3] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., & Garcia-Rodriguez, J. 2018. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, **70**, 41-65.

[4] Anubha Pearline, S., Sathiesh Kumar, V., & Harini, S. 2019. A study on plant recognition using conventional image processing and deep learning approaches. *J. of Intelligent & Fuzzy Systems*, **36(3)**, 1997-2004.

[5] Zhu, Y., Sun, W., Cao, X., Wang, C., Wu, D., Yang, Y., & Ye, N. 2019. TA-CNN: Two-way attention models in deep convolutional neural network for plant recognition. *Neurocomputing*, **365**, 191-200.

[6] Aggarwal, S., Madaan, R.,& Bhatia, M.2020Morphological based Optimized Random Forest classification for Indian Oxygen Plants.*Int. J. on Emerging Technologies* **11(3),**707-14.

[7] Pandey, M., Bhatia, M., & Bansal, A. 2016. IRIS based human identification: analogizing and exploiting PSNR and MSE techniques using MATLAB. In *Int.Conf. on innovation and challenges in cyber security (ICICCS-INBUSH)*, 231-35. IEEE.

[8] Bansal, A. 2013. Implementing edge detection for detecting neurons from brain to identify emotions. *Int. J. of Computer Applications*, **61(9)**.

[9] Bhatia, M., Bansal, A., & Yadav, D. 2017. A proposed quantitative approach to classify brain MRI. *Int. J. of System Assurance Engineering and Management*, **8(2)**, 577-84.

[10] Madaan, A., Bhatia, M., & Hooda, M. 2018. Implementation of image compression and

cryptography on fractal images. In *Advances in Data and Information Sciences* .49-61. Springer, Singapore.

[11] Bhatia, M., Bansal, A., Yadav, D., & Gupta, P. 2015. Proposed algorithm to blotch grey matter from tumored and non tumored brain MRI images. *Indian J. of science and Technology*, **8(17)**.

[12] Pandey, M. 2016. An amalgamated strategy for iris recognition employing neural network and hamming distance. *In Information Systems Design and Intelligent Applications* .739-47. Springer, New Delhi.

[13] Pandey, M., Bhatia, M., & Bansal, A. 2016, February. An anatomization of noise removal techniques on medical images. In *Int. Conf. on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*. 224-29. IEEE.

[14] Grinblat, G.L., Uzal, L.C., Larese, M.G. and Granitto, P.M., 2016. Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, **127**, 418-24.

[15] Pacifico, L., Britto, L., Oliveira, E., & Ludermir, T. 2019, October. Automatic Classification of Medicinal Plant Species Based on Color and Texture Features. In 2019 8th *Brazilian Conf. on Intelligent Systems (BRACIS)* ,741-46. IEEE.

[16] Aggarwal, S.,Bhatia, M.,Pandey, H. M. & Madaan, R. 2020Envisaging Variance amid Indian Floras owed to contaminates via SSIM Technique,*Indian J. of Env.Protection*,In Press.

[17] Aggarwal, S., Bhatia, M., & Pandey, H. M. 2020. A New Method to Classify Leaves Using Data Visualization: Spreading Awareness about Global Warming. In *Proc. of ICETIT 2019* Springer, Cham.p. 1129-39.

[18] Harris, C. G., & Stephens, M. 1988, August. A combined corner and edge detector. In *Alvey vision conf.* ,15(50), p.10-5244.