Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Comparative Disease Prediction Performance Analysis for Indian Brinjal Plant

[1] Balwant Gorad, [2] Sirbi Kotrappa

[1] Research Scholar, CSE, Dr. MSSCET, Belagavi, Karnataka & Lead AI Developer, Fittr, Pune MH, IN
[2] Professor, CSE, KLE Dr. MSSCET Belagavi, Karnataka, IN
[1]goradbj@gmail.com ,[2] kotrappa06@gmail.com

Abstract

As per NCRB Government of India's report on farmers, around 35 famers die every day. There are many reasons for this. Some of the major causes are low crop production after many efforts, increased cost of cultivation and management, debt on farmer, low crop prices and many other environment issues. Brinjal is one of most cultivated crop in major part of India. One of the major reasons of above issues is improper disease prediction and management on the crop. This paper presents comparative analysis of disease prediction on Indian Brinjal plant using various deep learning techniques. Deep learning has performed outstandingly in image classification problems. Various deep convolutional neural network (DCNN) models include VGG16, VGG-19, ResNet50, ResNet101, InceptionV3 and Dense Net 121 are implemented to predict the disease on Brinjal plant. The high-quality preprocessed dataset collected from real field with data augmentation of 40,336 images is used to conduct this research work. These implemented models are achieved the training accuracy of 90.92%, 90.69%, 45.80%, 52.21%, 89.17% and 92.35% respectively whereas validation accuracy of 97.01%, 99.66%, 45.92%, 55.86%, 95.10% and 97.89% respectively. These accuracy results showed that the DenseNet121, VGG19, VGG16 and InceptionV3 deep convolutional neural network models are promising way for effective implementation of such disease prediction systems in real time agricultural field.

Keywords: Deep Learning, Convolutional Neural Network, Disease prediction, Agriculture

Introduction

Deep convolutional neural networks have performed exceptionally well in many computing tasks, such as supervised, unsupervised, and semi-supervised learning. But unfortunately, many areas of research do not have a good quality dataset with a huge volume. One of these areas is Indian agriculture. Quality farming is based on collecting comprehensive information about both the environment and the plants. Information from the environment such as soil moisture, temperature, wind speed, precipitation and humidity must be collected and analyzed. While in the case of plants information such as plant diseases, plant growth and pests must be analyzed and monitored. Currently, many researchers working in the field of machine and deep learning use datasets available on the Internet, so satisfactory work is not being done in these fields due to the limited size of the datasets. To build a machine or deep learning model, one of the main tasks is data acquisition; it is the pre-processing and preparation of a good quality dataset to create models. If we don't have good data, these models won't have good performance metrics. Many well-known researchers in these fields suggest that it is always best to have a good quality pre-processed dataset before building the model. In the case of a set of plant data available on the Internet from different geographic locations and experiments are conducted or the deployment of research models are conducted in different geographic locations, leading to an impractical way due to environmental factors, soil properties, types of diseases and many other factors. This experiment is based on the research we did for the Indian Brinjal. In the first part of this work a good quality preprocessed dataset with nearly 40,336 images in size are generated. In our study, the data set used to generate a high-quality, high-volume is collected manually from various farms in the western region of Maharashtra and northern region of Karnataka. Data augmentation consists of several techniques that improve the size and quality of training datasets so that better deep learning models can be built. Since deep convolutional neural networks perform well in the task of image classification, several CNN models are implemented as part of this research. These models include VGG16, VGG19, ResNet50, ResNet101, InceptionV3/Google Net and Dense Net 121. Finally, the result of these research experiments is to have a comparative analysis of the performance of the above-mentioned architectures on a huge data set.

Literature Survey

Various research papers have been studied on machine learning algorithms on image classifications. This review is broadly classified into two areas - some of them related to algorithmic techniques used for image classifications and other study gives various types of diseases on Brinjal plant

Table 1 Literature Summary

Paper title	Journal / Conference	Published Year	Algorithms / Techniques	Gap Found
Real time detection of apple leaf diseases using DL approach	IEEE – advanced optical imaging for extreme	May 2019	CNN, DL	Work done in China on apple and Accuracy received is 78%

based on Improved Convolutional Neural Network	environment			
Deep learning models for plant disease detection and diagnosis	Elsevier – CEA	Feb 2018	Convolutional Neural Network (CNN)	Used Available data-sets from greece and does not focus on Indian crop (80%)
A robust deep learning based detector for real time tomato plant diseases and pest recognition	MDPI – Open Access journal, Switzerland	Sept 2017	Region Based faster CNN (R-CNN), VGG net	Work done on tomato plant species in korea
Automatic plant disease diagnosis using mobile capture devices applied on wheat use case	Elsevier – CEA	June 2017	Image Segmentation, Feature Extraction (AuC metric)	Image processing techniques are used, Analysis carried out in spain and germany using 7 cameras and 3500 images
Detection of plant leaf diseases using image segmentation and soft computing	Science Direct – Elsevier, Information Processing in Agriculture, China	June 2017	Image segmentation and genetic algorithms in soft computing	Done in the china on banana, beans, lemon and rose Environmental factors were different

techniques	Agriculture University			
Plantix- The App to detect and predict crop diseases	https://plantix.net/en	Jan 2015	CNN, VGG16, SVM	Its German based agtech Startup currently available in Germany, Brazil and India. Real time plant datasets are used for some plants

After studying and analyzing above research papers, used algorithms, techniques and their drawbacks it has been observed that, there is need of experimenting various deep convolutional neural networks architectures on this dataset. The gap between works mentioned in the research paper and farmers actual exceptions still not met. So this is the reason which motivated, to carry research in this area.

System Architecture

The steps that are encountered in image classification are depicted in following figure. There are 6 major diseases are found on the eggplant. These are Alternia leaf Spot, Cersocpera Leaf Spot, Damping off, Collar rot, Tobacco mosaic virus and Bacterial Wilt.

In this architecture 7 input classes (6 unhealthy + 1 healthy) are given to train the various models and validation is done with these 7 classes. The sequence of convolution layers and pooling layers followed by fully connected layer is been used to build this architecture.

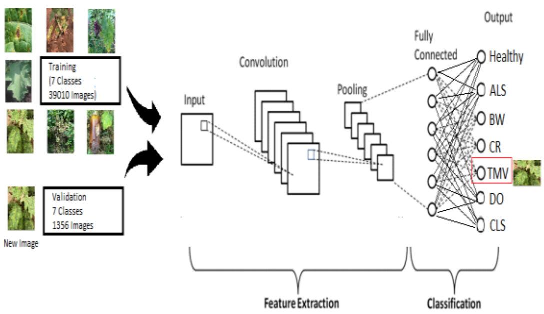


Figure 1 System Architecture

Let us understand how image is stored into pixel values. We have considered one of the healthy Brinjal image sample to understand how it is stored and processed in the computer.

Figure 2 Healthy Brinjal image represented in pixel values

Convolution is mathematically denoted as asterisk (*) sign. Input image is represented as X and filter or kernel is represented as f. Then expression would be

Z = X * f

Dimension of image = (n, n)

Dimension of filter = (f, f)

Dimension of output will be ((n-f+1), (n-f+1))

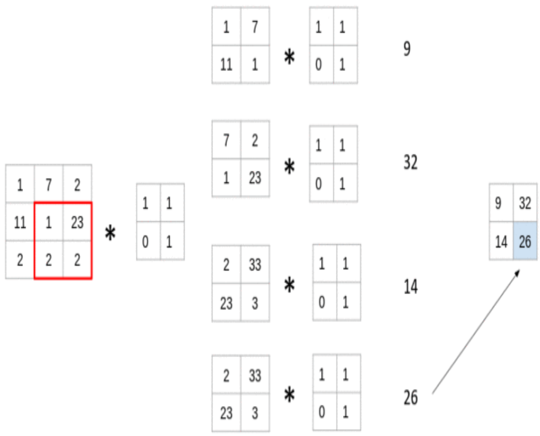


Figure 3 Convolution operation mathematical model

The equation of linear transformation is

Z = WT.X + b

Deep learning classifier involves five measure steps- dataset acquisition, preprocessing, model building, model evaluating and improving model. Following five sections elaborates each step-in brief.

Data Acquisition

In our research we need a training data set in image formats. In case data set is available in videos, video to image conversion is recommended. The sample data set used in this research for training and is manually collected from a farm in Athani Karnataka, India and Solapur, Maharashtra, India

Data Preprocessing

This technique includes cleaning, preparing and manipulating data in such way that it gets ready to apply these datasets to classifier.

Clean Dataset

The technique used to clean the data for a classifier is removed unwanted images from your dataset. Also, during dataset collection utmost care has taken about which image has to take for training and which image has to ignore. This way is useful to collect quality datasets.

Prepare Data

There are a number of preprocessing steps we might want to do before using dataset in the deep learning project.

Uniform aspect ratio - One of the first steps is to ensure that the images have the same size and aspect ratio.

Image Scaling- Once we've ensured that all images are of same dimensions; it's time to scale each image appropriately. We have decided to have images with width of 224 and height of 224 pixels as it has shown good performance in Image Net challenge.

Data Manipulation

For manipulation of data, we have considered data augmentation and dimensionality selection operation

Dimensionality reduction: We could choose to collapse the RGB channels into a single gray-scale channel. To achieve high accuracy in image classification color is one of the important features, so instead of going for single gray channel we kept RGB channels as it.

Data augmentation: Another common preprocessing technique involves expanding an existing data set with modified versions of existing images. Scaling, rotation, zooming, channel shifting and other transformations are typical. This is done so that the neural network is subject to a wide range of variation. This reduces the likelihood that the neural network will recognize undesirable features in the data set.

Training Model

Algorithms are used to extract features from the input data and create rules which are called as model. For most the unstructured data, deep learning algorithms specifically CNN architectures have shown better performance. So, to train models we have selected visual geometry group -16(VGG-16), visual geometry group -19 (VGG-19), Residual Network-50 (ResNet50), Residual Network -101 (ResNet101), InceptionV3 and Dense Net-121

Evaluating Model

Always it is better to evaluate model by keeping part of dataset which is unknown to model. Dataset is divided into training and validation part. Training part is fed to model during training and models are created. It is totally depending on us keep train-test ratio. Once models are built, evaluate models by passing validation data and observe the performance metrics.

Improving Model

It is always good to improve any machine and deep learning model once it is built and if you not satisfied with its performance metric. There are two types of parameters model parameters and hyper parameters. Model parameters are calculated by the model like weights in NN, m and c in linear regression etc. whereas hyper parameter are the parameters who value is set by human before training the model like number of epochs, batch size, test size, random state, alpha, max depth, number of layers, number of neurons etc. These hyper parameters can be adjusted in such way that, error of model will be optimized.

Experimental work

In this experimental work, dataset acquisition, preprocessing, model building, evaluating model and comparative analysis of various architecture are done. Let us focus on every step-in brief.

Dataset acquisition-

The actual data set used in this research to train and the model is collected manually from farm located in Athani Karnataka India and Solapur, Maharashtra India

Figure 4 Dataset and disease info collection

Data Preprocessing-

This technique includes cleaning, preparing and manipulating data in such way that it gets ready to apply these datasets to classifier

Uniform Aspect Ratio- This technique ensures, all the images are of same size and same aspect ratio.

Scaling - Images are scaled with width of 224 and height of 224 pixels.

Data Augmentation- With the help of data augmentation, it is possible to create more data. It performs various operations on an image and returns number of new and unique images, all based on input image, by doing flipping, rotating or cropping it.

Following table shows the how many samples actually we have collected from real fields and with the help of data augmentation how many images we have newly generated for training the models.

Table 2 Training - No of samples used vs. No of augmented images

Type of Brinjal disease	No. of samples used	No. of Augmented Images
Alternia leaf spot	40	1907
Bacterial Wilt	45	3136
Collar Rot	35	2621
Damping Off	32	3608
Tobacco Mosaic Virus (TMV)	55	3302
Cersocpera leaf spot	27	1835
Healthy samples	116	22601
Total	**350**	**39010**

Following table shows the how many samples actually we have collected from real fields and with the help of data augmentation how many images we have newly generated for validating the models.

Table 3 Validation - No of samples used vs. No of augmented images

Type of Brinjal disease	No. of samples used	No. of Augmented Images
Alternia leaf spot	40	152
Bacterial Wilt	45	252
Collar Rot	35	202
Damping Off	32	100
Tobacco Mosaic Virus(TMV)	55	104
Cersocpera leaf spot	27	99
Healthy samples	116	447
Validation Total	350	1356
Training Total		39010
Training + Validation		40336

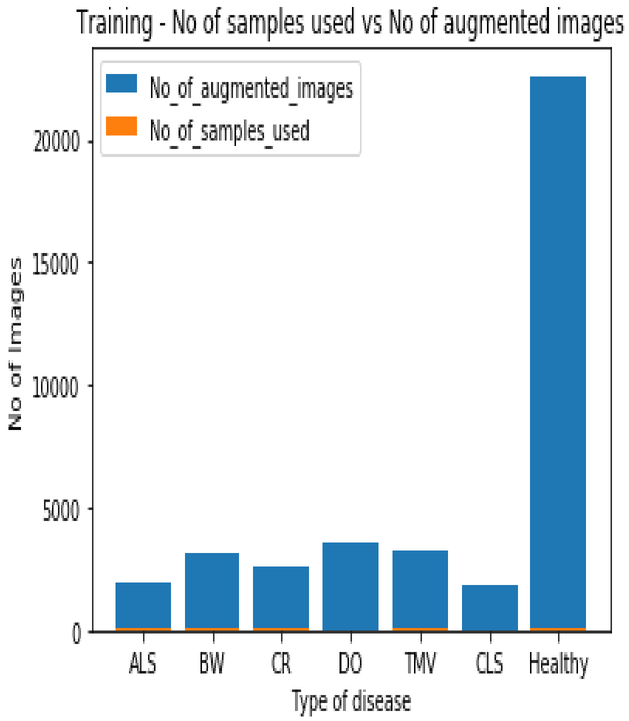


Figure 5 Training Augmented Images vs. No of Samples-Bar plot

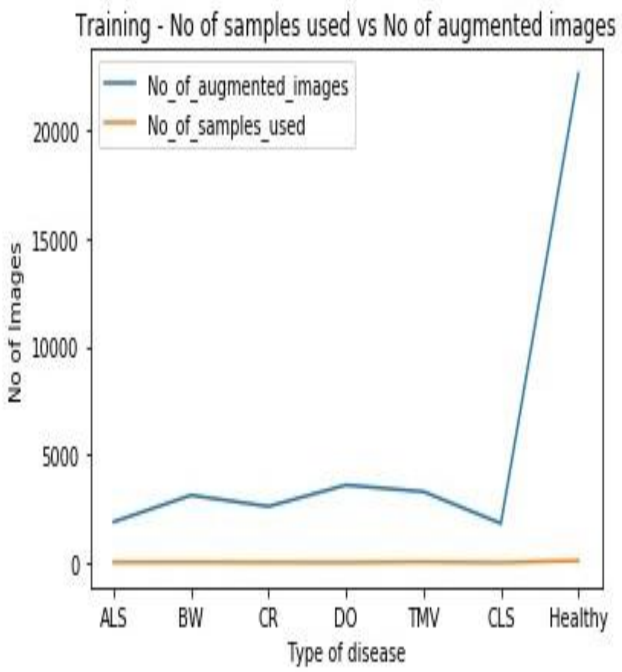


Figure 6 Training Augmented Images vs. No of Samples - Line Plot

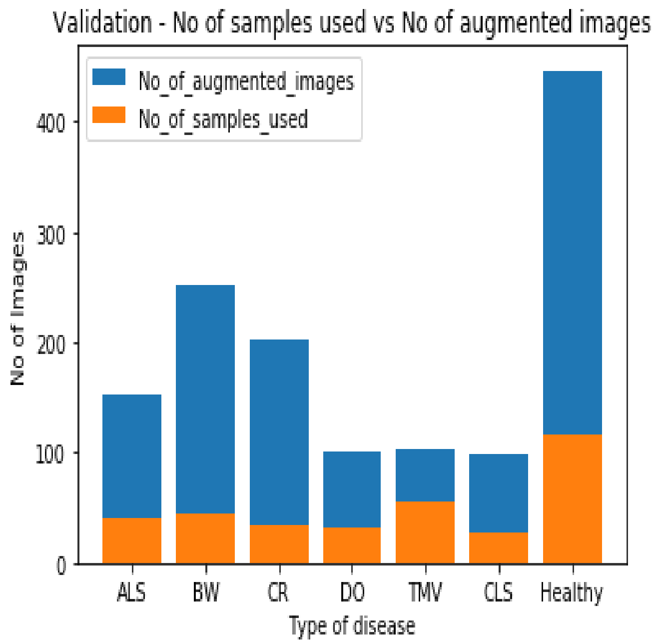


Figure 7 Validation Augmented Images vs. No of Samples – Bar Plot

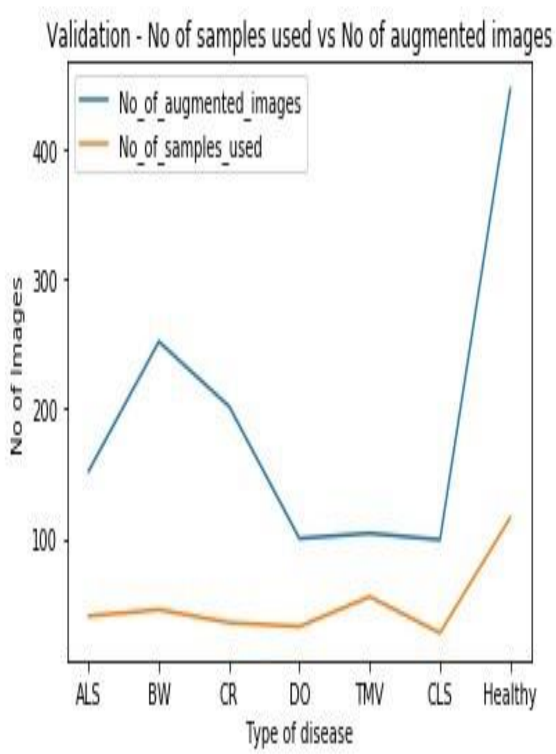


Figure 8 Validation Augmented Images vs. No of Samples used Line Plot

Training Model

To train models we have selected visual geometry group -16(VGG-16), visual geometry group -19 (VGG-19), Residual Network-50 (ResNet50), Residual Network -101 (ResNet101), InceptionV3 and Dense Net-121

Table 4 Different CNN Architectures

CNN architecture		Year	No. of Layers	Model Description Con + fc layers	No of Parameters
VGG-16		2014	16	13+3	138,357,544
VGG-19		2014	19	16+3	143,667,240
ResNet-50		2015	50	49+1	25,636,712
ResNet101		2015	101	100+1	44,707,176
InceptionV3		2016	48	45+3	23,851,784
DenseNet121		2017	121	120+1	8,062,504

Evaluating Model

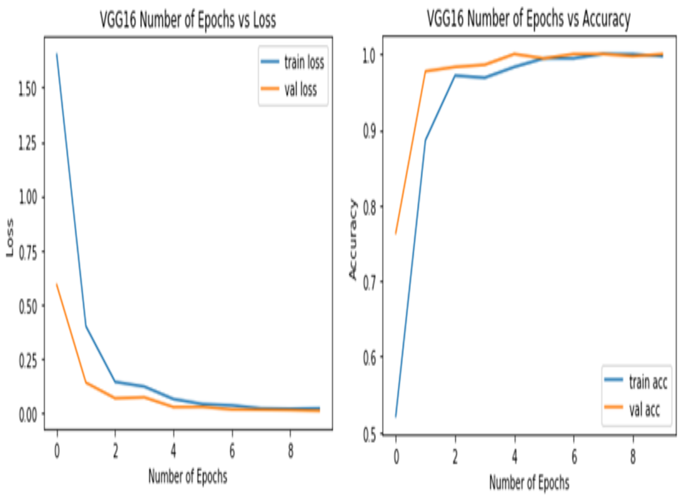
Always it is better to evaluate model by keeping part of dataset which is unknown to model. Dataset is divided into training and validation part. Training part is fed to model during training and models are created. It is totally depending on us keep train-test ratio. Once models are built, evaluate models by passing validation data and observe the performance metrics. (Training Set – 39010 and Validation Set- 1356)

VGG-16:

Table 5 Performance Analysis of VGG16

Model	Epochs	Training Loss	Training Accuracy %	Val Loss	Val Accuracy %
VGG-16	1	2.1001	37.08	0.589	76.35
	2	0.604	81.4	0.138	97.72
	3	0.155	96.84	0.066	98.29

	4	0.114	97.35	0.070	98.58
	5	0.078	97.84	0.026	100
	6	0.041	99.52	0.026	99.43
	7	0.031	99.66	0.014	100
	8	0.024	100	0.013	100
	9	0.020	100	0.012	99.72
	10	0.027	99.58	0.009	100
Average		0.319	90.92	0.0963	97.009

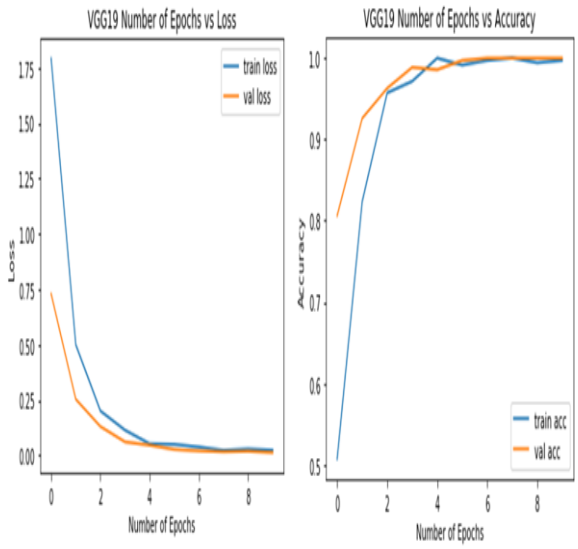


VGG-19:

Table 6 Performance Analysis of VGG19

Model	Epochs	Training Loss	Training Accuracy %	Val Loss	Val Accuracy %
VGG-19	1	2.371	35.79	0.729	80.63
	2	0.573	81.74	0.254	92.59
	3	0.235	94.65	0.131	96.30

	4	0.119	96.83	0.062	98.86
	5	0.052	100	0.0471	98.58
	6	0.052	99.03	0.0278	99.72
	7	0.035	99.67	0.0214	100
	8	0.023	100	0.0174	100
	9	0.0274	99.35	0.0196	100
	10	0.0224	99.93	0.0140	100
Average		0.3509	90.69	0.1323	99.66

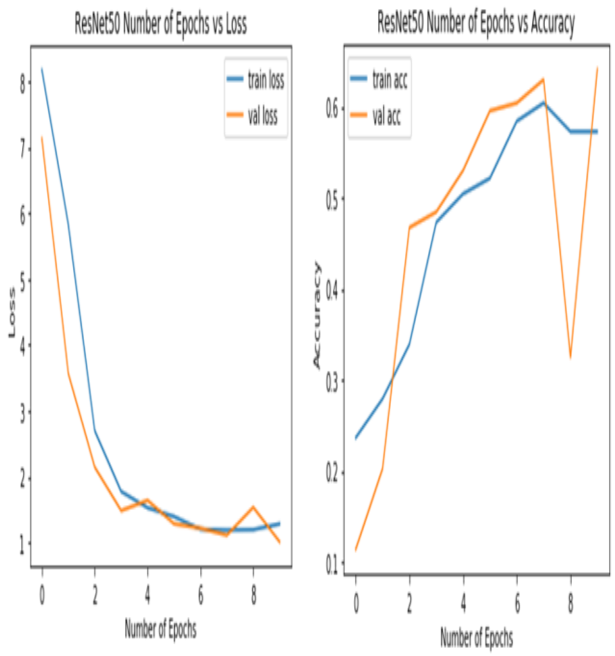


ResNet50:

Table 7 Performance Analysis of ResNet50

Model	Epochs	Training Loss	Training Accuracy %	Val Loss	Val Accuracy %
Res	1	7.694	18.82	7.125	11.40

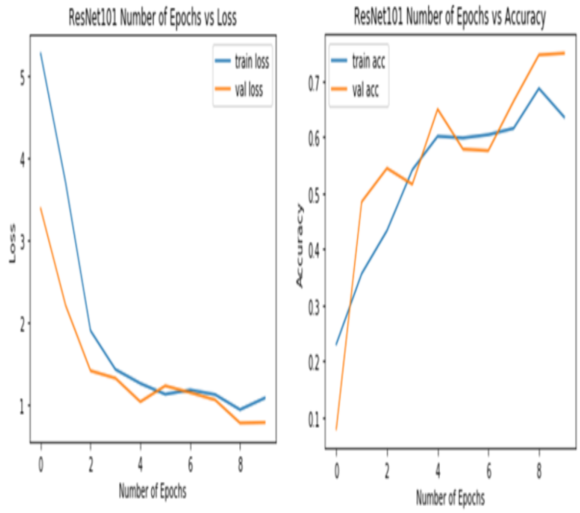
Net 50	2	7.125	20.04	3.563	20.23
	3	3.033	28.94	2.142	46.72
	4	2.009	43.44	1.482	48.43
	5	1.458	54.11	1.638	52.99
	6	1.457	53.65	1.283	59.54
	7	1.125	59.72	1.215	60.40
	8	1.207	60.10	1.110	62.96
	9	1.021	64.22	1.532	32.48
	10	1.288	55.04	1.008	64.10
Average		2.741	45.808	2.209	45.925



Res Net 101:

Table 8 Performance Analysis of ResNet101

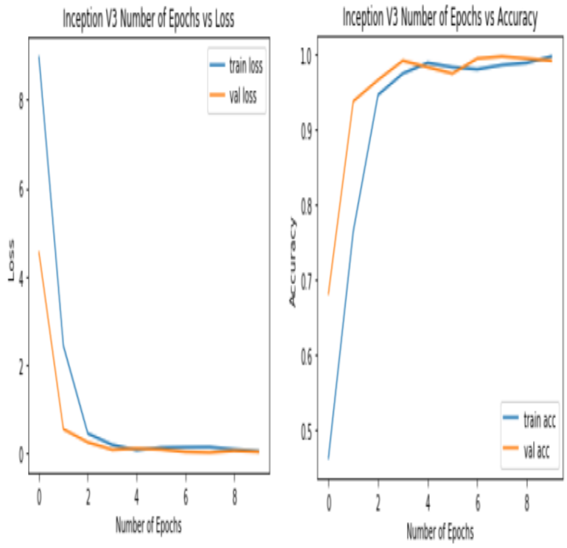
Model	Epochs	Training Loss	Training Accuracy %	Val Loss	Val Accuracy %
ResNet 101	1	5.435	20.33	3.388	7.98
	2	3.352	28.38	2.211	48.43
	3	1.948	43.96	1.413	54.42
	4	1.441	54.10	1.322	51.57
	5	1.382	53.51	1.036	64.96
	6	1.105	62.33	1.230	57.83
	7	1.220	61.66	1.148	57.55
	8	1.172	60.45	1.059	66.38
	9	0.9622	67.45	0.7780	74.64
	10	0.9582	70.02	0.7844	74.93
Average		1.897	52.21	1.436	55.86



InceptionV3:

Table 9 Performance Analysis of Inception V3

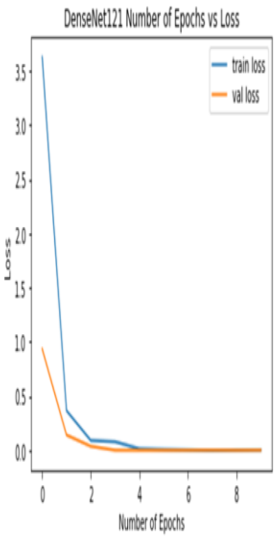
Model	Epochs	Training Loss	Training Accuracy %	Val Loss	Val Accuracy %
InceptionV3	1	8.592	36.66	4.534	68.09
	2	3.331	70.44	0.539	93.73
	3	0.393	95.28	0.239	96.58
	4	0.223	96.99	0.072	99.15
	5	0.040	99.21	0.100	98.29
	6	0.107	98.46	0.077	97.44
	7	0.134	97.56	0.023	99.43
	8	0.088	99.00	0.009	99.72
	9	0.093	98.52	0.046	99.43
	10	0.053	99.58	0.024	99.15
Average		1.305	89.17	0.566	95.101

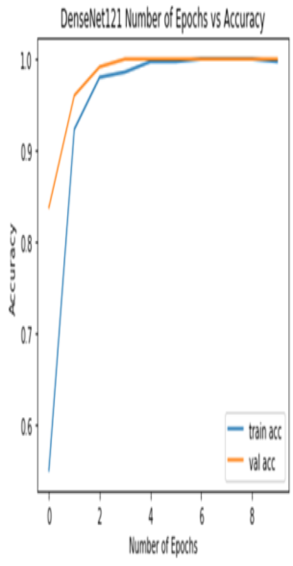


Dense Net 121:

Table 10 Performance Analysis of DenseNet121

Model	Epochs	Training Loss	Training Accuracy %	Val Loss	Val Accuracy %
DenseNet121	1	4.408	39.55	0.934	83.76
	2	0.5571	90.03	0.141	96.01
	3	0.175	97.12	0.037	99.15
	4	0.132	97.90	0.001	100
	5	0.0302	99.19	0.0003	100
	6	0.0066	99.78	0.0007	100
	7	0.0079	100	0.0011	100
	8	0.0033	100	0.0016	100
	9	0.0028	100	0.0006	100
	10	0.0016	99.93	0.0019	100
Average		0.532	92.35	0.111	97.89





Comparative Analysis

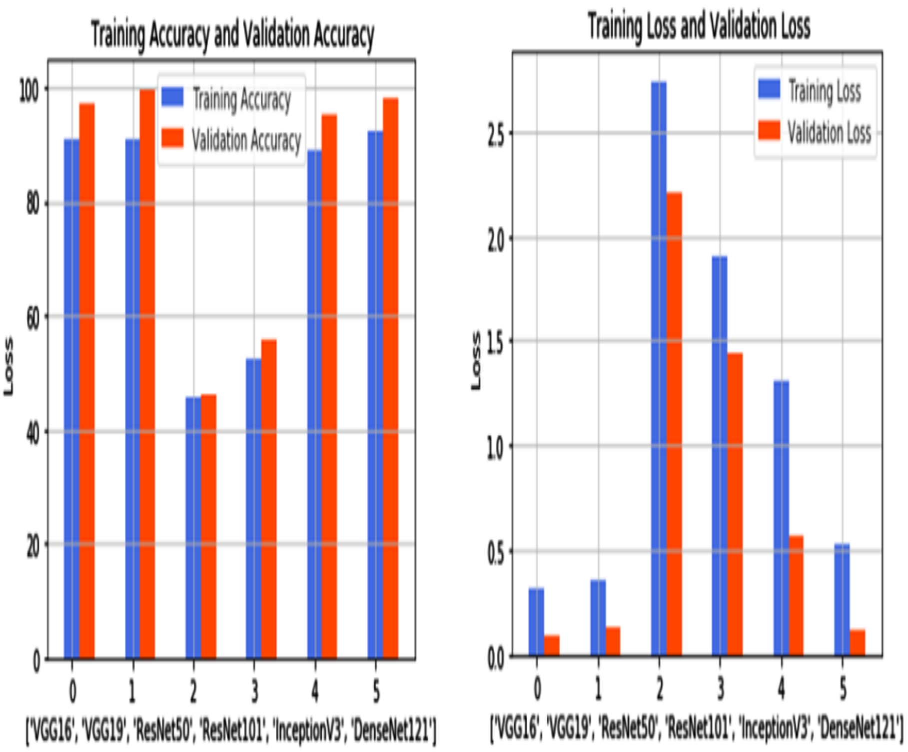
Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Following table shows the comparative analysis includes training accuracy, training loss, validation accuracy and validation loss of above mentioned 6 CNN architectures

Table 11 Comparative Performance Analysis

CNN Architecture	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
VGG16	0.319	90.92	0.0963	97.009
VGG19	0.3509	90.69	0.1323	99.66
ResNet50	2.741	45.808	2.209	45.925
ResNet101	1.897	52.21	1.436	55.86
InceptionV3	1.305	89.17	0.566	95.101
DenseNet121	0.532	92.35	0.111	97.89

Figure 9 Comparative Performance Analysis – Bar graph



Results and Discussions

Various deep convolutional neural network (DCNN) models include VGG16, VGG-19, ResNet50, ResNet101, InceptionV3 and Dense Net 121 are implemented in this research paper to predict the disease on Brinjal plant. The high-quality preprocessed dataset collected from real field with data augmentation of 40,336 images is used to conduct this research work. These implemented models are achieved the training accuracy of 90.92%, 90.69%, 45.80%, 52.21%, 89.17% and 92.35% respectively whereas validation accuracy of 97.01%, 99.66%, 45.92%, 55.86%, 95.10% and 97.89% respectively. These accuracy results showed that the DenseNet121, VGG19, VGG16 and InceptionV3 deep convolutional neural network models are one of the promising ways for effective implementation of such disease prediction systems in real time agricultural field. DesNet121 followed by VGG19 performs well in comparison to other convolutional neural network architectures.

Conclusions

As deep convolutional neural networks performed well at image classifications task, various CNN models are implemented as part of this research. Finally, outcome of this research experimentation is having a comparative performance analysis of above-mentioned architectures. DesNet121 followed by VGG19 performed well in comparison to other convolutional neural network architectures on real time huge preprocessed dataset. These can be one of the best promising architectures which can be used to build such disease prediction systems for Indian farmers in real time environment.

Acknowledgement

We would like to sincerely thank many stakeholders for helping us to carry out the work. Some of them are KLE Dr. MSSCET, Belagavi, Karnataka; VTU, Belgaum, Karnataka and MPKV Rahuri, Maharashtra. Opinions, observations, findings, conclusions, or recommendations in this paper are solely of us and these are not views of above-mentioned institutions.

References

1. Konstantinos P. Ferentinos, Deep learning models for plant disease detection & diagnosis – Science Direct, Elsevier Journal Sept 2018

2. Fuentes A, Kim S C, Yoon S, Park DS, 2017. A robust deep learning based detector for real-time tomato plant diseases & pest recognition, http://dx.doi.org/10.3390/s17092022.
3. Johannes A, Picon A, Alvarez-Gila A, Echazarra J, Rodriguez-Vaamonde S, Navajas A D, Ortiz-Barredo A, 2017. Automatic plant disease diagnosis using mobile capture devices, applied on wheat. Compute Electron. Agric. 138, 200–209.
4. Vijai Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques – Science Direct, Elsevier Journal, Nov 2016
5. Yun Hwan Kim, Seong Joon Yoo, Yeong Hyeon Gu, Jin Hee Lim, Dongil Han, Sung Wook Baik - Crop Pests Prediction Method using Regression and Machine Learning Technology, Nov 2014
6. T. Rumpf, A K Mahlein, U Steiner, E C Oerke, H W Dehne, L. Plümer . Early detection and classification of plant diseases with Support Vector Machines Jun 2014
7. Tejal Deshpande, Sharmila Sengupta, K.S.Raghuvanshi. Grading & Identification of Disease in Pomegranate Leaf and Fruit, IJCSIT, Vol. 5 (3), 2014, 4638-4645
8. Rakesh Kaundal, Amar S Kapoor and Gajendra PS Raghava. Machine learning techniques in disease forecasting: a case study on rice blast prediction, BMC Bioinformatics, doi: 10.1186/1471-2105-7-485
9. Diptesh Mujumdar, Deepak Kole, Aruna Chakraborthy, Dwijesh Mujumdar. Detection and diagnosis of plant leaf disease using integrated image processing techniques: a review. IJCEA, Volume-VI, Issue- III, June 2014
10. Dhaygude Sanjay B, Kumbhar Nitin P. Agricultural plant leaf disease detection using image processing. Int J Adv Res Electron Instrum Eng 2013; 2(1)
11. Plantix – Mobile Application to detect and cure plant disease http://plantix.net/en/ Balwant Gorad, Sirbi Kotrappa, "Novel Dataset Generation for Indian Brinjal Plant Using Image Data Augmentation", IOP Conf. Series: Materials Science and Engineering, 1065 (2021) 012041, doi:10.1088/1757-899X/1065/1/012041
12. Balwant Gorad, Sirbi Kotrappa, "Comparative Study and Review for Development of Disease Prediction System for Indian Crops", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 2, Ser. I (Mar - Apr 2019), PP 68-73
13. Balwant Gorad, Sirbi Kotrappa, Jibi Abraham, "Development of Intelligent Monitoring IoT Solution for Indian Agriculture Fields", NCRET, SITCOE, India

14. Peng Jiang, Yuean Chen, Bin Lu, Dongian he and Chaunquan Liang, May 2019, Real time detection of apple leaf disease using deep learning approach based on improved CNN, IEEE – section on advanced optical imaging
15. MPKV Rahuri, Maharashtra, Karnataka and Tamilnadu Agritech Portal
16. https://ncrb.gov.in/sites/default/files/adsi_reports_previous_year/table-2A.1.pdf
17. https://www.towardsdatascience.com , https://www.data-flair.training ,
18. https://www.analyticsvidhya.com ,
19. https://kaggle.com , http://kdnuggets.com , http://medium.com