Plant Disease Identification uses Deep Learning Methods

K. Rahul, K. Sandeep Varma, M. Venkata Sreeram, G. Goutham Subhash Nirmal Kumar, J. Raghu

Abstract**:** ***The Farmers in the agriculture sector facing a difficult task in identification of plant diseases. Depending on the humannakedeyeitisdifficulttoclassifyplantdiseasesbecauseit changes the control policy of one disease to another. Finding of disease effect leaf and healthy leaf is a tricky task. It requires knowledge in the plant diseases and technology for processingthe input images. In this scenario Image processing uses a Deep Learning method like Convoulation Neural Network (CNN) for predicting the disease affected leaf or not. In this paper CNN is compared with some machine learning, classification methods like K-nearest neighbor (KNN) , Decision tree, Random Forest, Linear Discriminant Analysis (LDA),support vector machine (SVM) and LogisticRegession.***

Index Terms**:** ***Image processing, Machine learning and Deep Laerning.***

I. INTRODUCTION

Theincreaseofcropdiseasesandpestoutbreaksarecommon nowadays,dependingonclimateandenvironmentconditions. Furthermore the majority of small holder farming household individuals suffering from hunger. Fortunately, diseases can be managed by identifying the diseases as soon as it appears on the plant. The current advances in computer vision like deep learning made possible the way for disease diagnosis in plants.In addition, with the development of the internet worldwide, it is easy to access diagnostic information on a particular type ofdiseases.

The purpose of this project is to identify diseases of a differentplantspeciesinalesstime.Fortrainingthemodelof classifier consider a public dataset of 14,400 images of healthy and diseased plant leaves collected under ideal conditions, To train a convolutional neural network to identify 3 crop species and 15 (absence thereof) or diseases.The trained model achieves a highest accuracyof 97.20 percent generated by CNN and least accuracy of 90.30 percent on a test set, Representing the viability of this approach.

Revised Manuscript Received on June 01, 2019.
K. Rahul, Dept of IT, SRKREC, Bhimavaram, India.
K.Sandeep Varma, Dept of IT, SRKREC, Bhimavaram, India.
M. Venkata Sreeram, Dept of IT, SRKREC, Bhimavaram, India.
G. Goutham Subhash Nirmal Kumar, Dept of IT, SRKREC, Bhimavaram, India.
J. Raghu, Dept of IT, SRKREC, Bhimavaram, India..

towards crop disease identification on a enormous global scale. The set of images collected from trusted online sourcesandtestingmodelisappliedtothem.Suchthatunder different conditions images are used for training the model achieves the accuracy 90 4%. For more accuracy it needs to train the model on a various set of datasets. The approach of training deep learning models on increasingly large datasets and available public image datasets presents a clear path

II. LITERATUREREVIEW

F. Argenti et al [2] in this paper,we get the great importance fortheImageclassificationthatistherecognitionofdiseased plant image from the other images. The problem ofboundary recognitionofthediseasedplantuptowhatextentitisspread also supervised under thisclassification.

Anand et al [1] in the paper Applying image processing technique to detect plant diseases, proposed that Gabor filter is used for Segmentation and filtering. The color features are extractedfromtheSegmentationprocessandArtificialneural network (ANN) is used forclassification.

P. Revathi et al [3] in this paper, we found symptoms of leaf spot images and find the diseased leaf using our techniques and algorithms. The color image segmentation is carried out to get diseased spots. Techniques in image processing and deep learning can be used to find the edge features whichare further used to identify diseasedspots.Tushar H. Jaware et al [4] in this paper, it detects that plant disease through some stages. Which uses color space transformation for color transformation structure and the segmentation and find the required features and finally extracted features are passed through Neural network.

Prof.Sanjay B. Dhaygude [5] proposed how to transform RGB(Red, Green and Blue) image into Hi-Spectral Image(HSI)andalsogreenpixelvaluesaremaskedtoremove theunwantedsegmentsandsomestatisticalmethodsareused forclassification.

Mokhled S. Al-Tarawneh[6] proposed Image resolution was transformed from RGB to L*a*b color space.The image was cropped and classified by using fuzzy c- means(FCM) clustering and K-means clustering(KCM) methods.

Yan-Cheng Zhang[7] suggested fuzzy feature selection approaches like Fuzzy Curves(FC) and Fuzzy Surfaces(FS) but these two approaches works on a small set of significant features.

III. PROPOSEDMETHODOLOGY

Infected or healthy leaves are to be selected and uploaded ,then different types of image processing techniques are appliedtotheminordertoprocessthoseimagesandidentifies thediseases.

A. Algorithm for image processing using CNNillustrated the step by step approach

Step-0: Considerdiseasedorhealthyleafandgavethisleafas aninputtothehiddenlayersinordertopredictthediseaseof plant leaf. This stage is also known as InputLayer.

Step-1: In this hidden layer we have five inner layers. First Layer is Convolution layer, this layer gets the input from input layer and transform the leaf image into a Feature Map and provide this output as an input to the second layer that is Activation layer.

Step-2: In this Activation Layer uses Rectified Linear Unit(ReLU) function.f(x) = max (0, x), where x is the input to a neuron. TheReLUfunctionreplacesthenegativevalueswithzeroand positive values are to be kept same on the Feature Map.Now the Feature Map has only zeros and positive values. Provide this map as an input to the next layer that is Batch Normalization.

Step-3: In this Batch Normalization layer, we are going to normalize the values by adjusting and scaling. By using this normalization technique, the training speed will be more. Now the output produced in this layer is the given as inputto the next layer i.e. MaxPooling.

Step-4: Max Pooling layer gets the input from the Batch Normalization layer. In this layer, it reduces the number of parameters when the images are too large. From the entire matrix it considers the maximum values in a sub matrix.

Step-5: The above sub matrix from max pooling layer is given as an input to Dropout layer. In this layer we use the softmax function. This function activates one of the outputs having moreaccuracy.

Step-6: Now from step-1 to step-5 the entire process will continue for five times and then we get one output which is having moreaccuracy.

B. Architecture

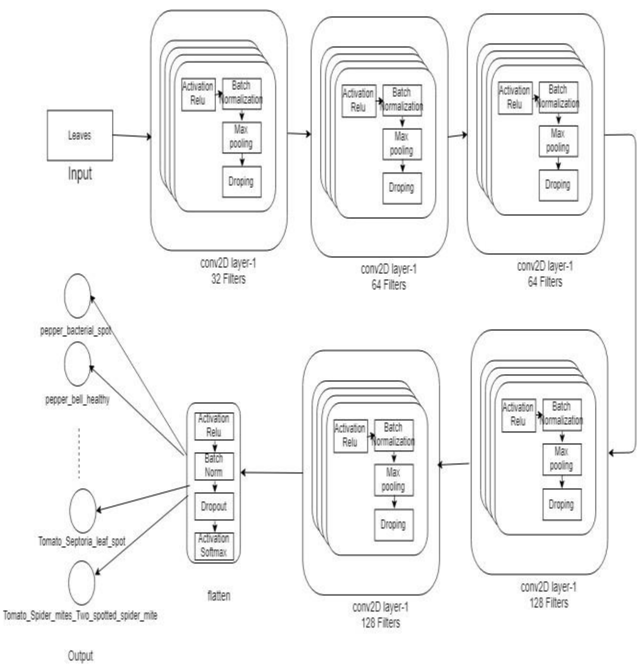


Fig.1.Architecture of CNN Layers

Machine Learning uses algorithms encouraged by the function of the brain's neural network. Deep Learning is an internalbranchofMachineLearning.Itanalysesthedataand thenmakeadeterminationorpredictionovernewdata.Inthis detection on plant diseases architecture, we use Computational Neural Network. Neural networks are collection of neurons or artificial neurons. These neurons are arranged in three types of layers like input layer,hidden layer and outputlayer.

- Input Layer: In this layer user provides input to the hidden layers in order to determine or predict the disease of aplant.
- Hidden Layer: In this hidden layer we used total five convolutionallayerswithdifferentneuronsorfilters and for each filter some methods are to beapplied

Those are:

a) Convolution
b) Activation
c) BatchNormalization
d) Maxpooling
e) Dropout

a) Convolution:

Convolutionisthefirstlayertoextractfeaturesfroman input image. Convolution layer finds the relationships between the pixel or input of small squares by learning the image features.The mathematical it takes two input matrix such as image matrix and a filter or kernelmatrix.

- An image matrix (volume)is of threedimensions ($h * w * d$), Where h represents height, w represents the width and d represents the dimension.
- A filter matrix ($f_h * f_w * d$), Where f_h is a function of height and f_w is a function ofwidth.
- Outputsavolumedimension$(h-f_h+1)*(w-f_w+1)*1$

Consider a 5 x 5- Image Matix , whose image pixel values are binary and filter matrix 3 x 3 its image pixel values are also binary as shown inbelow.

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

5x5-ImageMatix

*

1	0	1
0	1	0
1	0	1

3x3- FilterMatrix

Fig.2. 5x5 Image matrix multiplying with a 3x3Filter matrix

The convolution layer after multiplication of 5 x 5 binary image matrix multiplies by 3 x 3 binary filter matrix it produces an output matrix called "feature map".

4	3	4
2	4	3
2	3	4

Fig.3.Feature map

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

InthislayertheImageMatrixismultipliedbydifferentFilter Matrix can perform the operations like blur, edge detection and sharpen.

b) Activation

There are different types of activation functions such as ReLU, sigmoid,tanhetc..,butinthehiddenlayersofCNN,theoptimized activation function is Rectified Linear Unit(ReLU). The alias name for ReLU is also known as a ramp function and this function is formed of analogous converted to half-wave rectification. ReLU's purpose is to convert non-linearity into ConvNet. Since, ConvNet data is to learn from amodel.

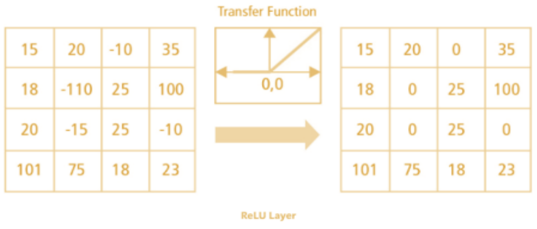


Fig.4.ReLu function operation

c) BatchNormalization:

The input layer in CNN can be normalized by using two operations like scaling and activations. Suppose consider the featuresfrom0to1andsomefeaturesfrom1to1000then normalize them by speed up learning. In this process the input layer has benefited from it. The same normalization can also be implemented in hidden layers. The training speed of the model gets improved.

d) MaxPooling:

If the image sizes are too large then Max pooling is applied. The alias names for Max pooling is spatial pooling, sub sampling or down sampling. These methods reduce the dimensionality of each map and retains the important information. Max pooling collects the largest element from the feature map such that it can also take an average pooling. Theotherpoolingtechniquesumpooling,whichcollectssum of all elements in the featuremap.

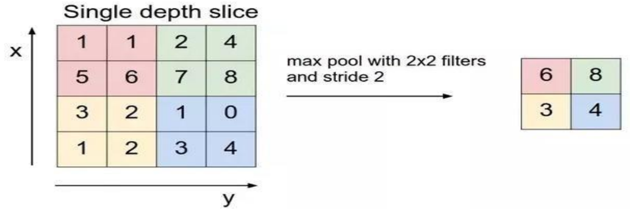


Fig.5: Max pooling layer reduces the number of parameters

e) Dropout:

Dropout is a process of shut down some neurons for each layer, which is a regularization technique. In this process randomlyselecteddon'tuseneuronsinbothforwardandback propagations. These neurons are dropped out randomly on each iteration. The learning algorithm will not have no idea which neurons will shut down. Therefore, learningalgorithm will not focus on some specific features. Finally the output layer uses the softmax activation function. Due to this the outputs are having highaccuracy.

IV.RESULTS

A. *Results during Training and Testing Period using CNN*

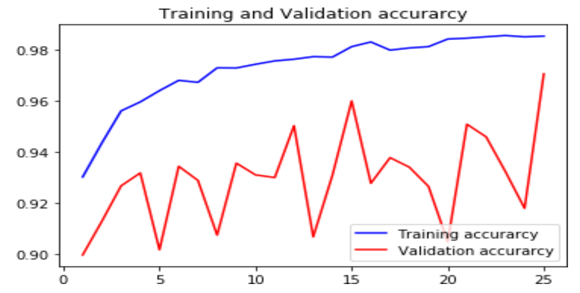


Fig.6.Accuracy of Training and Validation Datasets

Theabovegraphshowstheaccuracybetweenthetrainingand validation data sets. Training data set is nothing but we are goingtodividetheentiredatasetintotwopartsinaparticular ratio(4:1). 80 percent of data is allowed for training and the remaining 20 percent is allocated for testing purpose. During thetrainingperiod wearegoingtoprovidebothinputaswell as the output of the model. During validation period model considers some of the trained data as an input and predictthe output. Now we are going to find the accuracy basedon successful predictions made by the model. In the above graph blue color indicates the training accuracy and red color indicates validationaccuracy.

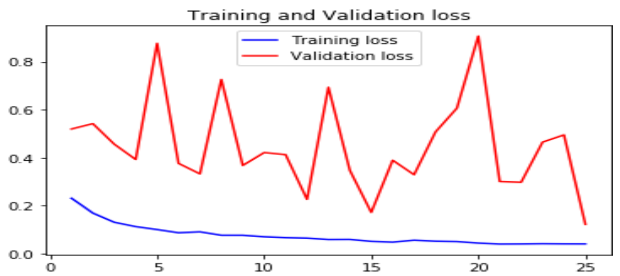


Fig.7: Loss occurs during Training and Validation

The above graph shows the loss between the training and validation data sets. During the training period we are going to provide both input as well as the output of the model. Duringvalidationperiodmodelconsiderssomeofthetrained data as an input and predict the output. Now we are going to find the loss based on unsuccessful predictions made by the model. In the above graph blue color indicates the training loss and red color indicates validationloss.

B. *The Output of the system during testing periodusing CNN*

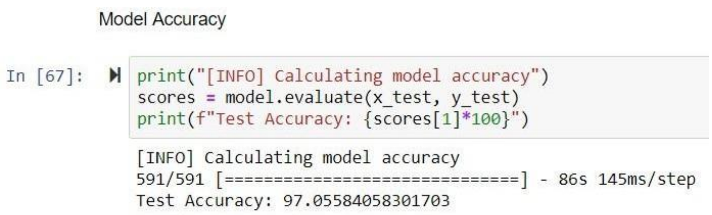


Fig.8 Calculating model accuracy for Training Datasets

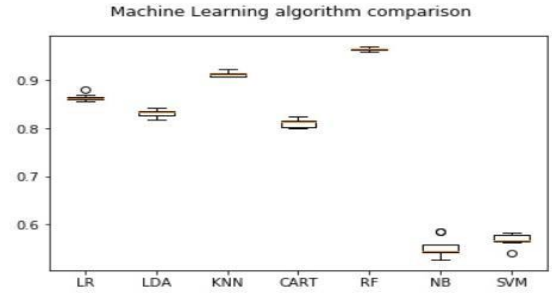
Theabovefigurerepresentstheaccuracyofthetestdatasets. Here we are going to provide the remaining 20 percent data setasaninputtothemodelinordertopredicttheoutput.The model achieved 97.05 percent accuracy of the test dataset.

C. Results during Training and Testing Periodusing Machine LearningAlgorithms

In the below figure it shows the comparison of machine learningalgorithm'saccuracyusingboxplotdiagrams.Inthis areausingdifferentalgorithmslikeLogisticRegression(LR), LinearDiscriminantAnalysis(LDA), K-Neighbors Classifier (KNN),Decision Tree Classifier(CART), Random Forest Classifier (RF), GaussianNB(NB), Support Vector Machine (SVM). The below figure shows the mean and standard deviation of theiraccuracies.

```
# boxplot algorithm comparison
fig = pyplot.figure()
fig.suptitle('Machine Learning algorithm comparison')
ax = fig.add_subplot(111)
pyplot.boxplot(results)
ax.set_xticklabels(names)
pyplot.show()
```

```
LR: 0.864272 (0.006225)
LDA: 0.832669 (0.007384)
KNN: 0.912458 (0.004798)
CART: 0.811618 (0.008058)
RF: 0.964844 (0.002773)
NB: 0.551468 (0.019050)
SVM: 0.569074 (0.011728)
```




```
Wall time: 57min 42s
```

Fig.9. Comparison of ML algorithm's Accuracy using Box plot diagrams

S.No	Algorithm	Mean	Standard Deviation
1.	Logistic Regression (LR)	0.864	0.0006
2.	Linear Discriminant Analysis (LDA)	0.832	0.007
3.	K-Neighbors Classifier (KNN)	0.912	0.004
4.	Decision Tree Classifier (CART)	0.811	0.008
5.	**Random Forest Classifier (RF)**	**0.964**	**0.002**
6.	Gaussian NB(NB)	0.551	0.019
7.	Support VectorMachine (SVM)	0.569	0.011
8.	**Convolution Neural Network(CNN)**	**0.9703**	**0.001**

Table 1. Different Algorithms with their mean and standard deviation of their accuracies

Retrieval Number: A10620681S319/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

From the table it is observed that Random Forest Classifier has greater mean value when compared to other algorithms. SoherewearegoingtoconsidertheCNNandRandomForest Algorithms in order to predict the PlantDiseases.

V. OUTPUTSCREENS

A. *Predicted Page is using CNN*

The below screen shows us that the leaf is Tomato and it is having the disease Tomato Mosaic Virus and the chances of having that virus is 59 percent.

Fig.10 Output Screen uses CNN

VI. CONCLUSION

This system shows us that the image processing techniques that have been used for recognizing plant diseases. This systemabletodetectandclassifyplantdiseases.Inthisweuse CNN technique. By using this technique, we can be able to analyses the healthy and diseased plant leaves. The results of thisevaluationshowclearlythatwecanimprovetheaccuracy using a CNN architecture which achieved anaccuracy of 97.30percent.Someofthechallengesinthesetechniquesviz. Effectofbackgrounddataintheresultingimage,optimization of the technique for a specific plant leafdisease.

Asalimitationofthissystem,wenoticedthattheaccuracy the data which is no of trained before is low when compared to the data which is trained before. So this system requires some more training on a large dataset containing more than 50,000 images. Therefore, our future work will focus on the preparation of a strong labeled dataset, which makes it possible to detect any image diseases perfectly which is not trained previously.

REFERENCE

1. Anand H. Kulkarni, Ashwin Patil R. K, "Applying image processing technique to detect plant diseases", International Journal of Modern Engineering Research. , Oct-2012,vol.2, Issue.5, pp: 3661-3664.
2. F.Argenti,L.Alparone,G.Benelli,"Fastalgorithmsfortextureanalysis using co-occurrence matrices radar and signal processing", IEEE Proceedings, Dec- 1990,vol. 137, Issue 6,pp:443-448.
3. P. Revathi, M. Hemalatha, " Classification of cotton leaf spot diseases using image processing edge detection techniques", IEEE International Conference on Emerging Trends in Science, Engineering and Technology, Tamilnadu,India, May-2013 , pp-169-173.
4. Tushar H. Jaware, Ravindra D. Badgujar and Prashant G.Patil,

HELPFUL HINTS

image segmentation" , National Conference on Advances in Communication and Computing, World Journal of Science and Technology, , Dhule, Maharashtra, India, April-2012. pp:190-194.
5. Prof.Sanjay B. Dhaygude, Mr. Nitin P. Kumbhar, " Agricultural plant leaf disease detection using image processing ", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,SandSPublication,Jan-2013,vol.2,Issue1,pp:599-602.
6. Mokhled S. Al-Tarawneh, " An Empirical investigation of olive leave spot disease using auto-cropping segmentation and fuzzy c-means classification" , World Applied Sciences Journal,Aug-2013, vol.23, no.9,pp:1207-1211.
7. Yan-Cheng Zhang, Han-Ping Mao, Bo Hu, Ming -Xi Li, " Feature selectionofcottondiseaseleavesimagebasedonfuzzyfeatureselection techniques ", Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, , Beijing, China, Jan-2008, pp:124-129.

AUTHORS PROFILE

Mr. K Rahul, SRKR Engineering College (rah199703ul@gmail.com)srkr engineering college, department of IT chinnaamiram bhimavaram, Andhra Pradesh 534204, India. K Rahul is currently persuing B.Tech in the Department of Information Technology at S.R.K.R. Engineering College, Bhimavaram, and Andhra Pradesh (INDIA). His research interests are in the field of Machine Learning, Deep Learning,, Artificial Intelligence.

Mr. K Sandeep Varma, Andhra university (sandeepvarma501@gmail.com) srkr engineering college, department of IT chinnaamiram bhimavaram, Andhra Pradesh 534204, India. I K.Sandeep Varma, working as Assistant Professor in SRKR Engineering College, Bhimavaram,AP,India. My interested areas are Machine Learning, Artificial Intelligence and Distributed Computing Technologies.

M. Venkata Sreeram, Andhra university (nani.mallula777@gmail.com)srkr engineering college, department of IT chinnaamiram bhimavaram, Andhra Pradesh 534204, India. M Sreeram is currently persuing B.Tech in the Department of Information Technology atS.R.K.R. Engineering College, Bhimavaram, and Andhra Pradesh (INDIA). His research interests are in the field of Machine Learning, Deep Learning,, Artificial Intelligence.

G. GouthamSubhashNirmalKumar,Andhra university(gudapati.gsnkumar@gmail.com) srkr engineering college, department of IT chinnaamiram bhimavaram, Andhra Pradesh 534204, India. G Subash is currently persuing B.Tech in the Department of InformationTechnologyatS.R.K.R.EngineeringCollege,Bhimavaram,and Andhra Pradesh (INDIA). His research interests are in the field of Machine Learning, Deep Learning,, ArtificialIntelligence.

J. Raghu, Andhra university (raghujunnuri@gmail.com) srkr engineering college, department of IT chinnaamiram bhimavaram, Andhra Pradesh 534204, India. J Raghu is currently persuing B.Tech in the Department of Information Technologyat S.R.K.R. Engineering College, Bhimavaram, and Andhra Pradesh (INDIA). His research interests are in the field of Machine Learning, Deep Learning,, Artificial Intelligence.

Retrieval Number: A10620681S319/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication