

Plant Disease Detection System : Based on Machine Learning (CNN)

Shubham Singh
CSE Department SRMCEM
LUCKNOW,INDIA
shubham11.07.02@gmail.com

Siddharth Singh
CSE Department SRMCEM
LUCKNOW,INDIA
be20cs107siddharthsingh@gmail.com

Er. Aayush Pratap Singh
(Assistant Professor) CSE
Department SRMCEM
LUCKNOW,INDIA

Abstract— One of the important and tedious task in agricultural practices is detection of disease on crops. It requires huge time as well as skilled labor. To mitigate this issue, computer aided disease recognition model is proposed. It uses leaf image classification with the help of deep convolutional networks. In this paper, VGG16 and Resnet34 CNN was proposed to detect the plant disease. It has three processing steps namely feature extraction, downsizing image and classification. In CNN, the convolutional layer extracts the feature from plant image. The pooling layer downsizing the image. The disease classification was done in dense layer. The proposed model can recognize 38 differing types of plant diseases out of 14 different plants with the power to differentiate plant leaves from their surroundings. The performance of VGG16 and Resnet34 was compared. The accuracy, sensitivity and specificity was taken as performance Matrix. It helps to give personalized recommendations to the farmers based on soil features, temperature and humidity.

Keywords—Plant disease, Convolutional neural network, VGG16, Resnet34, leaf image , Image Processing.

I. INTRODUCTION

Agriculture plays vital role in Indian economy. Agriculture contributes nearly 17-18% country GDP. Due to various factors, sometime yield of the agriculture is not as good as expected. The agriculture yield is affected by nature as well as disease. The challenges in agricultural field are seed selection, irrigation techniques, soil defects, weather condition and water treatment [1]. The virus, fungi and bacteria causes the plant disease. It is another challenge in the agriculture field. If the plant diseases are not correctly and timely identified, then it will lead to heavy loss in the yield of crops. The identification of plant disease needs experts and infrastructure. It is not available in all places. The computer aided identification plant disease identification helps the experts to find the disease accurately and timely manner. The machine techniques are used to identify, classify and predict the plant's disease. The traditional machine techniques such as Support Vector Machine, Decision Tree classification, Random forest, Naïve Bayes are used for plant disease classification and prediction. Nowadays convolutional neural networks are used for plant disease classification and prediction [2-4]. The main contribution of the article is to predict the plant disease using convolutional neural network. The CNN extracts the features from plant and it identifies the disease of the plant.

2. ORIGINALITY

The proposed methodology is predict the plant disease from the plant leaf using convolutional neural network. It extracts the features and predict the disease from the plant leaf.

3. SYSTEM DESIGN

The main objective of the proposed model is to predict the plant disease. The convolutional neural network is used to implement the proposed model. The processing steps of the convolutional neural network is shown figure 1.

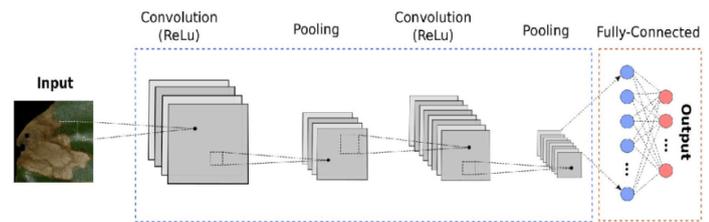


Figure 1. Processing steps in CNN

The convolutional neural network has four layers namely convolution layer, pooling layer, flatten layer and dense or fully connected layer.

3.1 Convolutional layer :

The important features are extracted from the leaf in the convolutional layer. The kernels or filters are used to extract the features from image. The kernels are operated on input image and the feature map is generated. The feature map is calculated using formula shown in equation 1 (1) where i is the input image and k is kernel. The n and m is the row and column indexes respectively. The hyper parameters of the convolutional layer are kernels, strides.

$$F(n, m) = (i * k)[n, m] = \sum_x \sum_y k[x, y] * i[n - x, m - y]$$

3.2 Pooling layer :

The images are downsized in the pooling layer to reduce the computational complexity. There are two types of pooling namely max pooling and average pooling. The output of the convolutional layer is divided into $N \times N$ sub matrices. The max pooling will take maximum value from each submatrix. The average value of each submatrix is calculated in the

average pooling. The number of convolutional and pooling layer is used based on the requirements.

4.3 Dense Layer :

The output of the convolutional layer or pooling layer is given as input to the dense layer. The dense layer will accept only flatten output. The inputs are flattening in the flatten layer. One or more dense layer is used to predict the plant disease. The output of the final dense layer is the network output. The whole process is called forward propagation. The input is multiplied with weighted and added all the inputs. It is represented in equation 2.

$$O_1 = \sum_{j=1}^m w_j x_j \tag{2}$$

Where w is weight of each input and x is the feature input. The output is given to activation function. It is represented in the equation 3.

$$output = A(O_1) \tag{3}$$

Where A is the activation function. The convolutional layers and dense layers have activation function. The ReLU (Rectified Linear Unit) activation function is used in convolutional layers and softmax activation function is used in the last dense layer. The performance of the network is not as good as expected. Then error is calculated and it is propagated back to dense layer and convolutional layer to adjust the weights. This process is called as backpropagation. After adjustments' weights, again forward and backward propagation is performed until the network reaches the expected performance.

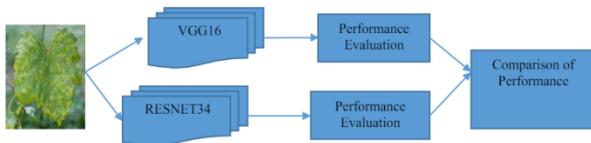


Figure 2. Plant disease prediction steps

The VGG16 and Resnet 34 predefined Convolutional Neural Networks are proposed to predict the plant disease. The prediction steps of the proposed model is shown in figure 2. The VGG16 and Resnet34 are kind of convolutional neural network. Both are used for image classification. In VGG16, the 16 refers number layers in the network. VGG16 used small filters. The input image size is 224X224 RGB. It has nearly 138 million parameters.

Resnet34 has 34 layers. Each layer has number of blocks. Each block has operations. The Resnet34 takes residuals of

each and every layer and these residuals are used in the subsequent layers. The optimization in Resnet34 is easy. The performance of the VGG16 and Resnet34 is compared.

5. EXPERIMENT AND ANALYSIS :

The plant leaf is given as input for disease prediction. The 150000 instances were taken from Kaggle database for experimentation. The leafs are taken from 14 different plants. The proposed model can recognize 38 differing types of plant diseases. Python language is used for implementation. The proposed CNN is implemented using Python. The transfer learning architectures such as VGG16 and Resnet34 were used for implementation. The filter size was 3×3 in VGG16 and resnet34 architectures. The size of the filter used in VGG 16 and resnet 34 ranges from 64 to 512. The number of filters used in VGG 16 and resnet 34 was shown table 1.

	Number Filters	Number of Chennal	Number Filters
VGG 16	64	2	3×3
	128	2	3×3
	256	3	3×3
	512	6	3×3
Resnet 34	64	3	3×3
	128	4	3×3
	512	3	3×3

Table 1. Filters in VGG 16 and Resnet 34

The maximum number filters used in CNN is not fixed. The maximum number filters are fixed based on the complexity of the problem. The configuration of VGG 16 and Resnet34 is given in figure 3.

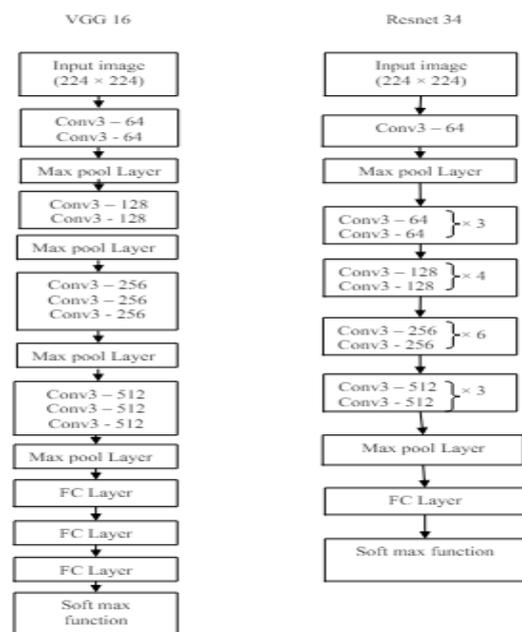


Figure 3. VGG16 and Resnet34 configuration

The data is divided into training data and validation data. 80% of the data is taken for training and 20% of the data is taken for validation. The Cross entropy error function is used for error calculation. It is shown equation 4.

$$E(X_i, T_i) = - \sum_{k=1}^n T_{ik} * \log(p_{ik})$$

Where T_{ik} is the target vector. p_{ik} is the probability of i th element in k th class. Accuracy, sensitivity and specificity was used as performance metrics to assess the performance of the proposed approach. using The accuracy is calculated using the equation 4 .

$$\text{Accuracy} = \frac{(\text{True positive} + \text{True negative})}{(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})}$$

The sensitivity is calculated using the formula shown in equation 5

$$\text{Sensitivity} = \frac{\text{True positive}}{(\text{True positive} + \text{False negative})}$$

The specificity is calculated using the formula shown in equation 6

$$\text{Specificity} = \frac{\text{True negative}}{(\text{True negative} + \text{False positive})}$$

The error rate of the Resnet34 Model is shown in figure 3. The error rate of the training data is reduced uniformly when the number epochs increased. But, the validation data error rate is not uniformly reduced because of uncertainty of the data.

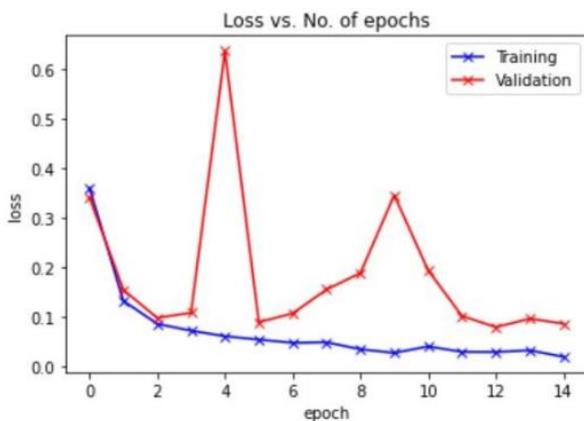


Figure 4. Resnet34 – Error rate of training and test data

The accuracy of the test data is shown in figure 4. The accuracy is increased after 10 epochs.

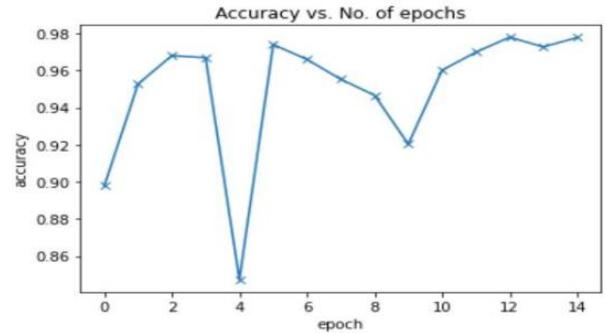


Figure 5. Resnet34 – Test data accuracy
The error rate of the VGG16 Model is shown in figure 5 . The error rate of the training data and testing is reduced gradually when the number epochs increased. The accuracy of test data of VGG16 model is shown figure 6. In this model after 12 epochs, the accuracy is decreasing.

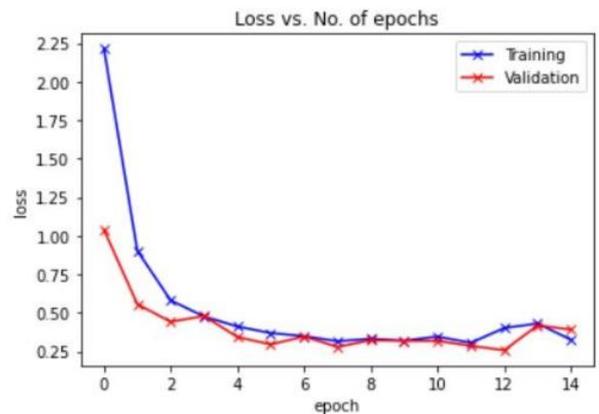


Figure 6. VGG16- Error rate of training and validation data

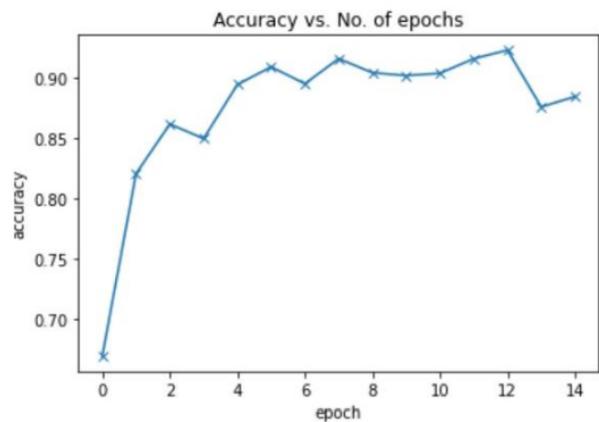


Figure 7. VGG16- Test data accuracy

The performance comparison of Resnet34 and VGG16 is shown in table 2. Both model performance is good. The Resnet34 performs better than VGG16 for plant disease prediction application because Resnet34 optimize the residual and it deals with vanishing gradient problem.

S.No.	Resnet34	VGG16
Accuracy	97.77	97.58
Loss	0.0857	0.0875
Sensitivity	1.0000	0.9741
Specificity	0.9921	0.9561

Table 2. Resnet34 and VGG16 comparison

6. CONCLUSION :

The prediction of plant disease using convolutional neural network was proposed and implemented. The Resnet34 and VGG16 CNN architectures were used for prediction of plant diseases. The convolutional layer was used for extraction of features, pooling layer was used for sample downsizing and dense layer was used for prediction. The ReLU activation function was used in convolutional layer. The softmax activation function was used in dense layer. Nearly 150000 plant leaves were taken from 14 different plants for experimentation. The proposed architectures were able recognize 38 differing types of plant diseases.

REFERENCES :

- [1]<https://www.sciencedirect.com/journal/measurement-sensors> Measurement: Sensors
[file:///C:/Users/win-10/Downloads/1-s2.0-S2665917422000757-main%20\(1\).pdf](file:///C:/Users/win-10/Downloads/1-s2.0-S2665917422000757-main%20(1).pdf)
- [2] [file:///C:/Users/win-10/Downloads/2106.10698%20\(1\).pdf](file:///C:/Users/win-10/Downloads/2106.10698%20(1).pdf) Research Paper on Plant Disease Detection Using Image Processing and Machine Learning .
- [3] [file:///C:/Users/win-10/Downloads/08437085%20\(1\).pdf](file:///C:/Users/win-10/Downloads/08437085%20(1).pdf) Research paper on Plant Disease Detection Using Machine Learning
- [4] Online AI : ChatGPT for reference and steps for working on this particular topic .
<https://chatgpt.com/>