Plant Disease Detection using Deep Learning Approach

Satyajit Khandu Khot[1], Dr.S.D.Bharkad[2], Vivek Barme[3]
[1,2,3]*Electronics and Telecommunication Engineering, Government Engineering College Aurangabad, Maharashtra*

Abstract **- The potential growth of developing countries like India depends on agriculture. The basic need of humans and animals is food. Disease plants directly affect the yield of crops and which leads to an imbalance in the economy of developing countries. So Plant Disease detection is very important. Traditionally diseases are detected by professionals or plant pathologists with an empty eye, but this approach is time-consuming and expensive also. Now in the digital era, machine learning and deep learning is widely used in various sector and agriculture is one of them. In this paper, we have created the model with the help of a convolution neural network for the detection of disease and deployed it on google cloud to use in the mobile app. The model can easily identify 11 different kinds of diseases which contain 13610 images of 3 plant species. For this, we have used the Kaggle dataset. with the Model, we have achieved 95% accuracy over different plants. This shows that the model achieved a good accuracy rate for plant disease detection.**

Index Terms **- Plant disease detection, Machine learning, Deep learning, Convolution neural network, Mobile application, Python, Google Cloud.**

I.INTRODUCTION

In developing countries like India, the main source of income is agriculture which provides raw ingredients to other industries. Many diseases come due to climate change and many other reasons. It is easy to cure the plant disease in the early stage but in some places, farmers don't have advanced facilities like plant pathology laboratory due to which detecting the disease become expensive and time-consuming.

Disease to the plant is concerned threat to global food security which results in 10-16% losses in the world harvest of crops annually [1].

Traditional methods for detecting the disease of plants and diagnosis depends on expert with the naked eye and symptoms of the disease. But there is a variety of symptoms and a large difference in the same symptom due to which experts may fail to identify the disease in the early stage so the traditional method fails in efficiency and accuracy [2].

The use of deep learning and machine learning approaches for the detection of disease opens the door to gaining high performance compared to traditional methods.

Nowadays, the use of Convolution neural networks (CNN) made a big achievement in the detection of the various plant diseases, But it needs to search the ways to reduce training time and increase accuracy [3]. Also, CNN is best for problems that involve image data as input and it also requires less amount of pre-processing.

Mobile application created with the use of machine learning and deep learning model is beneficial for farmers. The research aims to develop a mobile application that allows users to capture images of diseased plant leaves and get the name of the plant and its disease.

II. LITERATURE SURVEY

Deep learning in the field of Prediction of Disease detection plays a major role and increases the productivity of the country. Many types of research have been done in this field. Applying a neural network to a tomato plant leaf T.Prajawala Achieved an accuracy of 94% [4].

Using the Transfer learning Process on the dataset Wang et al achieved 90.4% accuracy [5]. The team of S. Sladojevic analyzed the same dataset and showed an average accuracy of 82.3% with photographic conditions [6].

Monalisa Saha et.Al used CNN with clustering to classify the images as healthy and unhealthy [7].

Qiaokang Liang et.al Used ResNet50 to solve the problem and got good accuracy [8], and Mohit Agrawal et.al Used the same algorithm for tomato plants to achieve good accuracy
[9].

A.Lakshamanrao et.al used a random forest algorithm on a rice dataset which has good accuracy [10].
Badage et.al solved the problem in two steps, first they applied the EDA process for Data pre-processing and cleaning and in the second step applying the machine Learning algorithm achieved good results [11].
Monji Kherallah used CNN and ResNet Architecture and compared the result in which they found with the method of CNN and without Data Augmentation accuracy of the model is 94.80%, CNN with Data Augmentation 97.2% and ResNet 98.96% [13].
Prof.A K Misra and Vijay Singh applied image processing techniques with genetic Algorithms on Banana, beans, jackfruit, etc. their technique showed efficiency in the recognition and classification of diseases of the leaf. This technique is very different than ML and DL [14].
Sammy V. Militante and Bobby D. Gerardo applied CNN architecture on 35000 images which include 9 different diseases of tomato,4 grape, 4 corn, 4 apples, and 6 sugarcane disease. The model achieved 96.5% accuracy rate [15].
Xulang Guan first balanced data with Up-Sampling Process and then used four models Inception, ResNet, Inception ResNet and DenseNet and added to a voting Mechanism. This method is also known as stacking achieved 87% accuracy over 10 plant diseases [12].
M.Raja Babu, T.Srinivasa Ravi Kiran experimented with 4 models which are Random forest achieved 97% accuracy, ANN achieved 94% and CNN achieved 95% [16].

III. PROPOSED METHODOLOGY

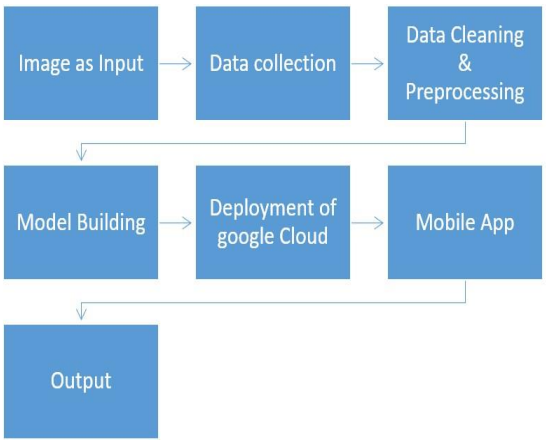


Fig. 1. Process to execute Task above figure 1 shows how we are proceeding for this project

A. Image as Input
The user will scan the image of the diseased plant leaf and pass it to the application.
B. Data Collection
Dataset of Plant leaf of disease is a prime requirement for training the model and evaluating the model performance and accuracy. There are three ways to collect the data to solve our problem.
1) Ready-Made Data: Data can be downloaded from opensource websites like Kaggle etc.
2) Capture images and Annotate : This process is expensive and time-consuming because it requires capturing photos of disease plants and annotating them with expert.
3) Web scraping: With the API like selenium, Beautiful soap we can write web scraping code and collect the images of disease plants.
We have collected Dataset from Kaggle. Dataset has three plants tomato, Potato and pepper with their plant leaf images.

Table 1. Details of Dataset

Plant Name	Plant Disease Name	No of Images
Pepper	Bacterial spot	997
	Healthy	1478
Potato	Healthy	152
	Late blight	1000
	Early blight	1000
Tomato	Bacterial spot	2127
	Early blight	1000
	Healthy	1591
	Late blight	1909
	Leaf mold	952
	Target spot	1404

Table 1 shows no of images for each plant disease .Healthy plant leaves are there to distinguish between healthy and disease plant leave. Dataset is divided into 80% images in a train ,10% images in the test ,and 10% images in the validation dataset. which contain 10884 train images ,1369 test images ,and 1357 images for validation.

C. Data Cleaning and Preprocessing
Xulang Guan applied the EDA process to their dataset and found data is imbalanced so he applied the UpSampling method to fix the number of images in each of the classes. We also applied the EDA process to find if dataset imbalance or not? but we found our dataset is balanced so there is no requirement of Up-Sampling for dataset [12].
Caching and Prefetching are used to improve the performance of the Model and save time, Prefetch helps for using both CPU and GPU while training, and Cache save the time of reading images for every

epoch. Fig 2 shows how Prefetch works and Fig 3 shows how cache works on data.

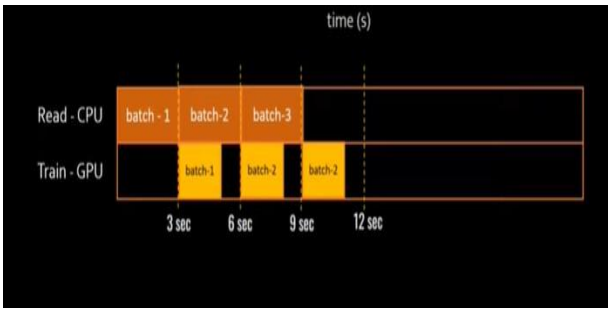


Fig. 2. Prefetch

Rescaling helps for scaling the image to 255 and resizing helps for resizing the image to 256x256 pixels because when a working mobile app going to capture images it may not have 256x256 image size so by using resizing we can change the size of the image to 256x256 pixels.

The data augmentation Process is used to increase the volume of the dataset, distort the image, and also help to reduce the problem of overfitting. By applying horizontal flip, and vertical flip we performed data augmentation.

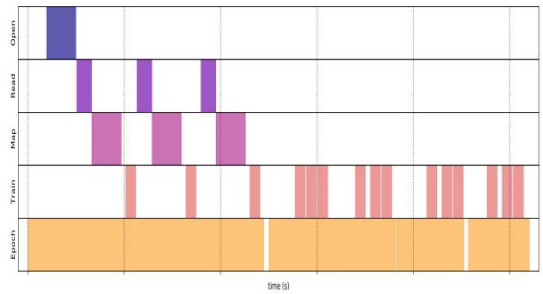


Fig. 3. Cache

Images should be of the same size to feed the model. We resized the all images in 256x256 pixels which also helped to improve the accuracy of the dataset.

D. Model Building

1) CNN : Convolutional Neural Network

A Convolutional Neural network is a very famous type of architecture for image classification problems. Fig 4 shows the architecture of CNN. CNN consists of several neural layers of different types. Which begins with the input image, convolution layer, pooling layers, and activation layers and goes on connecting with more fully connected layers and output.

Feature extraction is performed by the convolution layer and then passed to the activation layer. Several steps involve to obtain global features from input. In the final step extracted features are passed to a full deep neural network that performs classification tasks.

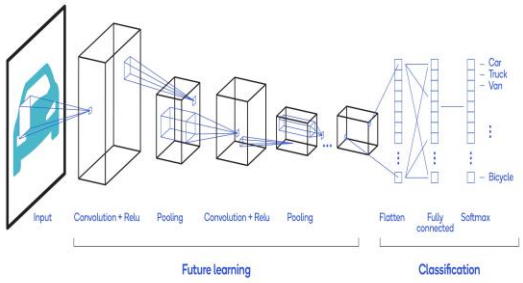


Fig. 4. Architecture of CNN

2) Convolution layer

This layer stores the output generated from the previous layers which have biases and weights for learning to the next layer. Also in this layer mathematical pipeline exists which used to extract the feature map of the image. Filters are used to reduce the size of the image ex: a 5x5 image can be reduced to a 3x3 size image. The new matrix is generated from the input image. Fig 5 shows how filter reduce the images size.

3) Pooling Layer

In Model training sometimes the model shows 100% accuracy on the training dataset and 50% on the test dataset

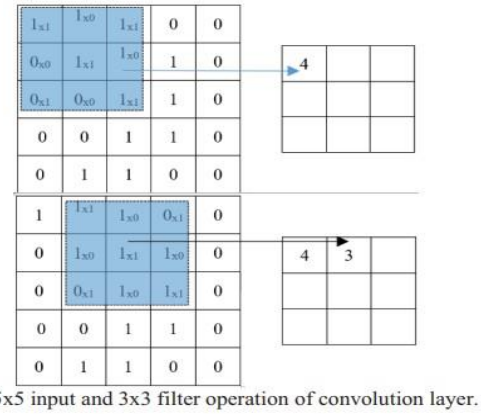


Fig. 5. Filters

this problem is known as Overfitting. ReLU and maxpooling are used to lower the dimensions of the feature map. Also, this layer can reduce the parameters, Time required for Training and most important is controlling the overfitting.Fig 6 shows working of pooling layer.

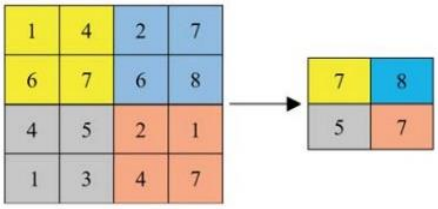


Fig. 6. Pooling Layer

4) Activation Layer

This layer is used to introduce non-linearity on the model also dropout layers are used to prevent the overfitting. Mainly ReLU is used in every convolution layer.

5) Fully Connected Layer

This performs the classification of plant leaves. Softmax is used for recognition and classification. This layer contains a deep neural network of many neurons which gets input from the flattened array size of the feature map.

6) CNN Model History

CNN Model History		
Layer	Output shape	Param
Conv2d_6	(None,254,254,32)	896
Max_pooling2d_6	(None,127,127,32)	0
Conv2d_7	(None,125,125,64)	18496
Max_pooling2d_7	(None,62,62,64)	0
Conv2d_8	(None,60,60,64)	36928
Max_pooling2d_8	(None,30,30,64)	0
Conv2d_9	(None,28,28,64)	36928
Max_pooling2d_9	(None,14,14,64)	0
Conv2d_10	(None,12,12,64)	36928
Max_pooling2d_10	(None,6,6,64	0
Conv2d_11	(None,4,4,64)	36928
Max_pooling2d_11	(None,2,2,64)	0
Flatten_1	(None,256)	0
Dense_2	(None,64)	16448
Dense_3	(None,13)	845

Fig. 7. CNN Model History Number of trainable parameters are shown in fig 7.

7) Deployment on Google Cloud Platform

Google Cloud Functions is a server less execution environment for building and connecting cloud services. With Cloud Functions, you write simple, single-purpose functions that are attached to events emitted from your cloud infrastructure and services. Your function is triggered when an event being watched is fired. Your code executes in a fully managed environment. There is no need to provision any infrastructure or worry about managing any servers.

It can be written in Java, python on the Google Cloud Platform. (GCP website)

8) Mobile App

We proposed a low-cost mobile application that is made by using react js. Mobile is communicated with the the model which is deployed on google cloud performs the classification task.

Users can capture an image of a diseased plant and get the result of which disease of that plant.

E. Result

Model Achieved 95-97% rate by using 50 epochs in the Training model. And 89-95% on testing. The plot of Accuracy and loss are shown below fig 8. which shows the model is efficient in detecting plant disease.

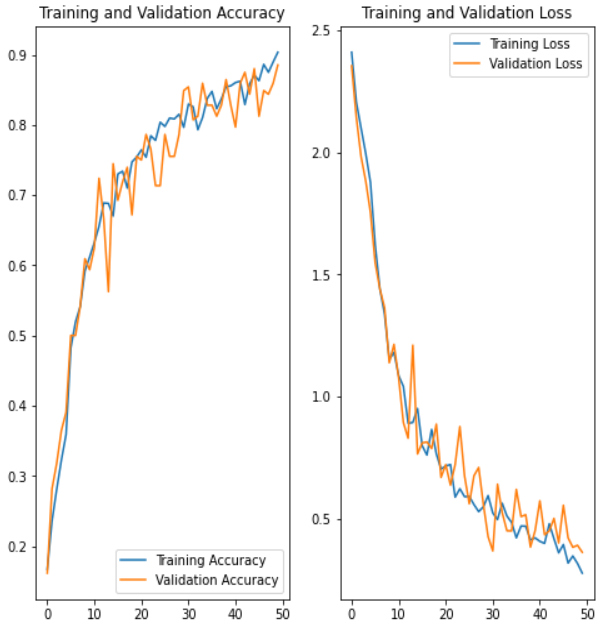


Fig. 8. Graph of Accuracy and loss

fig 9 and 10 shows the results of developed application on mobile phone.

F. Conclusion

As we all know the agriculture sector is one of the most important sectors. Agriculture plays a crucial role in everyone's life, as food is the basic necessity of everybody. To yield the surplus amount of food we need to protect the plant from insecticides and recognition of diseases plant. So to detect the

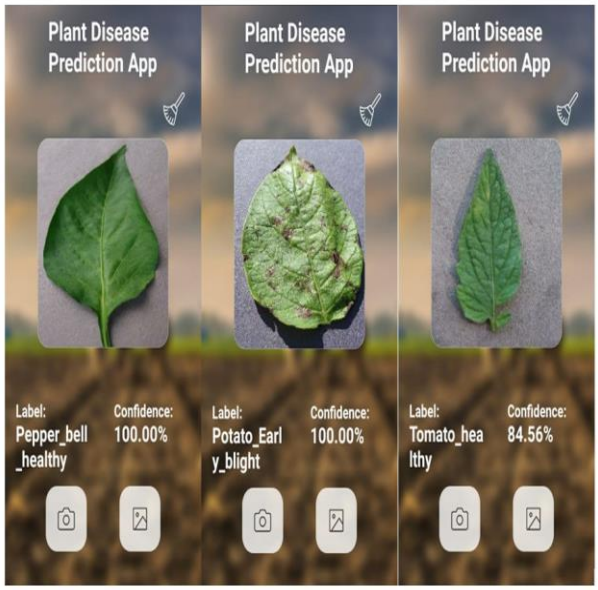
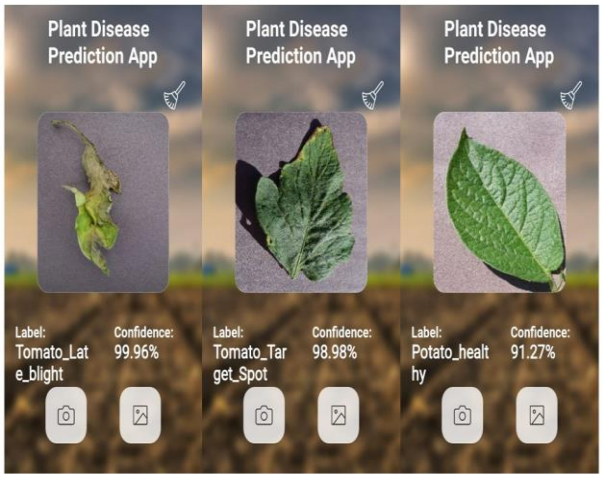


Fig. 9. Result 1



Results of Mobile App.

Fig. 10. Result 2

plant disease we developed an application like a mobile app. that identifies the plant disease.

We used the plant dataset for Kaggle furthermore we applied CNN architecture on the dataset and we achieved an accuracy of 95-98% for plant disease detection

REFERENCE

[1] Plant Disease: A Threat to Global Food Security Richard N. Strange Peter R. Scott Annual Review of Phytopathology 2005 43:1, 83-116

[2] Riley, M.B., M.R. Williamson, and O. Maloy. 2002. Plant disease diagnosis. The Plant Health Instructor. DOI: 10.1094/PHI-I-2002-102101

[3] Toda Y,Okura F. How Convolutional Neural Networks Diagnose Plant Disease. AAAS Plant Phenomics. 2019;1-14 https://doi.org/10.34133/2019/9237136.

[4] T.Prajwala,"Tomato Leaf Disease Detection using Convolutional Neural Networks" , International Conference on Contemporary Computing – IC3,2018,pp.1-5.

[5] G. Wang, Y. Sun, and J. Wang, "Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning," Computational Intelligence and Neuroscience, pp. 1-8, 2017.

[6] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," Computational Intelligence and Neuroscience, vol. 2016, pp. 1-11, 2016.

[7] M Saha,Sasikala, " Identification of plants leaf diseases using Machine Learning algorithms", International Journal of Advanced Science and Technology, 29(9s),2900-2910,2020.

[8] Q.Liang, Xiang,Y Hu, G.Coppola, " Computer-assisted plant disease diagnosis and severity estimation network", Computers and Electronics in Agriculture , Vol -157 Pages 518-529, ISSN 0168-1699,2019.

[9] M Agarwal, A.Singh , S.Arjaria , " Tomato Leaf Disease Detection using Convolutional Neural Network", Procedia Computer Science, Volume 167,2020 , ISSN 1877-0509

[10] A.Lakshamanarao,G.Prasanna , S. Ramasai, " Machine Learning Methods for Plat-Disease-Detection", Journal of Information and Computational Science, Vol -10 Iss- 6,2020.

[11] A.Badage , "Crop Diseaase Detection using Machine Learning : Indian Agriculture" , IRJET,Volume:05,Issue 09/Sep 2018.

[12] X. Guan," A Novel Method of Plant Leaf Disease Detection Based on Deep Learning and Convolutional Neural Network," 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), 2021, pp. 816-819, doi: 10.1109/ICSP51882.2021.9408806.

[13] F. Marzougui, M. Elleuch and M. Kherallah," A Deep CNN Approach for Plant Disease Detection," 2020 21st International Arab Conference on Information Technology (ACIT), 2020, pp. 1-6, doi: 10.1109/ACIT50332.2020.9300072.

[14] V. Singh, Varsha and A. K. Misra, ”Detection of unhealthy region of plant leaves using image processing and genetic algorithm,” 2015 International Conference on Advances in Computer Engineering and Applications, 2015, pp. 1028-1032, doi: 10.1109/ICACEA.2015.7164858.
[15] A. Lakshmanarao, M. R. Babu and T. S. R. Kiran, ”Plant Disease Prediction and classification using Deep Learning ConvNets,” 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670918.