International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Plant Disease Detection using Deep Learning

Omkar Mindhe[1], Omkar Kurkute[2], Shrutika Naxikar[3], Prof. Nikhil Raje[4]

[1-4]Department of Computer Engineering, PHCET, Rasayani, Maharashtra, India

---***---

Abstract - *Modern agriculture has become far more than simply a method to feed ever growing populations. A country's economy is dependent on agriculture in some way. With the population increasing each day there must be a rise in focus on the primary sector. World Bank report states that 3 out of 4 people in developing countries survive in countryside and take home as less as Rs.200 a day. Agricultural refinement is necessary for improving the quality of agro-product industries especially in emerging nations. Therefore, the timely detection of plant infection might just be the key to stopping agricultural losses. We built 'Plant disease detection' because we truly believe that all knowledge that helps people grow good quality of food should be openly accessible to everyone. Developing algorithms that can accurately diagnose a disease based on an image is the next big disease diagnostics tool which assists in turning a vision of productive agriculture into reality. The aim of this project is to create an AI app that will detect and classify plant diseases. We will be using the public dataset PlantVillage with 54,444 images and PyTorch as our deep learning platform. The plant leaf images will be used to check for plant diseases. Hence, we believe that the detection of plant infections in early stages will surely help sustain agricultural stability and progress a country's growth.*

***Key Words*: Deep learning, PlantVillage, PyTorch, Disease diagnostics tool**

1. INTRODUCTION

For decades agriculture was thought to be the engineering of basic food crops. Agriculture and farming were considered the same until much later when farming was actually commercialized. With the boom of industrialization, suddenly people understood the scope of high profits due to economic development and hence many other occupations related to farming came to be recognized as a part of agriculture. Currently, agriculture in addition to farming includes forestry, fruit cultivation, dairy, poultry, etc. The science, technology, engineering that goes behind production, processing and distribution is now considered a part of modern agriculture. Thus, agriculture can be elaborated as: "*the art of cultivating crops including the production, processing, marketing and distribution of crops and livestock products*".

Modern tech has assisted humans in producing enough food for over 8 billion population. Even then, food security stays vulnerable by several factors like plant diseases, climate fluctuations, etc. Plant diseases threaten food security as everyone depends on the availability of healthy crops.

1.1 Fundamentals

In a developing nation, majority of the agro-production is manufactured by small-wage farmers and the results of the crop losses due to pests and infection have become far too common. Furthermore, it is the small farmers who suffer from food scarcity, hence making them the most vulnerable to bad quality of food. Plant related infections and illness are a major threat to food security, but it is their rapid identification that still remains difficult due to the lack of necessary infrastructure.

Plant diseases are generally classified into the following 3 categories: -

1) Viral- Viruses are very small infectious RNA parasite. They are made of pieces of genetic material, enclosed in a fur of protein. Viruses take over the cells in the plants and manipulate them to multiply the virus itself. This process often damages or kills the plant's infected cells.

2) Bacterial- Bacteria is a member of group of microscopic single celled organisms which can only be seen through a microscope. They are found everywhere on surface of the Earth, from living in soil, volcanic regions to thriving in deepest trenches of the ocean. Bacteria are both beneficial and pathogenic. Beneficial bacteria helps with digestion in animals, nitrogen fixation in roots, breaking down of animal and plant leftovers and more. Bacteria often overwhelm the immune system and results in severe and harmful diseases in living organisms.

3) Fungal- Fungi are organisms that lack chlorophyll and thus they do not have the ability to photosynthesize their own food. They obtain energy by absorbing nutrients from others through tiny threads called hyphae. Collectively, fungi and fungi-like-organisms cause more plant diseases than any other group. Fungi are especially harmful during preharvest and postharvest of crops. They produce highly toxic, hallucinogenic chemicals that have affected millions of animals including humans and still continue to do so.

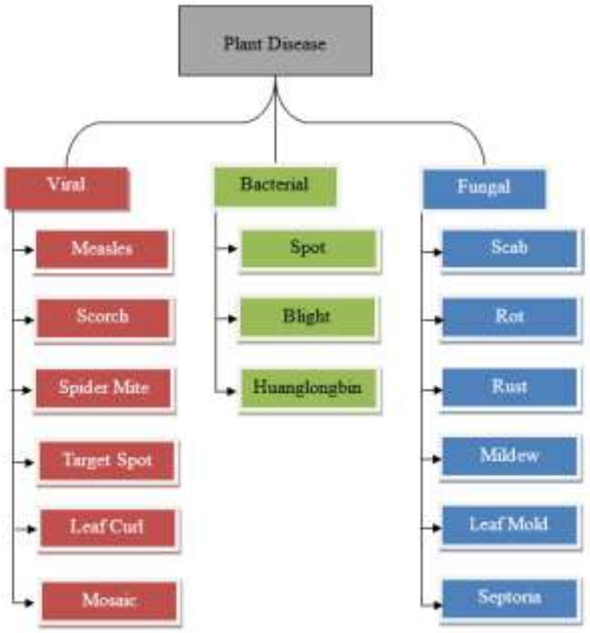


Fig. 1: Basic Classification of Plant Diseases

2. LITERATURE SURVEY

In this chapter the existing technologies and research done on similar topic is reviewed along with various techniques used and also identifies the current literature on plant disease detection. We will be identifying the techniques that have been developed and present the various advantages and limitation of these methods used.

The system proposed by Abirami Devarai et al. will solely know the type of infection which damage the plant while providing the solution in less time. The project starts by capturing of plant leaf images. Healthy and unhealthy pictures are taken. Pictures of leaves are then segmented to create clusters. Options are removed before applying clustering from K-means and then random forest classifier is employed for training and recognition. Finally, infections are classified.

Shima Ramesh et al. [2] starts their project by creating dataset images of ailing and fit non-ailing leaves. They are trained using random forest to classify the input images till the terminal of the image is reached. For drawing out patterns in an image they have used Histogram of an Oriented Gradient (HOG). To train on the public datasets Machine learning is used. This way the detection of infections in the plants is done in a sophisticated manner.

In Yuan Tian et al. [3] research, color sequence pattern are represented in RGB to HSI (hue, saturation, intensity) by using generalized linear model (GLCM). They have also used Support vector machine which has Multiple-Cell Size (MCS), used for finding infection in wheat plant.

Aakanksha Rastogi et al. [4] uses k-means clustering to separate out the parts which are not healthy; present in the leaf image. GLCM is deployed to get the texture and features, and later fuzzy logic is used to get the grading of the diseases. Artificial neural network(ANN) is later employed to check severity in the infected leaf.

In the paper by Sanjay B. Dhaygude et al. [5] there are four steps to get the fully functional system. In step one, a color transformation skeleton is created from the input RGB image. The RGB image is utilized for color production and transformed HIS image is utilized for color descriptor. In the second step, a threshold value is used and green pixels are removed by masking them . In step three, by using a pre-computed threshold level, from all the useful segments from step one, the green pixels are removed and masked. In the final step, the segmentation of image is done.

3. METHODOLOGY

We have approached the given problem by using facebook's deep learning framework PyTorch. For our end goal we decided to develop an AI application using this deep learning model and transfer learning technique.

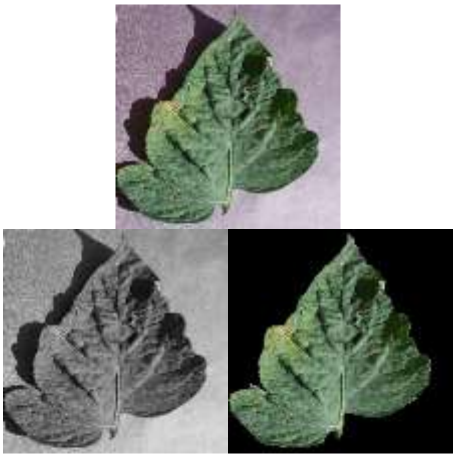
Computer Vision: -It can be defined as a scientific field of study that harness computer capabilities so that they yield high magnitude of learning and understanding from files like images and non-static images like videos. It aims to automate jobs that the current visual system cannot do. It is a field within deep learning that is getting better every day. There are many areas where computer vision can help which include examining and understanding digital files like images and videos, drawing out of data from the active and stationary real world, etc.

The flow of our model is as below,

1) Importing Dataset: The model starts with importing of the downloaded data set. The data images are segmented into RGB, Black & White, Segmented folders. The RGB folder contains leaf images in RGB color while the Black & White folder contains the images in black and white or grayscale. The segmented folder contains the leaves segmented or cut from the background (please refer fig. 3). This later helps the model in obtaining crucial features that could possibly be missed when the image is analyzed with background. Fig. 2 shows the 38 types of plant leaf categories folders that we will be using in this project.
2) Organize the data set: Next the data set is organized into Train data, Validate data, Test data. The images in Train data will be used for training our neural network, while the validate data will be used to validate results obtained from our trained model. The test data is used to test our model. The trained neural network will be put to test on these set of

images and we will know if the model works as expected or if there are any flaws in it.

Fig. 2: Disease Set



a) Color b) Greyscale c) Segmented

Fig. 3: Data set image types

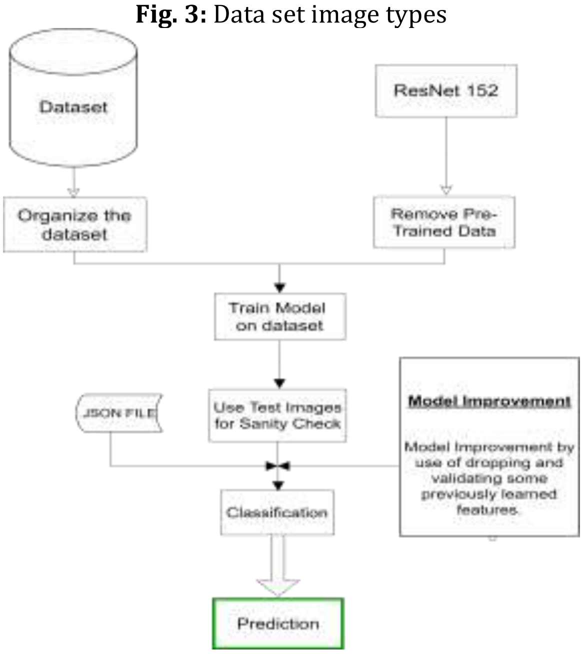


Fig. 4: Architecture of Implemented Model

3) Why ResNet?

ResNet short for 'Residual Networks' is a neural network. It can have a very deep network and it is a subclass of convolutional neural networks. According to [8] It does this by understanding the residual representation functions rather than learning and understanding the signal representation right away. The new concept introduced in ResNet was shortcut connections or skip connections, to fit the preceding layer input to next following layer without changing it. This shortcut attachment allows it to have an in-depth network. ResNet won in image recognition, classification, and localization at ILSVRC [7] 2015. So, we had to pick ResNet as our model.

Below are the benefits of using ResNet:

a) Problems with Plain Network: Usually Conventional deep learning networks contain convolutional layers interconnected with fully connected layers for the classification job, without skipping or changing any connection. Due to deeper layers, the complication of vanishing or exploding gradients may appear in the plain network.

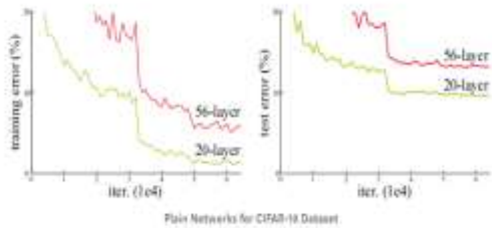


Fig. 5: Vanishing Gradient Problem; Image source: [6]

As seen from the above figure the deeper networks suffer more from vanishing/exploding gradient problem than shallow networks.

b) Skipping Connection in the ResNet: To solve the complication in the areas of vanishing/ and exploding gradients, a skipping interconnection is joined so that the raw input x to the next layer is the output given by the previous layer after few weight layers.

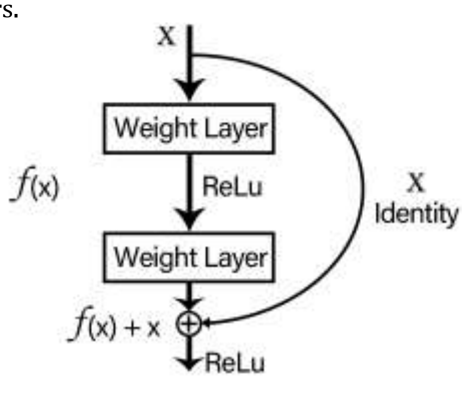


Fig. 6: ReLu

The skip connection in the diagram above is labeled "identity." It allows the interconnection of networks to learn the identity function, which facilitate it to pass the input to the block needed without passing or sending it through the other weight layers. Hence, the output is given as: H(x) = F(x) + x. The weight layers are there to understand the types of residual mapping like F(x) = H(x)−x. This allows us to stack additional layers and build a deeper network, offsetting the vanishing gradient by allowing the network interconnections to skip between some layers which it feels are less needed for training. Even if vanishing gradient occur in the weight layers, we will still have the feature x to get back to the preceding layers.

c) ResNet vs Plain Networks: When a plain network is used, a low layer network is always better. For eg. It is better to use plain network on an 18-layer network than a 34-layer network. For a high layer network Resnet performs better because in a deep network it beats plain networks by introducing skip connections, hence eliminating vanishing gradient problem.

	Plain NN	*ResNet*
18 Layers	27.94	27.88
34 Layers	28.54	25.03

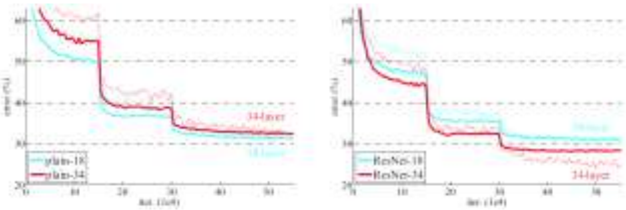


Fig. 7: Plain Networks v ResNet; Image source: [6]

If we compare18-layer plain network and18- layer ResNet, the difference isn't much. This is because vanishing gradient problem does not appear for shallow networks. However, when ResNet is used on 34-layer network, it performs way better. Here vanishing gradient problem has been solved by using skip connections.

4) Feed Forward Mechanism: Feed forward means to transfer a signal in a control system from source to destination through a pathway. Feed forward neural networks is the essential model used here. The main primary aim of a feed forward network is to approximate some mathematical formula. [9]
These models are called feedforward because information flows through the function being evaluated from input to the output. There are no feedback back link interconnection connections in which output of the model is fed back into the system again.

5) Data Training: Here the previously clean and transformed data is trained on the training set. The images in 'Train data' folder will be used for training our neural network, while the 'Validate data' will be used to validate results obtained from our trained model. During training the model will analyze the input data set and find its own meaning. Later on the 'Test data' will be used to test our model. Our trained neural network will be put to test on a set of images and we will know if the model works as expected or if there are any flaws in it.

6) Sanity Checking: Sanity test is a done to check and examine if a output of the calculated computations are possibly true or to eveluate if the produced matter is rational enough. The point of a sanity test is to figure out classes of false results and not to fetch every possible fault. If the result from the NN is irrational, then the result is sent back to a feed forward function. If the result passes the sanity test then it is marked as a acceptable and added to the results.

RESULTS

The Jupyter Notebook has a simple and elegant UI. The home page of the project looks like as in fig 8:

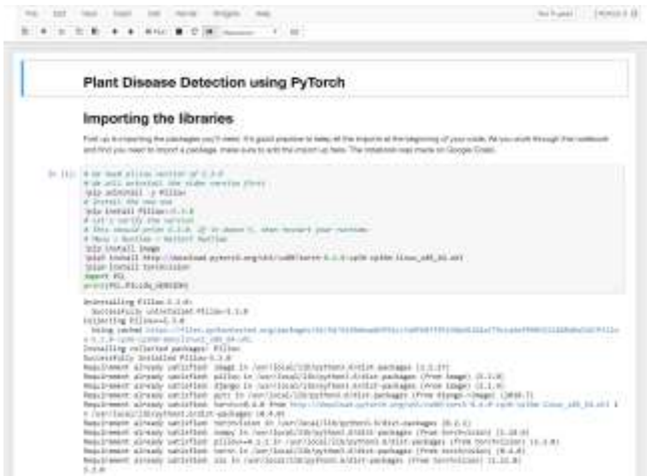


Fig. 8: Home Screen of Jupyter Notebook

```
Epoch 10/10
----------
train Loss: 0.2196 Acc: 0.9290
val Loss: 0.1169 Acc: 0.9621

Training complete in 109m 37s
Best valid accuracy: 0.962105
```

Fig. 9: 10 Epoch Accuracy

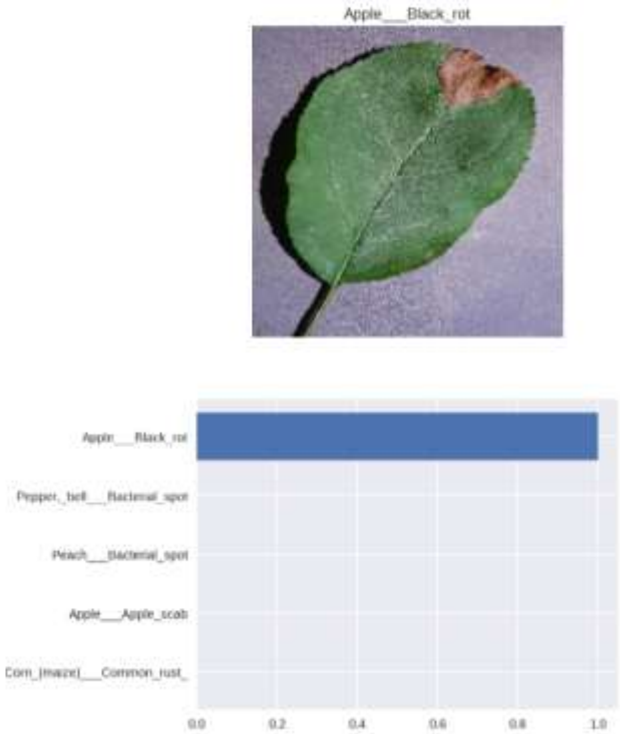


Fig. 10: Healthy Apple Leaf

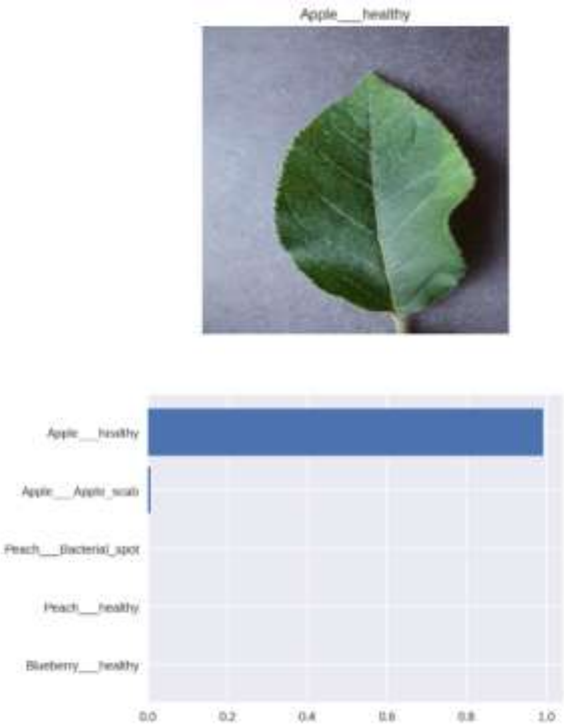


Fig. 11: Apple with Rot disease

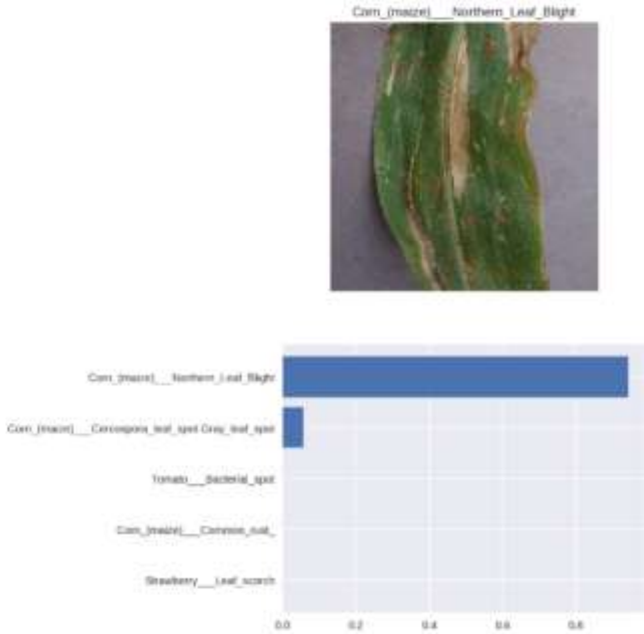


Fig. 12: Corn Leaf with Blight

The results of our classifier gives us an accuracy of 96.211 when trained on 10 epochs (fig 9). The images in fig. 10 and fig. 11 are of apple leaves. One is a healthy leaf and one has black rot disease. In the first case it gives a 98.5% prediction that it is a healthy leaf, but it also tells us that there is a 1.5% chance that it has a apple scab disease. For the diseased leaf, the classifier gives a 1.0 output which means a 100% prediction that it has a 'Apple Black rot' disease which it does have.

The leaf in figure 12 is of corn(maize). When passed through the classifier, it gives a 96.5% prediction that it has blight and a 3.5% chance that it may have gray leaf spot. It does not give us a 100% accurate prediction but an average of 96.21% isn't bad for this approach. The results highly depend on the number of epochs the model is trained on and also on the amount of testing dataset.

We have also developed a web application which gives the project a simple and clean look. It works by hosting a local server. A user just has to select a plant leaf image and the application will display the disease with the highest predicted percentage.

1) Homepage of Web-App

Fig. 13: Web-App Home

2) Selecting a plant leaf image

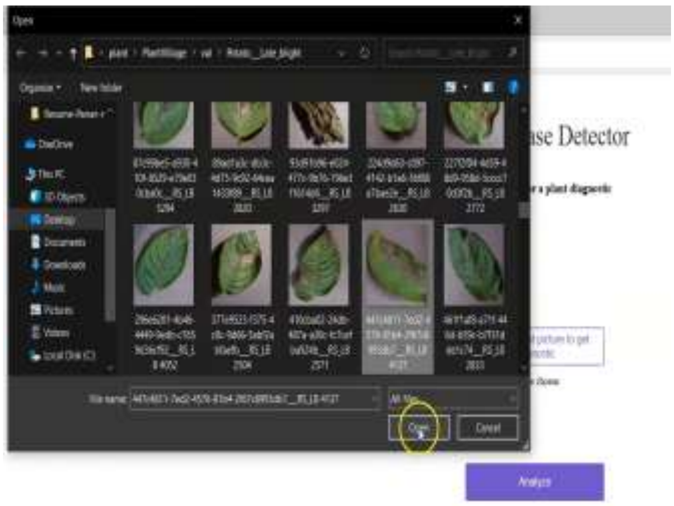


Fig. 14: Selecting Image

3) Analyzing the image

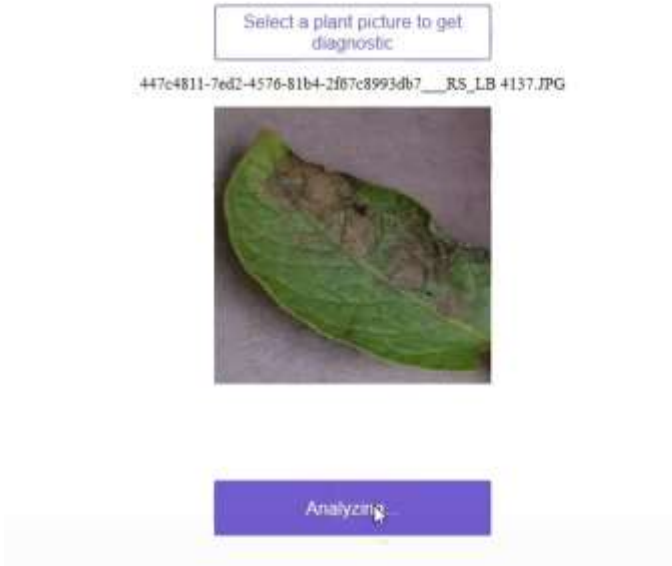


Fig. 15: Analyzing Image

4) Displaying Result

Fig. 16: Displaying Result

CONCLUSION

Humans for centuries have evaluated and produced plant-based food products for fiber, medicine, home, etc. Diseases in plants are just one of the many hazards that must be considered while cultivating crops. Thus, it is important that we enhance the food quality and look to stable agricultural sector as it ensures a nation of food security. The project "Plant Disease Detection using Deep Learning" is aimed at building a neural network capable of detecting 14 crop species and 26 common diseases. Using ResNet 34 as our neural network, the model has given us an accuracy of 96.21%. A complimentary web application was also developed by us to make this project more accessible worldwide. We have hoped to create a project that will help in early detection of plant diseases and we sincerely hope this project poses as a base for further plant disease detection techniques.

REFERENCES

[1] A. Devaraj, K. Rathan, S. Jaahnavi and K. Indira, "Identification of Plant Disease using Image Processing Technique," 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0749-0753.

[2] S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, 2018, pp. 41-45.

[3] Yuan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo, "SVM-based Multiple Classifier System for recognition of wheat leaf diseases," World Automation Congress 2012, Puerto Vallarta, Mexico, 2012, pp. 189-193.

[4] A. Rastogi, R. Arora and S. Sharma, "Leaf disease detection and grading using computer vision technology & fuzzy logic," 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, 2015, pp. 500-505.

[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 Vol. 2, Issue 1, January 2013.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition" arXiv:1512.03385v1, 10 December 2015.

[7] Large Scale Visual Recognition Challenge.

[8] Sik-Ho Tsang, 2018, Review: ResNet — Winner of ILSVRC 2015 (Image Classification, Localization, Detection), https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015image-classification-localization-detection-e39402bfa5d8.

[9] Tushar Gupta, 2017, Deep Learning: Feedforward Neural Network, https://towardsdatascience.com/deep-learning-feedforward-neuralnetwork-26a6705dbdc7.

[10] Hughes, D. and Salathe, M., 2020. An Open Access Repository Of Images On Plant Health To Enable The Development Of Mobile Disease Diagnostics. [online] arXiv.org. Available at: <https://arxiv.org/abs/1511.08060>.

[11] CroudAI, PlantVillage Disease Classification Challenge, https://www.crowdai.org/challenges/plantvillage-disease-classification-challenge.