MODIFIED IMAGE BASED PLANT DISEASE DETECTION USING DEEP LEARNING MODEL

NIKITHA MARAM, DR R. K. SELVAKUMAR, DR, M. JAIGANESH

STUDENT (M. TECH), PROFESSOR, ASSOCIATE PROFESSOR

COMPUTER SCIENCE ENGINEERING
CVR COLLEGE OF ENGINEERING, HYDERABAD, INDIA.

Abstract: Plant disease detection performs a significant role in impressive results in agricultural product yield and quantity. On average smallholder farmers generates more than 80% of agriculture production, for sustainable agriculture production, plant health observation, and disease detection plays an important role. This paper shows the related works on plant disease detection using several algorithms and techniques such as CNN, ANN, SVMs. Here this research proposes a computer vision deep learning model for accurate plant disease detection to overcome the problem. The experimental result shows the 95.2% of model accuracy for the proposed plant disease detection based on the Convolutional neural network. The proposed deep learning model of this research provides a better solution for plant leaf disease detection at an early stage in agriculture production.

Key Words: Deep learning, convolution neural network, disease detection**.**

1. INTRODUCTION

In Telangana [8], according to the 2018 census the state confined 39.644 million of the population. In continues growing of the enormous population of Telangana and cultivation land is declining every day by converting them to commercial lands, this leads to a large number of smallholder farmers or farming homes. So, plant health and disease detection play a significant part in the effective farming of plants within a limited area. It requires an expert person to monitor the plant health observation because the detection of plant disease manually makes difficult and time sophistication which leads to great loss for smallholder farmers. Early diagnosis and accurate detection of plant diseases can control the spread of disease and ensure the healthy development of agriculture production.

This research proposes a computer vision deep learning model to overcome the problem. With the popularity of machine learning algorithms in computer vision, researchers have studied automated plant disease diagnosis based on traditional machine learning algorithms such as random forest, k-nearest neighbor, and support vector machine (SVM) to improve the accuracy and speed of diagnostic results. But since these methods have been selected and applied on the basis of human experience, the recognition level is not as high as expected.

Deep Learning (DL) [9] is a new learning multiple-layer model designed by using Convolutional neural networks with artificial intelligence that intimates the functioning of the human brain in information processing and generating patterns for making a decision. A deep learning CNN works in certain ways as an end-to-end pipeline which can automatically identify the differentiate image classification features. The DL is the application of a nonlinear input conversion and the creation of a statistical model as output. Deep Learning algorithm that is capable of getting an input image with a standard size called Input Layer, the subsequent layers are used to convolution, pooling, fully connected and dropout. A Convolutional neural network was one of the best ways of classifying patterns. Using a dataset of 3651 images of diseased and healthy plant leaves, the proposed deep Convolutional neural network model is trained to identify the 13 diseases of three plant varieties.

PLANT DISEASE:

Plant diseases [13] are the impairment of a plant's ordinary state that interrupts or alters its essential functions. In particular, when a plant is continually troubled by some causative agent that results in an abnormal physiological mechanism that disrupts the ordinary structure, development, function, or other operations of the plant. Figure 1 shows the different types of plant disease images.

Figure 1. Leaf Images from The Plant Disease Dataset

In this research, plant disease is identified through the leaf image. The distinct color of the leaf reflects its functionality, the color may be different with different functionality, among those functionalities; some of the functionalities recognize the disease. The color which split overall leaf that indicates the disease affected, here leaf images go through some deep learning techniques, in this way we can estimate the plant disease name.

Kaggle [12] is an online community of data scientists and machine learners owned by Google LLC. Kaggle enables users to discover and publish data sets, explore and build models in a web-based data science setting, work with other data scientists and machine learning technicians, and join contests to address data science problems.

This paper shows the related works on plant disease detection using individual techniques and algorithms. This paper discusses deep learning CNN methodology and its techniques, this paper also shows the 'kaggle' tool which is an online community for data scientist and machine learners**.**

2. Related Work

This section discussed various methods of image processing for the detection of plant disease.

Sharada P. Mohanty, et.al. [1] developed a Convolutional neural network to identify 14 crop species and 26 diseases, Using a public dataset of 54,306 images of diseased and healthy plant leaves, the dataset has three different versions of images such as Color, Grayscale, Leaf Segmented. The proposed model has been trained and achieves an accuracy of 99.35%.

Sachin D. khirade, et.al. [2] addressed different methods for segmenting the plant's disease portion. It also addressed some methods for extracting the characteristics of infected leaf and classifying plant diseases. Using ANN techniques effectively to classify disease in crops like other techniques self-organizing feature maps, back propagation algorithm, SVMs, etc. Using image processing techniques, they can correctly recognize and classify different plant diseases.

Bin Liu, et al. [3] developed a deep convolutional neural network model to identify the four common apple leaf diseases, using a dataset of 13,689 images of diseased apple leaves, the proposed model is under the hold-out test set, the experimental results show that the proposed apple leaf disease identification approach based on the convolutional neural network achieves an overall accuracy of 97.62%.

Islam et.al. [4] presented a segmentation approach and utilization of SVM demonstrated disease detection in over 300 images, which integrated image processing and machine learning to allow the diagnosis of diseases from leaf images. This automated method classifies diseases on potato plants from Plant Village, which is a publicly available plant image database, and obtained an average accuracy of 95%.

These studies show how CNN, ANN, SVM techniques were widely implemented on various plant disease identification and achieved good results, but in other hand, these studies only implement the CNN based models by externally updating the model. On the other hand, our research used an online tool 'kaggle'.

3. Proposed Model

In the deep learning model, the convolution neural network tries to identify the patterns in the images, by detecting edges first and then goes through filters by changing all parameters. Keras High-Level API handles how we create models, define layers, or set up various models for input-output. First Convolutional neural network layers are used to edge detection through those edges to find other CNN layers. We can create a Sequential model by passing a list of layers.

Multiple Filters:

Convolutional neural networks do not learn a single filter; for a specified input, they learn various parallel characteristics. For instance, learning from 32 to 128 filters in parallel for a specified input is prevalent for a Convolutional layer. This provides the model 32, or even 128, distinct ways to extract characteristics from an input, or many distinct ways to "learn to see" and many distinct ways to "see" input information after practice.

Methodology:

At first, we take an RGB image which learns 32 filters by Conv 2d_1 Conv2D(256,256,32), where convolution is the first layer in an input picture to extract characteristics. Figure 2 shows the CNN architecture of plant disease detection.

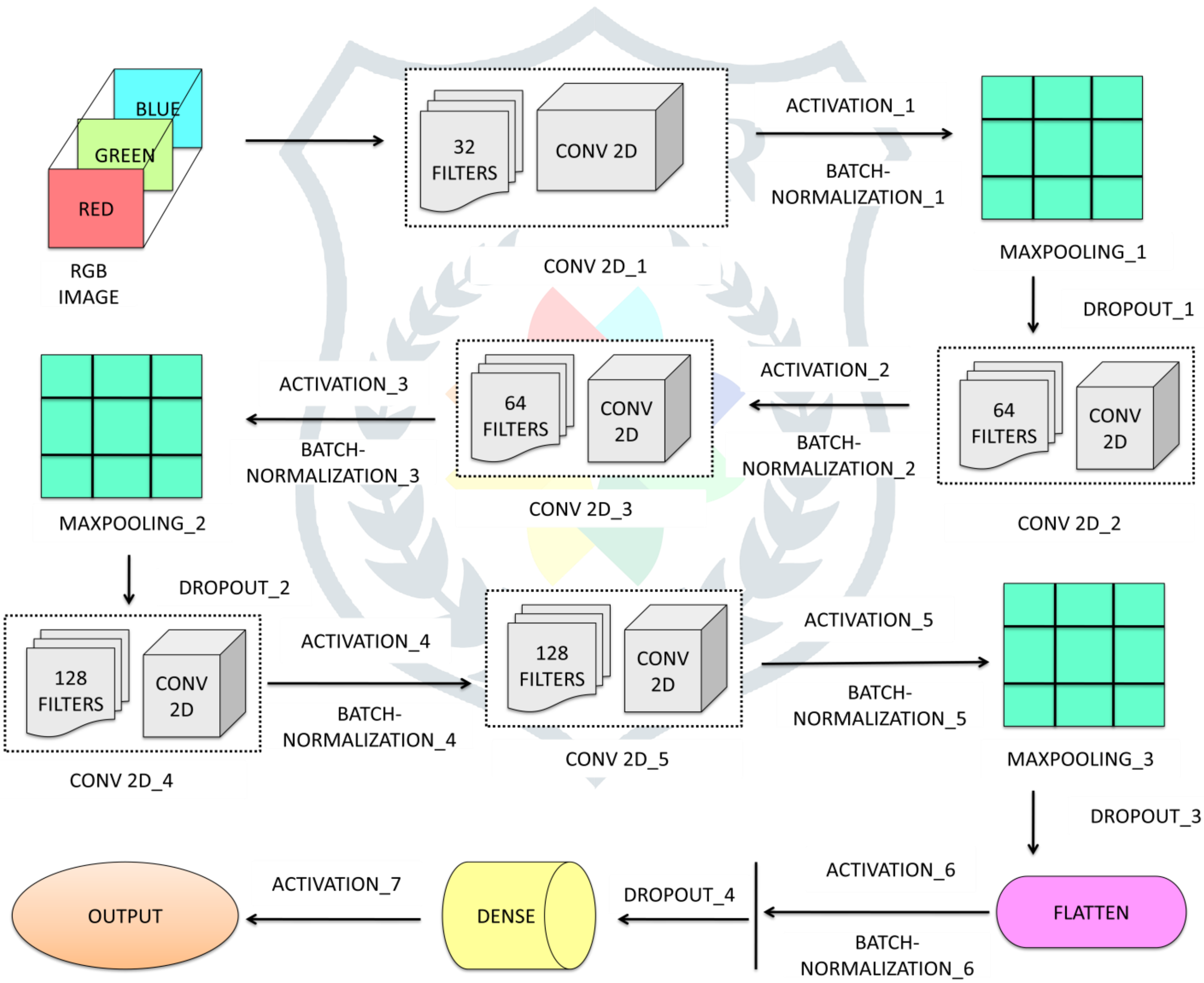

Figure 2. Proposed Deep Learning CNN Architecture

Convolution Neural Network: Convolution maintains the connection between pixels through the use of tiny squares of input information to learn picture characteristics. Here equation (1) and (2) shows the mathematical operation of convolution neural network.

$$\text{Output width} = \frac{W - F_W + 2P}{S_W} \quad (1)$$

$$\text{Output height} = \frac{H - F_H + 2P}{S_H} \quad (2)$$

Where
W = *Width of input image,*
F_W = *filter width,*
S_W = *Horizontal stride,*
P = *Padding,*
H = *Height of input image,*
F_H = *Filter height,*
S_H = *Vertical stride.*

Here it is a mathematical operation that requires two inputs like the image matrix and a filter or kernel. Through activation_1 Activation(256,256,32) and batch normalization_1 Batch(256,256,32), it is used to determine neural network performance such as yes or no. It maps values between 0 and 1 or between-1 and 1 etc. Here Figure 3 shows the system flow diagram of plant disease detection.

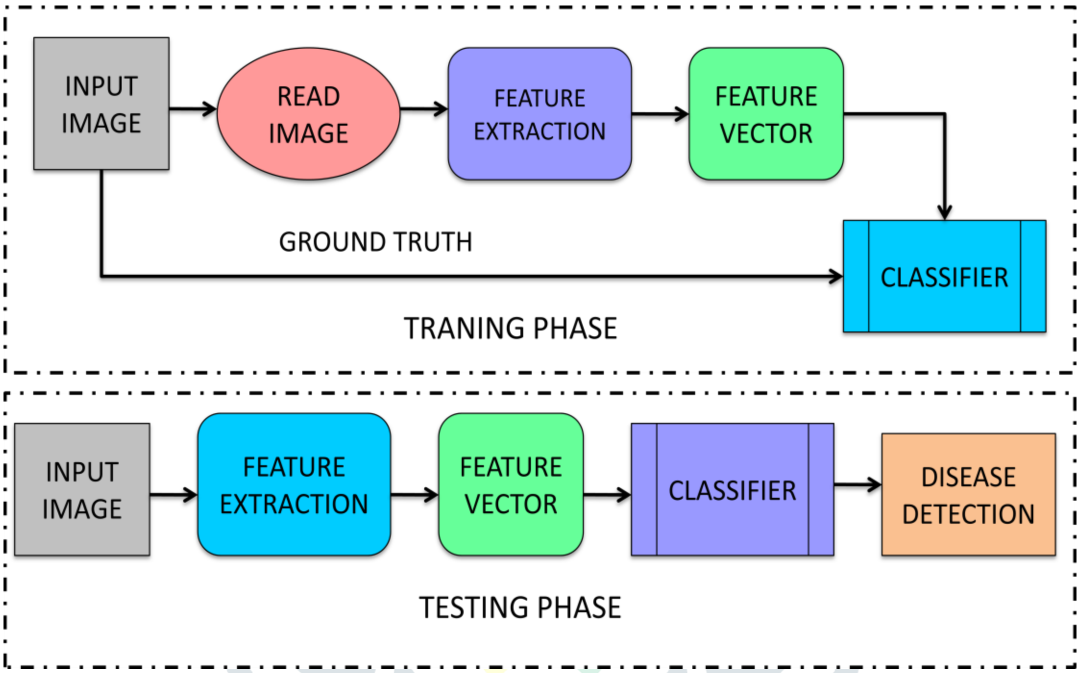

Figure 3. System Flow Diagram of Plant Disease Detection

Dropout: Dropout_1 Dropout(85,85,32) is a method for dealing with this issue. Then conv 2d_2conv2D(85,85,64). And then it goes several layers of CNN where filters been increased up to conv2D_5(42,42,128), Max pooling decreases the representation's spatial size to Maxpooling_3(21,21,128). Dropout is a method where neurons that are randomly chosen during training are ignored. They are randomly dropped-out to Dropout_3(21,21,128).

Flatten: Flatten_1 (Flatten) method transforms a two-dimensional feature matrix into a vector that can be fed into a neural network classifier that is fully attached. **Dense:** Dense_1 (dense) layers, which classify the characteristics obtained by the layers of convolution and down sample by the layers of pooling. Each node in the layer is linked to each node in the previous layer in a thick layer.

4. Experiments and Discussion

Kaggle Experimental Setup and Design
Step 1: open a new account in kaggle.
Step 2: Upload my dataset to kernel.
Step 3: Design my deep learning model.

a) Create a new notebook.
b) Insert the layer model and connect the dataset.
c) For example:

```
model. Add (Convolution2D(32, 3, 3, activation='relu'))
model. Add(MaxPooling2D(pool size=(2,2)))
model.add(Dropout(0.25))
```

d) It will reflect the output.
e) Repeat the steps (b) and (c) until my deep learning design is completed.
Step 4: Save or commit the model.

Comparison of Score Progression across Some Dataset Split as Train-Test Set

The Table 1 discusses about the total number of images in dataset is 3651 which we had three different types of plants with distinct types of disease, at particular dataset consists tomato(2482), potato(657), pepper(512) images, so here we are splitting dataset into two categories they are training and testing, here we defining a sequential model, function which process the image layer by layer by taking kernel size and stride i.e., how many slides it has to move, and activation as parameters and performs several convolutions and it stops executing by using Dropout().

Table 1. Comparison of Score Progression across Some Dataset Split as Train-Test Set

DATA SET	NUMBER OF IMAGES	RUN 1		RUN 2		RUN 3	
		TRAIN	TEST	TRAIN	TEST	TRAIN	TEST
		60%	40%	70%	30%	80%	20%
TOMATO	2482	1489	993	1737	745	1986	496
POTATO	657	394	263	460	197	526	131
PEPPER	512	307	205	358	154	410	102
TOTAL IMAGES	**3651**	**2190**	**1461**	**2555**	**1096**	**2922**	**729**
TOTAL COUNT		**3651**		**3651**		**3651**	

After that, the task of a generator is to generate the images and discriminator to label the images which are generated by the generator, and here we have defined train loss, train accuracy, validation loss, and validation accuracy variables to find the similarity between the healthy planted image and plant diseased image. Once the model is loaded the process requires several epochs (iterations) to produce the efficient output we save the model and run it for more time to save time.

Where we implemented three model stage run with different percentages allocations, such as (60% - 40%), (70% - 30%), (80% - 20%) dataset on a train and test set distribution. In order to address the problem of overfitting, we adapt the test set and train set ratio to find that the model achieves an overall accuracy of 95.2% (model accuracy score 95.2026) even in the extreme case of training on only 20% of the data and evaluating the trained model on the remaining 80% of the data.

Graph Comparison of Score Progression across Some Train-Test Set Split Data

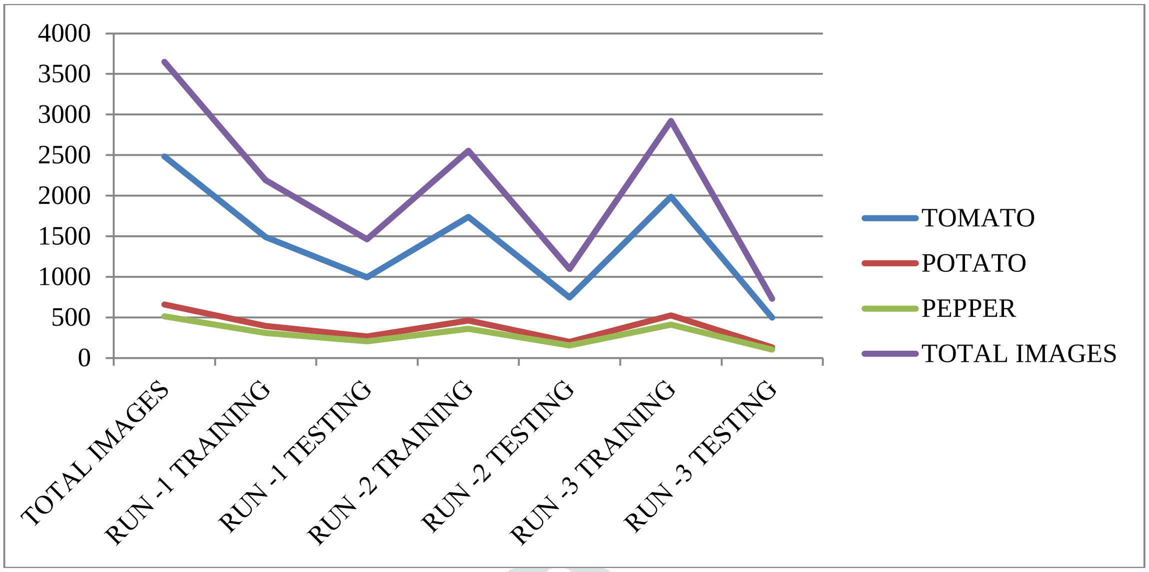

Figure 4. Graph Comparison of Score Progression across Some Train-Test Set Split Data

The Figure 4 shows the comparison of score progression across the total number of images, tomato images, potato images and pepper images where three model stage run with different percentages allocations dataset split as train and test set.

Here Figure 5 and 6 shows the validation accuracy and loss, training accuracy and loss of the model plant disease detection. The result shows that the validation and training process of the proposed model converged almost after 25 epochs and finally reached 95.2 percent accuracy. The reason behind this phenomenon is that the learning rate decreases gradually to almost the invariant, which significantly reduces the updated amplitude of parameters. Moreover, the learned weights of the Convolution Neural Network based model were updated to almost the state of convergence. Later, the learned weights only had a secondary update. As a result, the training process was stable after 25 epochs.

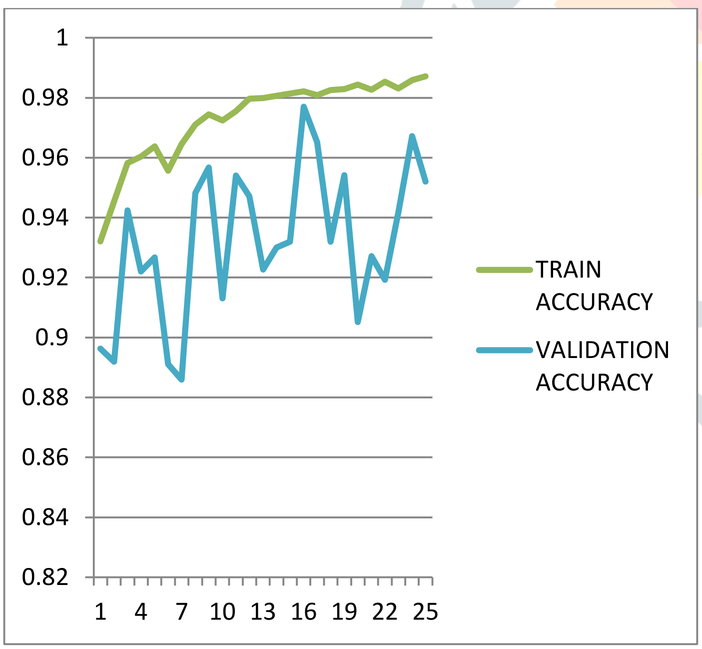

Figure 5. Graph of Model Training Accuracy And Validation Accuracy

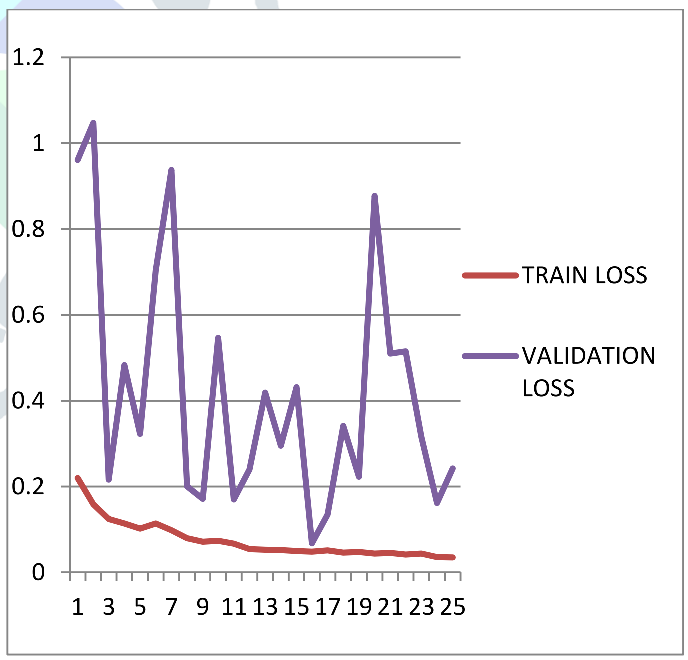

Figure 6. Graph of Model Training Loss And Validation Loss

Here Table 2 shows the images of sample output of plant disease detection. In Table shows at particular different diseased images such as Tomato Septoria leaf spot, Pepper Bacterial Spot, Potato Early Blight, Tomato Late Blight, and Potato Late Blight.

Table 2. Plant Disease Detection

RUN	IMAGE	DISEASE NAME
1		Tomato Septoria Leaf Spot
2		Pepper Bacterial Spot
3		Potato Early Blight
4		Tomato Late Blight
5		Potato Late Blight

5. Conclusion and Future Work

For an effective agriculture production, plant disease detection is very essential for smallholder farmers, accurate plant disease detection and classification is very essential and this can be achieved using computerized deep learning convolution neural network, where CNN which decrease computation time and makes manageable the project through those CNN layers such as convolution, max pooling, fully connected layer for extracting the characteristics of infected leaf and classifying plant diseases using deep learning. The experiment result of plant disease detection shows the 95% of model accuracy for the proposed deep learning model.

Future wok to develop deep learning model with our own tools in python and test the algorithms with various conventional filters like Sobal, Robert etc. And also increase the data set in number of plants with different diseases for increasing the accuracy.

6. Reference

[1] Sharada P. Mohanty,et.al. "UsingDeepLearningforImage-BasedPlantDiseaseDetection", frontiers in plant science, 22 SEPT, 2016.

[2] SachinD.khirade,et.al "plant disease detection using image processing" IEEE explore digital library,2015.

[3] Bin Liu and Yun Zhang, et.al. "Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks", MDPI Symmetry, 29 December 2017.

[4] Islam, M.; Dinh, et.al. "Detection of potato diseases using image segmentation and multiclass support vector machine". IEEE Canadian Conference on Electrical and Computer Engineering, Windsor, ON, Canada, 30 April–3 May 2017.

[5] Ms. Monika Gupta, et.al. "Plant Disease Detection using Digital Image Processing"., International Journal of Innovations & Advancement in Computer Science, Volume 7, Issue 5 May 2018.
[6] Kiran R. Gavhale, and U. Gawande, et.al. "An Overview of the Research on Plant Leaves International Journal of Pure and Applied Mathematics Special Issue Disease detection using Image Processing Techniques," IOSR J. of Compu. Eng. (IOSRJCE),vol. 16, PP 10-16, Jan. 2014.
[7] SandeshRaut and AmitFulsunge, et.al. "Plant Disease Detection in Image Processing Using MATLAB" International journal of Innovative Research in Science , Engineering and Technology, Volume 6 ,2017.
[8] www.indiapopulation2019.com/population-of-telangana-2019.html
[9] https://www.investopedia.com/terms/d/deep-learning.asp
[10] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
[11] http://cs231n.github.io/convolutional-networks/
[12] www.en.wikipedia.org/wiki/Kaggle
[13] www.britannica.com/science/plant disease