Potato Leaf Disease Detection Using Deep Learning

Ms. Deepanshi Samant
Department of Computer Science and Engineering
Graphic Era Deemed to be University
Dehradun, India
deepanshisamant@gmail.com

Mr. Vaibhav Bora
Department of Computer Science and Engineering
Graphic Era Deemed to be University
Dehradun, India
vab.bora@gmail.com

Ms. Rangoli Dhawan
Department of Computer Science and Engineering
Graphic Era Deemed to be University
Dehradun, India
rangolidhawan007@gmail.com

***Abstract*—This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. **CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.* (*Abstract*)**

Keywords—component, formatting, style, styling, insert (*key words****)***

I. INTRODUCTION

Agriculture is undoubtedly the backbone of a nation and the developed status of a country directly or indirectly depends upon the plantation and agriculture. Potatoes are one of the most widely grown and consumed crop in the whole world. They provide us great nutrition values such as carbohydrates, Minerals and vitamins. It shields us from various diseases such as spike in blood pressure, cardiovascular diseases, cancer etc. They are an essential part of our balanced diet. It does not get perished easily. Also growing them is also not a cumbersome task. India is one of the largest producer of potato crop in the whole world. Potato is responsible for the significant crop production statistics in India. It's contribution is over 27.8 percent of the total crop which is grown in our country. We collectively cultivate around 47.9 percent million tons in a year. In a study it was found out that Uttar Pradesh produces the largest amount of potatoes in India which is estimated around 31.4 percent. Potatoes are grown over more than 22 states in India. The highest producer of Potato in the whole world is China which contributes over 24.5 percent , then comes India with 15 percent followed by Russia with 7 percent. Day by day, the population of planet Earth is increasing at a very fast pace, so it is very important for us to find out the corrective measures to decrease effect of diseases on potatoes so that we can yield the maximum output and also to meet up with the future requirements of the human race. But due to some plant diseases the health of the potato plant gets adversely affected which leads to great economic losses which directly affects the nation's economy as well. Due to illiteracy and poor knowledge of botany and plants farmers used to suffer terrific amount of economic losses. The traditional methods used by farmers were not up to the mark. In earlier times the farmers used to detect if the plant is healthy or it has some sort of disease. This method is obviously not very productive and also very cumbersome. So with the help of new technology we can detect the problem very easily within fraction of seconds and it also provide very optimized results. It is very important to find out the ailment and it's root cause so that the amount of losses can be lessen or prevented, so that we can yield the maximum output. The earlier the problem is detected the less will be the economic losses. The well-being of the potatoes can be affected by various things such as climatic conditions, worms and specific ailment such as early blight and late blight. Ailment called early blight is generally caused by fungus. It's species is known as Alternaria Solani. This fungus can survive under the temperature of 25 to 30 degree Celsius.

Fig. 1 Potato Leaf Disease Classification Input image.

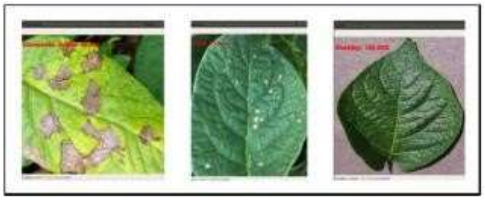


Fig. 2 Potato Leaf Disease Classification Output image.

The symptoms of this ailment are brownish dark colored circular rings. It has comparatively less destructive effects than the late blight ailment. The root cause of this type of infection is the wetness which prevail around the plant. On the other hand oomycetes is the root cause of late blight. It's species is known as Phytophthora infestans. It prefers a cold temperature of around 11 to 17 degree Celsius. The symptoms of this ailment are formation of tiny spots of color ranging from light to dark green color and also cotton like white structure is easily seen on the leaves which is fungus. We have a dataset called Plant village which has 4304 images of different leaves ranging from early blight, healthy and late blight disease leaves. In our project we will give our input image and our model will then predict weather the potato leaf is disease free or not. And if there is any disease then it will figure out whether it is early blight or late blight. We will use Deep learning in our project for optimum result. We will also use various algorithms such as ANN, CNN, and apply the one which yield us the most accurate results.

II. LITERATURE REVIEW

In recent times, many research papers have been published for identification of potato leaf disease and its classification. All of the works have used different techniques to achieve their respective task. M. Islam et al [6] detected the diseases in potatoes by using support vector machine and image classification. In the dataset there were

three hundred images in which there were two hundred leaves which has diseases such as early and late blight and there were hundred leaves which were healthy. It uses grey level co-occurrence matrix for the segmentation. The accuracy of their model was found out to be 95 percent. R Meena Prakash et al [7] used K- mean clustering for the removal of the background of the images of potato leaves. Then the grey level co-occurrence matrix was used for the extraction of the texture of the leaf with help of GLCM we can find out the abnormal leaves or the leaves which are not of color green. At last, SVM was used, and an accuracy of 90 percent was secured.
TT Mim et al [8] model has six types of diseases in the tomato leaves. In this problem, convolutional neural network was used. An accuracy of 96 percent was achieved against validation and 92 percent against training. Akhtar et al [9] used support vector machine on the foundation of discreate cosine transformation and with help of this the rate of efficiency became very high. AT Madhavi et al [10] model's main specialty was the use of Dense Net which is used for the classification of different diseases in the leaves of the tomato, capsicum, and potatoes respectively. Soma Ghosh et al [11] stated that for the exploration to build a classifier which can be used to classify the disease in the potato leaves, RGB images is used. In this an enormous sized dataset should be used. The ML model learns the various aspects and provide better results. Divyansh Tiwari et al [12] project's main attraction was the making of one automated system for finding out if the leaf is disease free or not. It is useful for the crop cultivators as it informs them about the disease early. The accuracy of the model was 98 percent. S Jeyalakshmi et al [13] model uses various machine learning algorithms such as decision trees and the neural network is further used and then the comparison takes place. Dr R Dhaya et al [14] proposed that for higher accuracy, more than one step was used for the identification of the Furasium Oxysporim in leaves of the tomato. Athanikar et al [15] checks on different aspects such as the place, color of the leaves of the potato. In this K-Man agglomeration is used. They achieved an accuracy of 92 percent. The main highlight of the project was the use of back propagation.

III. PROPOSED METHODOLOGY

A deep learning neural network model that is a member of the convolutional neural network family is the convolutional neural network (CNN). CNNs represent a major improvement in image quality. This technology is used for recognition. These are most usually used to look at visual images and are extensively used in picture classification. Convolutional Neural Network (CNN) is an example of a neural network that only accepts input in two dimensions . It becomes clear that using such neural networks in the pre-processing of the image is necessary because images can be represented in a similar way. The ability to produce certain features that can be seen in an image makes convolution special. Before creating a feature map that summarizes the presence of observed features, it filters the input. These neural networks learn the filters during training.

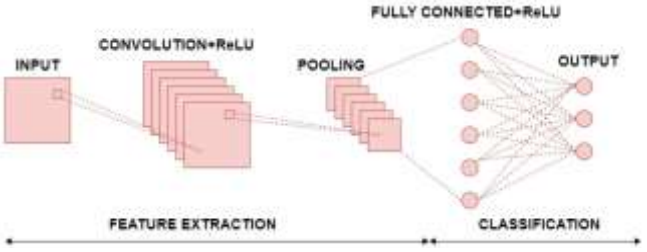
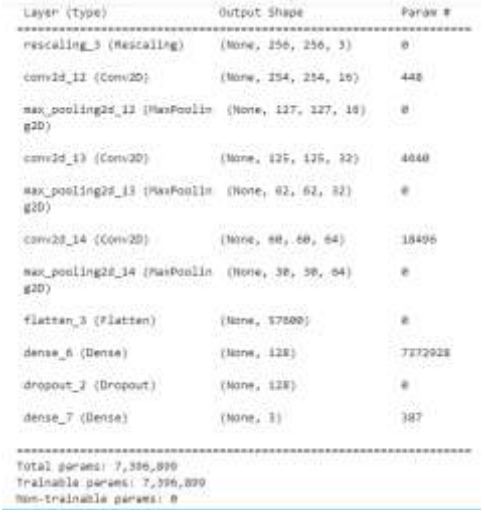


Fig. 3 Architecture of CNN

The structure of a CNN looks like this:
1. Convolutional (conv) layer
2. ReLU (Rectified linear unit) layer
3. Pooling layer
4. Pair of fully-connected (FC) layers

Convolution layer: Most of the computation happens at the convolution layer. These layers seek to extract image features using ReLU and pooling layers, as well as performs convolution operations using a kernel based on image features. To improve its ability to recognize patterns in photos, we add convolutional layers. No matter where a pattern emerges in the image, convolutional layers are capable of finding it. First, we'll call model.add to add the layer. The two types of convolutional layers available which are 1D and 2D. As we're working with images we'll need to include the 2D convolutional layers. In our CNN model we started from 16, Conv2D will be added, along with 16 filters, a three by three window size, and an activation function that uses relu once more and has the formula activation equals relu and it will move ahead by adding 32 and then by 64.
In this study's proposed CNN architecture for disease detection in potato leaves, three convolutional layers and three maxPooling layers are used.



Layer (type)	Output Shape	Param #
rescaling_3 (Rescaling)	(None, 256, 256, 3)	0
conv2d_12 (Conv2D)	(None, 254, 254, 16)	448
max_pooling2d_12 (MaxPooling2D)	(None, 127, 127, 16)	0
conv2d_13 (Conv2D)	(None, 125, 125, 32)	4640
max_pooling2d_13 (MaxPooling2D)	(None, 62, 62, 32)	0
conv2d_14 (Conv2D)	(None, 60, 60, 64)	18496
max_pooling2d_14 (MaxPooling2D)	(None, 30, 30, 64)	0
flatten_3 (Flatten)	(None, 57600)	0
dense_6 (Dense)	(None, 128)	7372928
dropout_2 (Dropout)	(None, 128)	0
dense_7 (Dense)	(None, 3)	387

Total params: 7,396,899
Trainable params: 7,396,899
Non-trainable params: 0

Fig. 4 CNN Model

ReLU: An activation function called a rectified linear unit (ReLU) addresses the problem of vanishing gradients and adds the property of nonlinearity to a deep learning model. ReLU(z)=max(0,x)

Mainly used for hidden layers.
Leaky Relu= max(0.1x,x)
The RELU layer expedites training by converting all negative values to zero.

The pooling layer: Because neurons in feature maps share weight, it helps to reduce the spatial resolution of feature maps. The operations that the pooling layer computes include average, maximum, multiscale order less, and stochastic pooling. A two-dimensional dataset is condensed into a single feature vector using the flatten layer. In max pooling, the output of the convolutional layers is scaled down by keeping only the greatest values and discarding the smaller ones.

Fully connected layer: Neurons in this layer are completely connected to those in the layers above. The fully connected layer takes a feature vector as input and recognizes an input image using the softmax function. Back propagation will be employed in this layer to lower the error value.

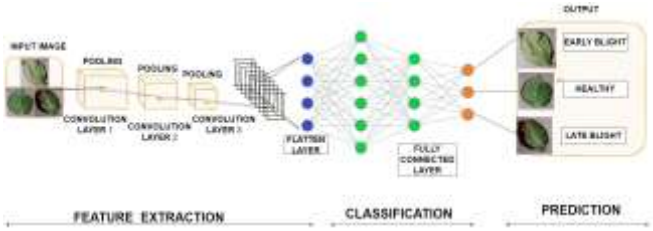


Fig.5 CNN working method in Plant leaf detection

The CNN working method in Plant leaf detection categorization is shown in the image above. It will accept an input image in the form of a 2D vector representation, do the above-mentioned operation on it, and then classify the supplied image.

The field of agriculture can benefit greatly from image processing. It can be used to detect disease, identify insects, and determine the size and shape of the fruits of various crops. Data collection, image preprocessing, image segmentation and feature extraction are some of the phases that make up the image processing technique. Each of the phases involved in the process of identifying disease from plant leaf photos can be addressed as follows:

i. Data Collection
ii. Image Pre-processing
iii. Image Segmentation
iv. Feature Extraction
v. Processed Data
Training Data
Testing Data
vi. Model
vii. Classification
viii. Evaluate Metrics

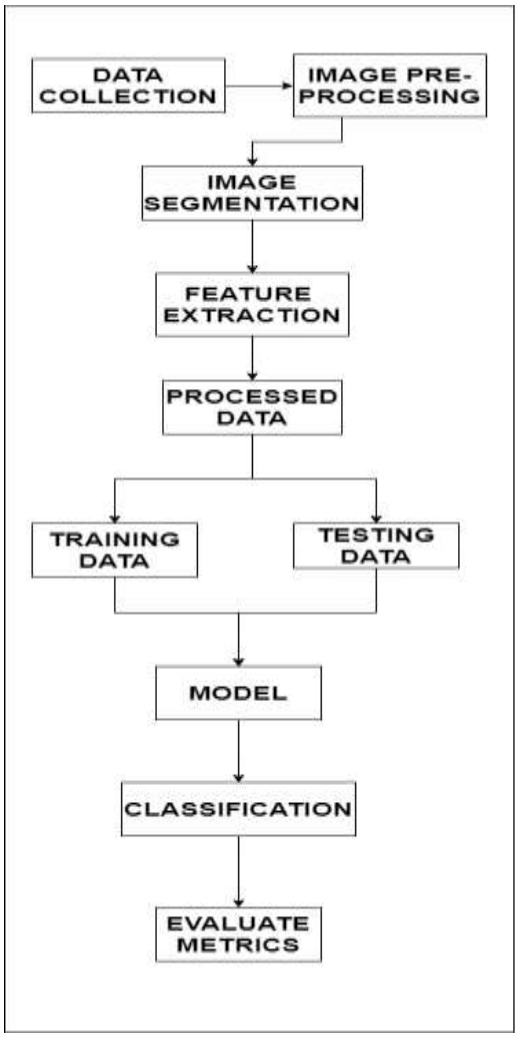


Fig. 6 Flowchart

Data Collection:
Data gathering is the first stage in evaluating and putting any algorithm into practice. The more data there are, the more accurately the result may be predicted. The gathering of data is the first step in the proposed strategy. This collection contains about 2150 photos that cover every sort of leaf pattern, including healthy leaves and leaves with early and late blight diseases.

Disease	Type of Disease	No. of Image
Early Blight	Fungal	1000
Late Blight	Fungal	1000
Healthy	No Disease	150

Fig. 7 No. of images and type of diseases

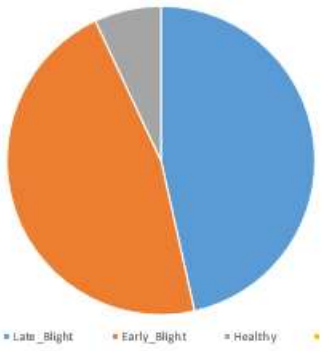


Image Pre-processing:
This stage of the framework is crucial. This phase largely focuses on picture improvement, denoising, and maintaining uniform image sizes across all photos. To segment the photos more effectively, denoising and augmentation of the images are necessary.

Image Segmentation:
The image will be divided into segments in this stage based on the area of interest. The areas on the leaf that are damaged by various diseases in this scenario are the ones that need to be distinguished from the previous photos as the region of interest.

Feature Extraction:
Determine the patterns that exist based on the obtained region of interest. In light of the fact that different areas of interest will exhibit distinct patterns, it is possible to extract features that are essential for both detection and classification. Using feature extraction, the dimension of an image can be reduced without losing any important data. Moreover, the unnecessary feature is removed.

Training Data:
From the data that has been processed, training data will be gathered. To train the classifier model, about 80% of the data with random indexing were taken into consideration.

Testing Data:
The processed data will also be used to collect the testing data. To test the classifier model, about 20% of the data with random indexing were taken into consideration.

Classification:
To categorize the photos into several categories, such as Late Blight, Early Blight, and Healthy, test data will be sent to the trained classifier. CNN is one neural network used for image classification as it is giving highest accuracy results when compared to all the other algorithm.

Evaluation Metrics:
Evaluation metrics including precision, recall, F1-score, and accuracy will be determined based on the classifier model's performance.

IV. Results and Evaluation

Epoch is a neural network training procedure that runs through all datasets in a single round before returning to the initial stage. Because the amount of data utilized in training a neural network model is rather enormous, it is necessary to divide the data rate per batch because using only one epoch will be too taxing on the training process (batch size). There were 32 batch sizes utilized in this model.
The image of a potato leaf that has been separated by the fit model must then be trained.

EPOCHS	DATA TRAINING		DATA TESTING	
	Accuracy(%)	Loss(%)	Val Accuracy(%)	Val Loss(%)
1	64.40	77.28	81.86	41.42
3	88.79	29.40	87.91	36.49
7	96.46	10.16	92.33	25.81
10	96.40	9.89	93.95	15.91
13	97.04	8.09	96.05	11.00
15	98.25	2.567	97.21	10.21
17	98.66	4.47	93.95	15.26
20	98.72	2.56	96.74	11.23

Fig. 8 Results of the model

Above table displays the classification results on the train data in a 80:20 ratio. The results of the first epoch show a train data accuracy value of 64.40% and loss value of 77.28%. This pattern continues until the results of the 20th epoch show an accuracy value of 98.72% with a loss value obtained at 2.56%. Although in the validation data, the accuracy value at the first epoch is 81.86%, with a loss value obtained of 41.42%, and so on until the 20th epoch displays an accuracy value of 96.74%, with a loss value obtained of 11.23%.

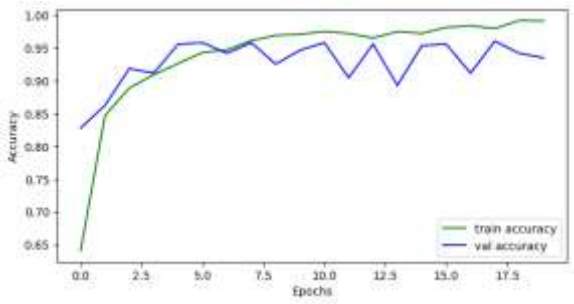


Fig. 9 Accuracy Plots

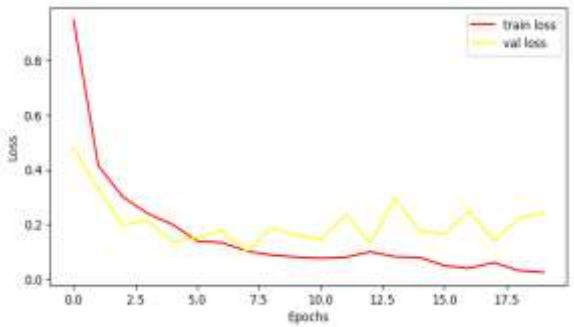


Fig. 10 Loss Plots

With the 80:20 data partition, the graph in Fig. 9 represents accuracy, whereas graph in Fig. 10 represents loss. The green line represents train accuracy, and the blue line represents validation accuracy in Fig. 9 and The red line represents train loss, and the yellow line represents

validation loss in Fig. 10 The improved accuracy and reduced loss on each epoch in this graph demonstrate that the fit model was successful because they are stable over time.

V. CONCLUSION

This paper analyzed many other papers, which came up with their research to detect the disease in potato leaves using deep neural networks. Convolutional neural network layers are used in the proposed method to determine the health status of the potato leaf. There are two types of diseases that can be found in a potato leaf, one is Early blight and second is Late blight. An adequate database with ample amount of images and subsequent image processing is required to increase the accuracy of CNN network. This study proposes a model that makes advantage of modest CNN network's resources. It is most suited architecture that can be used to judge plant diseases with an image database of more than 2000 images is used and less amount of resources. In addition, the accuracy of the judgement can reach upto 98.72%.

REFERENCES

Journal Paper

[1] Kumar Sanjeev, Narendra Kumar Gupta, W. Jeberson and Suneeta Paswan “Early Prediction of Potato Leaf Diseases Using ANN Classifier”,2020 https://www.computerscijournal.org/vol13no23/early-prediction-of-potato-leaf

Conference Paper

[1] Divyansh Tiwari, Mritunjay Ashish, Nitish Gangwar, Abhishek Sharma, Suhanshu Patel and Suyash Bhardwaj, "Potato Leaf Diseases Detection Using Deep Learning", 2020 https://ieeexplore.ieee.org/document/9121067

[2] S Jeyalakshmi and R Radha, "An effective approach to feature extraction for classification of plant diseases using machine learning", 2020. https://indjst.org/articles/an-effective-approach-to-feature-extraction-for-classification-ofplant-diseases-using-machine-learning

[3] Sindhuja Bangari, P Rachana, Nihit Gupta, Pappu Sah Sudi, Kamlesh Kumar Baniya “A Survey on Disease Detection of a potato leaf using CNN”,2022, https://ieeexplore.ieee.org/document/9742963

[4] Md. Khalid Rayhan Asif Md. Asfaqur Rahman Most. Hasna Hena “CNN based Disease Detection Approach on Potato Leaves”,2020 https://www.researchgate.net/publication/350077258_CNN_based_Disease_Detection_Approach_on_Potato_Leaves

[5] Md. Asif Iqbal, Kamrul Hasan Talukder “Detection of Potato Disease Using Image Segmentation and Machine Learning”,2020

[6] M Islam, A. Dinh, K. Wahid and P. Bhowmik, “Detection of potato diseases using image segmentation and multiclass support vector machine.”

[7] R Meena Prakash , G P Saraswathy , G Ramalakshmi , K H Managleshwari, and T Kaviya, “Detection of leaf diseases and classification using digital image processing.”

[8] T T Mim, M H Sheikh , R A Shampa , M S Reza, M S Islam “ Leaves Diseases detection of tomato using image processing.”

[9] Akhtar, Asma, et al. "Automated plant disease analysis (APDA): performance comparison of machine learning techniques."

[10] A T Madhavi, M M Yugesh, Srinidhi , C S Vijayasundar,”Detection of leaf disease using convolutional neural network”

[11] Soma Ghosh, Renu Rameshan and Dileep A. D.,” An Empirical Study on Machine Learning Models for Potato Leaf Disease Classification using RGB Images”.

[12] Divyansh Tiwari, Mritunjay Ashish, Nitish Gangwar, Abhishek Sharma, Suhanshu Patel, Dr. Suyash Bhardwaj, “Potato Leaf Diseases Detection Using Deep Learning”