Potato Late blight disease detection using Convolutional Neural Network (CNN)

1. Introduction:

Bangladesh is one of the most important producers of the tuber crop. It delivered a record high of 109 lakh tones a year, as per the Department of Agricultural Extension (DAE). With a yearly normal interest of around 70 lakh tones, the nation saw an excess of around 40 lakh tones. It is currently the third most significant food thing of Bangladesh by weight creation and broadly developed in all the areas during winter. The key factors that influence the plants and its items are arranged into two classifications: 1. Infections 2. Disorder. The infections are the biotic elements that are either brought about by the organisms, microbes or green growth. The issues are the abiotic factors brought about by the temperature, precipitation, supplement insufficiency, dampness and so forth [1].

There are several insect pests and diseases are related to potatoes in field. Late blight (*Phytophthora infestans)* disease is one of the major diseases caused by fungi in our country. The late blight causes damage to the potato plants at vegetative stage with high infection intensity, but the severity of damage is controlled by the farmers through routine application of fungicides in the potato field. It is required the automatic identification and diagnosis of potato disease in the field of agricultural information.

The traditional technique for recognizing potato infections requires heaps of specialists' experience and information which is frequently tedious, difficult. With the improvement of computer and web innovation, ranchers can look through the potato infections pictures information base or counsel the plant pathologist to pass judgment on potato sickness distantly. The disadvantage is that the wrong judgement and its service is inefficient. Consequently, a programmed and precise identification strategy for potato illnesses is expected in a speediest conceivable time for the cultivator's end. With the coming of Computer Vision (CV), Machine Learning (ML) and Artificial Intelligence (AI) advances, progress has been accomplished in creating computerized models engaging, exact, and in an ongoing recognizable proof of sickness from the tainted leaves. Over most recent couple of years, it has been perceived that Deep Learning (DL) has been predominately used to create, control, maintain, and upgrade horticultural creation. In this research, a Convolutional Neural Network (CNN) based strategy is proposed for sickness discovery utilizing pictures of sound and tainted leaves taken from primary and PlantVillage dataset.

2. Background study:

Various deep learning designs are accessible to use for the arrangement, design analysis, highlight extraction and change. Among these, CNN is a standout amongst other performing methods for object recognition. It can automatically take in proper features from preparing datasets with no intercession from the client end. It has been demonstrated to be a decent choice for plant disease recognition. Four cucumber infections named as anthracnose, downy mildew, powdery mildew, and target leaf sports are classified from the leaves in [2] utilizing a deep CNN model. Ferentinos [3] has proposed a VGG style engineering for the identification and classification of the plant leaves. Too et al. [4] have utilized four distinctive deep CNN models (VGG 16, Inception V4, RestNet, and DenseNets) for diseases characterization from a picture of plant leaf. The DenseNets network offers higher characterization precision and lesser computational time when contrasted and different designs.

In [5], Picon et al. have utilized profound CNN for three fungal illnesses grouping found in the wheat plant. Lu et al. [6] have proposed a profound CNN for the arrangement of ten distinct kinds of rice leave illness from the storehouse of around 500 pictures containing both the healthy and contaminated pictures. Creators have received the 10-crease cross validation system for accomplishing higher characterization results. Alexnet and then SqueezeNet profound learning network has been utilized by Durmus et al. in [7] for the grouping of plant leaf infections. The pictures are taken from the plantVillage information base for the tomato plant leaf pictures in ten distinct classes. Jain et al. [8] proposed a CNN model for the real-time arrangement of the plant disease detection system. Mango Leaves tainted by Anthracnose infection were arranged in [9] utilizing a Multilayer CNN. The writers examined other condition of-craftsmanship draws near and got the higher precision in order contrasted with Particle Swarm Optimization (PSO), Support Vector Machine (SVM) and Radial Basis Function Neural Network (RBFNN).

Thusly, Barbedo in [10] have present different issues and challenges in the arrangement of plant illnesses. The creator has researched this work with twelve unique plants having various credits and with various infections. GoogLeNet and Cifar10 network have been introduced by Zhang et al.in [11] for order of infections from the maize leaf pictures. The proposed model accomplishes higher exactness when contrasted and different networks like VGG and AlexNet for grouping nine unique kinds of maize leaves. Gandhi et al. [12] have worked with Generative Adversarial Network (GANs) and CNN for the recognizable proof of illnesses from the plant leaf pictures utilizing a portable application. As per the writing study, it illuminated to utilize a deep CNN model to detect the tainted one from a dataset of potato plant leaves.
Based on the study, it is found that the deep CNN model gained the best accuracy among the other state of the art technologies for image classification and recognition system. The CNN works in a hierarchical way which build network like a funnel and finally gives the output with a fully-connected layer where all the neurons are connected to each other. The CNN model handles billions of the data that's why it is more reliable in performance. This data driven model perform the best classification prediction among the state of the art technologies. In this research our goal is to detect Potato Late blight disease using Convolutional Neural Network (CNN).

3. <u>Objectives of the Project</u>:

The objectives of the research are to:

a) To collect real-time images of infected potato leaves and healthy leaves.
b) To classify the infected potato leaves from healthy potato leaves.
c) To detect the Late Blight disease from the leaves.
d) To implement the model in Android Phone using CNN model.
e) To compare the proposed model with ANN & SVM.

4. <u>Methodology:</u>

The methodology of the proposed research is divided into eight major parts:

a) Real Time Images Collection:

Machine learning techniques are able to predict based on the input data. First the model learn about the input data then it tries to generalize to predict based on the data. In this research, we will work on image based data. The neural network requires large amount of data to learn to work well. It requires minimum 10 times the data of the weight parameters. But simply, we can say the more data we have the better it will work. If we have limited number of data, there is a big change that the model will not work accurate and it will not be a reliable for use. There are several options to collect real time images. We have a wonderful government organization named Bangladesh Agricultural Research Institute (BARI), Gazipur that works for the agricultural research for the farmers. We will communicate with them to collect the images of both healthy and diseased images. We have to go to the potato field to collect images by myself.

b) Preprocessing and Training:

The model learns from the training data. We will fed the entire collected dataset to the model. Initially, the entire dataset is divided into three parts- training, testing and validation set. The whole data set splitted into three: 1) the training set comprises about 80% of the images, 2) the testing set constitutes about 10% of the images and 3) the validation set constitutes about 10% of the images. The training dataset will be augmented using operations of rotation, mirroring, addition of Gaussian noise, brightness adjustment and contrast enhancement. Thus, we will increase the size of the dataset to make the model more accurate, reliable. And the overfitting problems will reduce. The final accuracy for the network will obtain using a 10-fold cross-validation approach. After augmenting the entire dataset, we will fed the training set to the model, after learning we will tune the parameters of the model using validation set and finally we will test the model using the test set. We will calculate the accuracy of the model after testing.

c) Classify the diseased potato leaves:

To classify two different images, human brain extracts the features of the images and then match the variation of the features. Like our brain, CNN will work similarly. It will extract the features of the both images then it will classify the images according to the class name of the images that is defined as the label of the images. What is an image for neural network? Image is nothing but the numeric values of higher dimensional matrix stored as from 0 to 255. And for the color and grayscale image the dimension channel is different. Human can easily classify same grayscale and color image via just going through the image. But for the neural network they don’t have the capabilities like our brain. They generate the probabilistic result based on the input and its learnt feature at the output layer of the network that we interpret the final output. It will try to cluster the diseased potato leaves in the diseased class and healthy potato leaves in the healthy class.

d) Detect Late Blight disease from leaves:

During the training period, the model learns from the training data with its appropriate label. After training period, the model will be able to generalize which one is late blight and which one is healthy leaves. From its experience, it will try to match the input image to its learnt experience and it will output the probabilistic result according its availability.

e) Validation:

Initially, the dataset was divided in three sets– training, validation and validation set. After training period, the models need to tune its parameters to increase its accuracy and decrease the error. It is better to use an unseen set of data so that the model stops overfitting. For this reason, we will use the validation set for the validation. Using the validation set, we will tune the number of layers, number of neurons in each layer, number of epochs for training, learning rate, dropout rate etc. so that the model performs well during the testing period.

f) Testing:

There is a toolkit named TensorBoard- TensorFlow's visualization toolkit where the visualization is performed using several curves for training, validation & testing. It's better to use the Tensorboard to reach the local minimum using several error functions during the entire process. During the testing, several images are shown to the model from the testing set i.e. unseen for the model to check the generalization rate. If it has the higher rate of generalization, we will use the model.

g) Implementation in Android Phone:

In this digital era, Smartphones are available to everyone. It will be more effective for the farmers if it's possible to implement the model in Android Phones so that everyone can access it to detect the disease. There is a wonderful tool named Tensorflow Lite an open source deep learning framework that helps to run the model in smartphone. We will use this framework to run the model in smartphone.

h) Compare with the other state-of-art approaches

There are several modern approaches available for image classification and disease detection. We have to choose the best one for my research. During the background study, we already mentioned the importance of CNN in classification and detection. But that was theoretically proved. So, we have to compare our model with different modern state of art approaches like Artificial Neural Network (ANN) & Support Vector Machine (SVM) – specialist for classification, so that we can check whether our model is working best or not.

5. <u>Expected outcomes:</u> The expected results of the research are given below:

a) A novel potato disease identification method based on deep convolutional neural network.
b) Developing a model to provide the farmers an easy-to-use system to detect early-stage infections by using common smart phones.

References:

[1] S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, ''Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology,'' *IEEE Access*, vol. 6, pp. 8852–8863, 2018. doi: 10.1109/ACCESS.2018.2800685.

[2] J. Ma, K. Du, F. Zheng, L. Zhang, Z. Gong, and Z. Sun, ''A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network,'' *Comput. Electron. Agricult.*, vol. 154, pp. 18–24, Nov. 2018. doi: 10.1016/j.compag.2018.08.048.

[3] K. P. Ferentinos, ''Deep learning models for plant disease detection and diagnosis,'' *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018. doi: 10.1016/j.compag.2018.01.009.

[4] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, ''A comparative study of fine-tuning deep learning models for plant disease identification,'' *Comput.Electron. Agricult.*, to be published. doi: 10.1016/j.compag.2018.03.032.

[5] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, ''Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild,'' *Comput. Electron. Agricult.*, to be published. doi: 10.1016/j.compag.2018.04.002.

[6] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, ''Identification of rice diseases using deep convolutional neural networks,'' *Neurocomputing*, vol. 267, pp. 378–384, Dec. 2017. doi: 10.1016/j.neucom.2017.06.023.

[7] H. Durmus, E. O. Günes, and M. Kırcı, ''Disease detection on the leaves of the tomato plants by using deep learning,'' in *Proc. 6th Int. Conf. Agro-Geoinformatics*, Fairfax, VA, USA, Aug. 2017, pp. 1–5. doi: 10.1109/Agro-Geoinformatics.2017.8047016.

[8] L. Jain, M. A. H. Vardhan, M. L. Nishanth, and S. S. Shylaja, ''Cloud-based system for supervised classification of plant diseases using convolutional neural networks,'' in *Proc. IEEE Int. Conf. Cloud Comput. Emerg. Markets*, Nov. 2017, pp. 63–68. doi: 10.1109/CCEM.2017.22.

[9] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, ''Multilayer Convolutional Neural Network for the classification of Mango Leaves Infected by Anthracnose Disease'' in *IEEE Access*, vol. 7 pp. 43721 - 43729, Mar. 2019. doi: 10.1109/ACCESS.2019.2907383.

[10] J. G. A. Barbedo, ''Factors influencing the use of deep learning for plant disease recognition,'' *Biosyst. Eng.*, vol. 172, pp. 84–91, Aug. 2018. doi: 10.1016/j.biosystemseng.2018.05.013.

[11] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, ''Identification of maize leaf diseases using improved deep convolutional neural networks,'' *IEEE Access*, vol. 6, pp. 30370–30377, 2018. doi: 10.1109/ACCESS.2018.2844405.

[12] R. Gandhi, S. Nimbalkar, N. Yelamanchili, and S. Ponkshe, ''Plant disease detection using CNNs and GANs as an augmentative approach,'' in *Proc. IEEE Int. Conf. Innov. Res. Develop. (ICIRD)*, Bangkok, Thailand, May 2018, pp. 1–5. doi: 10.1109/ICIRD.2018.8376321.