A DEEP CNN-BASED APPROACH FOR DETECTING MAJOR DISEASE OF POTATOES: EARLY BLIGHT AND LATE BLIGHT

BY

NAIMUL HOSSAIN

ID: 172-15-10036

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Professor Dr. Md. Ismail Jabiullah

Professor

Department of CSE

Daffodil International University

Co-Supervised By

Zerin Nasrin Tumpa

Lecturer

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

SEPTEMBER 2021

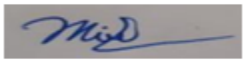
APPROVAL

This Project titled "**A Deep CNN-based Approach for Detecting Major Disease of Potatoes: Early Blight and Late Blight**", submitted by Naimul Hossain to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on September 11, 2021.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan
Professor and Head **Chairman**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Md. Sadekur Rahman
Assistant Professor **Internal Examiner**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Moushumi Zaman Bonny
Assistant Professor **Internal Examiner**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Shah Md. Imran
Industry Promotion Expert **External Examiner**
LICT Project, ICT Division, Bangladesh

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Professor Dr. Md. Ismail Jabiullah, Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

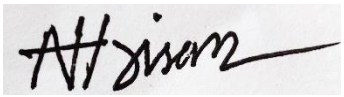
Supervised by:

Professor Dr. Md. Ismail Jabiullah
Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Naimul Hossain
ID: 172-15-10036
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project/internship successfully. I really grateful and wish my profound my indebtedness to **Professor Dr. Md. Ismail Jabiulla**. **Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "*Artificial Intelligence*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We are grateful to our honorable department head **Professor Dr. Touhid Bhuiyan**, Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

Within the advance of the world economy there's continuously a tremendous effect of farming. Considering the situation of Bangladesh 14.74 % of add up to GDP came from the farming segment, where the significance of crops to be more particular crops like potatoes is evident. To overcome misfortune of generation and stabilizing the nourishment chain identifying potato maladies and taking basic steps is required. In that case the mechanization of identifying infections plays a vital part particularly picture handling. Recognizing in a conventional way takes intemperate preparing time and needs skill. Minimizing the preparation time and detecting the infections in early stages is the most objective drawing closer to this issue. The starting step was collecting the information and building a well outfitted information for conveying this work. For conveying this investigative work, this particular issue is drawn nearer with a profound learning strategy. We executed Profound CNN engineering for classification which performs extraordinary in case of picture handling beneath the space of computer vision. The most excellent precision we got from our actualized design is 97.89%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

The impact of agriculture in developing countries' economies is pivotal and it's the main source of fulfilling the basic needs of rural peoples. Agriculture always played a vital role in Bangladesh's economy. 12.68 percent of Bangladesh's total GDP comes from the agriculture sector and not only that agriculture sector also creates lots of employment scope. From that, we can understand the importance of detecting plant diseases. Detection of plant diseases has become the top research area in computer vision over the past few years. To ensure the quality and quantity of agricultural product disease direction in the early stage is one of the most important things. Identification of disease combines few steps such as collecting the images of plants then pre-processing that collected images which is one of the most important parts of this section. Then segmentation of that preprocessed images and extracting features from that and finally classify them.

Different diseases have different symptoms and consist of combinations of individual colors. Over the past few years, a lot of research had done on this specific domain but in the case of detecting specific diseases where we have challenges. After studying the related works, I observed that for finding out specific diseases image processing is one of the best ways. In case of facing this kind of challenge and finding out the solution, Deep CNN gives us a lot of advantages. The detection of objects and the accuracy in Deep CNN comes more purified. This is one of the most important reasons for approaching this problem with Deep CNN and finding out the best accuracy so that people from this sector can be benefited and find their solution at the right time.

1.2 Motivation

For a developing country like Bangladesh agriculture is one of the main economical sources for rural peoples over the years. Globally we are evolving technologically in every sector. But in today's still, we have to find out the diseases with the traditional approach which experts' observation without any tools. Here is one of the main reasons why I tried

to find out the solution to this problem because people of the rural area always don't have the opportunity of getting expert's opinion and not only that the processing time with the traditional approach is pretty much time-consuming. There are so many expanding sectors trying to sustain in the phase of digitalization. No exception in the sector of agriculture it's also evolved digitally but there are also many more scopes to develop. Potato is one of the vast cultivated vegetables in the world.

In order to produce, Bangladesh is in the seventh position globally but there had been huge loss due to diseases named early-blight and late-blight. Losses not only come from damage happened by diseases but also from lacking's of farmer's productive education on agriculture. Farmers don't have that much knowledge on diseases of crops. Farmers don't have any hands-on solution or guideline according to their problems. Not getting the solution at the right moment cause huge loss. In Bangladesh one of the most cultivated vegetable is potato. That why I approached this problem to find out more effective and efficient solution.

1.3 Rationale of the study

In the area of agricultural research detection of vegetable disease is a more complicated technique. The benefit of this piece work will bring a long tram solation for the rural farmers as well as for those who loves to do gardening on their top roof in recent days. On this specific domain there have been some existing device and technique already but lacking in the effectiveness in remote areas this is where I worked on to make it as effective as possible for both rural farmers and also for modern gardeners. Smartphone is now so much handy that I can take advantage of it's by my making an mobile based application so that farmer can overcome a huge loss taking necessary step according their vegetable disease at the right moment.

1.4 Research Questions

Q1. What are the existing techniques of detecting Potato disease Early-blight and Late-blight?

Q2. What are the ways of detecting Potatoes Early-blight and Late-blight disease?

1.5 Expected Output

In this piece of work by implementing the architecture of deep CNN for vulcanization of not affected and affected leaves I approached to develop an automated system which will be reliable as well as effective at a time for detecting potato diseases Early blight and Late blight. The benefits of this research work able to solve the Hardship of farmers to identify such diseases and finding out the solutions. The architecture I implemented to solve this problem will also works problems on this domain.

1.6 Report Layout

Chapter 1: Introduction

Introduction of my research topic, Importance of doing the research on this specific topic, motivation of doing this research, rationale of the study, research questions on this topic and expected outcome of this research work explained into this chapter. The report layout of the research included into this chapter.

Chapter 2: Background

Exploration of the related work done on this specific topic introduced in this chapter. Other than that, the research summary of this piece of work, scop on this specific topic, challenges, and about tools and software described in this chapter.

Chapter 3: Research methodology

Chapter of research methodology covered with the explanation of my research subject and instrument, the work process of doing this research work, how I collected data and the preprocessing of the data, Model Architecture, Configuration of my model.

Chapter 4: Experimental result and discussion

In this chapter I discussed about the result of my experiment and analysis of the result.

Chapter 5: Conclusion and future scope

Chapter 5, covered with summary of this research work, conclusion and the future scopes on this specific topic and field.

CHAPTER 2
BACKGROUND

2.1 Introduction

According to getting more profits Potato is one of the most produced vegetable in Bangladesh [17]. We can see the use of potatoes in most of Bangladeshi dish and all over the world the number people consume potato is billions. It can be considered as one of the most important vegetable in global food chain. Every year the consumption potato in Bangladesh is 6.5-7.0 million tones and the produce 8.92 million tones (FY 2014). For balancing this production and consumption farmers always trying hard but due this kind of diseases' they also face a vast amount of loss in production which hampers them economically. Most of the farmers can't identify the disease. Understanding the disease at the right moment and taking action according to the disease is the only way to skip a huge loss and for that they need a reliable disease detecting option in real time and which can be solved by this piece of work. Detection of plant diseases is one of the most popular topics in the research field globally. Individual researchers implemented individual technique to identify plant diseases. To find out the most effective way lots of researchers giving their best efforts all the time. For continuing the research and new inventions in the field of plant disease detection government also taking some necessary steps.

The best way to detect plant disease is testing the leaf. Plant disease also can be identified by some other symptoms. But this is more difficult to find out the disease without applying repeated test on plants by farmers. Lot of things can be found out about a plant disease by testing and analyzing the leaf of the plant. It's not possible to understand the differences between individual symptoms for individual disease with open eye observation. Maximum of the diseases has some similar treatments, but it can be cause huge loss if the disease doesn't analyze properly the right moment. It's difficult to reach to an expert every time at the right moment specially in rural areas. Few research works was able to give farmers hands on salutation. But specifically in this research area there more scope to develop.

2.2 Vastly spread Diseases

Every year 8,06,394 acres area used for cultivating potato in Bangladesh (banglapedia). Farmers do cultivation of potatoes because of its demand in the market. From beginning to final products farmers has to face lot of challenges and one of the hardest challenges is to identify the potato disease. If anyone isn't able to analyze the disease at the right moment then they have to face a huge loss in the production. Two most dangerous potato disease is Early Blight and Late Blight which are descripted below.

- **Early Blight:** Early blight occurs when the leaf of potato got infected by a fungal pathogen named *Alternaria Solani*. It hampers the tubers the stems of potatoes, tuber size, loss in the tubers storability which cause the production of potatoes. Figure 2.2.1 is *Alternaria Solani* affected potato leaf.

Figure 2.2.1: Early Blight affected potato leaf

- **Late Blight:** Detection of potatoes Late Blight is possible by observing brown or light black lesions on stems and leaf. A fungal named *Phytophthora Infestans* is the reason why potato Late Blight occurs. Figure 2.2.2 is a snap of *Phytophthora Infestans* affected potato leaf.

Figure 2.2.2: Late Blight affected potato leaf

2.3 Related works

The description of this section will be according to the existing technique of detecting Potato Early-blight and late-blight. On this specific problem I didn't find any technique or approach to identify Potato disease Early-blight and Late-blight at the same time specially in Bangladesh but there had been very few works done on this domain using Deep CNN globally also in Bangladesh over the past few years.

Author Monzurlul Islam implemented a multiclass support vector machine and support vector machine on a dataset of potato leaves consist of 300 images and the dataset was collected from an online platform named plant village. The deployed architecture by author is able to detect potato disease with the validation accuracy of 93.7% and found testing accuracy of 95% [1]. A method of detecting potato virus Y was approached by Author Gerrotit Polder where on hyperspectral images deep learning technique was applied. Because of he didn't have a maximum number of infected planets while training his dataset the model predicted well. The predicted value represented here through recall values which wasn't up to the mark compared with a crop expert. Recall value was in between (75% - 92%) and the accuracy standard of a crop expert is 93% [2].

Author Dor Oppenheim introduced a model to detect tuber disease of potato where he applied deep learning technique on an image-based dataset. He applied CNN technique on

the dataset of 400 affected potato tubers images. Hence the dataset isn't that much the accuracy came out 96% [3].

A Machine learning and Unmanned Aerial Vehicles Technique Was introduced by author Julio M. Duarte for analyzing Potato Crops Late Blight Severity. He built a Dataset of 748,071 multispectral convergence patches where he applied multiple algorithms like Deep learning CNN, Support Vector regression, Random Forest and MLP [4]. Author Debabrata Samanta Introduced a machine learning technique for detecting scab diseases of potatoes. where she deployed image processing techniques to identify the scab diseases. It's a histogram-based system where she got the best classifier accuracy is 97.5% in this case the training sample was 235 and testing sample was 54 [5].

Author Priyadarshini approached a Performance comparison based on machine learning algorithms for detecting potato Blight diseases where she used leaf images. She built a dataset based on 892 images where 300 images were taken from an online resource named 'Plant village'. The number of normal leaf images and affected leaf images in the dataset follows 462 and 430.On her dataset she implemented SVM, RF and ANN classifiers and the best accuracy came in ANN classifier and accuracy of these classifiers as follows 84%, 79%, 92% [6].

Author Michael Barnes introduced machine vision technique for visually detecting potato blemishes. He used two sets of data for his model one for white potatoes and another for red potatoes. The number of images for white potatoes and red potatoes in the dataset as follows 102 and 22. on this data set he applied the Adaboost algorithm. The best success rate of white potatoes in selected features for different numbers is 89.8 % and for red potatoes is 89.5%. The best success rate for white potatoes in different subsets of feature categories is 90.2% and for red potatoes is 88.7% [7].

After observing the literature reviews, we can see that image processing is one of the best techniques to detect potato diseases in recent times.

To bring out more accurate result deep learning techniques are performing better then machine learning techniques. From this fact, I approached my dataset with deep CNN architecture. In case of prediction and classification this architecture gives outstanding results for problem sets like mine.

2.4 Research Summary

In this research work we tried to find out an efficient solation for Potatoes Early Blight and Late Blight diseases by deploying Deep CNN architecture and bring out a real time result. Here is my research work summer –

Step 1-

- Collecting Dataset
- Complete the preprocessing of Dataset

Step 2-

- Build the Deep CNN Model

Step 3-

- Train the Deep CNN model

Step 4-

- Testing the Model
- Cross Validation

Step 5-

- Find out the accuracy of prediction

2.5 Scope of the problem

On the basis of previous works, we can see that one the best way for detecting potato disease is processing the images of affected potatoes. The reason behind all this research has been done and why we are also approaching this problem is the same, trying to make an impact in the increasement of potato yield and also how farmers can be more benefited financially and get the solution in minimum time. We approached this problem with a deep learning technique to be more specific deep Convolutional Neural Network (CNN). Choosing this architecture to implement our study is an outstanding result of it and powerful in the field of image classification. In this study we will apply CNN repeatedly so that we can bring out the most accurate result. The layers of purification will bring out more accurate results which will be an advantage nationally and also for individuals.

2.6 Challenges

In the area of plant disease identification there are so many challenges to find out best outcome from the research work. Choosing model which can give an efficient solution for this specific problem is one the challenges. For understanding how my Deep CNN Model will work for this specific problem I had to learn working process of Deep CNN model. Classifying a particular disease from image is one the complex task in image processing and I also have to identify a particular disease from images and find out the solution. Building a quality full dataset was one of most challenging tasks for this research work. Due to pandemic situation real life data collection wasn't possible but after an extreme search on the internet I found a perfect dataset from a source named Kaggle. Most of the targeted user of this research work are farmers and people from rurales. In the country like ours most of the farmers are not that much literate. They are not that much familiar with smart devices like smartphones. The price of a smartphone is also not that reasonable from the prospective of rural farmers. Some particular disease has some common symptoms where identifying the specific diseases is big challenge.

2.7 Tools and Software

- Hardware- Laptop with 4GB RAM, Core i3, Broad Band Connection.
- Software- TensorFlow, Web Browser, Python 3, Google Colab.

Other than that, a Deep learning Architecture ana a well-furnished Dataset

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

In this section considering few steps like data collection, data pre-processing after that feeding the Deep CNN model and train the model finally detecting the potato disease. For training and evaluating the proposed model, TensorFlow 2.2.0 was used. We had done some prerequisites like data preprocessing before diving into training. The proposed model will be described after that. Fig. 3.1.1 shows the workflow of our proposed model.

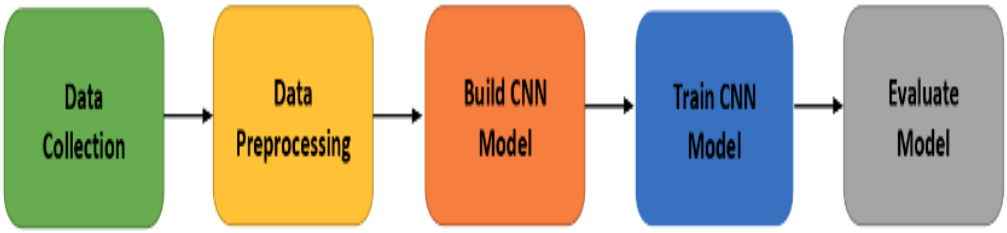


Figure 3.1.1 Work Flow of this Research work

3.2 Data Collection

When a system can act on the relevant things and also be able to generate an outcome on a specific variable which is possible by collecting data then measuring the information and finding out the insight, this whole process is under the vision of data collection. On the basis of collected data one can find out the proposition. The importance of data collection and validity complement the research work and its purity. We collected our data from a renowned website named kaggle [10] which is globally the biggest data science community with a vast number of resources to help researchers. Where we found a dataset, which is publicly available containing a dataset of 87k RGB images of healthy and diseased crop leaves which is categorized into 38 different classes and also the source validation was provided. From there we chosen our expected dataset consist of total image 5702 categorized in different classes and labeled as below-

- potato leaves affected with early blight
- potato leaves affected with late blight
- Healthy potato leaves.

TABLE 3.2.1: STATSTICS OF DATASET

Label	No. of Images	Format of Image	Size of Image
early blight	1939	jpg	Various
late blight	1939	jpg	Various
Healthy	1824	jpg	Various

Table 3.2.1 shows the number of images potato leafs affected with Early blight, potato leafs affected with Late blight and healthy potato leaf's as follows 1939, 1939, 1824. We choose Potato both healthy and affected leaf's because we will train our Deep CNN model with this dataset so that our model can differentiate between healthy and affected leaf's. Table 3.2.2 shows how I defined the ratio of my training and testing dataset as 80% and 20%.

TABLE 3.2.2: THE RATIO FOR TRAINIG AND TESTING DATA

Label	No. of Images	Training set (80%)	Testing set (20%)
early blight	1939	1550	385
late blight	1939	1550	385
Healthy	1824	1460	365

3.3 Data Pre-processing

In case of image processing implementation preprocessing is a common phenomenon with images which is the abstraction of base level. The main motto of preprocessing is

restraining the unexpected garble and exaggerate a few image attributes which are essential for other processing. For preprocessing our image dataset first of all we have done our analysis of images and then read the image. After that we resized our images and then we denoised our images. Completion of all stages effectively gives us the best shape of our dataset.

3.3.1 Labeling of Image

The interpretation of image is the work of interpreting the image with a label, which can be done by humans and also can be done using machine. For giving the machine a vision labels are predetermined by a deep learning engineer.

3.3.2 keras.preprocessing

Using keras.preprocessing we can bring the raw data on disk to a tf. Data. For training a model this dataset object can be used. Using this library, we can perform the augmentation module of deep learning library of keras and also the preprocessing of data.

3.3.3 ImageDataGenerator

The keras ImageDataGenerator is one of the most effective features. It's has that unique abelite where we augment our images during our model training. We can implement any kind of image transformations which can go through the model. It also can optimize our machine memory not only that it also makes sure our model as a robust model.

3.3.4 Kernel

Kernel is one kind of mask or convolution matrix in the section of image processing. This convolution matrix used for detecting edge, clearing, blurring, and many more thing can be done by making link between image and kernel.

3.4 Research Subject and Instrumentation

The subject of my research is "Detecting the Early Blight and Late Blight diseases of Potatoes using Deep CNN". In the area of research this kind of subject is bit challenging

and also interesting. Cause the performance and the accuracy of this kind of research work is the main key point. Bring out a reliable solution for rural people of Bangladesh is the main motto. Now internet and smart phones are more reliable then past few years. That's why I tried to find out a solution, where using our model in future we can build a mobile based application so that it becomes handier and give the right solution at the right moment. Using python, TensorFlow, and most purified Deep CNN model and also OpenCV being used for preprocessing task.

3.4.1 Statistical Analysis

Data Gathering and explanation for discovering the trends and patterns from it is known as statistical analysis. This is a part of data analytics. We use statistical analysis for collecting research translation, statistical designing surveys & also for studies.

3.4.2 sklearn.metrics

I used sklear.metrics as accuracy classification score. It's also used for computing the accuracy of subset in multilabel classification. The predicted set of labels for a sample have to as same as the y_true corresponding set of labels. sklearn.metrics package also able to generate a confusion matrix. From confusion matrix we can analyze the accuracy of our classification.

3.4.3 Generation of .h5 file

The process of storing vast amount of data within a single file in hierarchical structure which known as Hierarchical Data Format 5 in short .h5 file and it's also an open-source file. If we want to access a specific file rather than the hole file, we can do this using HDF5 format. For taking the advantage of HDF5 we imported numpy. One of the unique features of HDF5 it has capability of composing the metaset to individual data in the file along with more efficient accessing and searching.

3.4.4 Adam Optimizer

Adam optimizer operate as a substitution optimization algorithm for training a model of

deep learning using stochastic gradient descent. Adam optimizer can gather the best entity of RMSProp and AdaGrad algorithms in one place so that it can bring out an algorithm which can optimize along with having the capability of handling the noise problem. Configuring the Adam optimizer is not that much difficult and not only that in case of maximum problem the default parameters of configuration work really well.

3.4.5 ConV2D

2D convolution layer is one of the most commonly used convolution and short form is stands as conv2D. In conv2D layer kernel and filter has a width and height. Usually, the input image is bigger than the filters and kernels so that movement on image become easy.

3.4.6 Softmax function

For identifying the distribution of multinomial probability in neural network model softmax function works as a activation function which in the outer layer of neural network model. Not only that in case of multi-class classification complexity as an activation function softmax can be used.

3.4.7 ReLU activation

ReLU briefly known as rectified linear activation function. In deep learning ReLU function considered as one of the most important function because the ability to making the familiarity between the real-world entities which are nonlinear and neural networks. It also helps neural network to complex pattern from dataset. ReLU activation function switch the input to the output in other word we can say output of previous layer works as input in next layer. Without ReLU activation function the output of our model will transform into a linear function. If we don’t use activation function in our neural network then our model will not get that much learning power and it will look like a simple linear regression. This ReLU function works batter then the tanh or sigmoid.

3.5 Model Configuration

In this recent time there are so many options in case of choosing a model for a specific

problem. We choose the best one according to our problem. Some models give best performance for text classification like Support Vector Machine (SVM) and for sentiment analysis Naïve Bayes, Logistic Regression works well. We used deep CNN in this piece of work. The detail description of our model given below:

Deep Convolutional Neural Network (CNN) - Convolutional Neural Networks (CNN) is considered as Potentially most approved deep learning architecture because of its efficiency of convnets. That makes CNN unique from other architectures is the ability to automatically detect key features by itself and the effectiveness in computation.

CNN works with two significant parts first one is feature extraction and next one is classification. Convolutional layers used for feature extraction and fully connected layer used for classification.

Convolution: Convolutional layer is set of filters where they have some parameters and these parameters have to learned. The volume of input is bigger than the height and weight of the filters. For computing the activation map which build on neurons every filter should be convolved according to the input volume. The maps of activation and the depth dimension of the filters is a stack of the output of that layer of convolution. As it is defined that the height and width of filters will be smaller from the input, ever activation map of the neuron is associated with the input volume. In short, the main work of convolution layer is to sum up the features from an image which is given as input.

Pooling: In every CNN Model it's common that always there will be a pooling and convolution layer in a sequence one after another. For reducing dimensions of feature map, we use pooling layer. Pooling layer also minimize the number of parameters for learning and also reduce the amount of computation. When a convolutional layer generates a feature map then the pooling layer sum up the features from feature map which is generated by convolution layer. There are various types of pooling layer. In our research work we used Max Pooling.

Max Pooling: This pooling layer extracts the heights component from feature map which feature map has filter cover. When max pooling ends it's brought out the key features from

the earlier feature map.

Flattening: A two-dimensional matrix of features can be turn into the vector features using this flattening techniques. Which we can fed to a neural network classifier.

Full-Connection: Feeding the images to the neural network which is flattened before this process is known as full-connection in neural network.

Configuring the model for this research work is one of the most important. Now, it's time to defining the parameters for our model as like learning rate, class of potato leaf, batch size and many more parameters. In our proposed model we introduced 4 convolutional layer, 4 pooling layer followed by a flatten layer and two dense layers. Now it's perfect time for building our deep CNN model. The following figure 3.5.1 shows all the required dependencies are imported and figure 3.5.2 shows the defined prerequisites.

```
from keras.models import Sequential
from keras.layers import Conv2D,MaxPooling2D
from keras.layers import Activation, Dense, Flatten, Dropout,BatchNormalization
from keras.preprocessing.image import ImageDataGenerator
from keras.callbacks import ModelCheckpoint
from keras import backend as K

model = Sequential()
model=Sequential()
model.add(Conv2D(filters=8,kernel_size=(5,5),padding="Same",activation="relu",input_shape=(256,256,3)))
model.add(MaxPool2D(pool_size=(2,2)))
#model.add(Dropout(0.25))

model.add(Conv2D(filters=16,kernel_size=(4,4),padding="Same",activation="relu"))
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
#model.add(Dropout(0.25))

model.add(Conv2D(filters=32,kernel_size=(4,4),padding="Same",activation="relu"))
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(Conv2D(filters=48,kernel_size=(4,4),padding="Same",activation="relu"))
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(Dropout(0.25))
model.add(Flatten())
```

Figure 3.5.1: Imported all dependecies

```
model.add(Dense(512,activation="relu"))
model.add(Dropout(0.5))
model.add(Dense(3,activation="softmax"))
#defining optimizer
optimizer=Adam(lr=0.0001,beta_1=0.9,beta_2=0.999)
#compile the model
model.compile(optimizer=optimizer,loss="categorical_crossentropy",metrics=["acc"])

history = model.fit_generator(
    train_generator,
    steps_per_epoch=nb_train_samples // batch_size,
    epochs=30,
    validation_data=validation_generator,
    validation_steps=nb_validation_samples // batch_size)
```

Figure 3.5.2: Defineing the prerequisite

Following figure 3.5.3 shows how convulational neural network works internaly.

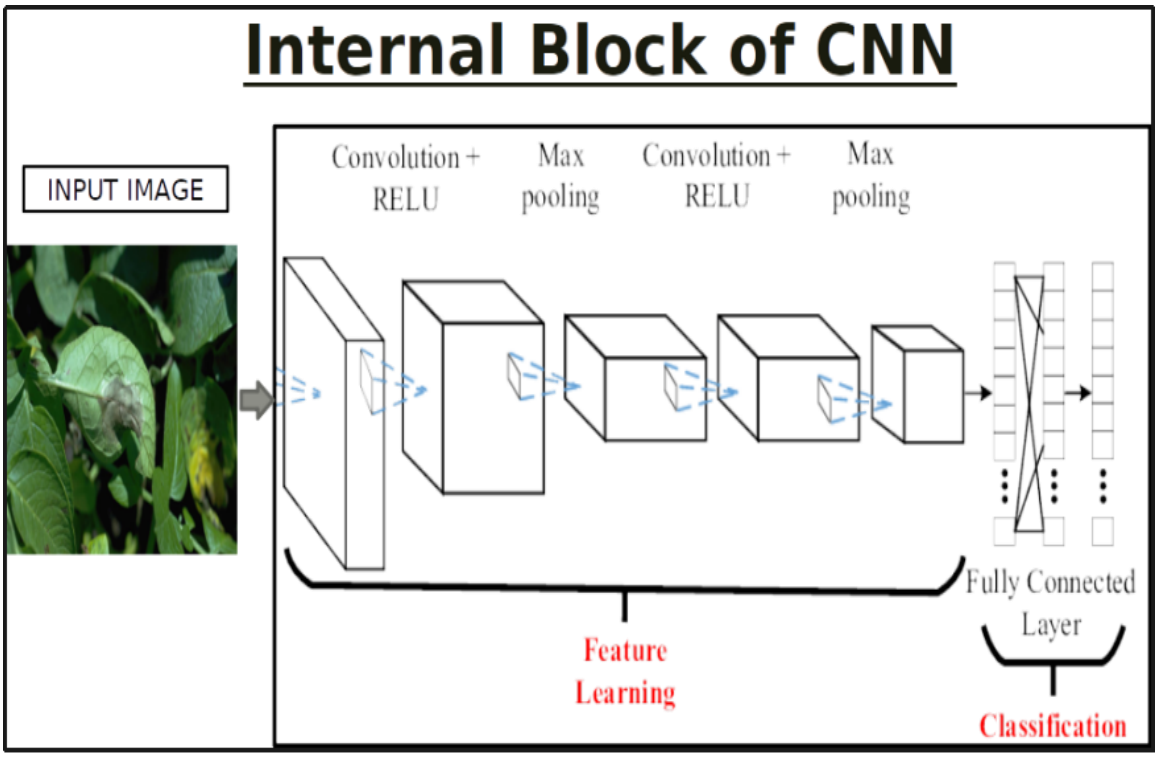


Figure 3.5.3: How CNN works

3.6 Architecture of our Model

For building deep learning models TensorFlow offers one of the best framework which is developed by Google. Deep learning is a domain of machine learning where we can use multipale layers of neural networks for bring out more accuracy of our prediction from the model. For processing our image dataset we used CNN Model.

In figure 3.6 at first, CNN takes input images dimension of 256x256x3. For getting better result RGB channel used here. In first convolutional layer filter size of 8 with 4x4 kernel introduced with a ReLU activation function to adding non-linearity. Later a max-pooling size 2x2 used for reduce the dimensionality. Filter size of 16 with 4x4 kernel used again with a ReLU activation in second convolutional layer. Again, also used a max-pooling size 2x2. The output of second layer is fed again with another third convolutional layer. 32 filter sizes with 3x3 kernel used with a ReLu activation in third layer. 2x2 max-pooling again used after third layer.

The output of all third layer then fed into forth convolutional layer with a filter size of 64, 3x3 kernel size and a ReLU activation. Later 2x2 max-pooling used. Reduce overfitting again a dropout rate of 0.35 used after forth layer. A flattening layer used for make 2D sequences into 1D matrices. Later a fully connected layer 512 nodes used with ReLU activation. Again, dropout rate of 0.25 used for reduce overfitting. Last a dense layer of 3 units used along with a softmax activation as for classification.

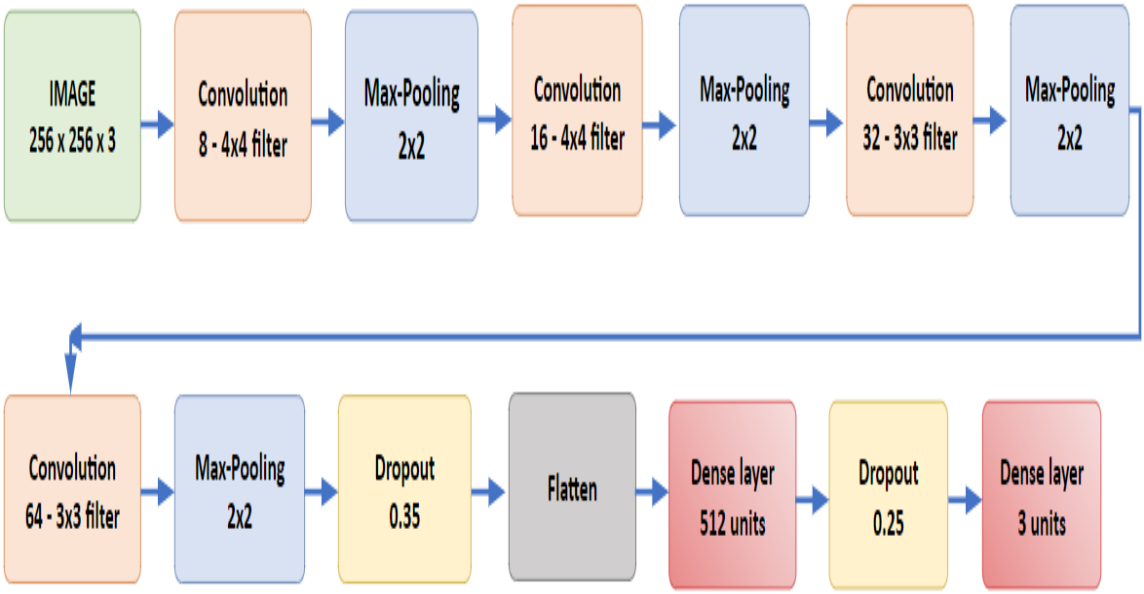


Figure 3.6: Proposed CNN Model

3.7 Parameter:

The weights of neural network connections is known as parameters. During the training period these parameters can learn themselves. In case of this kind of senario tunning of parameters done by the algorithms. Parameters are one of key thing to the model in case of pradiction. The efficiency of a choosen model on a specific problem depends on the parameters value. These parameters generally learned from the dataset and not only that setting of parameters can't done manually. Parameters are one of the key part of model which are estimated from training data and these data are historical.

3.8 Hyperparameter

We had to configure our hyperprarameters externaly to our model beacause values of hyperparameters can't be learned from data. These hyperparameters usaually helps us to learn the model parameters. Hyperparameters are genarally defiend by implementor. Setting of hyperparameters done on the basis of heuristics. Tuning of these parameters specified on the basis of a specific problem. It's not possible to knowing the best value of a hyperparameter for our model or our specific problem. Overcoming this challeng is possible by using ruels of thumb, from error and trial we can find the best value, or values from other problem can be used.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

I have done all the training and the experiment on my dataset. In this section I will test our model with real data. In this section, we will usually analyze the result of the proposed model of our work and explain the result of our work with proper logic. We trained our model using TensorFlow 2.0.1. Setting hyperparameters to our model was the first step. Define Batch size as 45. The ratio of our training and testing data as follow 80%,20%. After that we compiled our model using adam [9] optimizer and our learning rate was 0.0001. Then .fit() used for start the training. Our model got an accuracy of 97.89% after training 30 epoch.

4.2 Experimental Output

ModelGraph: Figure 4.2.1 which is the accuracy curve of our model where we observed our training accuracy 97.89% and validation accuracy 96.29%. Figure 4.2.2 represents the loss curve of our model, where we observed our training loss 0.0345 and validation loss is 0.897

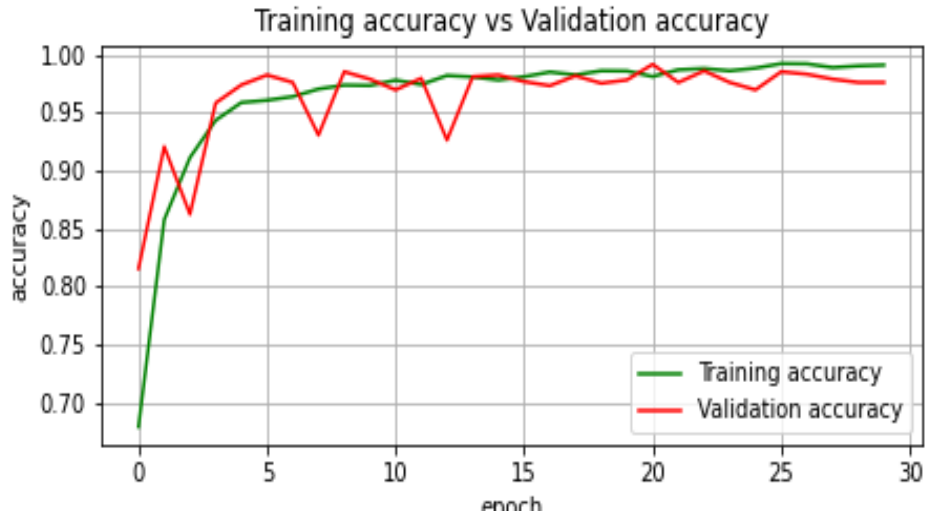


Figure 4.2.1: Model accuracy curve.

Figure 4.2.2: Model loss curve

After training we've also testing our model with unseen data. We got pretty much good testing accuracy of 97.19% and our testing loss was 0.751. Table 1 shows the classification report of our proposed model.

Confusion Matrix- The figure 4.2.3 represent the classifiers prediction result clear and explicit through Confusion Matrix. We summarized our classification algorithms performance through this confusion matrix. We have multiple class in our dataset so it's difficult judge just on the basis of accuracy of classification. That's why we introduced a confusion matrix so that we can find whether our classification model doing right or what kind of error it's doing.

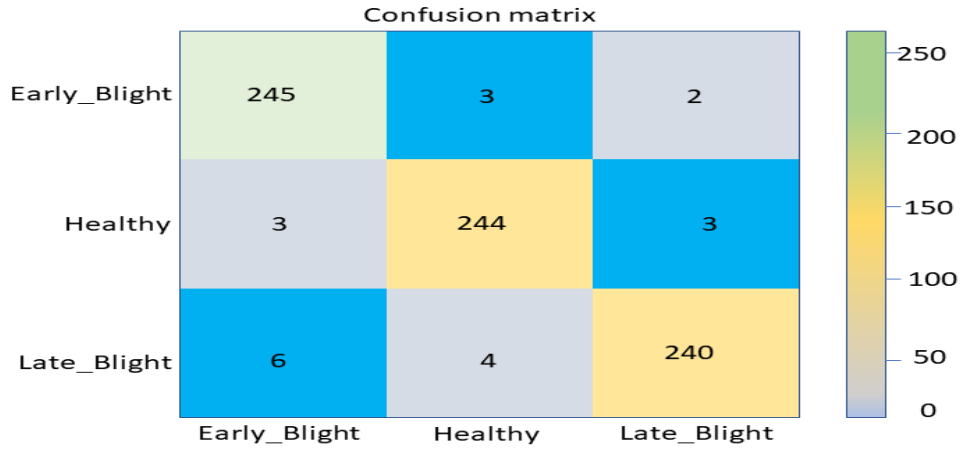


Figure 4.2.3: Confusion Matrix

Classification Report:

Classification report is the method of measuring classification algorithm performance and the quality of its prediction. We can know how many of our classification algorithms predictions are True or False through classification report. We can predict the matrix our classification report using True Negatives, True Positives, False Negatives and True Negatives. Table 4.2.1 shows the classification report of my model.

Table 4.2.1: Classification Report

Class	Precision	Recall	F1-score
Early Blight	0.95	0.96	0.95
Late Blight	0.94	0.93	0.94
Healthy	0.95	0.94	0.97
Average	0.97	0.96	0.96

Recall- Recall classifier is capable of find out all positive instances. The ratio of surely classified entities of all positive. The capability of detecting the positive entities of our model can be evaluated recall classifier. If we get high recall value then our model will detect more positive samples.

The only concern of our recall classifier is how our positive entities are classified. Negative samples are classified independently. If our model can classify all the positive entities as positive that will be counted as 100% recall.

Precision- In case of doing any scientific identification Precision counted as one of the most important measurement entities. Because it helps us to find out that our model giving the right prediction. In this piece of work when we divided the total of True Positives by the total of False positive and True Positive, we got our Precision values. The precision value also known as Positive Predictive value. Precision helps us to understand the correctness of our model's classifier measures. With Precision we can understand that though All Recurrence has lower accuracy than All No Recurrence but it's more efficient

than All No Recurrence and CART is more quality full model.

F1 Score- Find out the accuracy of classification of our model is not the end of our model performance evaluation. For further performance evaluation we used F1 score. Building a model for specific classification problem where we always focus on the model accuracy and only count the correct prediction among all the prediction done by our model. Which we also can define as the accuracy of classification. But only depending on the classification accuracy we can't make decision that model is giving best solution for our specific problem. F1 score also known as F Measure or F Score. The equity in-between recall and precision is done by F1 score. When we try select a model on the basis of equity between recall and precision, at this point F1 score help us to understand that the competitiveness of CART model isn't that much efficient but on the other hand All Recurrence model is the best.

Table 4.2.2: Accuracy Comparison with Other Existing System

Authors Name	**Category**	**Technique**	**Accuracy**
M. Islam [1]	Detection of potato diseases	SVM	93.7%
Samanta [5]	**Scab diseases detection**	**K-means**	**97.5%**
Oppenheim [3]	Detection of potato tuber disease	CNN	96%
Griffel [8]	Detection of potato virus Y	SVM	89.8%
Proposed model	**Detection of early blight & late blight**	**CNN**	**97.89%**

CHAPTER 5
CONCLUSION AND FUTURE SCOPE

5.1 Summary of this Research Work

In this piece of work, we tried to deploy a Deep CNN based technique to identifying the disease of potatoes Early Blight and Late Blight. Where we implemented Deep Convolutional Neural Network. A shorten overview of our research work given below.

Stage 1

- Collection of Dataset
- Preprocessing of Dataset
- Extraction of feature
- Separating data for training and testing

Stage 2

- Build model
- Train model
- Classify the diseases

Stage 3

- Model compiling

Final Stage

- Evaluating the model
- Find out the accuracy
- Prediction on test data

For getting best outcome tunning the parameters and Hyperparameters.

Most of the people from rural in Bangladesh depends on the agricultural economy. Development in agriculture sector technologically of Bangladesh isn't that much upgraded comparing to the global scale. In this piece of work, we specifically worked on two major potato diseases Early Blight and Late Bligh identification. In most of the case framers relay on the traditional technique in case of identifying any kind plant disease. That's why they don't able to find out the solution at the right moment.

With traditional approach any kind plant disease identification is time consuming. Specially in case of potato disease identification with tradition approach which is known as open eye observation is too much time consuming and not only that without help of an expert it's not possible. People of rural don't have the privilege getting an expert's opinion any time. So, minimizing the disease identification time and find out a solution on time is possible using our implemented model.

5.2 Conclusion

In recent times through image processing there are huge research opportunity in the sector of agriculture more specifically in case of plant disease detection. For getting the best out come from image processing Deep Convolutional Neural Network make a huge impact. In this work by deploying deep CNN architecture for processing the affected and healthy leaves we tried to develop an automated system which will be handy as well as efficient at the same time in case of detecting two major potato diseases Early blight and Late blight. Every year farmers who cultivate potato face a huge loss because of not able to identifying the diseases at the right moment. Because of that they can't take the necessary step on time. The benefits of our study will make an impact to resolve the difficulties of farmers in detecting such diseases and finding out the solutions. This architecture will also work for the problems within these domains.

5.3 Recommendations

Identification of plant diseases more specifically potato disease detection system has the ability to adding more value not only in agriculture sector but also in other field where we will be able to deploy Deep Convolutional Neural Network. As like in case of medical sector diseases which effect human body externally can be analyze and identified by implementing model like ours. In case of educational sector, it has huge opportunity to add more value in our educational system. Biologist are also can be benefited from this kind model. In case of identifying various types of species from nature.

5.4 Further Study

As this is the initial attempt of us there is more scope, we found that can be improved. We worked on a specific problem where is this field has much more diverse options to work on. The attachment of the ability to detect multiple diseases within a system will be a major goal to achieve in future. Other than that enrichment of the dataset to bring out more purity and accuracy also a part of future work. Deployment of other architectures to find out the best solution for this problem will step up our study to the next level in future.

REFERENCES

[1] M. Islam, Anh Dinh, K. Wahid and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, 2017, pp. 1-4, doi: 10.1109/CCECE.2017.7946594.

[2] AUTHOR=Polder Gerrit, Blok Pieter M., de Villiers Hendrik A. C., van der Wolf Jan M., Kamp Jan,"Potato Virus Y Detection in Seed Potatoes Using Deep Learning on Hyperspectral Images," JOURNAL=Frontiers in Plant Science DOI=10.3389/fpls.2019.00209

[3] Oppenheim, D., Shani, G., Erlich, O., & Tsror, L. (2018). *Using Deep Learning for Image-Based Potato Tuber Disease Detection. Phytopathology.* doi:10.1094/phyto-08-18-0288-r

[4] Duarte-Carvajalino, J.M.; Alzate, D.F.; Ramirez, A.A.; Santa-Sepulveda, J.D.; Fajardo-Rojas, A.E.; Soto-Suárez, M. Evaluating Late Blight Severity in Potato Crops Using Unmanned Aerial Vehicles and Machine Learning Algorithms. Remote Sens. 2018, 10, 1513.

[5] Samanta, Debabrata, Prajna Paramita Chaudhury, and Arya Ghosh. "Scab diseases detection of potato using image processing." International Journal of Computer Trends and Technology 3.1 (2012).

[6] P. Patil, N. Yaligar and S. M. Meena, "Comparison of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)

[7] Barnes, M., Duckett, T., Cielniak, G., Stroud, G., & Harper, G. (2010). Visual detection of blemishes in potatoes using minimalist boosted classifiers. Journal of Food Engineering, 98(3), 339–346.

[8]Griffel, L. M., Delparte, D., & Edwards, J. (2018). Using Support Vector Machines classification to differentiate spectral signatures of potato plants infected with Potato Virus Y. Computers and Electronics in Agriculture, 153, 318–324.

[9] https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c

[10] https://www.kaggle.com/vipoooool/new-plant-diseases-dataset

APPENDIX

Appendix A: Reflection of Research

Detection of potatoes two major diseases Early Blight and Late Blight by implementing Deep Convolutional Neural Network was the main goal of mine from the beginning of my final year and didn't had enough knowledge of this specific field. That's why I had start everything from basic. By the Grace of Allah and with the support from my dear supervisor, I was able to overcome the challenges. Accomplishment of this research work wasn't that much easy I thought because there is all read some existing system and multiple publication on this specific domain. So, accomplishment of this research work wasn't the only challenge the main challenge was bring out an outcome which is better than the existing systems outcome and more efficient.

Appendix B: Related Issues

This research work was a great learning experience for me from the beginning to till the end of this piece of work. Due to pandemic situation, it wasn't possible for me to collect images of affected and healthy leaves of potatoes. But from an online source I found a dataset which was absolutely perfect for my research work. I also got help from my well hearted teachers and classmates.

Library Functions

Here I imported all the needed packages

```
[ ] import numpy as np # linear algebra
    import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
    import matplotlib.pyplot as plt
    import seaborn as sns
    import glob
    import cv2
    import os

    import warnings
    # filter warnings
    warnings.filterwarnings('ignore')


    import os
    print(os.listdir("../content/drive/My Drive/Colab Notebooks/potato"))
```

```
[ ] import numpy as np # linear algebra
    import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
    import matplotlib.pyplot as plt



    import seaborn as sns
    import cv2

    import keras
    from keras.models import Sequential
    from keras.layers import Dense, Activation, Dropout, Flatten,GlobalAveragePooling2D
    from keras.layers import Conv2D
    from keras.layers import MaxPooling2D,MaxPool2D
    from keras.layers.normalization import BatchNormalization
    from keras.optimizers import Adam,RMSprop
    from keras.preprocessing.image import ImageDataGenerator
```

Imported necessary Packages

```
from keras.models import Sequential
from keras.layers import Conv2D,MaxPooling2D
from keras.layers import Activation, Dense, Flatten, Dropout,BatchNormalization
from keras.preprocessing.image import ImageDataGenerator
from keras.callbacks import ModelCheckpoint
from keras import backend as K

model = Sequential()
model=Sequential()
model.add(Conv2D(filters=8,kernel_size=(5,5),padding="Same",activation="relu",input_shape=(256,256,3)))
model.add(MaxPool2D(pool_size=(2,2)))
#model.add(Dropout(0.25))

model.add(Conv2D(filters=16,kernel_size=(4,4),padding="Same",activation="relu"))
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
#model.add(Dropout(0.25))

model.add(Conv2D(filters=32,kernel_size=(4,4),padding="Same",activation="relu"))
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(Conv2D(filters=48,kernel_size=(4,4),padding="Same",activation="relu"))
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(512,activation="relu"))
model.add(Dropout(0.5))
model.add(Dense(3,activation="softmax"))
#defining optimizer
optimizer=Adam(lr=0.0001,beta_1=0.9,beta_2=0.999)
#compile the model
model.compile(optimizer=optimizer,loss="categorical_crossentropy",metrics=["acc"])

history = model.fit_generator(
    train_generator,
    steps_per_epoch=nb_train_samples // batch_size,
    epochs=30,
    validation_data=validation_generator,
    validation_steps=nb_validation_samples // batch_size)
```

Train and Validation

Validation curve and plot the train

```
[ ] plt.figure(figsize=(7,4))
    plt.plot(history.history['acc'],'g')
    plt.plot(history.history['val_acc'],'r')
    plt.title('Training accuracy vs Validation accuracy')
    plt.ylabel('accuracy')
    plt.xlabel('epoch')
    plt.grid()
    plt.legend(['Training accuracy', 'Validation accuracy'], loc='lower right')
    plt.savefig("/content/acc.png")
    plt.show()
    # summarize history for loss
```

```
    plt.figure(figsize=(7,4))
    plt.plot(history.history['loss'],'g')
    plt.plot(history.history['val_loss'],'r')
    plt.title('Training loss vs Validation loss')
    plt.ylabel('loss')
    plt.xlabel('epoch')
    plt.grid()
    plt.legend(['Training loss', 'Validation loss'], loc='upper right')
    plt.savefig("/content/loss.png")
    plt.show()
```

```
[ ] from sklearn.metrics import confusion_matrix, classification_report
    valid = '/content/drive/My Drive/Colab Notebooks/potato/test'
    test_datagen = ImageDataGenerator(rescale=1 / 255.0)

    test_generator = test_datagen.flow_from_directory(
        test,
        target_size=(img_width, img_height),
        batch_size=batch_size,
        class_mode='categorical'
      )
```

Classification Report and Confusion Matrix

Generator of confusion matrix and classification report.

```
[ ] #Confution Matrix and Classification Report
    num_of_test_samples=1420
    steps = num_of_test_samples/batch_size
    Y_pred = model.predict_generator(test_generator, steps)
    y_pred = np.argmax(Y_pred, axis=1)
    print('Confusion Matrix')
    print(confusion_matrix(test_generator.classes, y_pred))
    print('Classification Report')
    target_names = ['0','1','2']
    print(classification_report(test_generator.classes, y_pred, target_names=target_names))
```

```
[ ] model.evaluate(test_generator)
```

```
[ ] from keras.utils import plot_model
    plot_model(model,show_shapes=True,to_file='model.png')
```

```
[ ] model.save("potato.h5")
```

```
    from keras.models import load_model
    from keras.preprocessing import image
    from keras.applications.mobilenet import preprocess_input
    import numpy as np
    model= load_model('potato.h5')
```

```
[ ] img = image.load_img('/content/drive/My Drive/Colab Notebooks/potato/valid/Healthy/00fc2ee5-729f-4757-8aeb-65c3355874f2___RS_HL 1864_180deg.JPG', target_size=(256,256))
    img
    x = image.img_to_array(img)
    x
```

```
[ ] x = np.expand_dims(x, axis=0)
    img_data = preprocess_input(x)
    classes = model.predict(img_data)
```

9%	8%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	www.mdpi.com Internet Source	1%
3	Camila Maione, Eloá Moura Araujo, Sabrina Novaes dos Santos-Araujo, Alexys Giorgia Friol Boim et al. "Determining the geographical origin of lettuce with data mining applied to micronutrients and soil properties", Scientia Agricola, 2022 Publication	<1%
4	Priyadarshini Patil, Nagaratna Yaligar, S.M. Meena. "Comparision of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images", 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2017 Publication	<1%
5	Submitted to CSU, San Jose State University Student Paper	<1%