# PLANT DISEASE CLASSIFICATION

# A

# Project Report

# *submitted in partial fulfillment of the requirements*

# *for the award of the degree of*

# BACHELOR OF TECHNOLOGY

# in

# COMPUTER SCIENCE

# Specialization in

# Big Data

By :

| **Name** | **Roll No** | **SAP ID** |
| --- | --- | --- |
| Bhavanshu Arora | R172218032 | 500067696 |
| Dipansh Mittal | R172218044 | 500067347 |
| Kartik Bhatia | R172218058 | 500068463 |
| Parth | R172218085 | 500067724 |

# Under the guidance of

Dr. Vijendra Singh

Department of Informatics



# Department of Informatics

# School of Computer Science

# UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

# Bidholi, Via Prem Nagar, Dehradun, Uttarakhand 2012-23



# CANDIDATES DECLARATION

We hereby certify that the project work entitled **CNN BASED PLANT**

**HEALTH IDENTIFICATION USING LEAF IMAGES** in partial

fulfillment of therequirements for the award of the Degree of Bachelor of

Technology in ComputerScience and Engineering with Specialization in

Big Data and submittedto the Department of systemics at

School of Computer Science, Universityof Petroleumand Energy Studies,

Dehradun, is an authentic record of our work carried out during

period from **JANUARY, 2022** to **MAY, 2022** under the supervision of

**Dr.Vijendra Singh, Assistant Professor, Department of Informatics**.

The matter presented in this project has not been submitted by us for the

award of any other degree of this or any other University.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **NamNO Of Students(s)** | **Bhavanshu Arora** | **Dipansh Mittal** | **Kartik Bhatia** | **Parth** |
| **Roll No.** | **R172218032** | **R17228044** | **R172218058** | **R172218085** |

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Date: 4th May 2022) **Dr. Vijendra Singh**

Project Guide

**Dr. Tiphendra Pal Singh**

Head

Department of Informatics

School of Computer Science

University of Petroleum and Energy Studies Dehradun - 248 001 (Uttarakhand)

# ABSTRACT

Identification of plant health is the new challenging area for the researchers. One of the most important steps in automatic identification of plant diseases is to extract the infected region from the normal portion of the plant. Studying the infected leaves it has been observed that the greenness of the infected portion of the leaves changes significantly with respect to the normal leaves. Vegetation indices (VI) are some metric used for the remote sensing images to measure the greenness. Images of tomato leaves of both categories healthy and diseased captured with digital camera and resolution of 256x256 pixels forms the dataset. CNN model is used for identifying the health status of plants.

**Keywords:** Convolutional Neural Networks, Plant Health classification.

**TABLE OF CONTENTS**

**Contents**

1. [**Introduction**](#_heading=h.30j0zll) **7**
2. [**Background Study**](#_heading=h.147n2zr) **7**
3. [**Problem Statement**](#_heading=h.3znysh7) **8**
4. [**Objective**](#_heading=h.tyjcwt) **9**
5. [**Methodology**](#_heading=h.1t3h5sf) **10**
6. [**Implementation**](#_heading=h.2s8eyo1) **11**
7. [**System Requirements (Software/Hardware)**](#_heading=h.qsh70q) **19**
8. [**Result Analysis**](#_heading=h.1pxezwc) **21**
9. [**Conclusion and Future Scope**](#_heading=h.2p2csry) **22**

# Introduction

Increasing population and rapid climate change have compelled the researchers to think about sustainable development. One of the major challenges of sustainable development is to re- duce the use of pesticides not only to reduce the cost, but to save the environment and increase quality. Precise, accurate and early diagnosis of plant diseases may reduce the use of pesticides.

Plant disease diagnosis through optical observation of the symptoms on plant leaves, incorporates a significantly high degree of complexity.

The existence of an automated computational system for the detection and diagnosis of plant diseases, would oer a valuable assistance to the agronomist who is asked to perform such diag- noses through optical observation of leaves of infected [plants[1].](#_heading=h.3o7alnk)

With the development of computational systems in recent years, and in particular Graphical Processing Units (GPU) embedded processors, Machine Learning-related Artificial Intelligence applications have achieved exponential growth, leading to the development of novel method- ologies and models, which now form a new category, that of Deep Learning. Deep learning refers to the use of artificial neural network architectures that contain a quite large number of processing layers, as opposed to swallower architectures of more traditional neural network methodologies. The now computationally feasible deep learning models have revolutionized sectors such as image recognition, voice recognition, and other similarly complex processes that deal with the analysis of large volumes of data, giving a huge boost to applications that use these processes, like, e.g., self-driving vehicles, machine translation and interpretation, etc. The introduction of these deep learning techniques into agriculture, and in particular in the eld of plant health diagnosis, has only begun to take place in the last couple of years, and to a rather limited e[xtent](#_heading=h.23ckvvd).

The basic deep learning tool used in this work is Convolutional Neural Networks (CNNs). CNNs constitute one of the most powerful techniques for modeling complex processes and per- forming pattern recognition in applications with large amount of data, like the one of pattern recognition in [images](#_heading=h.ihv636)

In this work, specific CNN architectures were trained and assessed, to form an automated plant health identification system, based on simple images of leaves of healthy and diseased plants. The available dataset contained images captured in [laboratory](#_heading=h.32hioqz)

Studying the literature it has also been found that, with the help of advance technologies like remote sensing, image processing, and Sensor technology, researchers are trying to develop some systems for automatic and early detection of [diseases](#_heading=h.1hmsyys)

Studying the infected leaves, it has been observed that the greenness of the infected portion of the leaves changes significantly with respect to the normal leaves.

# Problem Statement

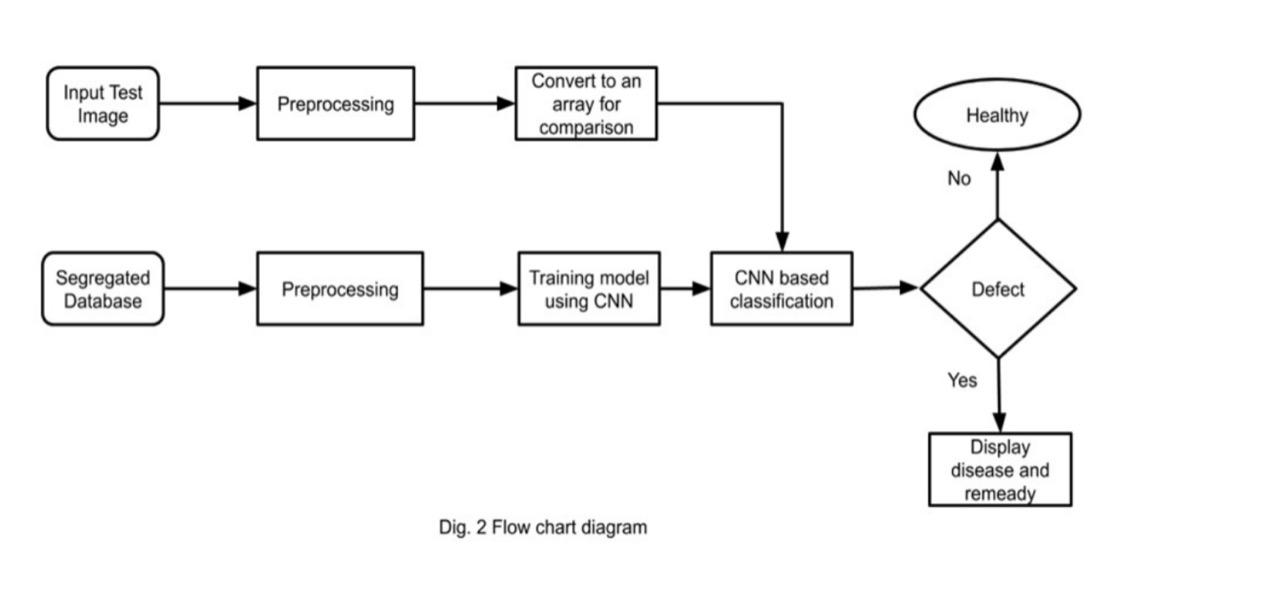
* Leaves, flowers, and fruits change throughout the season, indicating that they are always evolving. During the day, their appearance changes gradually as the amount and angle of incident solar radiation affect their spectral response.
* Manual detection of plant health is a tedious job. Hence, it is required to develop computational methods which will make the process of health identification and classification using leaf images.
* Plant diseases reduce both the quantity and quality of plant yield. Diseases interrupt vital functions of a plant such as photosynthesis, transpiration, pollination, fertilization, germination, etc.

# Objective

* The goal is to help novice plant owners determine if their plants are healthy or need some special attention. If the model can accurately classify an unhealthy plant, then users can take the next step to research and determine whether their plants are being over/under watered, receiving too much/too little sunlight, infested, etc.

# Methodology

* Step 1 - Image Acquisition: This is the process of acquiring the images through a camera by going to the site or from other available sources such as image databases or online repositories. The captured images are in three colors, that is, Red, Green, and Blue (RGB), for which a color transformation structure is created, and a device-independent color space transformation is applied on it.
* Step 2 - Image Pre-processing: To remove noise in an image, different pre-processing techniques are used. Clipping of leaf image is applied to extract the region of the image in which we are interested. The extracted plant leaf image is transferred to a digital system to remove the unnecessary areas. Some essential steps of pre-processing are: Resizing the image, Noise removal from the image, enhancement and smoothing of the image.
* Step 3 - Image Segmentation: This method of image processing is used to partition an image into significant components according to similar characteristics. Various methods are available for image segmentation such as boundary and spot detection algorithm, region and edge-based methods, Otsu’s method, thresholding techniques and k-means clustering, etc.
* Step 4 - Feature Extraction: It is a type of dimension reduction technique that effectively represent the useful part of the image. Various features such as texture, color, edges and morphology can be extracted for the detection of plant disease. Color co-occurrence method is used for feature extraction.
* Step 5 - Classifiers: Classifiers are used to identify and categorize the different diseases that occur on plant leaves based on obtained features. Several classifiers that have been used in earlier work to detect diseases in plants are K-nearest neighbors (K-NN), Support Vector Machines (SVM), Convolutional Neural Network (CNN) and Artificial Neural Network (ANN), etc.



# Implementation

# 

# Algorithm

# 

**CNN Model Steps:**

Conv2D: It is a 2D Convolution Layer, this layer creates a convolution kernel that's wind with layers input which

helps produce a tensor of outputs.

keras.layers.Conv2D(filters, kernel\_size, strides=(1, 1),

padding='valid', data\_format=None, dilation\_rate=(1, 1),

activation=None, use\_bias=True,

kernel\_initializer='glorot\_uniform',

bias\_initializer='zeros', kernel\_regularizer=None,

bias\_regularizer=None, activity\_regularizer=None,

kernel\_constraint=None, bias\_constraint=None

**Maxpooling:**

Max pooling may be a pooling process that choose the very best element from the region of the feature map coveredby the filter. Thus, the output after max-pooling level would be a feature map comprising the foremost important features of the previous feature map.

**Flatten:**

In between the convolutional layer and therefore the fully connected layer, there is a ‘Flatten’ layer. Flattening transforms a two-dimensional matrix of features into a vector which will be fed into a totally connected neural network classifier.

**Image Data Generator:**

Image Data Generator quickly found out Python generators which will automatically turn image files on disk into batches of preprocessed tensors.

**Training Process:**

Effective training begins well before a trainer delivers a private training session and continues then training session is complete. Training are often viewed as a process comprised of 5 related stages or activities: assessment, motivation, design, delivery, and evaluation.

**Epochs:**

An epoch may be a term utilized in machine learning and indicates the amount of passes of the whole training dataset the machine learning algorithm has completed. Datasets are usually grouped into batches (especially when the quantity of knowledge is extremely large).

**Validation Process:**

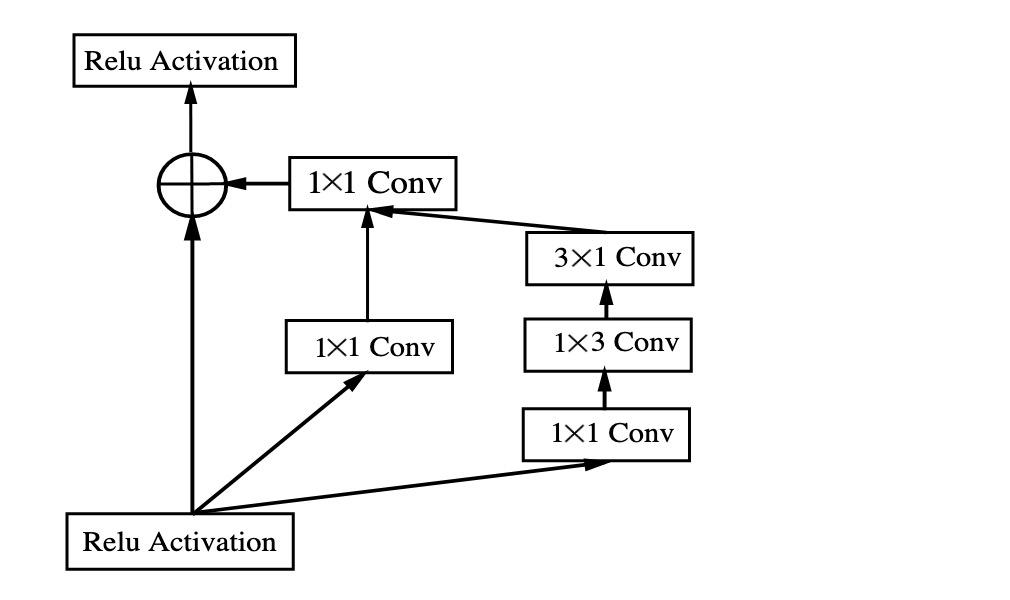
Validation is mentioned because the process where a trained model is evaluated with a testing data set. The testing data set may be a separate portion of an equivalent data set from which the training set springs. the most purpose of using the testing data set is to check the generalization ability of a trained model.

**Training and Testing Model:**

The dataset is preprocessed like Image reshaping, resizing and conversion to an array form. Similar processing is additionally done on the test image. A dataset consisting of about 38 different plant leaf diseases is obtained, out of which any image is often used as a test image for the software.

The train dataset is employed to coach the model (CNN) so that it can identify the test image and therefore the disease it is CNN has different layers that are Dense, Dropout, Activation, Flatten, Convolution2D, and maxpooling2d. After the model is trained successfully, the software can identify the disease if the plant species is contained within the dataset. After successful training and preprocessing, comparison of the test image and trained

model takes place to predict the disease.



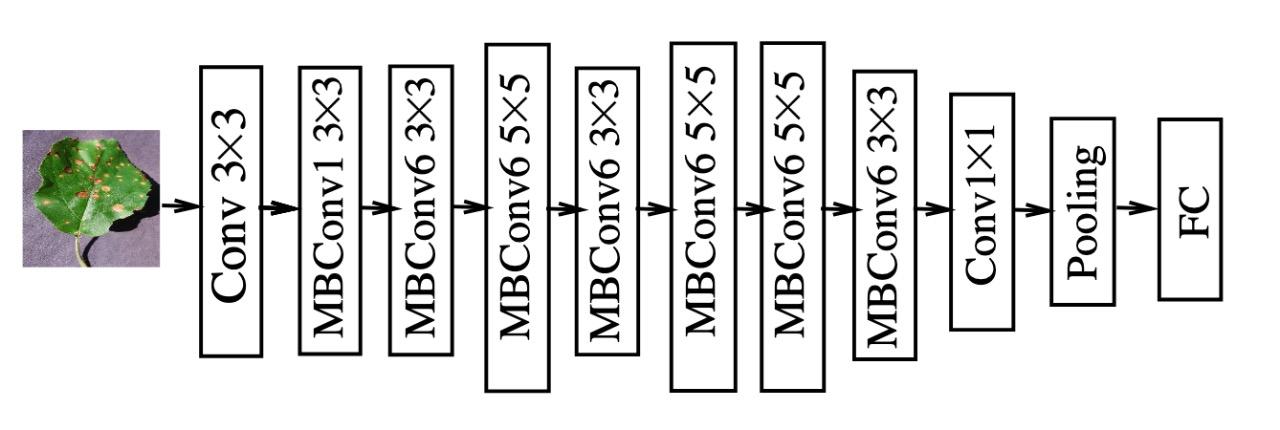


Figure: CNN Model

## Use case diagram

## 

Figure 1: Use Case Digram

## Class diagram

## 

Figure 2: Class Diagram

# Pseudo code

def get\_dataset\_partitions\_tf(ds, train\_split=0.8, val\_split=0.1, test\_split=0.1, shuffle=True, shuffle\_size=10000):

assert (train\_split + test\_split + val\_split) == 1

ds\_size = len(ds)

if shuffle:

ds = ds.shuffle(shuffle\_size, seed=12)

train\_size = int(train\_split \* ds\_size)

val\_size = int(val\_split \* ds\_size)

train\_ds = ds.take(train\_size)

val\_ds = ds.skip(train\_size).take(val\_size)

test\_ds = ds.skip(train\_size).skip(val\_size)

return train\_ds, val\_ds, test\_ds

model = models.Sequential([

resize\_and\_rescale,

layers.Conv2D(32, kernel\_size = (3,3), activation='relu', input\_shape=input\_shape),

layers.MaxPooling2D((2, 2)),

layers.Conv2D(64, kernel\_size = (3,3), activation='relu'),

layers.MaxPooling2D((2, 2)),

layers.Conv2D(64, kernel\_size = (3,3), activation='relu'),

layers.MaxPooling2D((2, 2)),

layers.Conv2D(64, (3, 3), activation='relu'),

layers.MaxPooling2D((2, 2)),

layers.Conv2D(64, (3, 3), activation='relu'),

layers.MaxPooling2D((2, 2)),

layers.Conv2D(64, (3, 3), activation='relu'),

layers.MaxPooling2D((2, 2)),

layers.Flatten(),

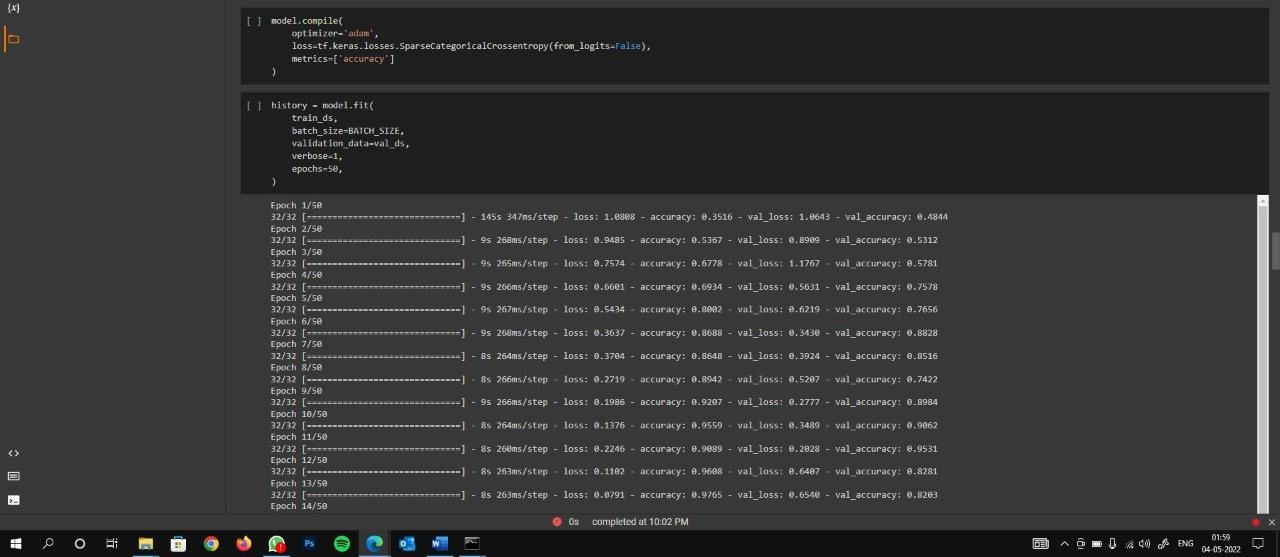
layers.Dense(64, activation='relu'),

layers.Dense(n\_classes, activation='softmax'),

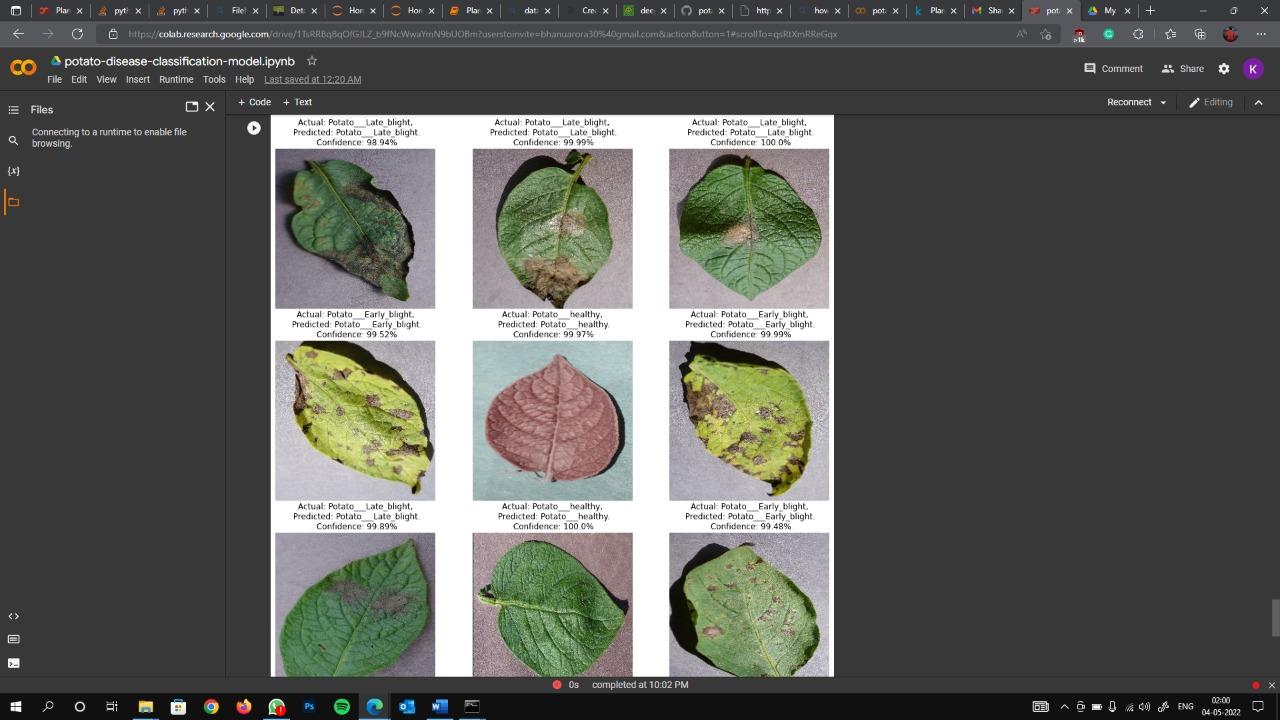
])

model.build(input\_shape=input\_shape)

## Output Screen



Output – 1



Output – 2

# 

# 

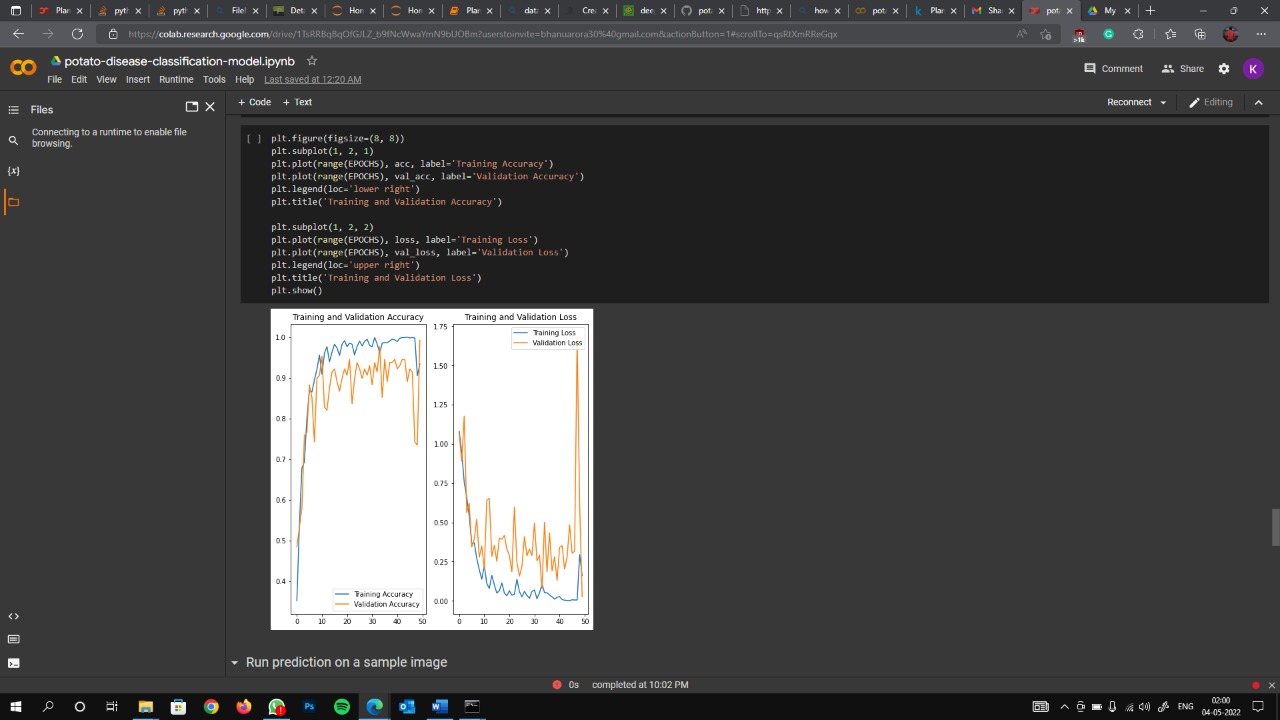
# 

# Software and Hardware Requirements:

* Python based Computer Vision
* Deep Learning libraries
* These will be exploited for the development and experimentation of the project.
* Tools such as Google collab and libraries such as Tensorflow, keras.

# Result Analysis

# We have successfully created a program to analyze whether the plant will be healthy or not.



# Conclusion and Future Scope

* To prevent losses, small holder farmers are dependent on a timely and accurate crop disease diagnosis.
* In this study, a pre-trained Convolutional Neural Network was fine-tuned, and the model was deployed online.
* Agricultural department wants to automate the detecting the yield crops from eligibility process (real time).
* To automate this process by show the prediction end in web application or desktop application. To optimize the work to implement in AI environment.
* The proposed system is based on python and provides an accuracy of around 98%.
* The accuracy and therefore the speed are often increased by use of Googles GPU for processing.
* The system is often installed on Drones in order that aerial surveillances of crop fields are often done.

# References

* Verma, “Precision Agriculture using Internet of Things: An Overview,” in International Journal of Management, IT and Engineering, vol. 9, no. 7, pp. 298-303, July 2019.
* <https://arxiv.org/ftp/arxiv/papers/1604/1604.03169.pdf>
* <https://www.turcomat.org/index.php/turkbilmat/article/download/7743/6139/13981>
* Parismita Bharali, Chandrika Bhuyan, Abhijit Boruah, “Plant Disease Detection by Leaf Image Classification
* Using Convolutional Neural Network” ICICCT 2019, pp.194-205.
* Rahman Tashakkori, Timothy Jassmann, R. Mitchell Parry, “Leaf Classification Convolutional Neural Network
* App” , IEEE Southeast 2015
* M. B. Riley, M. R. Williamson, and O. Maloy, “Plant disease diagnosis. The Plant Health Instructor,” 2002.
* Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk and Darko Stefanovic, “Deep Neural
* Networks Based Recognition of Plant Diseases by Leaf Image Classification”, 2016.
* H. Cartwright, Ed., Artificial Neural Networks, Humana Press, 2015.
* https://www.analyticsvidhya.com/blog/2017/06/architecture-of-convolutional-neural-networks-simplified-
* demystified/
* <https://en.wikipedia.org/wiki/Convolutional_neural_network>
* S. Ajantha, G. Mahapriya, J. Mispa, S. Nithya, Measurement of Multimodal Brain Tumor Segmentation using
* Neural Network in IJERT vol. 5, issue 13.
* Mohammad El-Helly, Ahmed Rafea, Salwa El-Gammal, “An Integrated Image Expert Processing System for
* Leaf Disease Detection and Diagnosis”, Central Lab. For Agricultural System(CLAES), Agricultural Research
* Center(ACR), Giza,Egypt
* https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/.