CONTENTS

Student’s Declaration.......................................................................................... i

Certificate by the Guide...................................................................................... ii

Acknowledgement.............................................................................................. iii

Abstract............................................................................................................... iv

List of Tables....................................................................................................... v

List of Figures..................................................................................................... vi

Chapter-1 INTRODUCTION ..................................................................... ix - xii

* 1. Machine learning ............................................................................................. ix
  2. Significance of Machine learning....................................................................... ix
  3. Types of machine learning .............................................................................. x
  4. Types of Supervised learning............................................................................xi
  5. Convolution Neural Networks........................................................................xi
  6. About COVID-19 and project………………………………………………xii

Chapter-2 DESIGN ................................................................................. xiii - xiv

* 1. Explanation of code and libraries used ...........................................................xiii

Chapter-3 ARCHITECTURE ................................................................... xv - xvi

* 1. Architecture diagram........................................................................................xv
  2. Explanation of architecture..............................................................................xvi

Chapter-4 Implementation ..................................................................... xvii – xxi

* 1. Program Code........................................................................................xvii
  2. Output ..................................................................................................xxi

Conclusion......................................................................................................xxii

References.......................................................................................................xxiii

**CANDIDATE’S DECLARATION**

I hereby, declare that the work presented in this report, entitled “ **Detection of COVID 19 VICTIMS** ”, in fulfilment of the requirement for the award of the degree Bachelor of Engineering in Computer Science and Engineering, submitted in CSE Department, Chandigarh College of Engineering and Technology (Degree Wing), affiliated to Panjab University, Chandigarh, is an authentic record of our own work carried out during my degree under the guidance of Dr. Ankit Gupta The work reported in this has not been submitted by me for award of any other degree or diploma.

AayushiAggarwal

(CO17502)

Abhijeet Baruah

(CO17302)

Samrendra Sagar Dwevedi

(CO17354)

Date: 14/04/2020

Place: CCET

**CERTIFICATE**

This is to certify that the Assignment work entitled “ **Detection of COVID 19 VICTIMS** ”, submitted by **Aayushi Aggarwal (CO17502), Abhijeet Baruah (CO17302) and Samrendra Sagar Dwevedi (CO17354)** infulfillment for the requirements of the award of Bachelor of Engineering Degree in Computer Science & Engineering at Chandigarh College of Engineering and Technology (Degree Wing), Chandigarh is an authentic work carried out by him/her under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project has not been submitted to any other University / Institute for the award of any Degree.

Dr. Ankit Gupta

Dept of CSE

CCET(DegreeWing), Chandigarh

Date: 14/04/2020

Place: CCET

DATE: 17/12/2019

PLACE: CCET

Date : 17/12/19

**ACKNOWLEDGEMENT**

“Any serious and lasting achievement or success, one can never achieve without the help, guidance and co-operation of so many people involved in the work.

I would like to express deep gratitude to Dr. Sunil K. Singh, Head of Department (Computer Science & Engineering), submitted in CSE Department, Chandigarh College of Engineering & Technology(Degree wing),and affiliated to Punjab University, Chandigarh, without whose permission the subject assignments on real-world problems would not be possible. I would also like to thank Dr. Ankit Gupta, CSE Department, who encourages me for taking up this assignment.

I have tried my best to keep report simple yet technically correct. I hope I succeed in my attempt.

**ABSTRACT**

Coronavirus disease (COVID-19), is an infectious disease, which was first identified in Wuhan, China ;COVID-19 is now a global pandemic. As per WHO, COVID-19 has so far affected 213 countries or territories or areas.  The disease causes respiratory illness (like the flu) with symptoms such as a cough, fever, and in more severe cases, difficulty breathing. This pandemic has infected 10,363 people in India, along with 1,036 recovered and 339 deaths. This number of victims is increasing so far. Thus, it becomes really important for concerned authorities to have proper detection equipments of this pandemic.

The following assignment is based on detection of COVID 19 by using image classification based on the X-Ray images of lungs of COVID 19 affected victims and non-affected people. Thus, the no. of corona positive cases is identified by distinguishing the X-Ray images of lungs of the COVID-19 affected victims from those of the non-affected people in the complete dataset.

**LIST OF TABLES**

**DATASET**

The main dataset contains all the following folders. Each of the folders shown below in the Fig-1 contains the images of X-Rays of lungs of various people corresponding to the name of the folder. For Example, ‘covid’ folder in ‘test’ contains the images of X-Ray of lungs people suffering from COVID-19. Whereas ‘normal’ folder in ‘test’ contains the images of X-Ray of lungs people who are not suffering from any such disease.

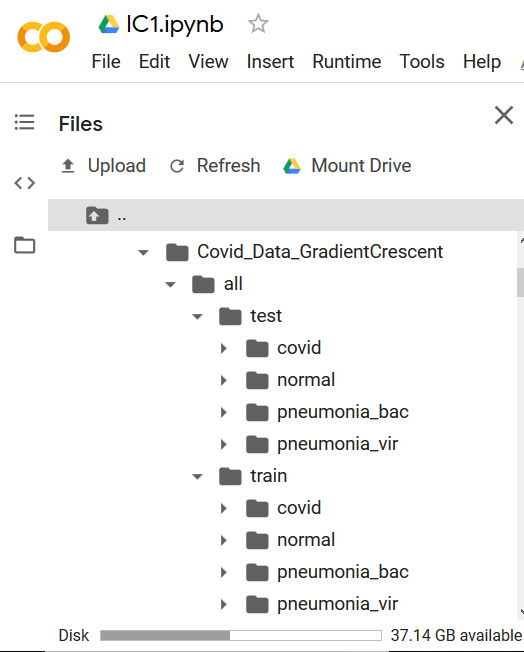
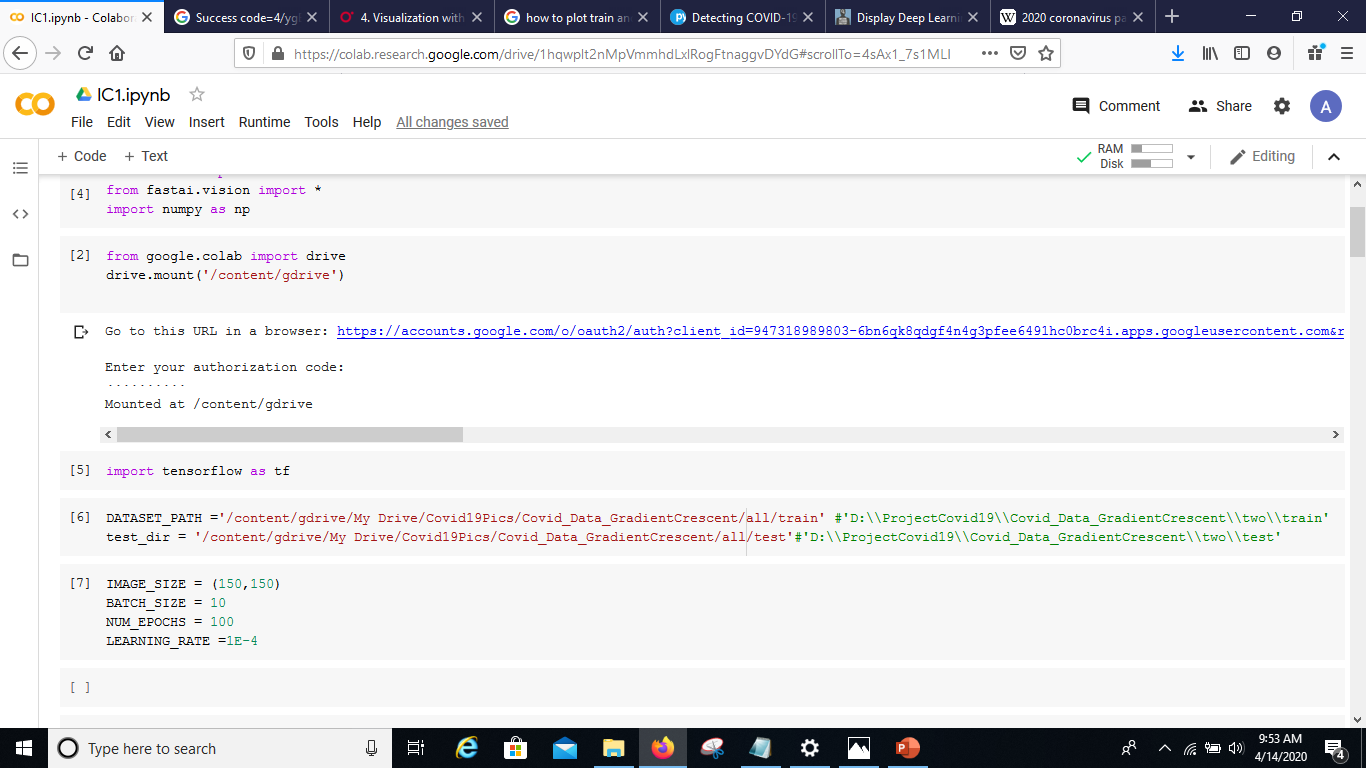
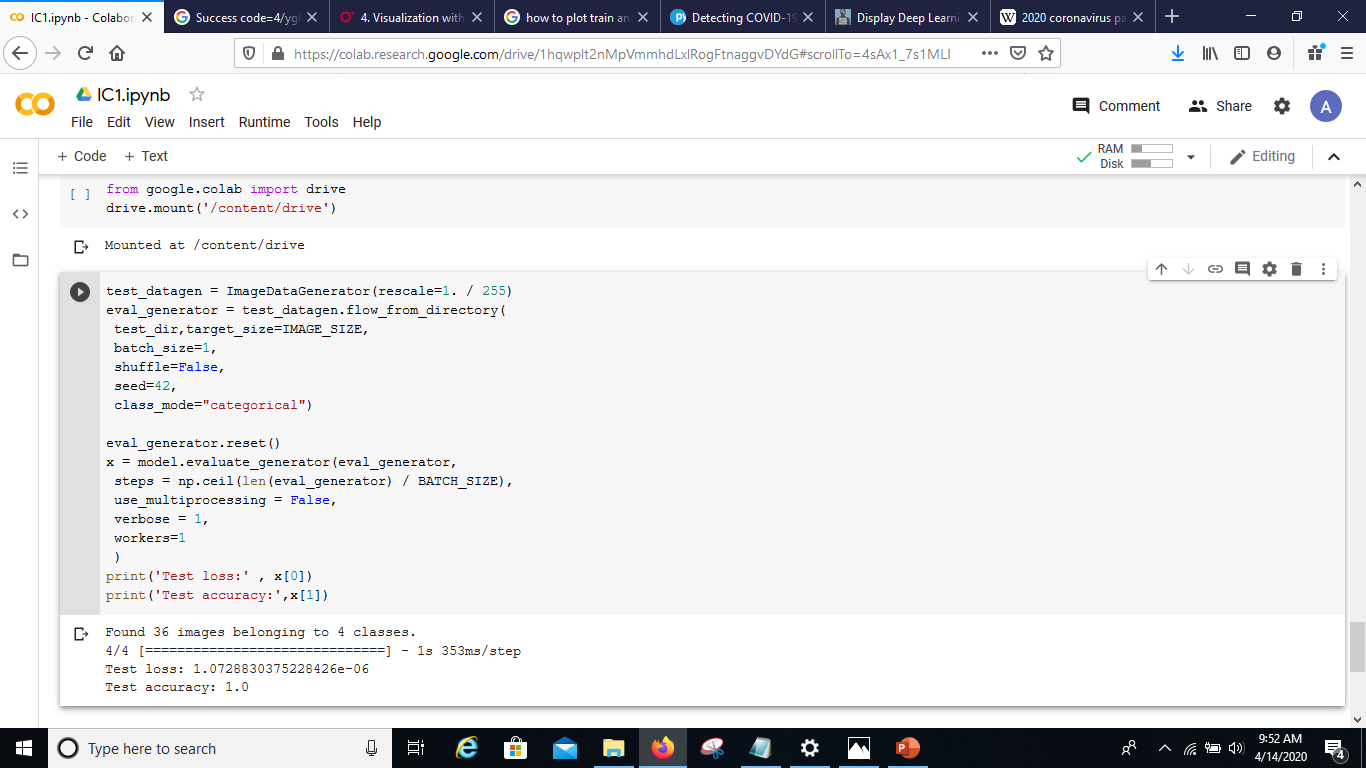


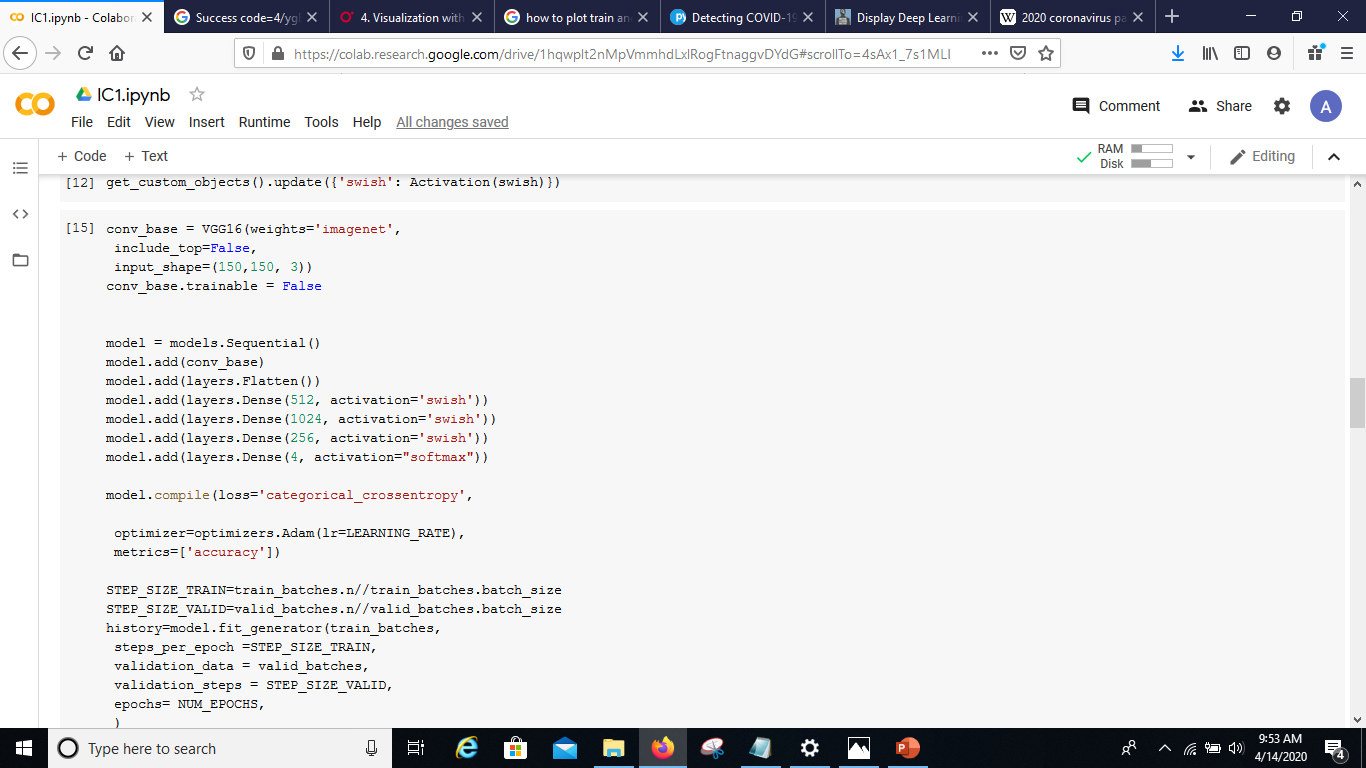
FIG-1 Folders in the Dataset used.

**LIST OF FIGURES**

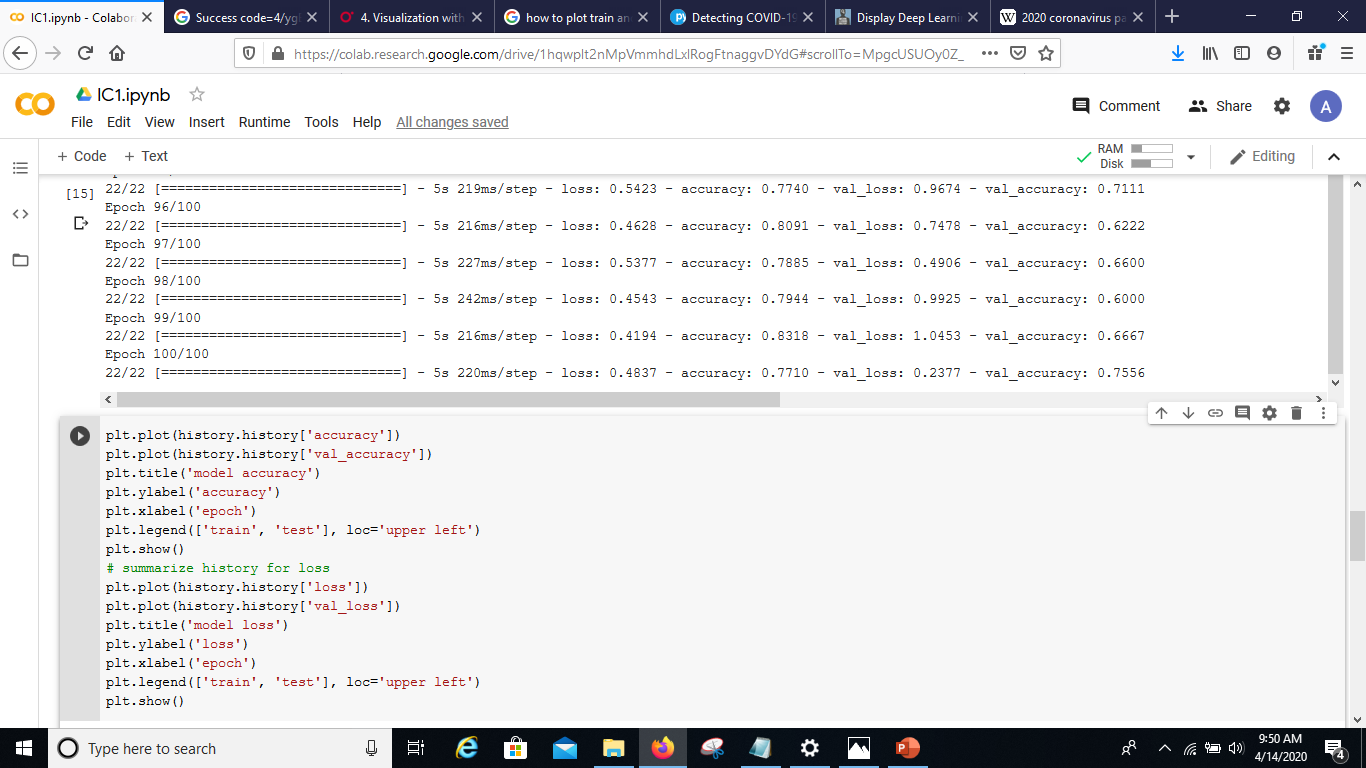
**Figure:- 1** Importing *fastai, numpy, tensorflow and google drive* to access the dataset folder. ****

**Figure:-2** Importing the images from the dataset folder stored in the google drive and assigning variables to them.****

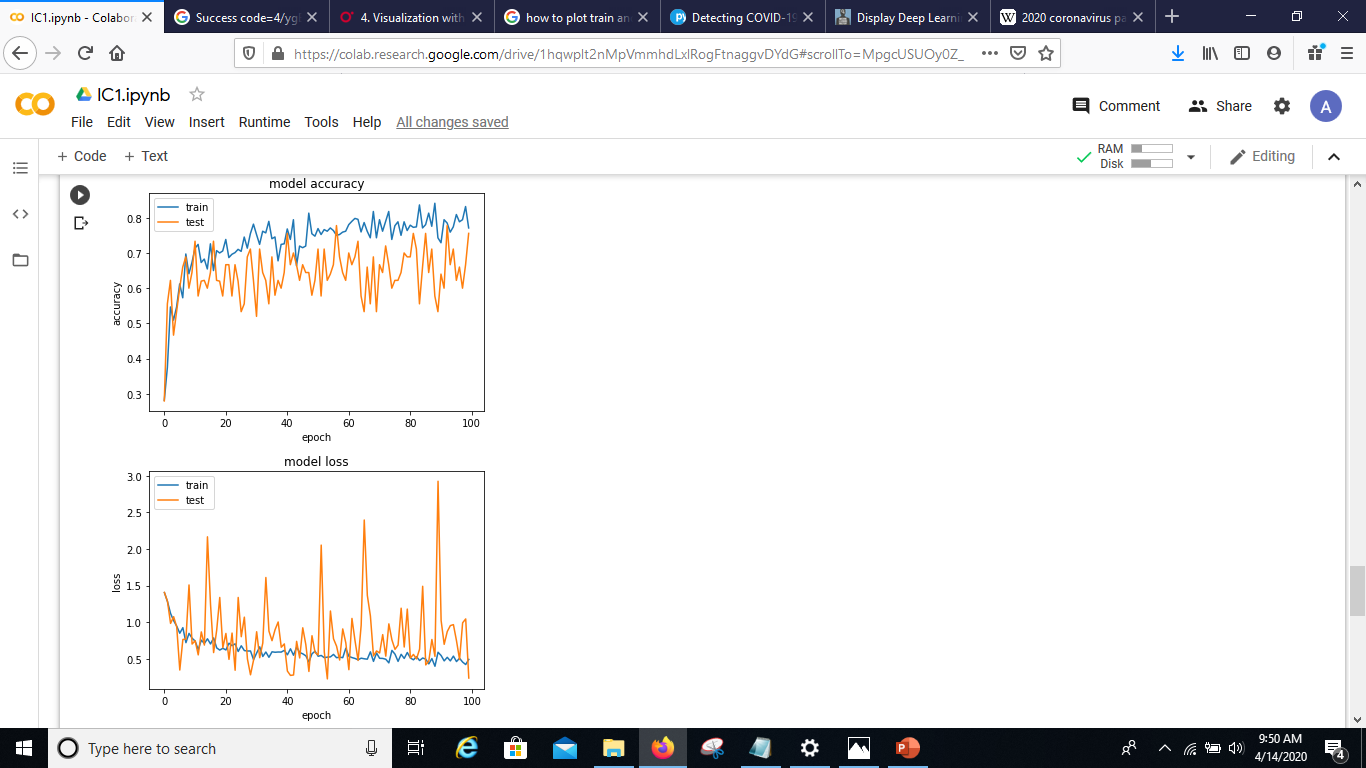
**Figure:-3** Training the model based on the training set images from the dataset folder.

****

**Figure:-4** Output in terms of accuracy value and loss value.

****

**Figure:- 5** Graphical output to depict model accuracy and model loss.

****

**Chapter – 1 INTRODUCTION**

**1.1 MACHINE LEARNING**

Machine Learning is now one of the most important topics around the world. Well, it can even be said as the new electricity in today’s world. But to be precise what is Machine Learning, well it’s just one way of teaching the machine by feeding the large amount of data. Machine learning is defined in 90’s by Arthur Samuel described as the,” *it is a field of study that gives the ability to the computer for self-learn without being explicitly programmed*”, that means imbuing knowledge to machines without hard-coding it. And also “*A computer algorithm/program is said to learn from performance measure P and experience E with some class of tasks T if its performance at tasks in T, as measured by P, improves with experience E.*” -Tom M. Mitchell.

Machine learning is mainly focused on the development of computer programs which can teach themselves to grow and change when exposed to new data. Machine learning studies algorithms for self-learning to do stuff. It can process massive data faster with the learning algorithm. For instance, it will be interested in learning to complete a task, make accurate predictions, or behave intelligently.

**1.2 SIGNIFICANCE OF MACHINE LEARNING**

Data is growing day by day, and it is impossible to understand all of the data with higher speed and higher accuracy. More than 80% of the data is unstructured that is audios, videos, photos, documents, graphs, etc. Finding patterns in data on planet earth is impossible for human brains. The data has been very massive, the time taken to compute would increase, and this is where Machine Learning comes into action, to help people with significant data in minimum time.

It also helps in predicting the outcomes that were not possible in earlier days. For example weather forecasting has become more accurate and also we can make humanoids that can work in case of disaster where humans cannot enter. The ability to process very large amount of data within seconds cannot be done by a human. Decision making has also been improved.

**1.3 TYPES OF MACHINE LEARNING**

1. **Supervised learning: -** Supervised Learning is the first type of machine learning, in which labelleddata used to train the algorithms. In supervised learning, algorithms are trained using marked data, where the input and the output are known. We input the data in the learning algorithm as a set of inputs, which is called as Features, denoted by X along with the corresponding outputs, which is indicated by Y, and the algorithm learns by comparing its actual production with correct outputs to find errors. It then modifies the model accordingly. The raw data divided into two parts. The first part is for training the algorithm, and the other region used for test the trained algorithm. Supervised learning uses the data patterns to predict the values of additional data for the labels. This method will commonly use in applications where historical data predict likely upcoming events.

Ex:- It can anticipate when transactions are likely to be fraudulent or which insurance customer is expected to file a claim. We will focus more in this in this project.

1. **Unsupervised learning :-**  Unsupervised Learning is the second type of machine learning, in which unlabeled data are used to train the algorithm, which means it used against data that has no historical labels. What is being showing must figure out by the algorithm? The purpose is to explore the data and find some structure within. In unsupervised learning the data is unlabeled, and the input of raw information directly to the algorithm without pre-processing of the data and without knowing the output of the data and the data cannot divide into a train or test data. The algorithm figures out the data and according to the data segments, it makes clusters of data with new labels. This learning technique works well on transactional data
2. **Reinforcement learning: -** Reinforcement Learning is the third type of machine learning in which no raw data is given as input instead reinforcement learning algorithm have to figures out the situation on their own. The reinforcement learning frequently used for robotics, gaming, and navigation. With reinforcement learning, the algorithm discovers through trial and error which actions yield the most significant rewards

**1.4 TYPES OF SUPERVISED LEARNING**

This project uses the supervised learning. Specifically Classification so we will focus more on this

There are two types, these are regression and classification: -

1. **Regression: -** Regression is the type of Supervised Learning in which labelled data used, and this data is used to make predictions in a continuous form. The output of the input is always ongoing, and the graph is linear. Regression is a form of predictive modelling technique which investigates the relationship between a dependent variable [Outputs] and independent variable[Inputs]. This technique used for forecasting the weather, time series modelling, process optimisation. Ex:- One of the examples of the regression technique is House Price Prediction, where the price of the house will predict from the inputs such as No of rooms, Locality, Ease of transport, Age of house, Area of a home.
2. **Classification: -** Classification is the type of Supervised Learning in which labelled data can use, and this data is used to make predictions in a non-continuous form. The output of the information is not always continuous, and the graph is non-linear. In the classification technique, the algorithm learns from the data input given to it and then uses this learning to classify new observation. This data set may merely be bi-class, or it may be multi-class too. Ex:- One of the examples of classification problems is to check whether the email is spam or not spam by train the algorithm for different spam words or emails.

We have various methods to perform classification; one of those methods is CNN, i.e. Convolution Neural Networks.

**1.5 CONVOLUTION NEURAL NETWORKS**

Convolution Neural Networks or CNN is a class of deep, feed-forward artificial neural networks ( where connections between nodes do *not* form a cycle) & use a variation of multilayer perceptrons designed to require minimal pre-processing. These are inspired by animal visual cortex. CNNs are generally used in computer vision, however they’ve recently been applied to various NLP tasks and the results were promising.

CNN is very useful where we have images and we want to train a model in it. Thus, CNN is generally used for image classification. CNN most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. They have applications in image and video recognition, recommender systems,image classification, medical image analysis, natural language processing, and financial time series.

The process of building a Convolutional Neural Network always involves four major steps.

Steps 🡪 Convolution 🡪Pooling🡪Flattening🡪Full Connection

**1.6 About COVID-19 and project**

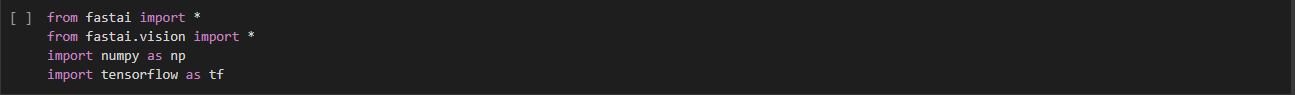
Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment.  Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness. The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. And currently no vaccine is available against this disease.

So its very important to detect those patients and separate them from healthy person. Detection can be done using the tests, tests can be swab test or blood test. Swab test results take approx. 5-6 hours whereas blood test can take 2-3 hours. These test can also be very costly until the manufacturing of low-cost methods of test. But the main issue is to decrease the amount of time for detecting the patients. Now, in this project we are using X-ray image of chest to detect corona virus patient. We are using the fact that corona virus attacks lungs and create suffocation, this is because this virus fills the lungs with fluids. This fluid can be observe by the image of a COVID-19 patient’s lung X-RAY image. According to a study, this virus attacks lower and upper part of lungs equally as compared to few other virus(like SARS virus) who attacks only lower part of lungs. So such differences can be used to differentiate corona virus from other virus infections. Humans currently may not quickly get these differences but complex networks like CNN can find these differences and help us in getting those differences.

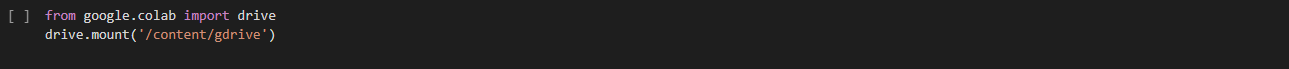
Thus, in the model described below , we are using the images of X-Ray of lungs (Chest) of people to identify the COVID-19 affected people. This has been done using CNN, which distinguishes the images of COVID 19 infected people’s chest X-Ray with those of the others (non-infected) people. This method is implemented as it is a quick corona positive detection method and also, it is cheaper enough to support many countries including the poorer, at the time of pandemic crisis. This model does not try to replace the existing lab testing methods but it is an attempt to help the existing infrastructure to detect more patients quickly. One can run this model on those suspected cases, if this model points out that a patient might be corona virus positive then he/she must undergo lab testing and while the results come, he/she must be quarantine.

**Chapter – 2DESIGN**

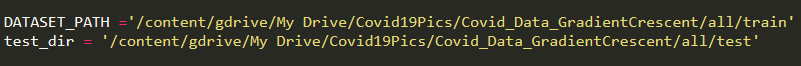
**2.1 EXPLAINATION OF CODE AND LIBRARIES USED**



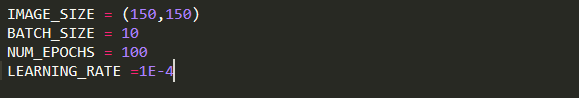
1. Here we are importing many of the helpful libraries which will help us in recognizing and modifing the file according to our need as:
2. **Fast.ai** - Fast.ai is synonymous to transfer learning and achieving great results in a short amount of time.
3. **Numpy** - NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these array
4. **Tensorflow** - TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks.



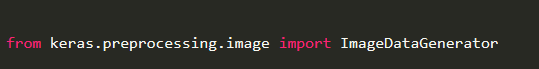
1. Here these code allow us to import the large dataset present in the google drive to the google colab main server which will use these library in learning itself and develop the basic relation between dependent and independent variables.



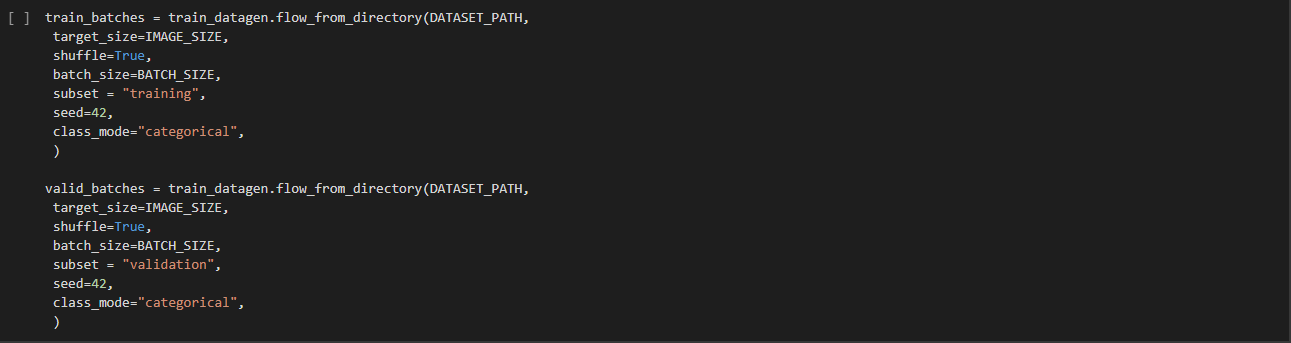
1. In the two lines above we are setting the path variables of the model for getting the Dataset and training set, DATASET\_PATH path is used to train the model and test\_dir path is used to test how much the model is accurate.

****

1. In this code part we have defined the dimensions of images that will be input to VGG16. 224 \* 224 is by default the image input size in VGG16 so we have changed the dimensions to 150\* 150. Batch size is 10, batch size is a term used in machine learning and refers to the number of training examples utilized in one iteration. Epochs defines the number of iteration. Learning rate defines how much the model will learn from every new learning or finding.

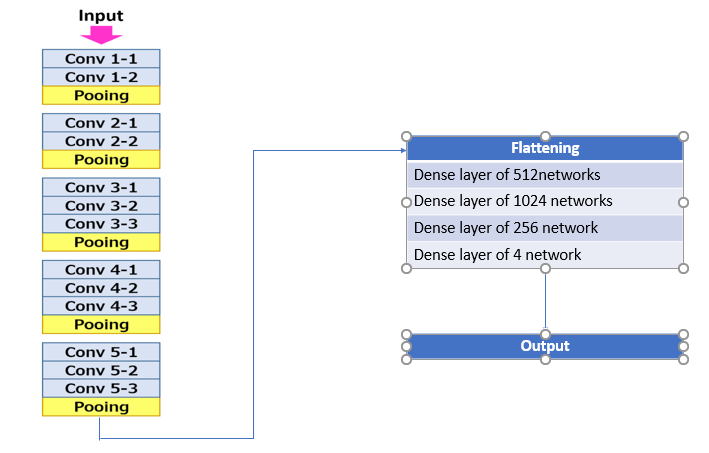
****

1. The ImageDataGenerator class is very useful in image classification as Generate batches of tensor image data with real-time data augmentation. The data will be looped over and trained over them.

****

**Chapter – 3 ARCHITECTURE**

* 1. **Architecture Diagram**

****

In this project we have used VGG16, since this model is very well pre-trained in millions of images. It is currently one of the best model along with VGG19, though both are quit the same the only difference is that VGG16 has less number of groups of layer as compared to VGG19. But for this project VGG16 good because we are adding our own layers in it. VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition”. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU’s.

* 1. **Explanation of architecture**

The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, center). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv.  layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 pixel window, with stride 2.

Usually after this we have other layers in VGG16 but we for this project we have removed these layers and added our own layers so that we can train our model better and can control the accuracy. So for this we have make first flatten the outputs that where received from above VGG16 layers then added a dense layer of 512 channels with activation function sigmoid , then another layer of 1024 channel is added with activation function swish. Here swish is non linear function and is recently found by google and it is better then relu and sigmoid. Another layer of 256 channel is added with activation function swish. Now since we have 4 classes in our model so we add a softmax layer of 4 channels to predict the final outcome.

**Chapter – 4 IMPLEMENTATION**

**4.1 PROGRAM CODE**

!curl -s https://course.fast.ai/setup/colab | bash

from fastai import \*

from fastai.vision import \*

import numpy as np

from google.colab import drive

drive.mount('/content/gdrive')

import tensorflow as tf

DATASET\_PATH ='/content/gdrive/My Drive/Covid19Pics/Covid\_Data\_GradientCrescent/all/train'

test\_dir = '/content/gdrive/My Drive/Covid19Pics/Covid\_Data\_GradientCrescent/all/test'

IMAGE\_SIZE = (150,150)

BATCH\_SIZE = 10

NUM\_EPOCHS = 100

LEARNING\_RATE =1E-4

from keras.preprocessing.image import ImageDataGenerator

train\_datagen = ImageDataGenerator(rescale=1./255,

 rotation\_range=50,

 featurewise\_center = True,

 featurewise\_std\_normalization = True,

 width\_shift\_range=0.2,

 height\_shift\_range=0.2,

 shear\_range=0.25,

 zoom\_range=0.1,

 zca\_whitening = True,

 channel\_shift\_range = 20,

 horizontal\_flip = True ,

 vertical\_flip = True ,

 validation\_split = 0.2,

 fill\_mode='constant')

train\_batches = train\_datagen.flow\_from\_directory(DATASET\_PATH,

 target\_size=IMAGE\_SIZE,

 shuffle=True,

 batch\_size=BATCH\_SIZE,

 subset = "training",

 seed=42,

 class\_mode="categorical",

 )

valid\_batches = train\_datagen.flow\_from\_directory(DATASET\_PATH,

 target\_size=IMAGE\_SIZE,

 shuffle=True,

 batch\_size=BATCH\_SIZE,

 subset = "validation",

 seed=42,

 class\_mode="categorical",

 )

from keras import models

from keras import layers

from keras.applications import VGG16

from keras import optimizers

from keras.layers.core import Flatten, Dense, Dropout, Lambda

from keras.backend import sigmoid

def swish(x, beta = 1):

  return (x \* sigmoid(beta \* x))

from tensorflow.keras.layers import BatchNormalization

from keras.utils.generic\_utils import get\_custom\_objects

from keras.layers import Activation

get\_custom\_objects().update({'swish': Activation(swish)})

conv\_base = VGG16(weights='imagenet',

 include\_top=False,

 input\_shape=(150,150, 3))

conv\_base.trainable = False

model = models.Sequential()

model.add(conv\_base)

model.add(layers.Flatten())

model.add(layers.Dense(512, activation='swish'))

model.add(layers.Dense(1024, activation='swish'))

model.add(layers.Dense(256, activation='swish'))

model.add(layers.Dense(4, activation="softmax"))

model.compile(loss='categorical\_crossentropy',

 optimizer=optimizers.Adam(lr=LEARNING\_RATE),

 metrics=['accuracy'])

STEP\_SIZE\_TRAIN=train\_batches.n//train\_batches.batch\_size

STEP\_SIZE\_VALID=valid\_batches.n//valid\_batches.batch\_size

history=model.fit\_generator(train\_batches,

 steps\_per\_epoch =STEP\_SIZE\_TRAIN,

 validation\_data = valid\_batches,

 validation\_steps = STEP\_SIZE\_VALID,

 epochs= NUM\_EPOCHS,

 )

plt.plot(history.history['accuracy'])

plt.plot(history.history['val\_accuracy'])

plt.title('model accuracy')

plt.ylabel('accuracy')

plt.xlabel('epoch')

plt.legend(['train', 'test'], loc='upper left')

plt.show()

# summarize history for loss

plt.plot(history.history['loss'])

plt.plot(history.history['val\_loss'])

plt.title('model loss')

plt.ylabel('loss')

plt.xlabel('epoch')

plt.legend(['train', 'test'], loc='upper left')

plt.show()

from google.colab import drive

drive.mount('/content/drive')

test\_datagen = ImageDataGenerator(rescale=1. / 255)

eval\_generator = test\_datagen.flow\_from\_directory(

 test\_dir,target\_size=IMAGE\_SIZE,

 batch\_size=1,

 shuffle=False,

 seed=42,

 class\_mode="categorical")

eval\_generator.reset()

x = model.evaluate\_generator(eval\_generator,

 steps = np.ceil(len(eval\_generator) / BATCH\_SIZE),

 use\_multiprocessing = False,

 verbose = 1,

 workers=1

 )

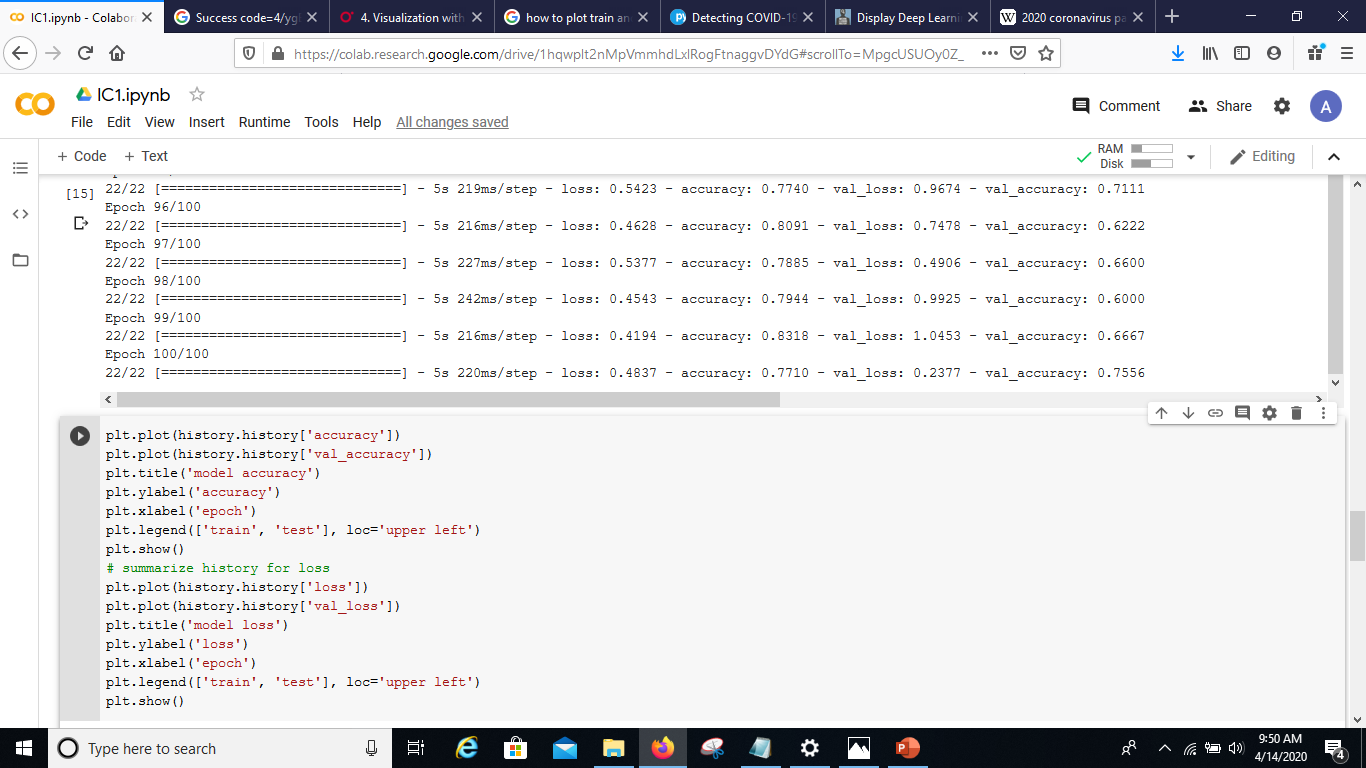
print('Test loss:' , x[0])

print('Test accuracy:',x[1])

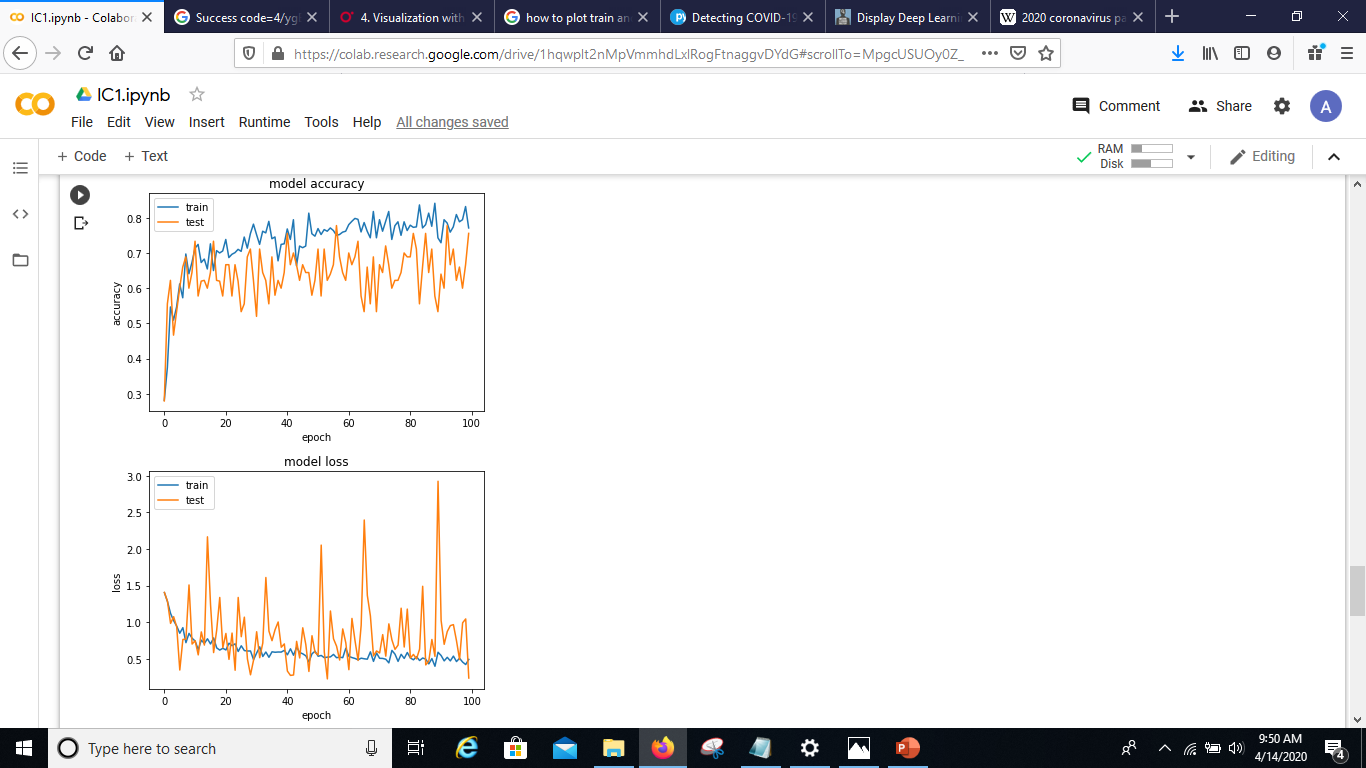
**4.2 OUTPUT**

**FIGURE-1** The following figure gives a textual results of the above implementation. **I**n the above code, ‘steps\_per\_epoch’ holds the number of training images, i.e the number of images the training\_set folder contains. **A**nd ‘epochs’, A single epoch is a single step in training a neural network; in other words when a neural network is trained on every training samples only in one pass we say that one epoch is finished. So training process should consist more than one epochs. In this case we have defined different no. of epochs, as shown in the figure.

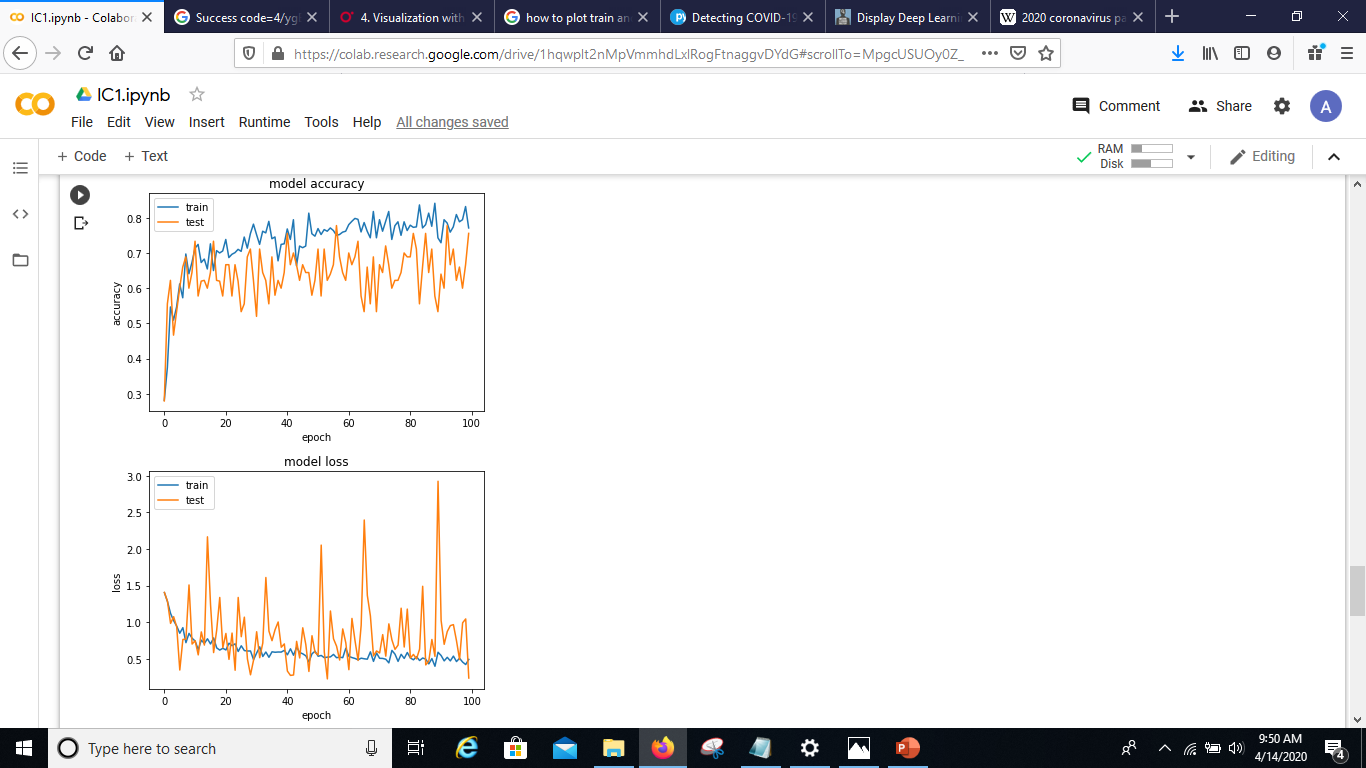
Also, we see the accuracy and loss values obtained at each execution of the above implementation.

****

**FIGURE-2** The following figure shows the plots of accuracy of the model.

****

**FIGURE-3** The following figure shows the plots of loss of the model.

****

**Conclusion**

This project basically tries to detect COVID19 patients and severity in condition of the patient using the Chest X-RAY images of patients. The model currently gives an accuracy of above 78%. This can further be improved if more number of images are added to the database and model is trained on more data. This project is an attempt to decrease the testing cost of covid19 patients also to detect those patients who need more care and monitoring also to make a model which can be useful in early detection of cases as well as making the procedure very fast. Currently different test takes several hours to give the results of a patient and in within this time, he/she can spread the disease so by this model we can detect the patient much quickly and quarantine the patient as soon as possible. In our test results this model has showed 100% accuracy though only about 50 images where used to test the model. As we will have more data we will make this model more accurate to detect the patients and help them in detecting covid19 patients in their early stages. Please note that this model does not try to replace the existing lab methods rather this model can be useful and help the health officials as their helping hand to filter the patients before the undergo lab testing.

**REFRENCES**

1. *https://machinelearningmastery.com/display-deep-learning-model-training-history-in-keras/*
2. *https://www.youtube.com/watch?v=L7qjQu2ry2Q*
3. *https://github.com/anujshah1003/Transfer-Learning-in-keras---custom-data*
4. *https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia*
5. *https://towardsdatascience.com/machine-learning-for-beginners-d247a9420dab*
6. *https://drive.google.com/uc?id=1coM7x3378f-Ou2l6Pg2wldaOI7Dntu1a*
7. *https://pubs.rsna.org/doi/10.1148/radiol.2020200463*