# PROJECT MANUAL

## **Covid-19 Detection from Lung X-rays**

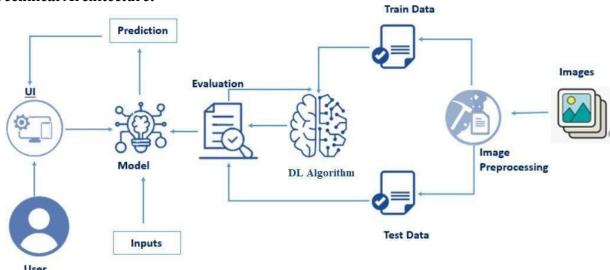
#### **Project Description:**

COVID-19 (coronavirus disease 2019) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is a strain of coronavirus. The disease was officially announced as a pandemic by the World Health Organization(WHO) on 11 March 2020. Given spikes in new COVID-19 cases and the re-opening of daily activities around the world, the demand for curbing the pandemic is to be more emphasized. Medical images and artificial intelligence (AI) have been found useful for rapid assessment to provide treatment of COVID-19 infected patients. The PCR test may take several hours to become available, information revealed from the chest X-ray plays an important role for a rapid clinical assessment. This means if the clinical condition and the chest X-ray are normal, the patient is sent home while awaiting the results of the etiological test. But if the X-ray shows pathological findings, the suspected patient will be admitted to the hospital for close monitoring. Chest X-ray data have been found to be very promising for assessing COVID-19 patients, especially for resolving emergency-department and urgent-care-center overcapacity. Deep-learning (DL) methods in artificial intelligence (AI) play a dominant role as high-performance classifiers in the detection of the disease using chest X-rays.

One of the biggest challenges following the Covid-19 pandemic is the detection of the disease in patients. To address this challenge we have been using the Deep Learning Algorithm to build an image recognition model that can detect the presence of Covid-19 from an X-Ray or CT-Scan image of a patient's lungs.

Transfer learning has become one of the most common techniques that has achieved better performance in many areas, especially in medical image analysis and classification. We used Transfer Learning techniques like Inception V3,Resnet50,Xception V3 that are more widely used as a transfer learning method in medical image analysis and they are highly effective.

### **Technical Architecture:**



### **Prerequisites:**

To complete this project, you must require the following software's, concepts and packages

- Anaconda navigator and PyCharm / Spyder:
  - Refer the link below to download anaconda navigator
  - o Link (PyCharm): <a href="https://youtu.be/1ra4zH2G400">https://youtu.be/1ra4zH2G400</a>
  - Link (Spyder): <a href="https://youtu.be/5mDYijMfSzs">https://youtu.be/5mDYijMfSzs</a>
- Python packages:
  - Open anaconda prompt as administrator
  - o Type "pip install numpy" and click enter.
  - o Type "pip install pandas" and click enter..
  - Type "pip install tensorflow==2.3.2" and click enter.
  - o Type "pip install keras==2.3.1" and click enter.
  - o Type "pip install Flask" and click enter.

### **Prior Knowledge:**

You must have prior knowledge of following topics to complete this project.

- Deep Learning Concepts
  - o CNN: https://towardsdatascience.com/basics-of-the-classic-cnn-a3dce1225add
  - o VGG16:

https://medium.com/@mygreatlearning/what-is-vgg16-introduction-to-vgg16-f2d63849f615

- o ResNet-50:
  - https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33
- o Inception-V3: <a href="https://iq.opengenus.org/inception-v3-model-architecture/">https://iq.opengenus.org/inception-v3-model-architecture/</a>
- o Xception:
  - https://pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/
- **Flask:** Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications.

Link: https://www.voutube.com/watch?v=li4I CvBnt0

### **Project Objectives:**

By the end of this project you'll understand:

- Preprocessing the images.
- Applying Transfer learning algorithms on the dataset.
- How deep neural networks detect the disease.
- You will be able to know how to find the accuracy of the model.
- You will be able to Build web applications using the Flask framework.

#### **Project Flow:**

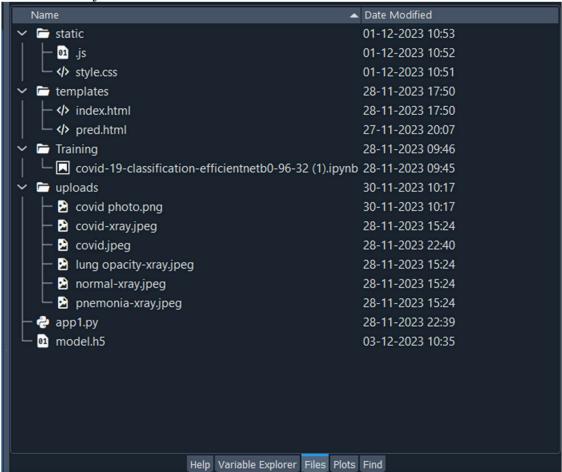
- The user interacts with the UI (User Interface) to choose the image.
- The chosen image analyzed by the model which is integrated with flask application.
- The Xception Model analyzes the image, then the prediction is showcased on the Flask UI.

To accomplish this, we have to complete all the activities and tasks listed below

- o Data Collection.
  - o Create a Train and Test path.
- o Data Pre-processing.
  - o Import the required library
  - Configure ImageDataGenerator class
  - o ApplyImageDataGenerator functionality to Trainset and Testset
- Model Building
  - Pre-trained CNN model as a Feature Extractor
  - Adding Dense Layer
  - o Configure the Learning Process
  - o Train the model
  - Save the Model
  - Test the model
- Application Building
  - Create an HTML file
  - Build Python Code

#### **Project Structure:**

Create a Project folder which contains files as shown below



- Flask folder consists of static, templates and app1.py
- IBM folder consists of trained model notebook
- Training file consist of inceptionv3 ct.ipynb, model training
- Uploads consists of images what we upload during start testing.

#### Milestone 1: Data Collection

There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.

#### Activity 1: Download the dataset

Collect images of Covid-19 Chest X-ray images then organized into subdirectories based on their respective names as shown in the project structure. Create folders of types of Covid-19 that need

to be recognized.

In this project, we have collected images of 2 types of Covid-19 images like Covid-19 positive and Covid-19 negative and they are saved in the respective sub directories with their respective names.

You can download the dataset used in this project using the below link

 $Dataset:-\underline{https://www.kaggle.com/code/rollanmaratov/covid19-detection-using-tensorflow-fr}\\ \underline{om-chest-xray/data}$ 

#### Note: For better accuracy train on more images

We are going to build our training model on Google colab.

Upload the dataset into google drive and connect the google colab with drive using the below code

Once mounted the drive create a folder with project name and move the dataset to that folder and mount to that folder.

### **Activity 2: Create training and testing dataset**

To build a DL model we have to split training and testing data into two separate folders. But In the project dataset folder training and testing folders are presented. So, in this case we just have to assign a variable and pass the folder path to it.

Four different transfer learning models are used in our project and the best model (Xception) is selected. The image input size of xception model is 299, 299.

```
# Concatenate data paths with labels into one dataframe ( to later be fitted into the model )
def define df(files, classes):
   Fseries = pd.Series(files, name= 'filepaths')
   Lseries = pd.Series(classes, name='labels')
   return pd.concat([Fseries, Lseries], axis= 1)
# Split dataframe to train, valid, and test
def split data(data dir):
   # train dataframe
   files, classes = define_paths(data_dir)
   df = define_df(files, classes)
   strat = df['labels']
   train_df, dummy_df = train_test_split(df, train_size= 0.8, shuffle= True, random_state= 123, stratify= strat)
   # valid and test dataframe
   strat = dummy df['labels']
   valid_df, test_df = train_test_split(dummy_df, train_size= 0.5, shuffle= True, random_state= 123, stratify= strat)
   return train df, valid df, test df
```

#### **Milestone 2: Image Preprocessing**

In this milestone we will be improving the image data that suppresses unwilling distortions or enhances some image features important for further processing, although perform some geometric transformations of images like rotation, scaling, translation, etc.

Link: <a href="https://thesmartbridge.com/documents/spsaimldocs/CNNprep.pdf">https://thesmartbridge.com/documents/spsaimldocs/CNNprep.pdf</a>

#### **Activity 1: Importing the libraries**

Import the necessary libraries as shown in the image

```
[2]: 1 # import system libs
      2 import os
      3 import time
      4 import shutil
      5 import pathlib
      6 import itertools
      8 # import data handling tools
      9 import cv2
     10 import numpy as np
     11 import pandas as pd
     12 import seaborn as sns
     13 sns.set_style('darkgrid')
     14 import matplotlib.pyplot as plt
     15 from sklearn.model_selection import train_test_split
     16 from sklearn.metrics import confusion_matrix, classification_report
     18 # import Deep Learning Libraries
     19 import tensorflow as tf
     20 from tensorflow import keras
     21 from tensorflow.keras.models import Sequential
     22 from tensorflow.keras.optimizers import Adam, Adamax
     23 from tensorflow.keras.metrics import categorical_crossentropy
     24 from tensorflow.keras.preprocessing.image import ImageDataGenerator
     25 from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Activation, Dropout, BatchNormalization
     26 from tensorflow.keras import regularizers
     28 # Ignore Warnings
     29 import warnings
     30 warnings.filterwarnings("ignore")
     32 print ('modules loaded')
```

#### **Activity 2: Configure ImageDataGenerator class**

ImageDataGenerator class is instantiated and the configuration for the types of data augmentation. There are five main types of data augmentation techniques for image data; specifically:

- Image shifts via the width\_shift\_range and height\_shift\_range arguments.
- The image flips via the horizontal flip and vertical flip arguments.
- Image rotations via the rotation range argument
- Image brightness via the brightness range argument.
- Image zoom via the zoom range argument.

An instance of the ImageDataGenerator class can be constructed for train and test.

```
tr_gen = ImageDataGenerator(preprocessing_function= scalar, horizontal_flip= True)
ts_gen = ImageDataGenerator(preprocessing_function= scalar)
```

#### Activity 3: Apply ImageDataGenerator functionality to Train set and Test set

Let us apply ImageDataGenerator functionality to the Train set and Test set by using the following code. For Training set using flow from directory function.

This function will return batches of images from the subdirectories

#### Arguments:

- directory: Directory where the data is located. If labels are "inferred", it should contain subdirectories, each containing images for a class. Otherwise, the directory structure is ignored.
- batch size: Size of the batches of data which is 32.
- target size: Size to resize images after they are read from disk.
- class mode:
  - 'int': means that the labels are encoded as integers (e.g. for sparse categorical crossentropy loss).
  - 'categorical' means that the labels are encoded as a categorical vector (e.g. for categorical crossentropy loss).
  - 'binary' means that the labels (there can be only 2) are encoded as float32 scalars with values 0 or 1 (e.g. for binary crossentropy).
  - None (no labels).

Total the dataset is having 19159 train images and 2006 test images divided under 4 classes

### **Milestone 3: Model Building**

Now it's time to build our model. Let's use the pre-trained model which is Xception, one of the convolution neural net (CNN) architecture which is considered as a very good model for Image classification.

Deep understanding on the Xception model – Link is referred to in the prior knowledge section. Kindly refer to it before starting the model building part.

#### Activity 1: Pre-trained CNN model as a Feature Extractor

For one of the models, we will use it as a simple feature extractor by freezing all the five convolution blocks to make sure their weights don't get updated after each epoch as we train our own model.

Here, we have considered images of dimension (299,299,3).

Also, we have assigned include\_top = False because we are using convolution layer for features extraction and wants to train fully connected layer for our images classification(since it is not the part of Imagenet dataset)

Flatten layer flattens the input. Does not affect the batch size.

#### **Activity 2: Adding Dense Layers**

A dense layer is a deeply connected neural network layer. It is the most common and frequently used layer. Let us create a model object named model with inputs as xception.input and output as dense layer.

#### **Generic Model Creation**

```
1 # Create Model Structure
2 img size = (224, 224)
3 channels = 3
4 img_shape = (img_size[0], img_size[1], channels)
5 class_count = len(list(train_gen.class_indices.keys())) # to define number of classes in dense layer
7 # create pre-trained model (you can built on pretrained model such as : efficientnet, VGG , Resnet )
8 # we will use efficientnetb3 from EfficientNet family.
9 base_model = tf.keras.applications.efficientnet.EfficientNetB0(include_top= False,
                               weights= "imagenet", input_shape= img_shape, pooling= 'max')
11
12 model = Sequential([
13
    base_model,
       BatchNormalization(axis= -1, momentum= 0.99, epsilon= 0.001),
14
       Dense(256, kernel_regularizer= regularizers.12(1= 0.016), activity_regularizer= regularizers.11(0.006),
15
                   bias_regularizer= regularizers.l1(0.006), activation= 'relu'),
17
       Dropout(rate= 0.45, seed= 123),
       Dense(class_count, activation= 'softmax')
18
19 ])
21 model.compile(Adamax(learning_rate= 0.001), loss= 'categorical_crossentropy', metrics= ['accuracy'])
```

The number of neurons in the Dense layer is the same as the number of classes in the training set. The neurons in the last Dense layer, use softmax activation to convert their outputs into respective probabilities. Understanding the model is a very important phase to properly use it for training and prediction purposes. Keras provides a simple method, summary to get the full information about the model and its layers.

```
22
23 model.summary()
```

Layer (type)	Output Shape	Param #
efficientnetb0 (Functional)	(None, 1280)	4049571
batch_normalization (BatchNormalization)	(None, 1280)	5120
dense (Dense)	(None, 256)	327936
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 4)	1028

#### **Activity 3: Configure the Learning Process**

The compilation is the final step in creating a model. Once the compilation is done, we can move on to the training phase. The loss function is used to find errors or deviations in the learning process. Keras requires a loss function during the model compilation process.

Optimization is an important process that optimizes the input weights by comparing the prediction and the loss function. Here we are using adam optimizer

Metrics are used to evaluate the performance of your model. It is similar to the loss function, but not used in the training process

#### **Activity 4: Train the model**

Now, let us train our model with our image dataset. The model is trained for 25 epochs and after every epoch, the current model state is saved if the model has the least loss encountered till that time. We can see that the training loss decreases in almost every epoch till 10 epochs and probably there is further scope to improve the model.

**fit generator** functions used to train a deep learning neural network

#### **Arguments:**

- steps\_per\_epoch: it specifies the total number of steps taken from the generator as soon as one epoch is finished and the next epoch has started. We can calculate the value of steps\_per\_epoch as the total number of samples in your dataset divided by the batch size.
- Epochs: an integer and number of epochs we want to train our model for.
- validation data can be either:
  - an inputs and targets list
  - a generator
  - an inputs, targets, and sample\_weights list which can be used to evaluate the loss and metrics for any model after any epoch has ended.
- validation\_steps: only if the validation\_data is a generator then only this argument can be used. It specifies the total number of steps taken from the generator before it is stopped at every epoch and its value is calculated as the total number of validation data points in your dataset divided by the validation batch size.

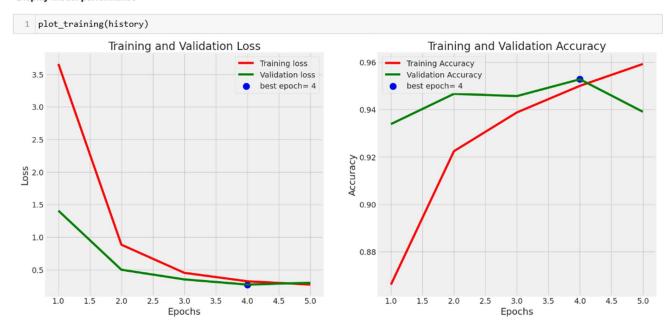
#### Set Callback Parameters

#### **Set Callback Parameters**

#### **Evaluate model**

```
[15]: 1 ts_length = len(test_df)
      2 test_batch_size = test_batch_size = max(sorted([ts_length
                       // n for n in range(1, ts_length + 1) if ts_length%n == 0 and ts_length/n <= 80]))</pre>
      4 test_steps = ts_length // test_batch_size
      6 train_score = model.evaluate(train_gen, steps= test_steps, verbose= 1)
      7 valid_score = model.evaluate(valid_gen, steps= test_steps, verbose= 1)
      8 test_score = model.evaluate(test_gen, steps= test_steps, verbose= 1)
     10 print("Train Loss: ", train_score[0])
     11 print("Train Accuracy: ", train_score[1])
     12 print('-' * 20)
     print("Validation Loss: ", valid_score[0])
     14 print("Validation Accuracy: ", valid_score[1])
     15 print('-' * 20)
     16 print("Test Loss: ", test_score[0])
     17 print("Test Accuracy: ", test_score[1])
     29/29 [============ ] - 1s 47ms/step - loss: 0.2537 - accuracy: 0.9612
     Train Loss: 0.20062509179115295
     Train Accuracy: 0.9849137663841248
     Validation Loss: 0.2536519765853882
     Validation Accuracy: 0.9612069129943848
     Test Loss: 0.25459566712379456
     Test Accuracy: 0.9641001224517822
```

#### Display model performance



From the above run time, we can easily observe that at 9<sup>th</sup> epoch the model is giving the better accuracy, which is initialized through call back parameter.

#### **Activity 5: Testing the Model**

Model testing is the process of evaluating the performance of a deep learning model on a dataset that it has not seen before. It is a crucial step in the development of any machine learning model, as it helps to determine how well the model can generalize to new data.

#### **Confusion Matrics and Classification Report**

```
g_dict = test_gen.class_indices
 2 classes = list(g_dict.keys())
 4 # Confusion matrix
 5 cm = confusion matrix(test gen.classes, y pred)
 6 plot_confusion_matrix(cm= cm, classes= classes, title = 'Confusion Matrix')
 8 # Classification report
 9 print(classification_report(test_gen.classes, y_pred, target_names= classes))
Confusion Matrix, Without Normalization
[[359 0 3 0]
 [ 5 561 36
               0]
 [ 8 18 992
               1]
 [ 1
      0 4 129]]
                precision
                           recall f1-score support
         COVID
                     0.96
                               0.99
                                        0.98
                                                   362
                     0.97
                               0.93
                                        0.95
                                                   602
  Lung_Opacity
        Normal
                     0.96
                               0.97
                                        0.97
                                                  1019
Viral Pneumonia
                     0.99
                               0.96
                                        0.98
                                                  134
                                        0.96
                                                  2117
      accuracy
     macro avg
                     0.97
                               0.96
                                        0.97
                                                  2117
  weighted avg
                     0.96
                               0.96
                                        0.96
                                                  2117
```

#### Milestone 4: Save the Model

The model is saved with .h5 extension as follows

An H5 file is a data file saved in the Hierarchical Data Format (HDF). It contains multidimensional arrays of scientific data.

#### Save model

```
image: model_name = model.input_names[0][:-6]
    subject = 'Covid-19'
    acc = test_score[1] * 100
    save_path = ''

# Save model
    save_id = str(f'{model_name}-{subject}-{"%.2f" %round(acc, 2)}.h5')
    model_save_loc = os.path.join(save_path, save_id)
    model.save(model_save_loc)
    print(f'model was saved as {model_save_loc}')

# Save weights
    weight_save_id = str(f'{model_name}-{subject}-weights.h5')
    weight_save_id = str(f'{model_name}-{subject}-weights.h5')
    weight_save_id = os.path.join(save_path, weight_save_id)
    model.save_weights(weights_save_loc)
    print(f'weights were saved as {weights_save_loc}')
```

model was saved as efficientnetb0-Covid-19-96.41.h5 weights were saved as efficientnetb0-Covid-19-weights.h5

#### **Milestone 5: Application Building**

#### **Milestone 5: Application Building**

In this section, we will be building a web application that is integrated to the model we built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI.

This section has the following tasks

- Building HTML Pages
- Building server side script

#### **Activity1: Building Html Pages:**

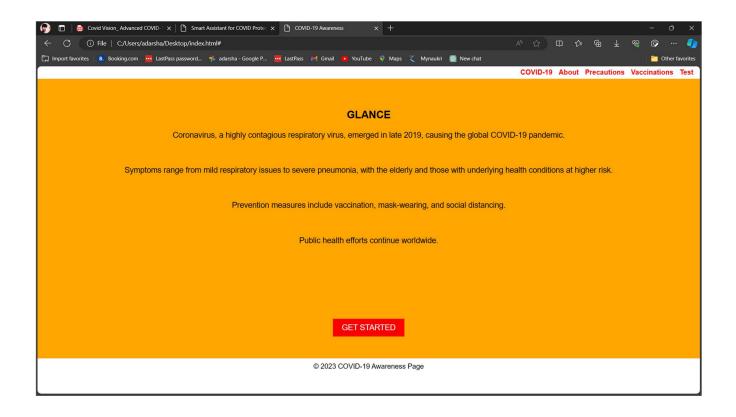
For this project create one HTML file namely

Index.html

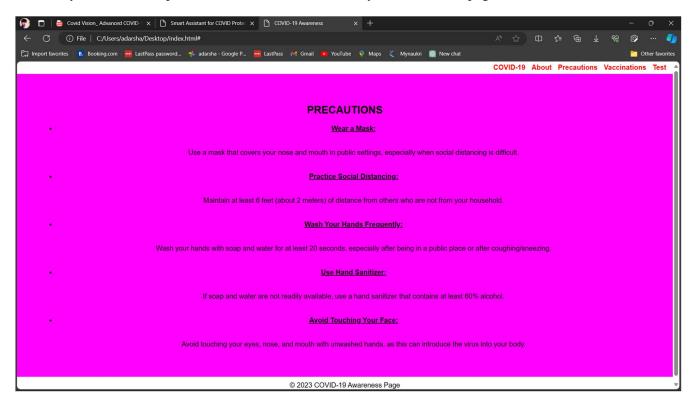
Let's see how our index.html page looks like:



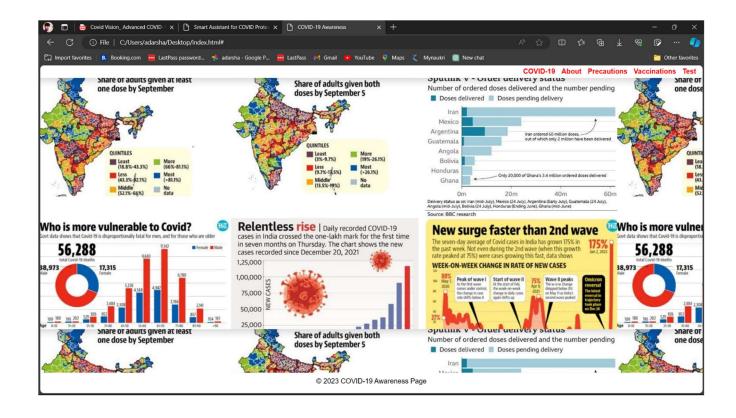
When you click on about button on the top, you will be redirecting to the following page



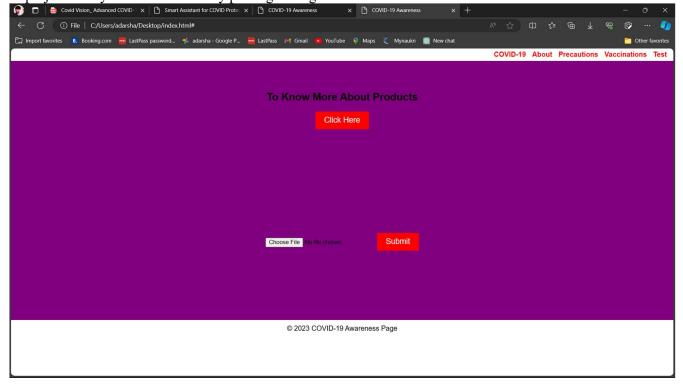
When you click on the precautions button, it will redirect you to the below page



When you click on vaccinations, it will direct you the below page, where you can find the complete static information about the cases and vaccinations



When you click on the test button ,it will display the below page, where you can know more about our projects and you can the model by passing a image



#### **Activity 2: Build Python code:**

Import the libraries

```
import re
import numpy as np
import os
from flask import Flask, app,request,render_template
from tensorflow.keras import models
from tensorflow.keras.models import load_model
from tensorflow.keras.preprocessing import image
from tensorflow.python.ops.gen_array_ops import concat
from tensorflow.keras.applications.inception_v3 import preprocess_input
import requests
from flask import Flask, request, render_template, redirect, url_for
#Loading the model
```

Loading the saved model and initializing the flask app

```
# Load the model
model = load_model("model.h5", compile=False)
app = Flask(__name__)
# default bore page as paute
```

#### Render HTML pages:

```
# default home page or route

@app.route('/')
def index():
    return render_template('index.html')
@app.route('/home')
def home():
    return render_template('index.html')
```

Once we uploaded the file into the app, then verifying the file uploaded properly or not. Here we will be using declared constructor to route to the HTML page which we have created earlier.

In the above example, '/' URL is bound with home.html function. Hence, when the home page of the web server is opened in browser, the html page will be rendered. Whenever you enter the values from the html page the values can be retrieved using POST Method.

```
@app.route('/result', methods=["GET", "POST"])
def upload():
    if request.method == "POST":
        f = request.files['image']
        print("current path")
        basepath = os.path.dirname( file )
        print("current path", basepath)
         filepath = os.path.join(basepath, 'uploads', f.filename)
         f.save(filepath)
        img = image.load_img(filepath, target_size=(299, 299))
        x = image.img_to_array(img)
        print(x)
        x = np.expand_dims(x, axis=0)
        print(x)
        y = model.predict(x)
        preds = np.argmax(y, axis=1)
         # preds = model.predict_classes(x)
        print("prediction", preds)
        index = ['COVID', 'Lung_Capacity', 'Normal', 'Viral Pneumonia']
text = "The person suffer from : " + str(index[preds[0]])
        print(text)
    return render_template('pred.html', y=text)
if __name__ == '__main__':
    app.run(debug=False)
```

Here we are routing our app to res function. This function retrieves all the values from the HTML page using Post request. That is stored in an array. This array is passed to the model.predict() function. This function returns the prediction. And this prediction value will rendered to the text that we have mentioned in the index.html page earlier.

#### **Main Function:**

#### **Activity 3: Run the application**

- Open the Anaconda prompt from the start menu.
- Navigate to the folder where your Python script is.
- Now type the "python app.py" command.
- Navigate to the localhost where you can view your web page.
- Click on the predict button from the top right corner, enter the inputs, click on the submit button, and see the result/prediction on the web

```
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.

* Debugger is active!

* Debugger PIN: 260-300-224

* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

The home page looks like this. When you click on the button "Drop in the image you want to validate!", you'll be redirected to the predict section

click on test button

Input 1:



Once you upload the image and click on submit button, the output will be displayed in the below page

Output:1



COVID-1	) About	Precautions	Vaccinations	Test
To Know More About Products				
Click Here				
Choose File Lovid-xray jpeg				
© 2023 COVID-19 Awareness Page				

