COVID-19 DETECTION USING CHEST X-RAY IMAGES

Project Phase-I report submitted
in
partial fulfilment of requirement for the award of
degree of

Bachelor of Engineering
in
Computer Science and Engineering

by

Mr. Amey S.Didolkar

Mr. Jatin P.Pahuja

Ms. Yamini G.Bhandarkar

Guide
Prof. Prashant Khobragade

Department of Computer Science and Engineering

G.H. Raisoni College of Engineering, Nagpur

(An Autonomous institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

November 2020

Declaration

We, hereby declare that the project report "COVID-19 Detection Using Chest X-Ray Images" Submitted here in has been carried out by us in the partial fulfilment of requirement for that award of degree Bachelor of Engineering in Computer Science and Engineering. The work is original and has not been submitted earlier as a whole or in part for the award of any degree / diploma at this or any other Institution / University.

We also hereby assign to G.H.Raisoni College of Engineering Nagpur all rights under copyright that may exist in and to the above work and any revised or expanded derivatives works based on the work as mentioned. Other work copied from references, manuals, etc. are disclaimed.

Place: Nagpur
Date:

Amey S.Didolkar
Jatin P.Pahuja
Yamini G.Bhandarkar

(Name and Signature)

Certificate

The project report titled "**Covid-19 Detection Using Chest X-ray Images"** submitted by **Amey S.Didolkar, Jatin Pahuja and Yamini Bhandarkar** for the award of degree of Bachelor of Engineering in Computer Science and Engineering has been carried out under our supervision. The work is comprehensive, complete and fit for evaluation.

Dr. Prashant Borkar
Head
Department of Computer Science and Engineering G.H.R.C.E, Nagpur

Prof. Prashant Khobragade
Assistant/Associate Professor
Department of Computer Science and Engineering G.H.R.C.E, Nagpur

Acknowledgement

This is to declare that this report has been written by us. No Part of the report is plagiarized from other sources. All information included from other sources have duly acknowledgement. We aver that if any part of report is found to be plagiarized, we shall take responsibility for it.

Signature of Author
Amey Sanjay Didolkar

Signature of Author
Jatin Pankaj Pahuja

Signature of Author
Yamini Gajanan Bhandarkar

Abstract

The spreading increase in covid-19 patients is overwhelming healthcare systems all over the world. With limited testing kits, every patient with respiratory illness cannot treat using conventional techniques. Deep Learning has boost multi-fold in recent years, and it has played a significant role in image classification, including medical imaging. Convolutional Neural Networks (CNNs) have performed well in detecting many diseases, including coronary artery disease, malaria, Alzheimer's disease, and different dental diseases. The test also has a long turn-around-time and limited sensitivity. The study reveals that infected patients exhibit distinct radiographic visual characteristics, fever, dry cough, fatigue, dyspnoea. Diagnosing possible covid-19 infections on chest X-ray may help high-risk quarantine patients while test results await. X-ray machines are readily available at all the healthcare centers, with no transportation time involved for samples. This project proposes using a chest x-ray to classify the patient's selection for further testing and treatment. The detection of critical acute respiratory syndrome coronavirus responsible for coronavirus disease 2019 (COVID-19), using chest X-ray images, has life-saving importance for both patients and doctors. Also, in countries that cannot purchase laboratory kits for testing, this becomes even more vital. This work shows an analysis of how a change in convolutional layers and an increase in dataset affect classifying performances.

List of Figures

Index

CHAPTER - 01

Introduction

1.1 Problem Statement

The Coronavirus (COVID-19) is unique among humans, which has high transmissibility, notable fatal deaths in some high-risk groups, and the ability to cause enormous societal and economic disruption in the nation. A daily increase in the number of patients of COVID-19, a new respiratory virus, has put a load on healthcare organizations all over the world. There are limited kits for diagnosing, limited beds in hospitals to admit patients, limited Personal Protective Equipment (PPE) for medical staff, and limited ventilators in many countries. Testing a large number of patients is difficult and costly. Thus, it is crucial to decide which respiratory patient could have covid-19 and be treated efficiently with limited resources. It is necessary to differentiate whether a patient is covid infected or having other diseases admitted in the covid ward, and non-infected patients admitted in different wards. But covid test results took at least two days to be diagnosed, and the patient's body shows significant symptoms in a span of 14 days from the day he got in contact with the virus. By the time patient's lungs get more affected, and he comes in contact with many more people spreading the virus to many.

1.2 Overview

The COVID-19 virus is unique among human corona viruses, which has a capacity of high transmissibility, substantial fatal deaths in some high-risk groups, and the ability to cause huge societal and economic disruption in the nation. The Number of a patient that is a daily increase of COVID-19, a new respiratory virus, has put a load on healthcare organizations worldwide. There are limited kits for diagnosing, limited beds in hospitals to admit patients, limited Personal Protective Equipment (PPE) for medical staff, and limited ventilators in many countries.

1. Testing a large number of patients is difficult and costly. Thus, it is important to decide which respiratory patient could have covid-19 and be treated efficiently with limited resources.
2. It is necessary to differentiate whether a patient is covid infected or having other diseases so that the infected patients can be admitted in covid ward, and non-infected patients are admitted in other wards.
3. Covid test results took at least two days to be diagnosed, and patient's body shows significant symptoms in the span of 14 days from the day he got in contact with the virus. By the time the patient's lungs get more affected, and he comes in contact with many more people spreading viruses to many.

Artificial Intelligence which is inspired by the human brain's working procedure. DL has the capability of unsupervised learning, i.e., to learn from the examples with unlabeled data. The features like unlabeled data utilization, working without feature engineering, and prediction with high accuracy and precision make DL very popular with Artificial Intelligence (AI) and Big Data analysis Convolutional Neural Network (CNN) is a DL algorithm which has been performing very well in solving problems Idea of an artificial neuron is first conceptualized.

This model is considered to be the base and predecessor of CNN. CNN is form of artificial neurons that have the property of self-optimization with learning like the brain neurons. Due to this self-optimizing property, it can extract and classify the features extracted from images more precisely than any other algorithm. Moreover, it needs very limited pre-processing of the input data through it yields highly accurate and precise results.

1.3 Objective

The main objective of "COVID-19 Detection using Chest X-Ray Images" is to facilitate a user-friendly environment for all users and decrease the manual effort. In past days detection is conducted manually, but in a further resolution of technology, we can generate the score and pose the queries automatically. Chest X-Ray imaging technique that plays an important role in the diagnosis of COVID-19 disease.

CHAPTER - 02

Literature Survey

2.1 Introduction

As of on-going Pandemic covid-19 is a devastating effect all over the world. Day-by-Day cases are going at a higher level breaking its previous day records. Thousands and Lakhs of tests are being conducted, and its cure is yet to be discovered. In structure to control the spread of covid-19, many suspected cases need to be screened for proper isolation and treatment. To ease doctors' jobs, radiographic images come into the picture, i.e. X-ray images of the human chest (Lungs) can be used to diagnose the infection. Machine Learning and Deep learning classification techniques can be used to classify whether the suspected patient is infected with the coronavirus (covid-19). Corona Infected Lungs get yellow in colour, which makes the X-ray of the chest fade.

Numbers of X-ray images are used to train the Machine Learning and Deep Learning Model and then the model can give more than 90% accuracy in the prediction.

This method can be adopted to decide the patient should be shifted to the corona ward with all infected patients or in the separate ward with the non-infected patients as the symptoms take a minimum of 10-14 days to get on the reports. Radiographic Images would help to diagnose and make a decision on a patient's health.

2.2 Summary

In conclusion, the selection of accurate technique is very important to make sure that system is successfully implemented and achieved the objective.

CHAPTER - 03

Methodology

3.1 Introduction

A methodology is a collection of methods, practices, processes, techniques, procedures, and rules. In project management, methodologies are specific, strict, and usually contain a series of steps and activities for each phase of the project's life cycle.

While performing the project, we follow agile model processes. The agile model refers to a software development approach based on iterative development. Agile methods break tasks into smaller iterations or parts that do not directly involve long-term planning. The project scope and need are laid down at the starting point of the development process. Plans in regard to the number of iterations, the duration and the scope of each iteration are clearly define in advance. There are some steps of model.

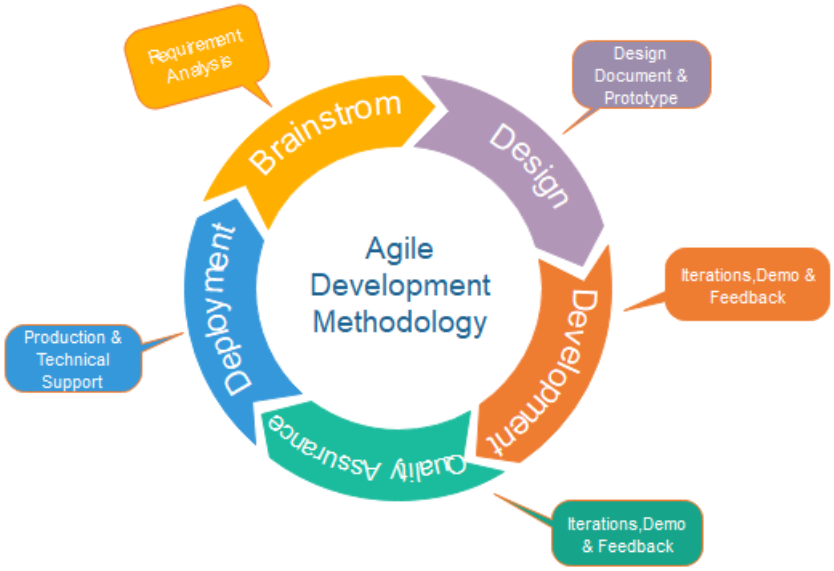


Fig - 1: Agile Model

- **Requirements gathering**: You should explain business chance and plan, the time and effort require to build the project. Based on this particulars, you can evaluate scientific and economic feasibility.
- **Design the requirements**: When you have recognize the project, work with stakeholders to define requirements. You can use the users flow digram or the high-level UML diagram to show the work of new feature and show it will apply to your existing system.
- **Construction** / iteration: When the team define the requirements, the work begins, Designers and developers start working on this project, which aims to deploy a working product. The product will undergo various stages of improvement, so it includes simple, minimal functionality.
- **Testing:** In this phase, the quality assurance team examines the product's performance and looks for the bug.
- **Deployment:** In this phase, the team issues a product for the user's work environment.
- **Feedback:** After releasing the product, the last step is feedback. In this, the team receives feedback about the product and work through the feedback.

3.2 Advantages of Agile Method are as follows

1. Frequent Delivery
2. Face to face communication with client.
3. Efficient design and fulfils the business requirement.
4. Anytime changes are acceptable.
5. It reduces total development time.

3.3 Disadvantages of Agile Method are as follows

1. Due to the shortage of formal documents, it creates and crucial decisions taken throughout various phases can be misinterpreted at any time by different team members.
2. Due to the lack of proper documentation, once the project completes and the developer allotted to another project, maintenance of the finished project can become a difficulty.

3.4 Application of Agile Model

- Forecasting the effort – The starting point for most of our project is forecasting the effort we believe to be involved in producing the software.
- Defining Sprints – Sprint is a pre-defined period of time with consistent duration through the development effort.
- Demonstrating Progress – Potentially reducing the risk of wasted time.

3.5 Summary

The methodology and technique to be used to develop the system has been described in details in this chapter. The methodology has the appropriate phase to facilities the smooth development of the software system. In addition, system requirement including the hardware and software requirements used in development this system are described.

CHAPTER - 04

Data Collection

4.1 Introduction

We use a dataset from Kaggle Dataset to detect Covid-19 Chest X-Ray that consists of 200 X-ray images, which are 100 of Covid Positive and 100 of normal for a front-view chest x-ray. We use the photos, which contain some bacterial, viral pneumonia, along with normal lungs. Through this dataset, we perform the analysis operation of the covid-19 chest x-ray.

4.2 Tools / Platforms

- **Windows 10 –** Windows 10 released for the general public on July 29, 2015. Windows 10, developed by Microsoft Company. Microsoft initially aimed to have Windows 10 installed on over a billion devices. The Windows Runtime app ecosystem. These universal apps made to run across multiple platforms and device classes, including smart phones, tablets etc.
- **Anaconda Navigator** – Anaconda Navigator is a desktop graphical user interface (GUI). It is free and open source distribution of the python programming language. Anaconda Navigator is available for Windows, macOS and Linux.
- **Google Colab –** Google Colab is free and open source platform in which we can perform operation like executing code and text in a single document, along with images.
- **Jupyter Notebook –** Jupyter Notebook is a open source platform which is having three core programming languages support (Julia, Python, R). We use Jupyter Notebook to execute the code.

- **VS Code –** VS (Visual Studio Code) is a free source code editor made by Microsoft. Instead of a project system, it allows users to open one or more directories, which can then be saved in workspaces for future reuse. We use Visual Studio for HTML, CSS and JS to code.
- **Python3 –** The Programming language Python started in December 1989. Python is an interpreter, high-level and general-purpose programming language. Python is an object-oriented programming language, and many of its features support functional programming.
- **Flask –** Flask is a mini web framework in python, which was release **on** April 1, 2010. Flask does not require any particular tools and libraries. Flask depends on the jinja template engine and the Werkzeug WSGI toolkit. We use flask in our project to show the front-end. Through flask, we can connect to the web view.
- **Html –** HTML (Hypertext Markup Language) the standard markup language for documents designed to display in a web browser. HTML describes the structure of a web page semantically and originally included cues for the document's appearance.
- **CSS –** CSS (Cascading Style Sheets) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS file to be cached to improve the page load speed between the pages that share the file and its formatting.
- **JS –** JS (JavaScript) is a programming language. It is a high-level often just in time compiled, and multi-paradigm language. HTML, CSS, JavaScript is one of the core technologies of the World Wide Web. JS and bootstrap.min.js contain all plugins in a single file.

CHAPTER - 05

Design / Implementation

5.1 Introduction

Since the recent sudden surge of Covid-19 infections across the world, many alternatives screening approaches have been developed to identify suspended cases of covid-19. There are few open source application which we can use chest X-Ray images. Publicly available data of chest X-ray are limited.

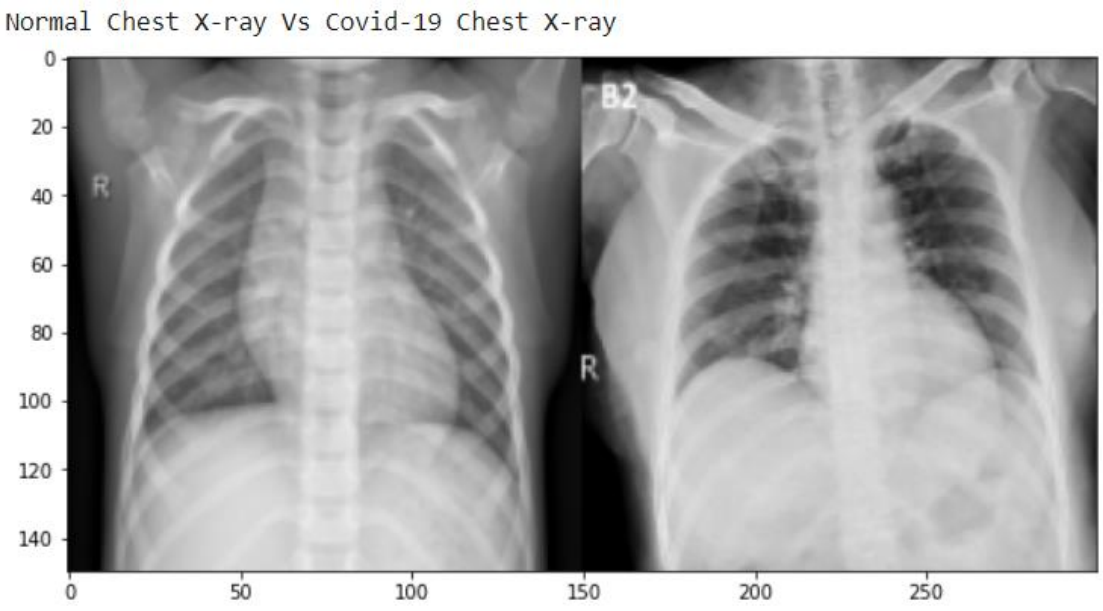
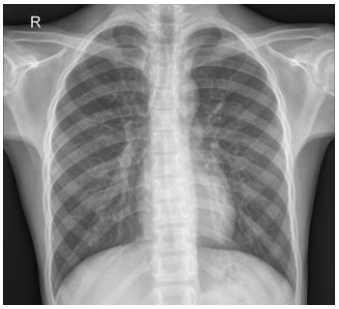


Fig - 2: Normal Chest X-Ray VS Covid-19 Chest X-Ray

While having the limited amount of data available of X-ray samples (including Normal and Covid-19), to train a deep neural network we are using pre-trained backbone on a dataset.



Normal

COVID-19

5.2 System Architecture

The whole system for detection of COVID-19 from chest X-Ray images comprised of few important steps – collection of dataset, pre-processing the data, categorization of dataset, training the models and evaluation and analysis of the model. The complete system architecture of the a detecting COVID-19 with CNN Model. The dataset needed for training and validating the model is collected and stored out. After this step, all the data are categorize according to the classification of the model. Then all the models are trained and validated with the same dataset and same environment.

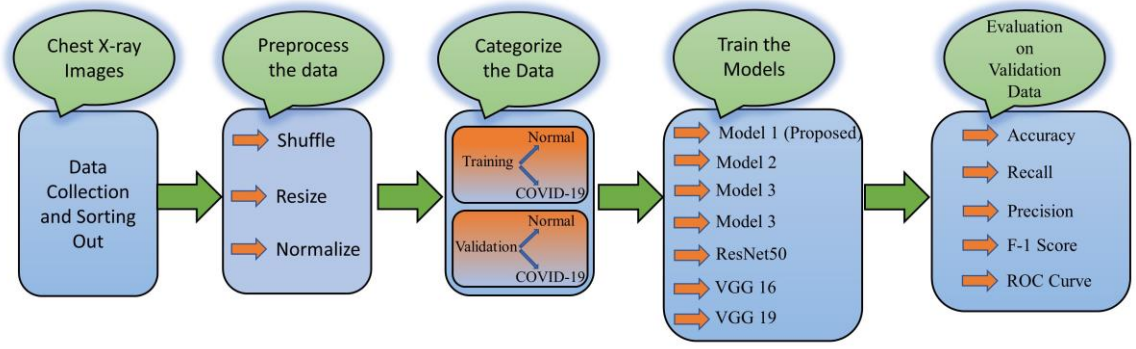


Fig - 3: Complete System Architecture of COVID-19 Detection

5.3 Model Architecture

Convolutional Neural Network has been playing a great role in categorize images, in specific medical images. This has opened new casement of opportunities and made the disease diagnosis much more convenient. It also successfully detects recent novel Corona virus with higher accuracy. One of the restrictions that researchers encounter is a little dataset for training their model. Actuality a novel disease, the chest X-ray dataset of COVID-19 positive patients is also limited. Therefore, to avoid over fitting, a sequential CNN model is proposed as in authors' earlier work of [53] for classifying X-ray images. CNN model for COVID-19 detection. This model has 4 main components: input layers, convolutional layers, fully connected layers and output layers.

The tuned data set is sustained into the input layers of a model. It has four CNN layers; first one is the 2D convolutional layer with 3 _ 3 kernels and ReLu activation function. Rectified Linear Unit is one of the most popular and successful activation functions that are being extensively used in DL. Rectified linear activation does not actuate all the neurons at the same time building it computationally systematic in comparison to other activation functions like tan. The next three layers are 2D CNN layer along with the ReLu activation function and Average pooling layer. Average pooling layer assemble the features of the CNN layer by convolving filters over it. It reduces the arithmetic cost as it minimizes the number of parameters thus it helps to avoid over fitting. In each of three layers a Average pooling layer is added after the CNN layer to avoid over fitting and to make the model arithmetic efficient. In the next step, the output of the CNN layers is converted to a long feature vector by a flatten layer. This output from the flatten layer is feed to the fully connected layer with dropout. In a fully join layer, every input neuron is connected to every activation unit of the next layer.

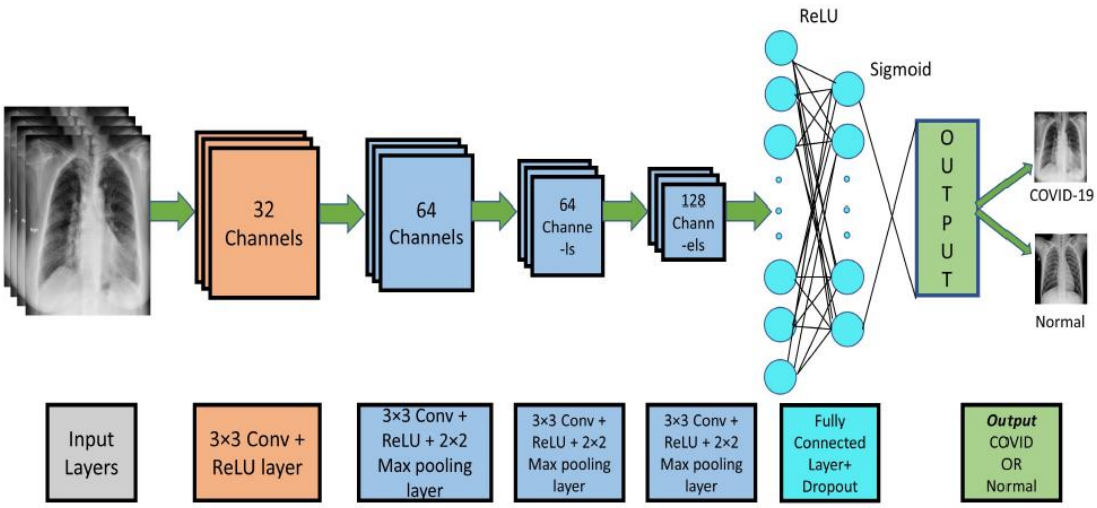


Fig - 4: Work-Flow Diagram of CNN model for COVID-19 Detection

The input neuron is connected to every activation unit of the next layer in a fully connected layer. The entire input feature is passing through the rectified linear activation function, and this layer categorizes the images to the assigned labels. The sigmoid activation function makes the categorization decision depending on the classification label of the neurons. Finally, the output layer, it is declared if the input X-Ray images.

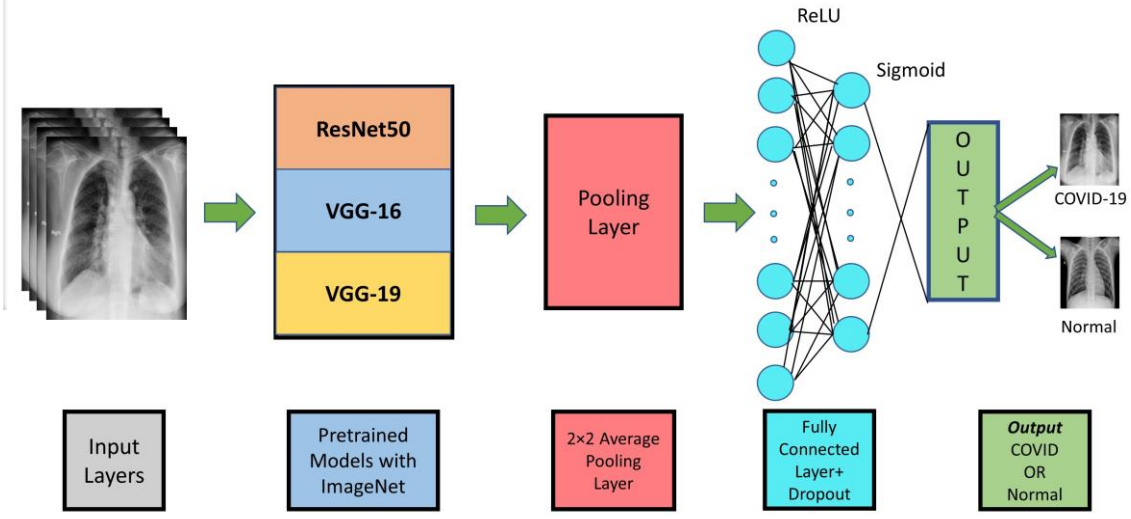


Fig - 5: Work – Flow Diagram of Pre-Trained Model of COVID-19 Detection

5.4 Work-Flow

For training we initialize out model with pre-trained weight form cheXNet implementation

1) In the first step, we have import important package after the importing is completed then we have given model learning rate. In what time model will run the learning rate is decide.
2) In the second step, we have a limited amount of dataset. Splitting then into two different lists. First is the data and second are the labels of the images. All the images are converting into RGB channel and after that they are resizing them.
3) In the third step, we are performing a data analysis part using CV package. Images transform size will be (150, 150, 3) and the mode is reflect. While doing the analysis we are getting two types of images Covid-19 and Normal. So, using labelbinarizer we are fitting the images and printing them will the categorical like (0, 1). Means (0) is a covid and (1) is a normal.
4) In the fourth step, we are training the data using train_test_split. Data and labels will be train and test size will be (0.20), random_state will be (42). We are using ImageDataGenerator will train and testing.
5) In the Fifth step, we are placing CNN model VGG16. In the model the weight are imagenet and including the shape is (224, 224, 3).
6) In the sixth step, using a Deep Neural Network a CNN model we are removing some dense layer and customs some new layer in them. In that we are adding Average Polling 2D layer, Flatten Layer, Dense Layer, Dropout layer, Dense Layer and the last layer is the output layer.
7) In the seventh step, using Adam Optimizer algorithm we are doing classical stochastic gradient descent procedure to update the weights iterative based in training data.
8) In the eighth step, after completing the running (10) epochs prediction is done. How many percentage it is of covid-19 and normal. The probability is set to be (0.5). Then it predicts Covid-19 or Normal.

9) In the ninth step, we proved some report classification report. In classification report we are getting precision, recall, f1-score. While the precision is call as predictive value and recall is call to be sensitivity. Both Precision and recall are based on understanding measure of relevance and f1-score is model accuracy.

	precision	recall	f1-score	support
Covid	1.00	0.95	0.97	20
Normal	0.95	1.00	0.98	20
accuracy			0.97	40
macro avg	0.98	0.97	0.97	40
weighted avg	0.98	0.97	0.97	40

Fig - 6: Classification Report

10) In the tenth step, we calculate the confusion matrix. The confusion matrix is a performance classification model on a set of test data for which the true values are know.

```
[[19  1]
 [ 0 20]]
acc: 0.9750
sensitivity: 0.9500
specificity: 1.0000
```

Fig - 7: Confusion Matrix

11) In the eleventh step, Plotting some graph related to train loss, Val loss and train acc, Val acc. After the plotting graph save the Deep Learning Model using (.h5) file. So, that we can use this file in flask framework

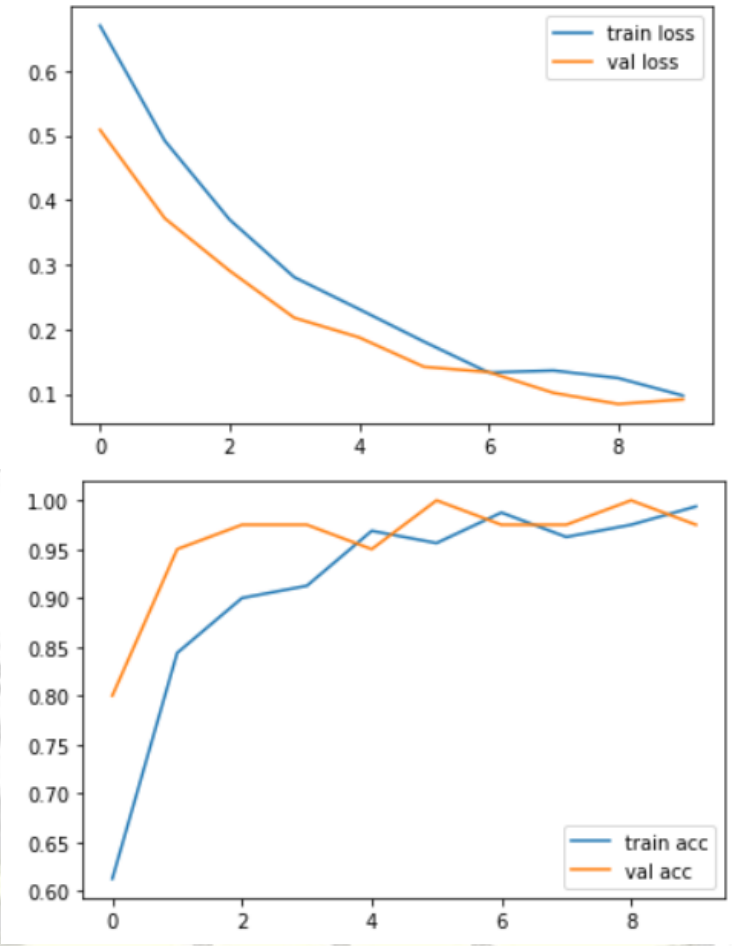


Fig - 8: Graph of train (loss, acc), Val (loss, acc)

The main part of this project is completed successfully. Now, we will move towards the web development side. User didn't understand the backend code. It required the frontend part. For the Frontend we use Flask Framework. Flask is the web base framework in python. The HTML, CSS, JS, and PHP file know we will connect to flask. Flask is the lightweight framework to develop the web application with flask we can create a simple framework that will show the prediction and result. Flask does not require any type of different tools and libraries. In flask we can use third-party libraries to develop the web view.

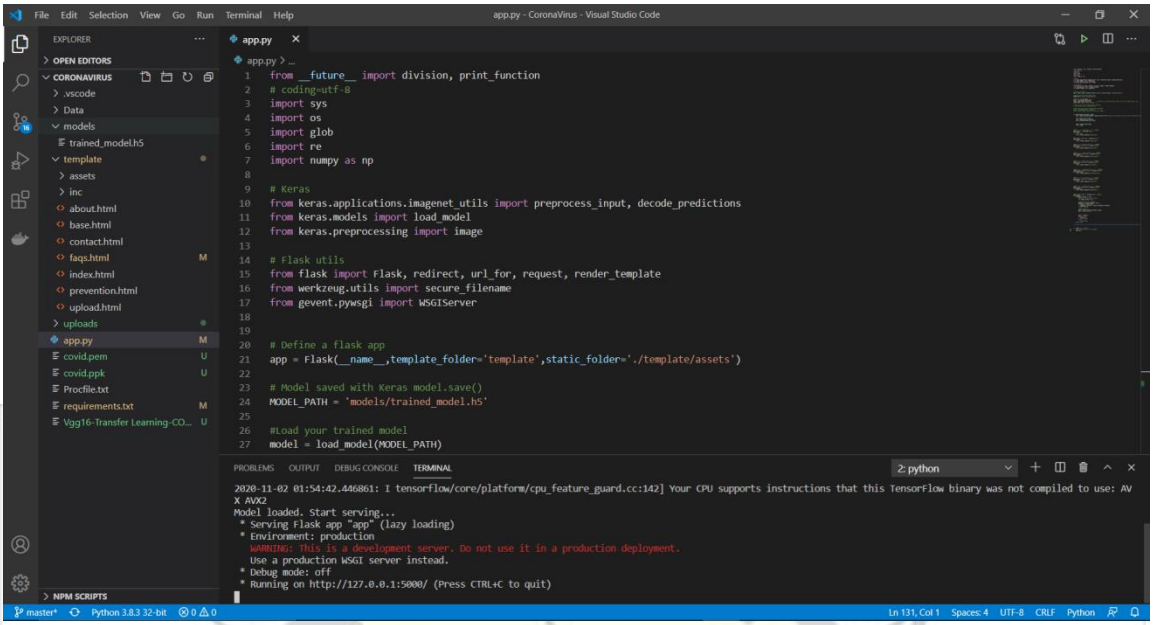


Fig - 9: Running Flask

To use flask framework we have to install it anaconda prompt by saying pip install flask. We are performing flask in VS code. VS (visual studio) by Microsoft. To use flask import some library of flask, Werkzeug and WSGIserver. Flask is having simple app code to run and get the output. Using (.h5) model file in flask we have connected to API. Our model is known connect with flask. We have connect some off our HTML, JS, PHP, CSS file with flask. Now, we are running to flask by writing (python app.py) the local host is creating with the (http://127.0.0.1:5000/). The flask lightweight so that it can run fast and create a local host.

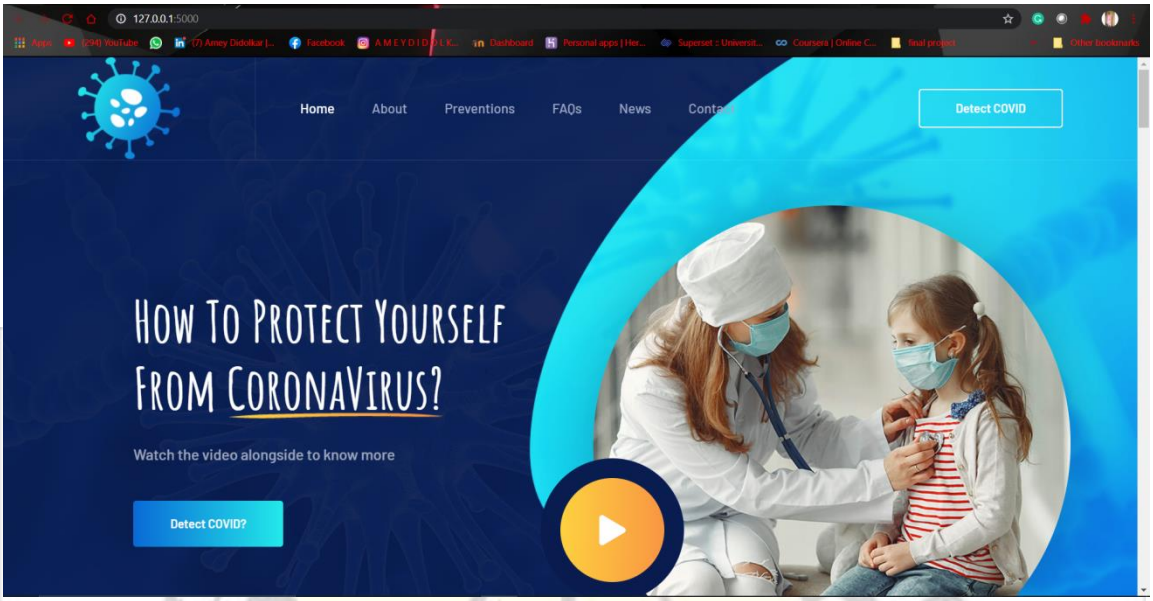


Fig – 10: Home Page

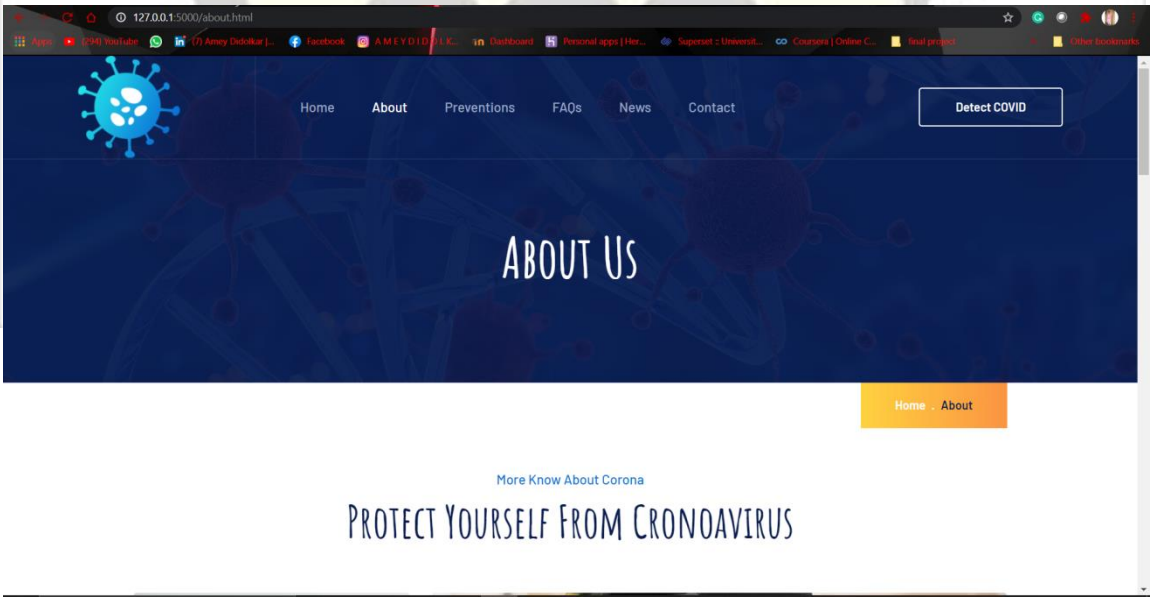


Fig – 11: About us Page

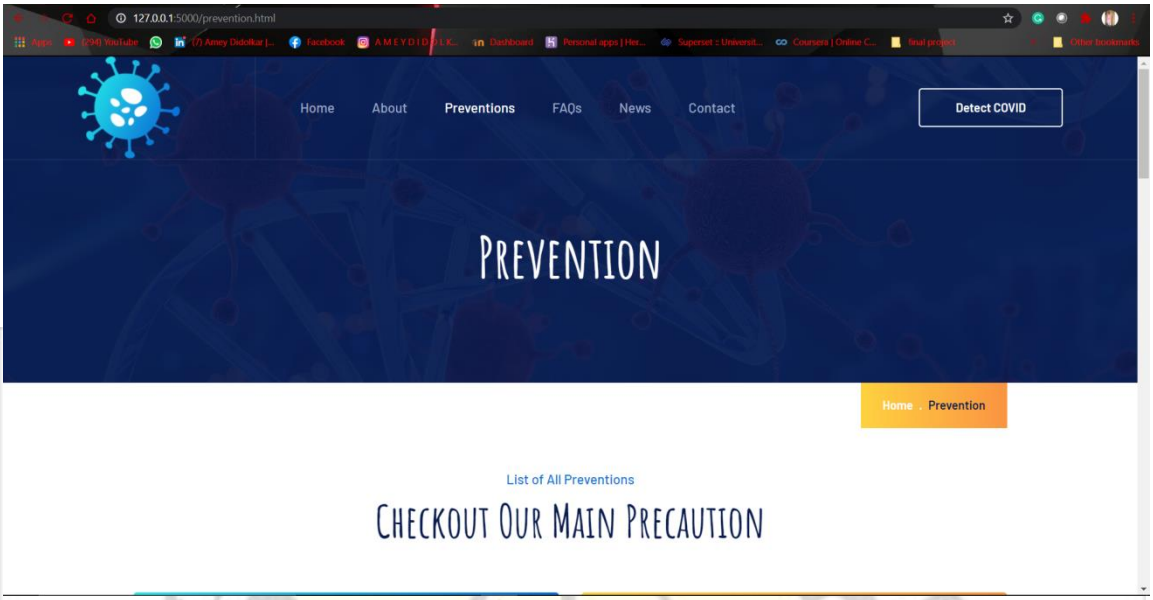


Fig – 12: Prevention Page

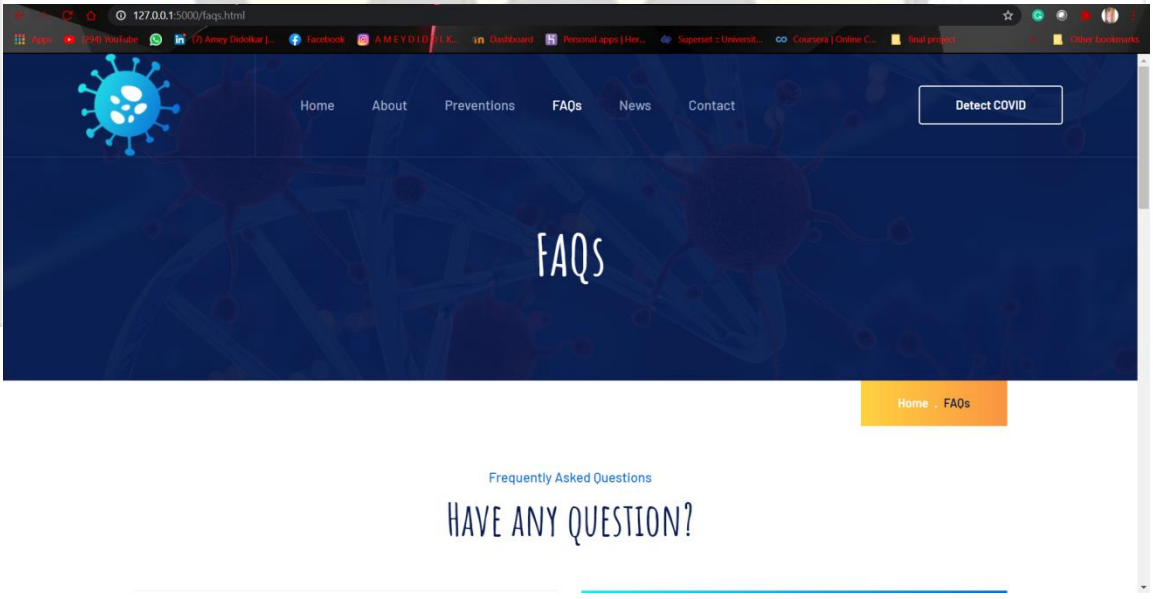


Fig – 13: FAQS Page

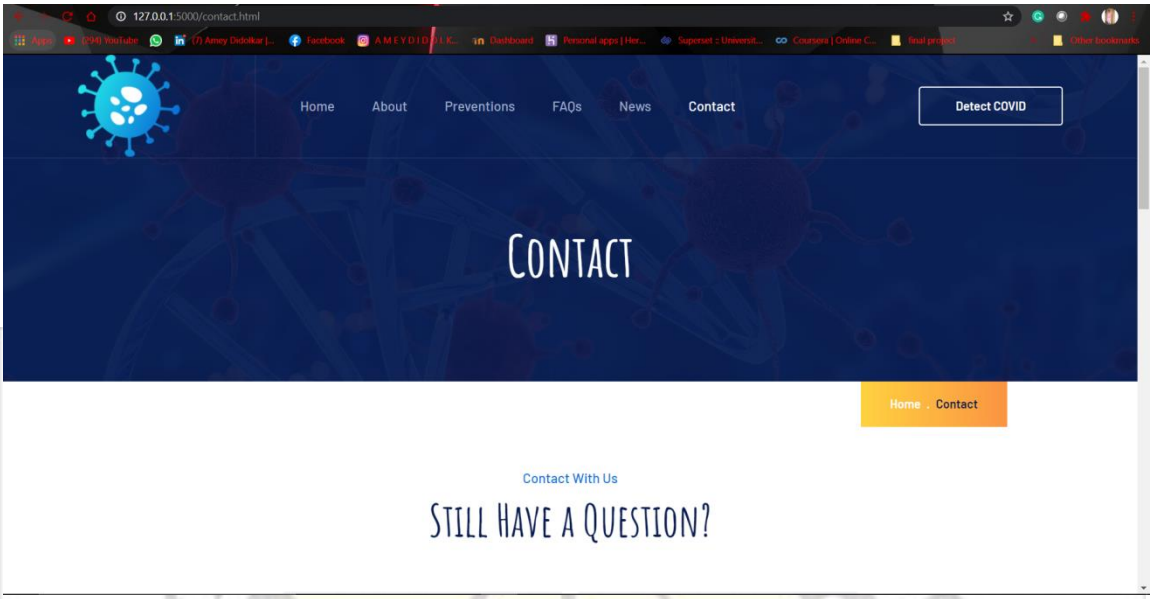
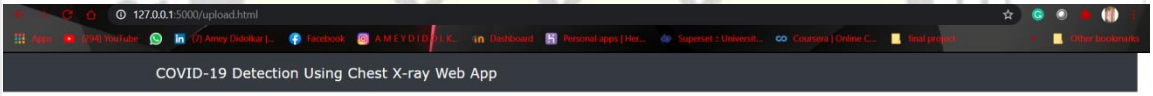


Fig – 14: Contact Page



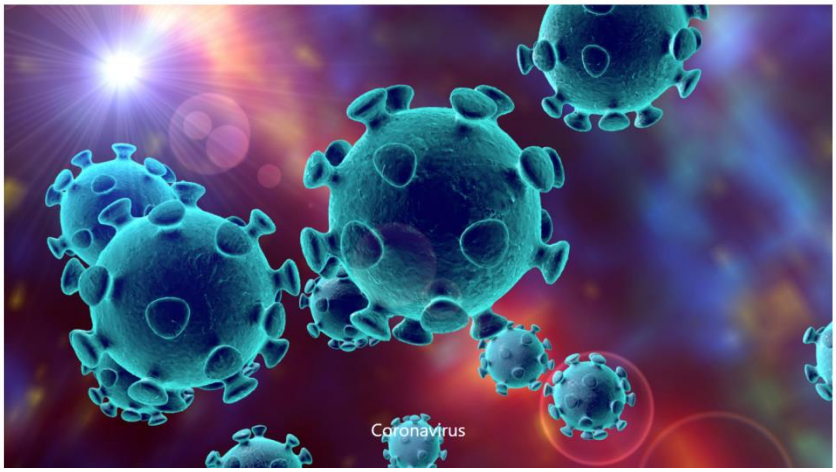


Fig - 15: Covid-19 Detection Page

Our Result is mainly of two type COVID-19 and NORMAL. Through the web representation of the flask app we can know clearly showing that the person is COVID-19 or NORMAL. In this flask app we have to select a image of chest x-ray from our system and insert it into flask app. After inserting the image deep neural network file will check the image and gives the result to flask and flask app will display the result. This process is time saving and not so costly. In figure 11 and 12 we can see the result of flask application.

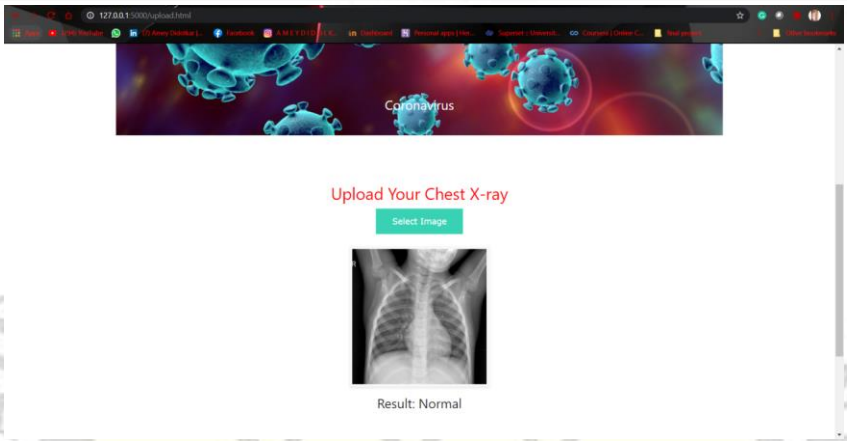


Fig - 16: Flask Graphical Representation (1)

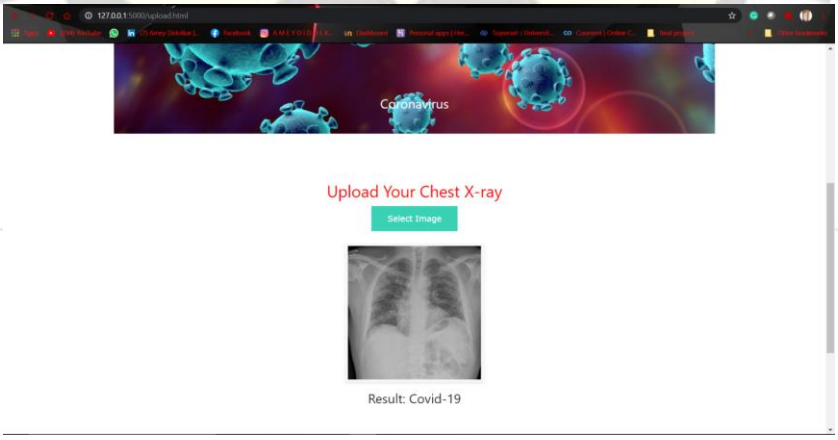


Fig - 17: Flask Graphical Representation (2)

CHAPTER - 06

Testing & Summary of Results

6.1 Introduction

For the dataset, the overall accuracy is 97% for a model, whereas the pre-trained model is VGG16. It clearly shows that the proposed model performs better than the other model's accuracy. The performance of the models is evident from the metrics like precision, recall and F1-Score. These execution metrics are calculated from the possible outcomes of the validation dataset obtained by the confusion matrix. A confusion matrix has four different outcomes. They are as – True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN). In this case, TP denotes the number of Corona virus-positive patients detected as positive, TN denotes the number of negative cases detected as negative, FP presents the number of negative but detected as positive.

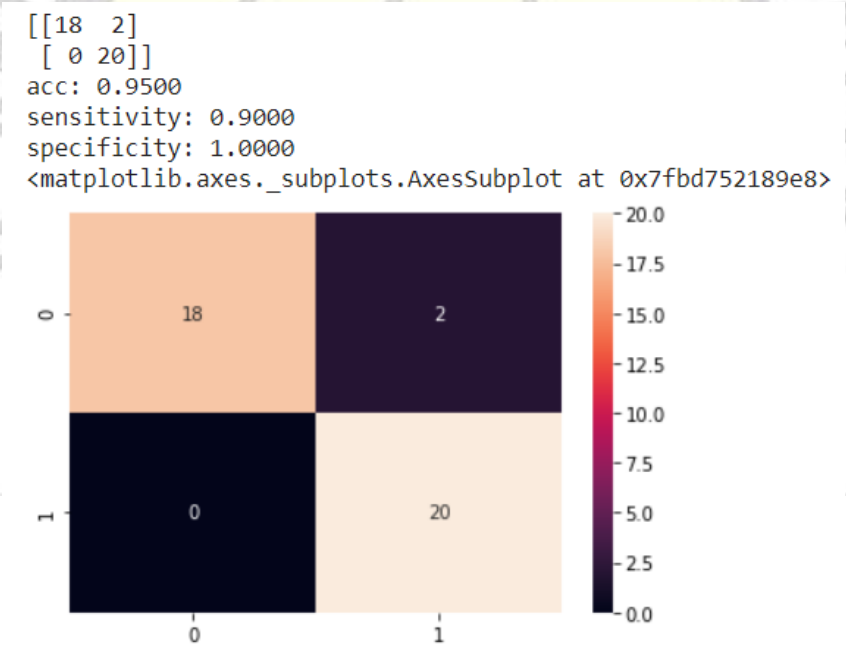


Fig - 18: Confusion Matrix for analytical analysis of Dataset

FN gives the cases which are positive but detected as negative. Receiver Operating Characteristics (ROC) curve represents the classifier's performance at different threshold values that plot the TP rates vs. FP rates.

1) Accuracy = (TP + TN) / (TP + TN + FP + FN)
2) Precision = TP / (TP + FP)
3) Recall = TP / (TP + FN)
4) F1 – Score = 2 x (Precision x Recall / Precision + Recall)

6.2 Result

Accuracy defines how close the cause result is close to the actual value, whereas precision calculates the percentage of the relevant results. The recall is another important factor for evaluating a convolutional neural network model. It gives the percentage of the total applicable results that a model can right classify. The F1-score combines both precision and recall, and it is designated as the weighted average of these two. Even though this is a fine result for the proposed model, a few analysts could realize better results than this with binary classification. However, this work shows how the number of convolutional layers and the number of images in the dataset play a role in the models' performances. As a convolutional neural network classifies images by extracting features from the images, it is practicable to differentiate between images with very minute and subtle changes. The chest X-ray of a COVID-19 patient from the early stage would show distinction from an X-ray of the same patient at the middle and late stages. It provided the necessary dataset. It would be possible to detect the stages of COVID-19 patients. As it is a new disease and a lack of classified data according to different stages, classifying stages are not labeled here in this work, but it is hoped to address this challenge in future work.

6.3 Conclusion

This study has preferred a deep learning-based model to detect COVID-19 cases from chest X-Ray images. The detection of COVID-19 plays an essential role in preventing the spread of this global pandemic. The results look promising as such from the size of the publicly available dataset is small. We use the Convolutional neural network model to perform the detection that we trained with the dataset with the total images of 100 COVID-19 and 100 Normal patients. We use VGG16 to predict accuracy. The accuracy and F1-Score of the model is of 97.5%. Through the small dataset, we get the F1-Score 97.5%. If we improve further with a multi-class classification and availability of a large dataset, we can get the best result. Finally, Convolution Neural Network has excellent success in detecting COVID-19 with minimal time, Resource and cost. Such a highest accuracy, the success of a model it is not clinically tested. Such a high accuracy will play an essential role in detecting fast COVID-19 patients, thus reducing the people's testing time and cost.

6.4 Project Link

To see a successfully result project go to the link

(https://github.com/ameyhub/Covid-19-Detection-Using-Chest-X-Ray-)

CHAPTER - 07

References

1. Kaggle. Covid-19 X rays (Accessed on 2 June 2020) Available from: https://www.kaggle.com/andrewmvd/convid19-x-rays?select=X+rays
2. Keras Transfer learning & fine tuning. (Accessed on 5 June 2020). Available from: https://keras.io/guides/transfer_learning/
3. VGG16 Model (Accessed on 10 August 2020). Available from: https://neurohive.io/en/popular-networks/vgg16/
4. CDC Centre of disease control and prevention (Accessed on 17 August 2020). Available from: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covid-net/purpose-methods.html
5. Ieee xplore Computer Vision (Accessed on 22 September 2020). Available from : https://ieeexplore.ieee.org/document/9154088
6. Ieee xplore Auto Detection (Accessed on 25 September 2020). Available from: https://ieeexplore.ieee.org/document/9205687
7. Flask web development (Accessed on 18 October 2020). Available from: https://flask.palletsprojects.com/en/1.1.x/