A PROJECT REPORT

ON

COVID-19 DETECTION USING CHEST X-RAYS

REPORT SUBMITTED TOWARDS PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF

BACHELOR OF TECHNOLOGY COMPUTER SCIENCE & INFORMATION TECHNOLOGY

SUBMITTED BY

Akul Bhardwaj - 17070124007

Aditya Ranjan - 17070124004

Abhas Gupta - 17070124003

Deeptanshu Panthi - 17070124019

UNDER THE GUIDANCE OF

Prof. POORVA AGRAWAL
Prof. GAGANDEEP KAUR

SYMBIOSIS INSTITUTE OF TECHNOLOGY PUNE - 412115

A CONSTITUENT OF SYMBIOSIS INTERNATIONAL

(DEEMED UNIVERSITY)

2020-21

Declaration

We hereby declare that the project work entitled **— COVID-19 DETECTION USING CHEST X-RAYS** is an authentic record of our work carried out at as requirements for final year project for the award of B.Tech. degree in Computer Science and Information Technology Engineering at Symbiosis Institute of Technology Pune, affiliated to Symbiosis International University Pune, under the guidance of **Prof. Poorva Agrawal**, during January 2021 to June 2021.

Akul Bhardwaj	**Aditya Ranjan**	**Abhas Gupta**	**Deeptanshu Panthi**
17070124007	17070124004	17070124003	17070124019

Date: 25 May, 2021

Certified that the above statement made by the student is correct to the best of our knowledge and belief.

Dr. Ambika Pawar
Department Head
CS & IT Department

Dr. Poorva Agrawal
Assistant Professor
CS & IT Department

Acknowledgment

This project's progress and final result involved a lot of support and assistance from many people, and we are incredibly grateful to have this during the completion of our project work. What we did is primarily because of such support and assistance and we wouldn't fail to thank them.

First and foremost, we express our gratitude and praise to Symbiosis Institute of Technology, Pune and Department of Computer Science and Information Technology for giving us this wonderful opportunity to undergo B. Tech Project Work, helping us to learn and attain great experience.

We would like to thank our HOD, **Dr. Ambika Pawar** for her valuable guidance, supervising this work and helpful suggestions.

We owe our profound gratitude and special thanks to **Prof. Gagandeep Kaur** for cooperating with us and giving us her valuable time and information and **Prof. Poorva Agrawal,** who took time out to listen, advise and keep us on the right track despite being incredibly busy with his duties. Thanks to both for empowering us with untiring assistance, direction, encouragement, continuous support, valuable ideas and constructive criticism throughout this project work.

At last, we are grateful to our respected teachers Department of Computer Science and Information Technology SIT, Pune, family and friends for their help, encouragement and co-operation during the project work.

Contents

List of Figures

1. Abstract

"Covid-19 Detection using Chest X-Rays" helps to filter out covid positive and covid negative patients. Covid-19 is a disease affecting the whole country concerning health-related problems. Today approximately around thirty lakh people have died due to this disease worldwide. The majority of people have died because fast and efficient detection techniques of covid-19 were not present. Because of late finding of covid-19 in patients, proper treatment was not possible. The majority of the studies suggest that this disease infects the lungs primarily and the infection caused due to the virus could spread in the lungs results in a lack of oxygen in the blood. For prediction of infection in the lungs, X-ray images and CT scans are mainly used. However, CT scans are expensive and not so easily accessible to every individual in this world. But X-rays are less costly and widely accessible around the world. Deep learning and artificial intelligence are integral components of machine learning methods for classifying images, text, and other types of objects into different categories and classes. Deep learning provides features like feature extraction in images and classifying them into categorical form. It takes input as a dataset of images which will be processed in a way that the common feature in all images could be used to predict the presence of a feature in the testing dataset. After data cleaning and pre-processing of images, the CNN model is implemented to train the dataset and evaluate the prediction. Further in this paper we have mentioned different strategies and methods to implement prediction of covid-19. We reviewed 44 research papers related to Covid-19 to evaluate the conclusion with references in this field of work. For implementation of project, we used 2000 photos and we created our own customised dataset which holds everything. After creation of custom dataset, we performed pre-processing of images and hyperparameter tuning with resizing and reshaping images. Wen also normalised image between 0 and 1 for better accuracy and evaluation of our model. We converted our image into grayscale and in four dimensional for performing model evaluation as well as creating our own CNN model for predicting covid-19 disease. After creating CNN model consisting of several convulational layers we achieved accuracy of 96.5% on 10% testing dataset and 90% training dataset. We used SoftMax for finding the probability distribution of our output nodes which are covid-19 and normal. After building of CNN model, we created a web application using flask and html where we are taking chest-ray images as input from user and performing classification based on our own CNN model.

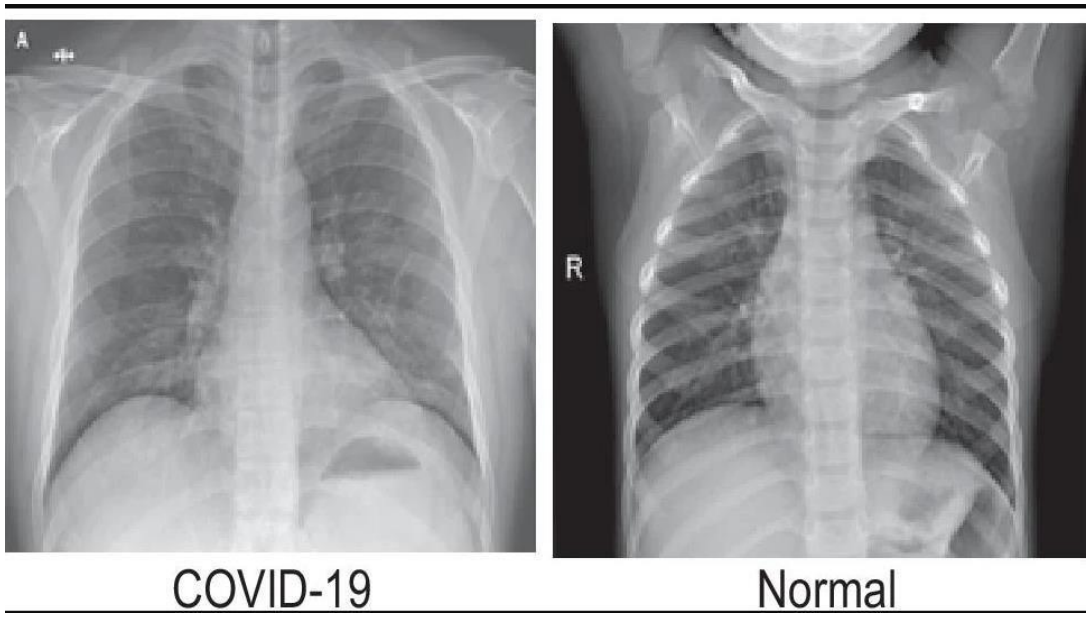
2. Literature Review

In [17], the author mentioned the usage of a deep neural networks framework for creating a capsule-based system called COVID-CAPS. This framework is made for working with a small dataset due to sudden emergence by the virus and presence of few dataset modules online at the time. For creating this model very few parameters were used for optimizing the accuracy of the model. In [18] author mentioned the public database that was created using other available databases and from articles that are published recently. Image augmentations were performed on the dataset to evaluate the classification into normal, covid-19 pneumonia, and normal pneumonia with help of pre- trained models used for validation and testing. In [19], the author mentioned three cases in Wuhan in the country of China where the first emergence of covid-19 was observed. In [20], the author mentioned shortcomings of the "RT-PCR" test which is utilized for testing the presence of coronavirus in the body. The author pointed out the low positivity rate of the RT-PCR test and focused on image processing techniques used for solving the classification problem. SoftMax activation function was used for classifying images and producing output as a probability distribution of output nodes. Then features extracted from the output of the SoftMax layer were sent to the classifier which then produces output with an accuracy of 98%. In [21], the authors presented a medical diagnosis of the covid-19 virus in the human body which included the study of symptoms like fever, throat infection, a stomach infection, and MERS. The author also mentioned the features extracted from the chest scan where different findings were observed. The author proposed the treatment. [22]. They built the CNN model and trained it on the collected dataset and achieved higher precision correctly finding true and false cases. In [23], authors reviewed the emergence of covid-19 disease in Wuhan, China, and still unknown source of the virus. The author also mentioned viruses entering different countries because of people traveling through the air as the disease is highly transmittable from individual to individual. In [24], the importance of image segmentation is mentioned along with the image feature extraction in abnormal parts of the lungs. The author also emphasizes the importance of the quality of the image for analyzing all features that are not normal or healthy patients. Starting with the usage of SVM for classifying pneumothorax and eliminating all image background abnormalities and noise for ensuring good quality of the image. Feature extraction in images was implemented using the local binary pattern technique. In [25], the author describes different findings in infected patients compared to normal patients using CT scans. These findings consisted of consolidative pulmonary opacities. These findings were further used to classify. In [26], author used data with the frontal view with 123 parameters. In [27], Covid-19 detection through deep learning algorithms is discussed by the author and reviews the process. In [28], the author implemented research on the impact of Covid-19 disease in the kidney and concluded that Covid-19 directly affects the kidney. In [29], research was established on 50 patients infected with Covid-19 and further divided them into good and bad groups based on the recovery rate of patients. It was observed that 50% of patients infected with Covid-19 had a very bad recovery rate and had symptoms for a longer duration of time.57 days of ongoing symptoms were observed in some patients which were highest among the group. In [31], Vector gadget classifier used to detect infected patients was used as a preliminary test for the virus. Multiple parameters were optimized for acquiring an accuracy of 97.48% through the deep learning method. In [32], author describes a new way that can be used to classify and detect covid-19 patients with good accuracy and precision. The author implemented the COVIDX-NET model which will be tested on a dataset of 50 images where 25 images are true Covid-19 positive cases. After implementing this model, the author was able to get an accuracy

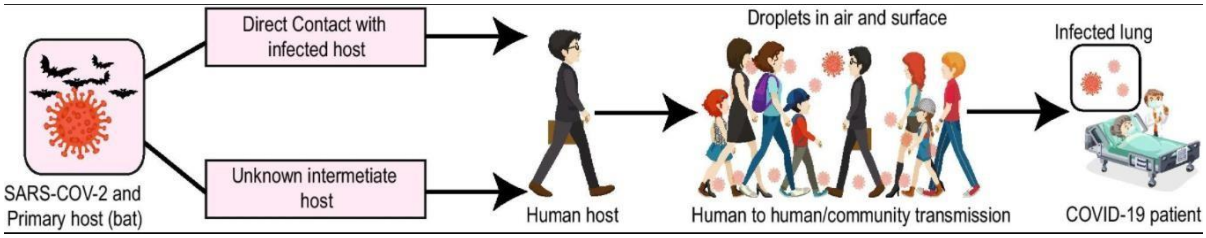
between 89% to 91%. Further, in [33], the authors conversed about the impact of Covid-19 in 196 countries around the globe and proposed a solution as an artificial intelligence method that focuses on models used such as Google net, ResNet, and Inception. In [34], the authors state all limitations of using CT scans for detecting abnormalities in the lungs such as portability of CT scans, availability of CT scan centers, and cost of CT scan around the globe. The author suggests the use of an alternative CT scan which is CXR for easy detection of Covid-19. The author also mentions the relationship between pneumonia and Covid-19. Further, in [35], the role of artificial intelligence is examined for detecting Covid-19 with the help of a simple CNN model. Implementation of AlexNet which is a pre-trained model is used for classifying Covid-19 is initiated with an accuracy of 98% whereas accuracy with some changes and modification in the model is observed as 94.1%. In [36], the authors implemented a customized dataset with the use of two sub-patches with dimensions 16*16 and 32*32 by using 150 images dataset. Further, these patches were used for training the model, and a total of 3000 image patches were used. In [38], the author emphasizes the performance of the RT-PCR test which is currently used for detecting Covid-19 disease, and alternative medical methods to diagnose this disease in the field of biological science. In [39], the study of the evolvement of Covid-19 throughout the time interval is described by the author. In [40], the author mentioned the most important concern of the virus mutating itself over some time and the presence of 4,300 strains around the globe, so early detection of the virus in the human body is very important to contain the disease from spreading throughout the world. Therefore, the author proposed a KE Sieve Neural Network model for detecting covid-19 without the usage of any instrument inside the human body. This model achieved an accuracy of 98.49% which is still very relevant for society. In [41], the mean age of Covid-19 infected patients was calculated at a hospital in Rawalpindi with the percentage of symptoms affecting the patients. The study concluded with the mean age of patients as 44 years and symptoms of Covid-19 as cough with the highest percentage followed by fever and difficulty in breathing. Loss of smell and throat infection showed in minimal patients. In [42], the author describes the importance of image classification in the field of medical sciences and implemented a CNN-based model for classifying diseases such as pneumonia. Three methods were used for classifying and compared to compute the best framework available for a small dataset. These three methods comprise SVM, VGG16, and InceptionV3. While performing these three experiments, the author observed that deep learning frameworks are better than SVM because of their better performance on small datasets. This comparison is very helpful for researchers working in this field as a sufficient dataset is not available every time. In [43], the author proposed something where 100 and out of 100, 70 were recognized as true positive cases. The author also performed this model on 1431 pneumonia patients. Evaluated accuracy is 96% for detecting Covid-19 and 70.65% for non-Covid-19 cases. Lastly, in [44], the author proposed an interesting study on 101 Covid-19 patients and rolled out all symptoms and effects of Covid-19 on the human body. This study can be used for treating different symptoms with different medications available in the world. this disease and over 3,370,000 deaths are reported till now. Late detection of Covid-19 is one of the main reasons for the high number of deaths worldwide. With the collaboration of medical experts and deep learning techniques, a proper model can be created for testing humans with the covid-19 disease. Pretrained models which performed accurately in previous researches such as the Xception model, Resnet model, and VGG16 model can be used to detect Covid-19 with some modifications such as hyperparameter tuning and image augmentation. Optimization of the model can lead to a more relevant and real-world solution for detecting any kind of lungs related disease. In the future, the advancement of machine learning and deep learning can lead to major changes in medical diagnosis methods and help build a better and disease-free world together.

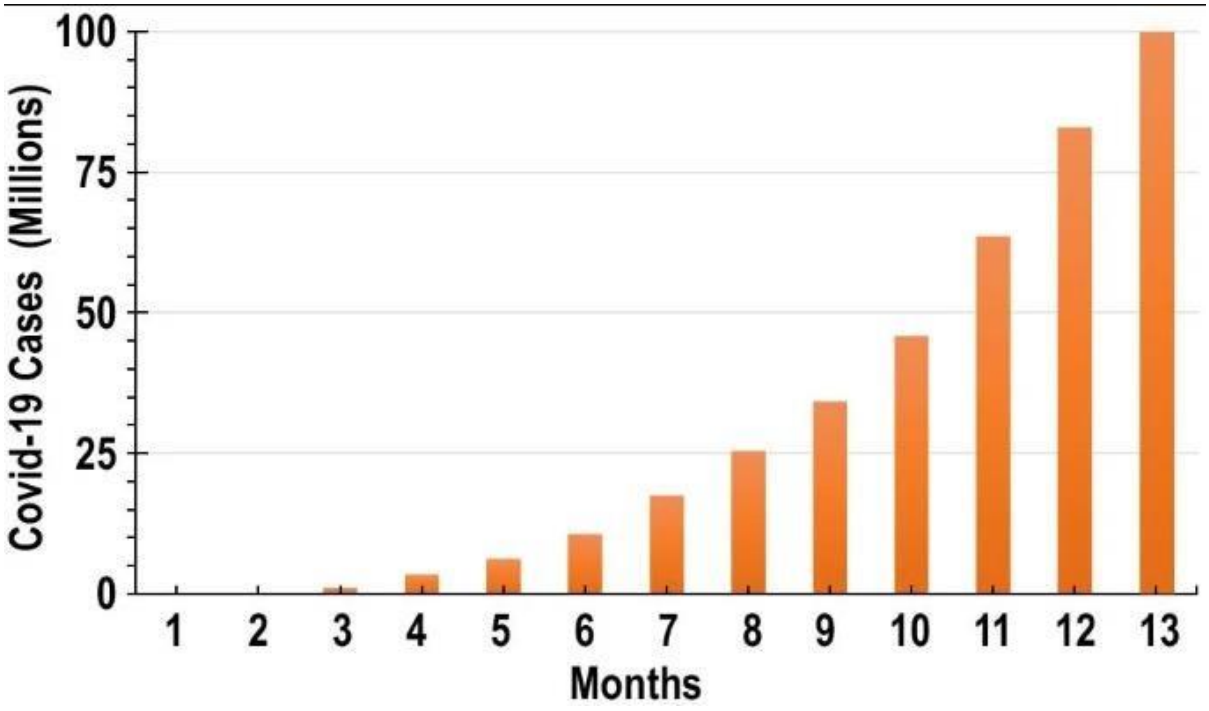
3. Introduction

Covid-19 has proven to be one of the most severe diseases with viruses mutating every two to three months. Viruses (like MERS, SARS, Flu, etc.) in the earlier time interval [2, 3] came into the picture, the viruses only existed for acute time intervals [4]. Today around the world, scientists and researchers from every country are participating in diagnosing Covid-19 with the information available on the development of vaccines. In the current situation, when the entire world is suffering from this pandemic [5], the development of the vaccine is taking a lot more time than expected because of the unidentified behavior of the virus and its changing pattern with time. Methods such as plasma therapy and lung photos were there ro predict and prevent[6, 7], but a permanent solution to this disease is still not available for citizens of the world. With people dying every day due to this disease [8], and Covid-19 detection or testing cost is very high which is not affordable to every individual in this world and due to the large population in some countries, testing is becoming very difficult [9]. From March 2020, the dataset of Covid-19 infected and non-infected cases were available on platforms like Kaggle and GitHub [10] which can be further used to identify patterns in chest x-ray images and extract common features in all Covid-19 infected people. Covid-19 detection has become a priority in the whole world because of its serious threats to the human species affecting the lungs of people and damaging it to the extent that can cause death [11]. Throat infection and breathing problems became one of the common symptoms among all Covid-19 infected patients. The covid-19 disease is one of the most dangerous threats to the world impacting the world financially as well as medically. Covid-19 infected patients should be kept in isolation [12] where proper treatment could be provided keeping all safety measures in check, involvement of all medical staff is highly important for contaminating the disease. The rate of transmission in Covid-19 disease [13] is very high resulting in a formation of a chain where every individual is infecting many other individuals, so to stop this transmission Covid-19 affected patients should be kept away from non-covid-19 patients. Medical imaging [14] is also a technique of finding and predicting the impact of Covid-19 on the human species. CT scans and x-ray image analysis [15, 16] could to performed for computing the classification. Researches are reviewed and arguments are presented in this paper with our point of view and conclusion.



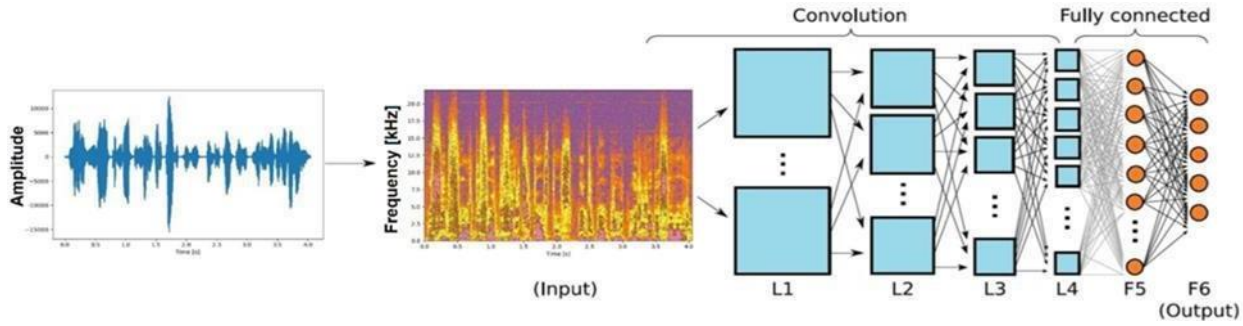
.


Direct Contact with infected host
Unknown intermetiate host
SARS-COV-2 and Primary host (bat)
Human host
Droplets in air and surface
Human to human/community transmission
Infected lung
COVID-19 patient


Covid-19 Cases (Millions)
100
75
50
25
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Months

I

3.1 Convolutional Neural Network



3.2 Scope

This project can be used extensively in situations where you have to detect if the patient is covid or non-covid. Some hospitals cannot afford expensive tests so this project will help them to detect covid and non-covid infections in patients. Covid has spread everywhere not in India but in so many countries. It has killed and is killing lot of people these days. People are not finding any solutions to the problem of this pandemic. So, we decide to create this project to help those people.

4. Platforms

Figure 5: Tools and frameworks

4.1 PostgreSQL

PostgreSQL, also known as Postgres, is an open-source object-relational database which supports SQL queries. It includes features like table inheritance and function overloading.

5. Software Requirements Specification

It is the first document that is created before developing a software. The SRS contains an in-depth review of the all the functional and nonfunctional requirements, resources and design details about the software.

5.1 Introduction

Detection of covid is a challenging problem for the hospitals who can't afford expensive tests.

- We first generate a distribution of emotional state probabilities for each section of the speech using CNNs.
- We then construct response-level features from distributions of probabilities.
- We then experiment to find out how the suggested solution successfully learns to detect covid 19 detection and what we can do to improve our accuracy for the results.

5.2 Purpose

The objective is to detect covid 19 infection with the help of chest X-Rays**.** There are ways to determine covid infection. This project can be used extensively in situations which deals with

5.3 Software Tools

Operating System	Windows 10
Web Server	XAMPP
Scripting Language	Python 3.7
Deep Learning Library	Tensorflow 1.15.2
Web Framework	Django 3.0.3
Testing Framework	PyTest 5.4.1
Server Administration	Secure Shell (ssh)
Shell	Bash
Text Editor	Google Colab and Visual Studio Code
Version Control	Git and Github
Collaboration	Trello
Cloud Services	Google Colaboratory
Storage Services	Google Drive

All the members of the team used Windows 10 for coding. As the availability of **SSH** and **Bash** commands make it easier to collaborate.

Python preferred **Gunicorn** server is used to run the **Django** project**. PyTest** is used to write unit test cases for the Django app. The web app development is done on **Visual Studio Code** text editor for its extensive features like built in terminal and git support.

Google Colaboratory is used to prepare the model while using **Google Drive** to store the dataset.
Jupyter Notebook is used to write the python code to generate the model on Google Colaboratory.

Managing development on the codebase with multiple developers in the team is made easier by the Distributed Version Control System (DVCS), **Git** and **Github**. In addition, **Trello** is used as the collaboration tool.

5.4 Hardware Tools

On home PC

We first attempted to train the model on our computers.

Processor	Intel i4
GPU RAM	GTX 1060Ti with 1280 CUDA cores and 6 GiB GDDR5 VRAM 8 GiB
Storage	DDR 4
	256 GiB SSD

The low configuration PC proved very time consuming hence we shifted to Google Colaboratory.

Google Colaboratory

Google Colaboratory provides high computing power with Tesla K80 GPU which reduces the training time significantly.

Processor	Xeon Processor @ 2.3Ghz having 1 Core and 2 Threads
GPU	12 GIB GDDR5 RAM
RAM	12.72 GiB DDR 4
Storage	107.77 GiB

5.5 Work Division

P: Akul Bhardwaj

Q: Aditya Ranjan

R: Abhas Gupta

S: Deeptanshu Panthi

See table 2 for the work-breakdown structure.

I	Analysis	P,Q,R,S
II	Requirement gathering	P,Q,R,S
III	Documentation.	Q,S
IV	Designing	P,Q
V	Developing functional specifications.	R,S
VI	Implementation	P,Q,R,S
VII	Testing and Debugging	P.S
VIII	Report	Q,S
IX	Experiment study	P,Q,R,S

5.6 Functional Requirements

Functional requirements refer to the services that software has to offer to the user of the software. It doesn't refer to the quality of those services.

Some functional requirements of our software are:

5.6.1 Retrieving Inputs

Purpose: The software will receive inputs in the form of x-ray images Input:

image file

Output: Classification of disease.

5.6.2 Processing-Chest x-ray images

Purpose: Processing of the 'x-ray image file'.

Input: The user uploads the image file for covid-19 detection.

Output: The software processes the image file and detects the covid-19 of the patient as an output.

5.7 Non-Functional Requirements

:

1. **Usability**
2. **Reliability**
3. **Performance**:.

6. High-Level Design

The high-level design can be explained with system architecture plan and data flow diagram which are mentioned below.

6.1 System Architecture Plan

The system architecture plan (see figure 6) is to train the samples x-ray files on its CNN model. Then SVM classifier of Keras is applied on with the proposed neural network, as the dataset is a classification type of data. Once the model is trained with CNN, it is ready to test sample image files and predict its result.

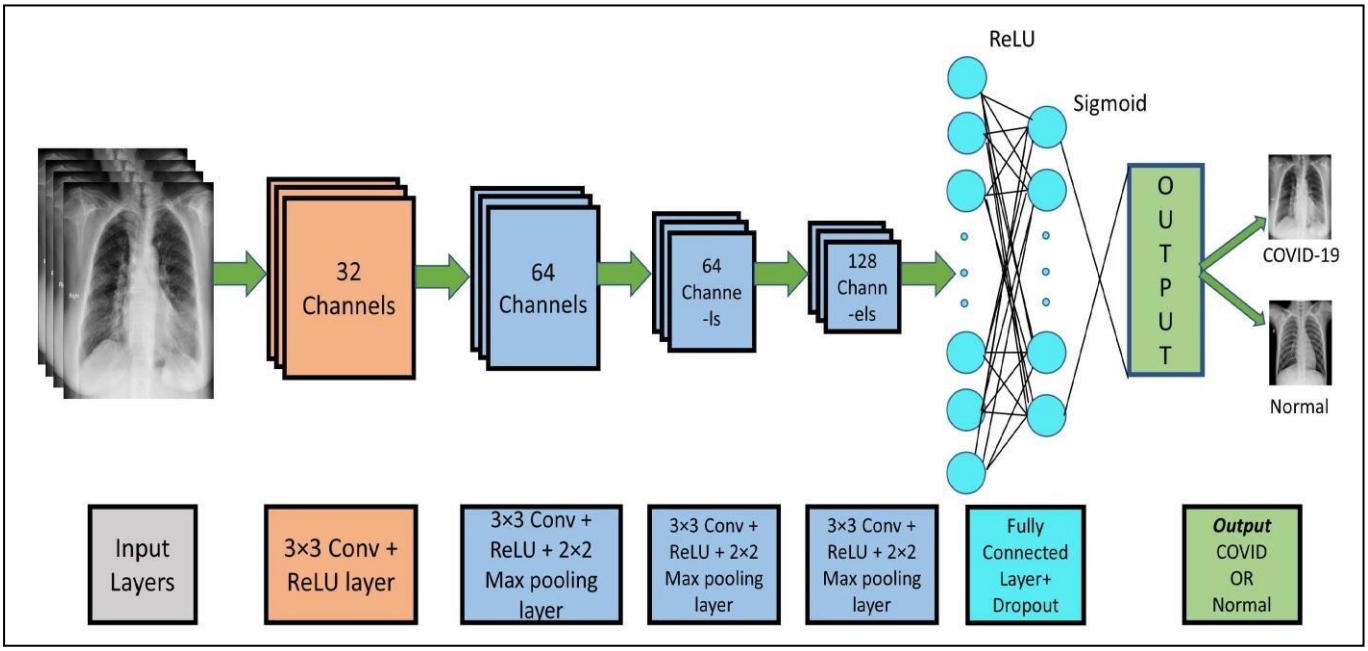


Figure 6: System Architecture Diagram

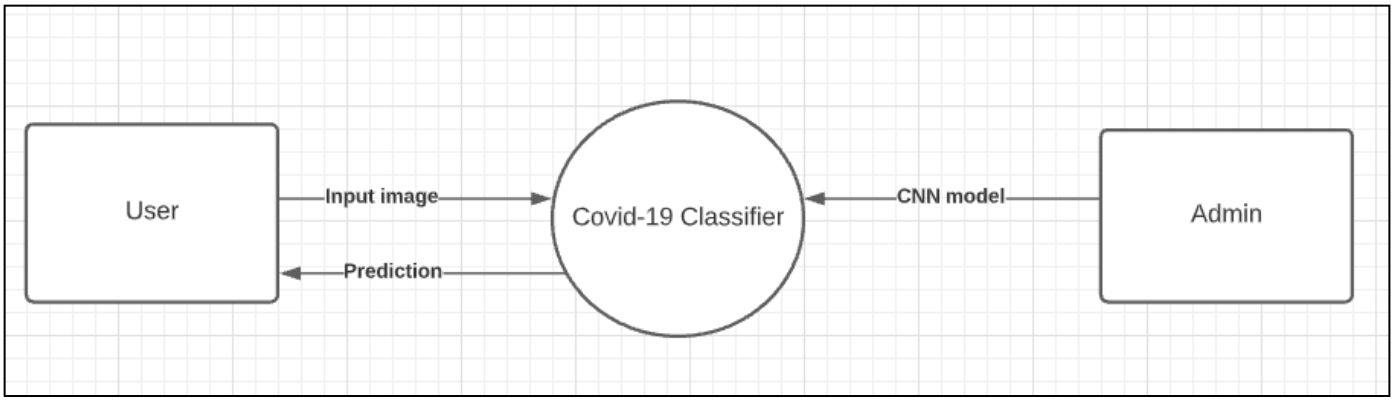
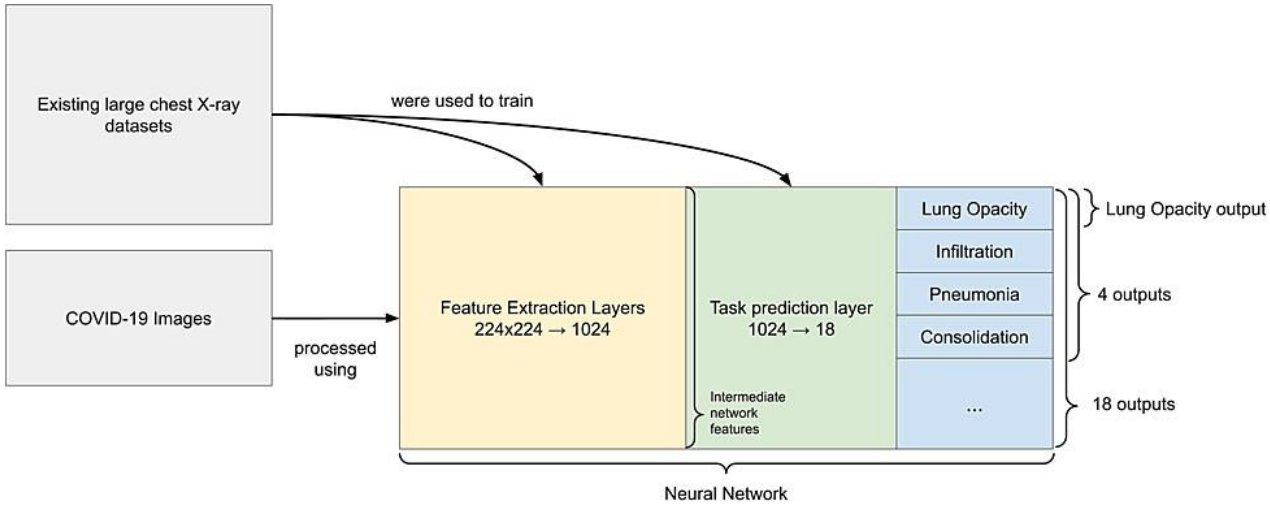


Figure 7: Data Flow Diagram (a):

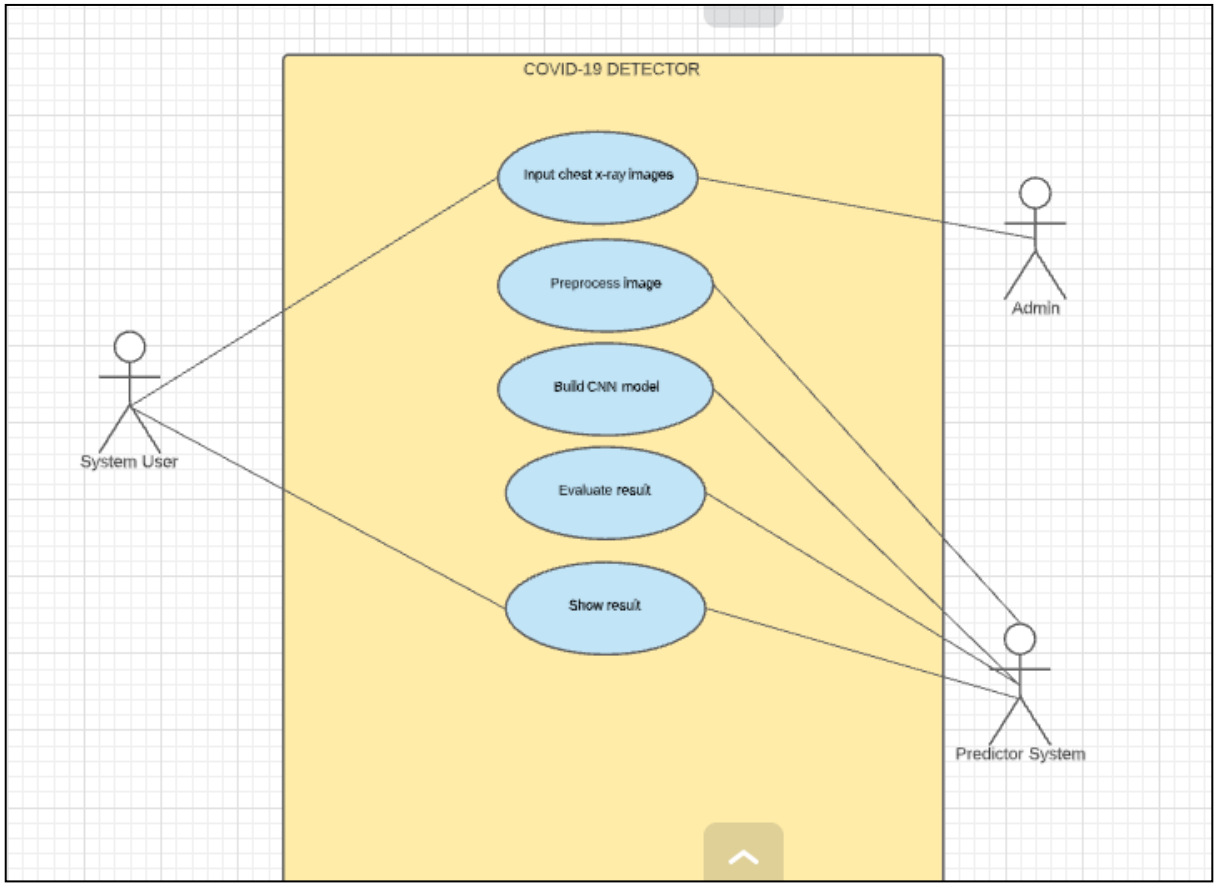
Level 0



(b) Level 1

7. UML Diagrams

Figure below shows our illustration



State Chart Diagram

See figure for our state chart diagram.

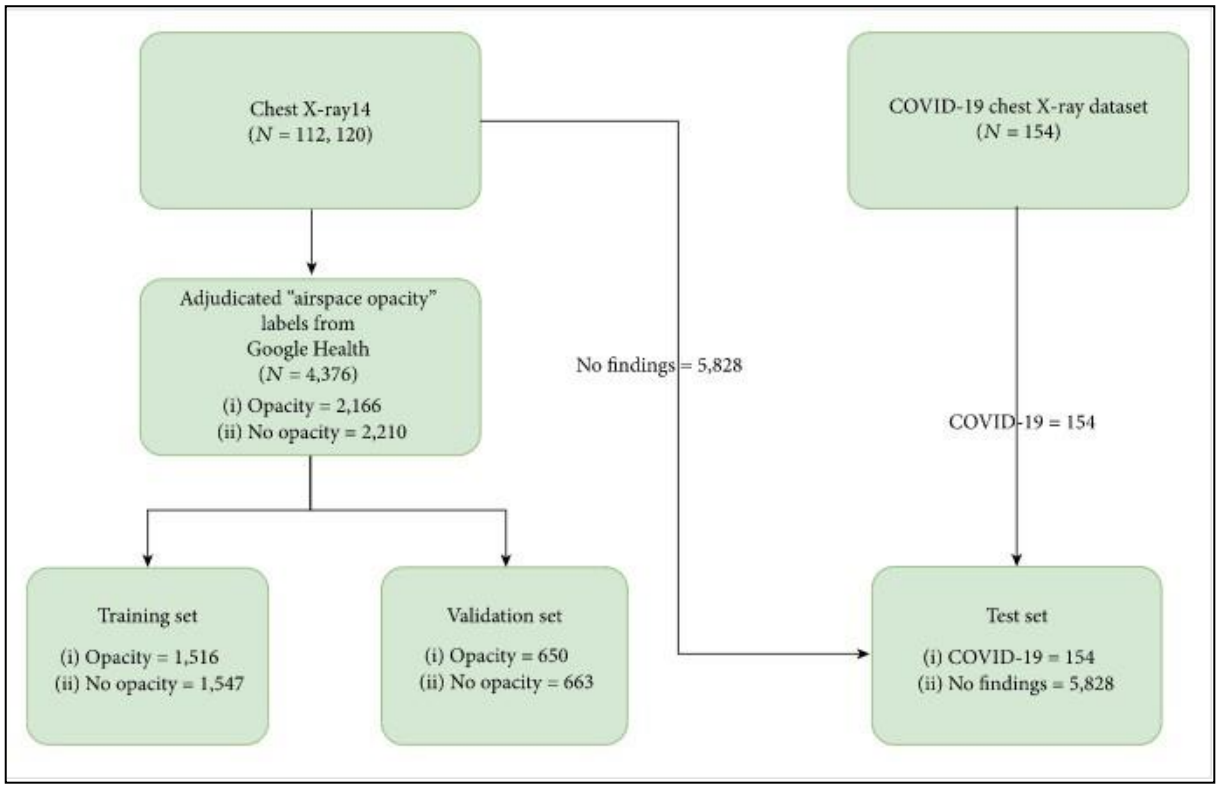


Figure 9: State Chart Diagram

Activity Diagram

Activity diagram show the state of activities by showing the sequence of activities performed. See figure 10 for our activity diagram.

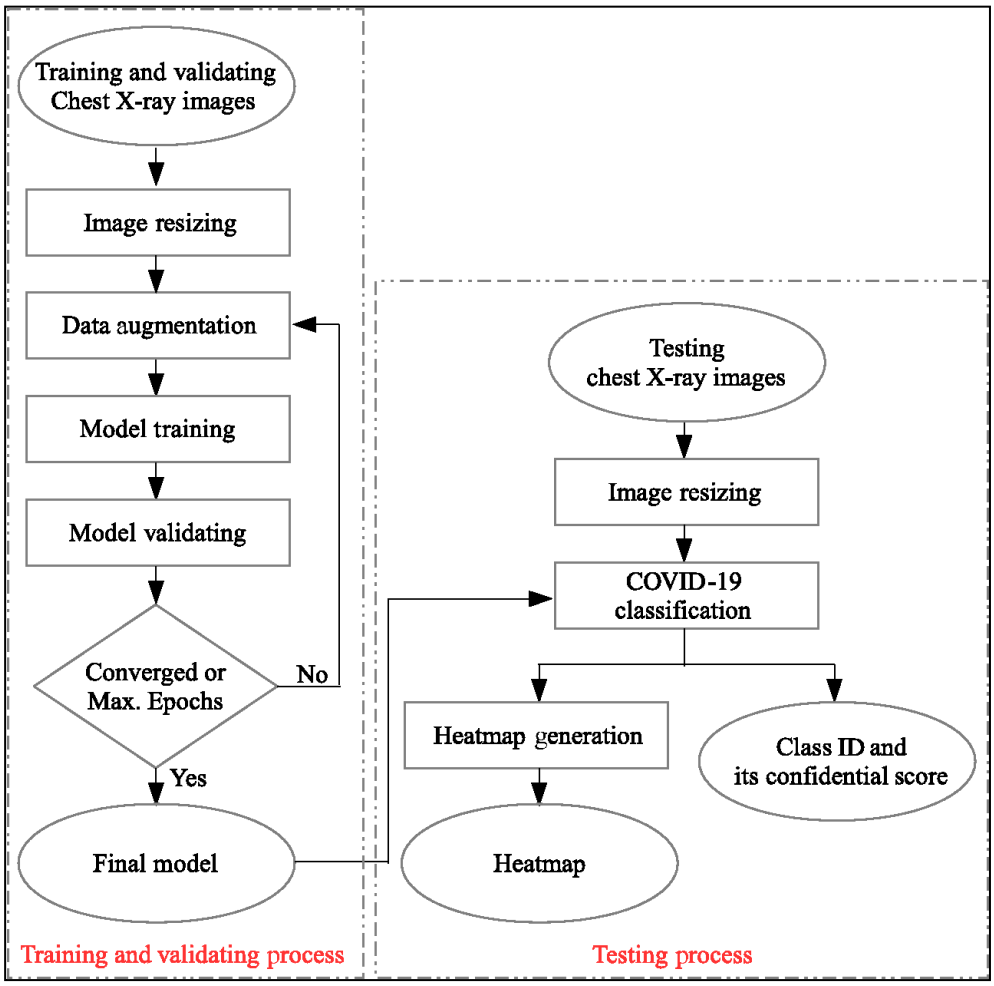


Figure 10: Activity Diagram

.

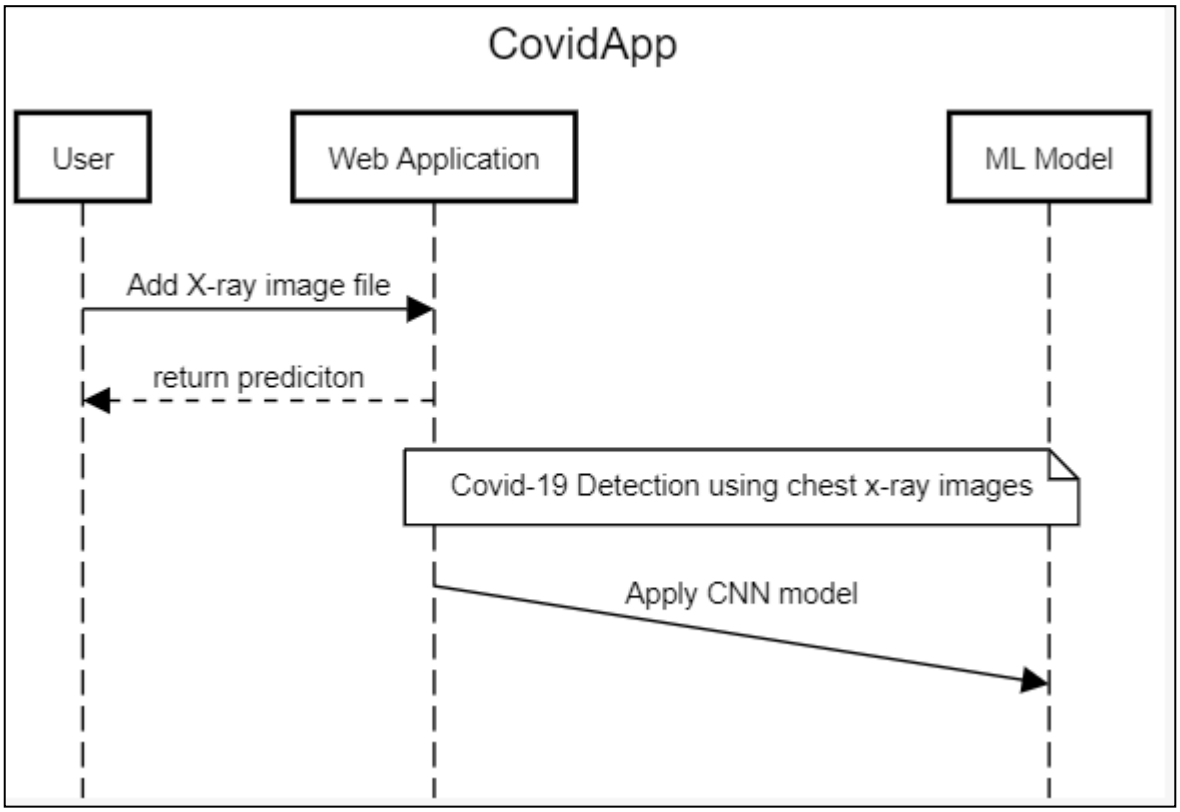
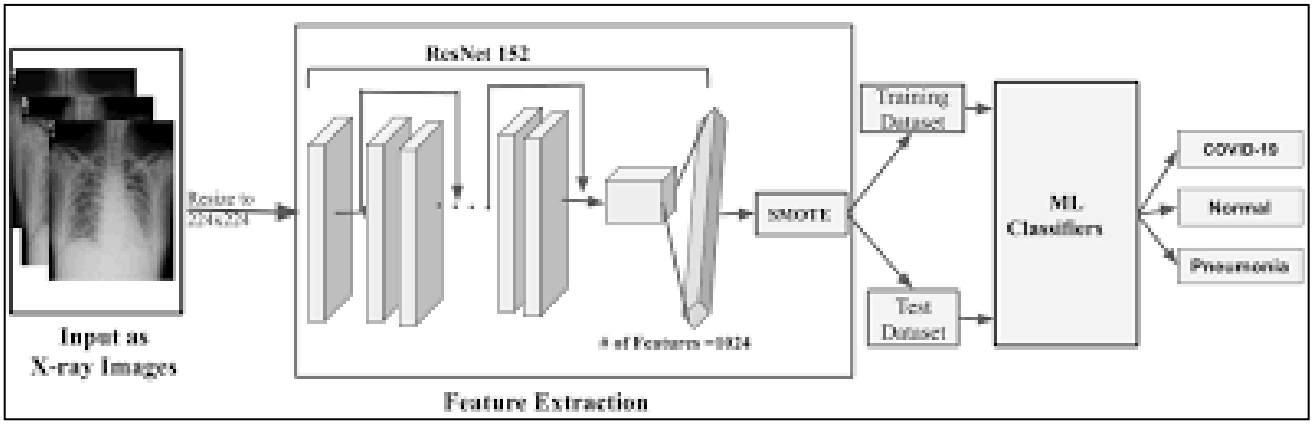


Figure 11: Sequence Diagram

8. Project Plan

The project plan can be explained with the development plan and a gantt chart which tells the schedule of the project execution.

8.1 Development Plan



Development Diagram

8.2 Gantt Chart

.

Refer to figure 13 for our Gantt chart.

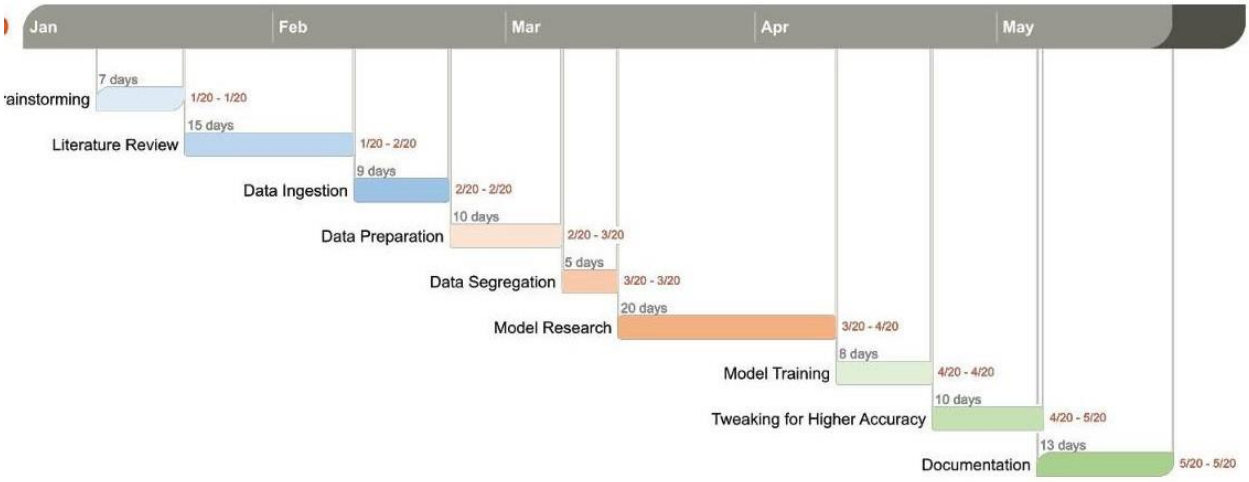


Figure 13: Gantt Chart

9. Risk Mitigation, Monitoring and Management (RMMM)

9.1 Human Error in the Dataset – A Major Risk

Initially we used Kaggle dataset which is a new and labeled dataset with in 1000 images.
However, it has a low human accuracy rate of 67% which makes classifying Covid-19 a difficult task.

Risk Mitigation: We used another labeled dataset with 2000 new chest x-ray images having same features in csv file but far better readings and more accurate findings. This dataset has a significantly better human accuracy rate which makes the model easy to train upon.

Risk Monitoring: We closely observed and recorded different results while using GITHUB dataset only and then combining KAGGLE and GITHUB dataset to ensure that the impact is significant.

Risk Management: We carefully monitored that adding GITHUB dataset is making an impact at all stages and is not leading to a case of overfitting or underfitting.

10. Implementation and Explanation

10.1 Dataset-Creating Custom dataset

We have created our customized dataset using 2000 pics and marked our dataset with two categories as 1 and 0 for respectively. These categories can later be used to display the result or prediction in the form of output nodes. Dataset cleaning was performed for better accuracy like clearing out all the null values from dataset. Exception handling was also performed for removing all corrupted images from dataset and only processing images that are perfect.

```
import pandas as pd
import os
import numpy as np

datapath1='covid-chestxray-dataset-master'
dataset_path='dataset'

categories=os.listdir(dataset_path)
print(categories)

dataset=pd.read_csv(os.path.join(datapath1,'metadata.csv'))
findings=dataset['finding']
image_names=dataset['filename']
```

```
['Covid19 Negative', 'Covid19 Positive']
```

```
datapath2='562468_1022626_bundle_archive'

dataset=pd.read_csv(os.path.join(datapath2,'Chest_xray_Corona_Metadata.csv'))
findings=dataset['Label']
image_names=dataset['X_ray_image_name']
```

```
negative_index=np.where(findings=='Normal')[0]
```

```
negative_image_names=image_names[negative_index]
```

```
for negative_image_name in negative_image_names:
    image=cv2.imread(os.path.join(datapath2,'images',negative_image_name))
    try:
        cv2.imwrite(os.path.join(dataset_path,categories[0],negative_image_name),image)
    except Exception as e:
        print(e)
```

10.2 Image Pre-processing

Image processing was performed for fetching better accuracy from our model. The more we provide image transformations more fined will be the quality of image. Then we are fixing the size of our chest Xray images in our dataset so that every image is equally is processed and evaluated with customized dataset. Then we created two empty arrays target and data for further use in our modelling. Next, we checked that the image is grayscale or not and if image is not grayscale, we are converting those images into grayscale using cvtColour method. In the end of pre-processing, we are normalizing pixel values of our dataset by dividing image pixel by 255 so we get the value between 0 and 1.

```
for category in categories:
    folder_path=os.path.join(data_path,category)
    img_names=os.listdir(folder_path)

    for img_name in img_names:
        img_path=os.path.join(folder_path,img_name)
        img=cv2.imread(img_path)

        try:
            gray=cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
            #Coverting the image into gray scale
            resized=cv2.resize(gray,(img_size,img_size))
            #resizing the gray scale into 100x100, since we need a fixed common size for all the images in the dataset
            data.append(resized)
            target.append(label_dict[category])
            #appending the image and the label(categorized) into the list (dataset)

        except Exception as e:
            print('Exception:',e)
            #if any exception rasied, the exception will be printed here. And pass to the next image
```

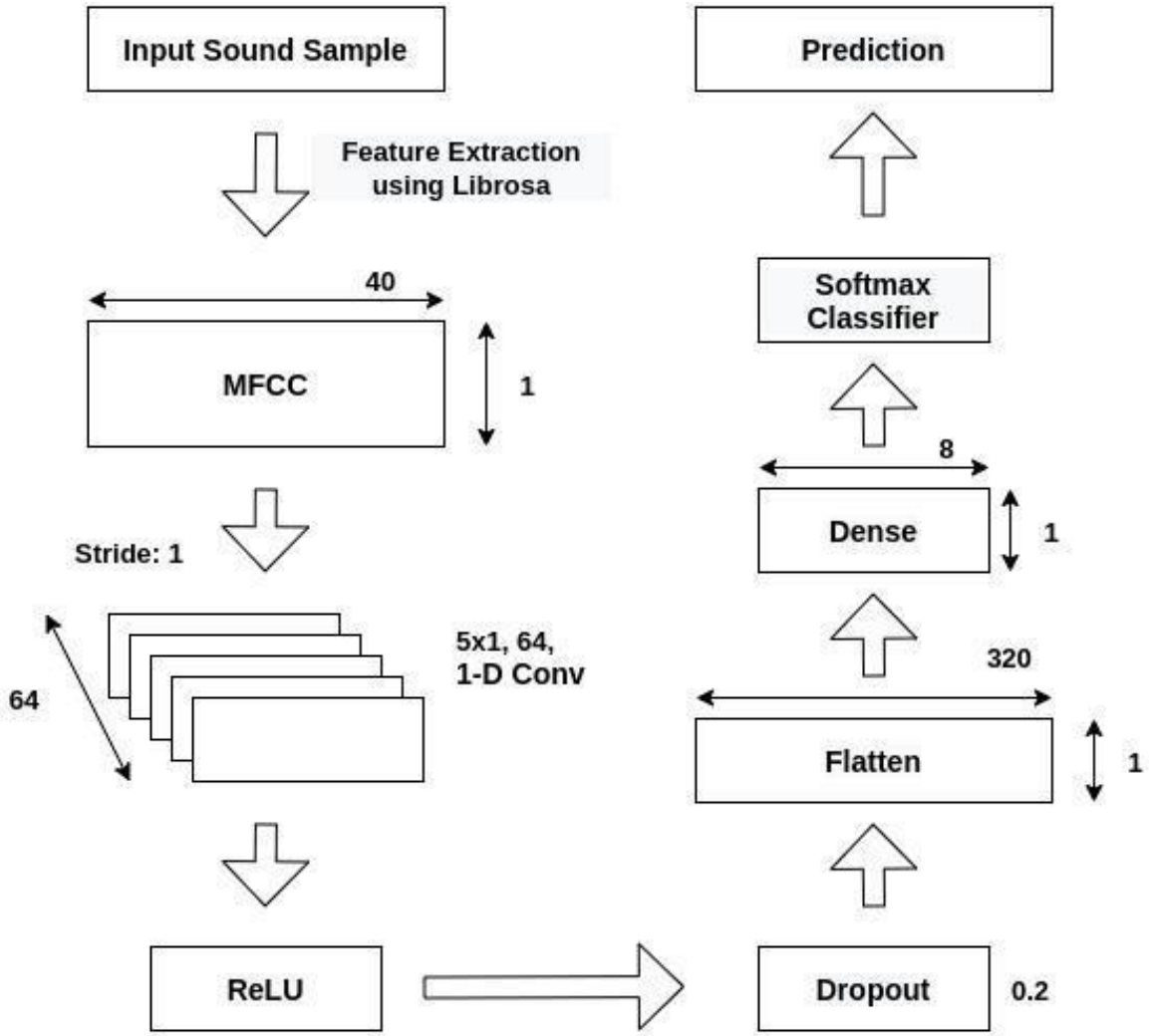
```
[ ]  import numpy as np

     data=np.array(data)/255.0
     data=np.reshape(data,(data.shape[0],img_size,img_size,1))
     target=np.array(target)

     from keras.utils import np_utils

     new_target=np_utils.to_categorical(target)
```

```
     np.save('data',data)
     np.save('target',new_target)
```

```
     np.save('data',data
     np.save('target',ne
```

```
[ ]
```

10.3 Building CNN model



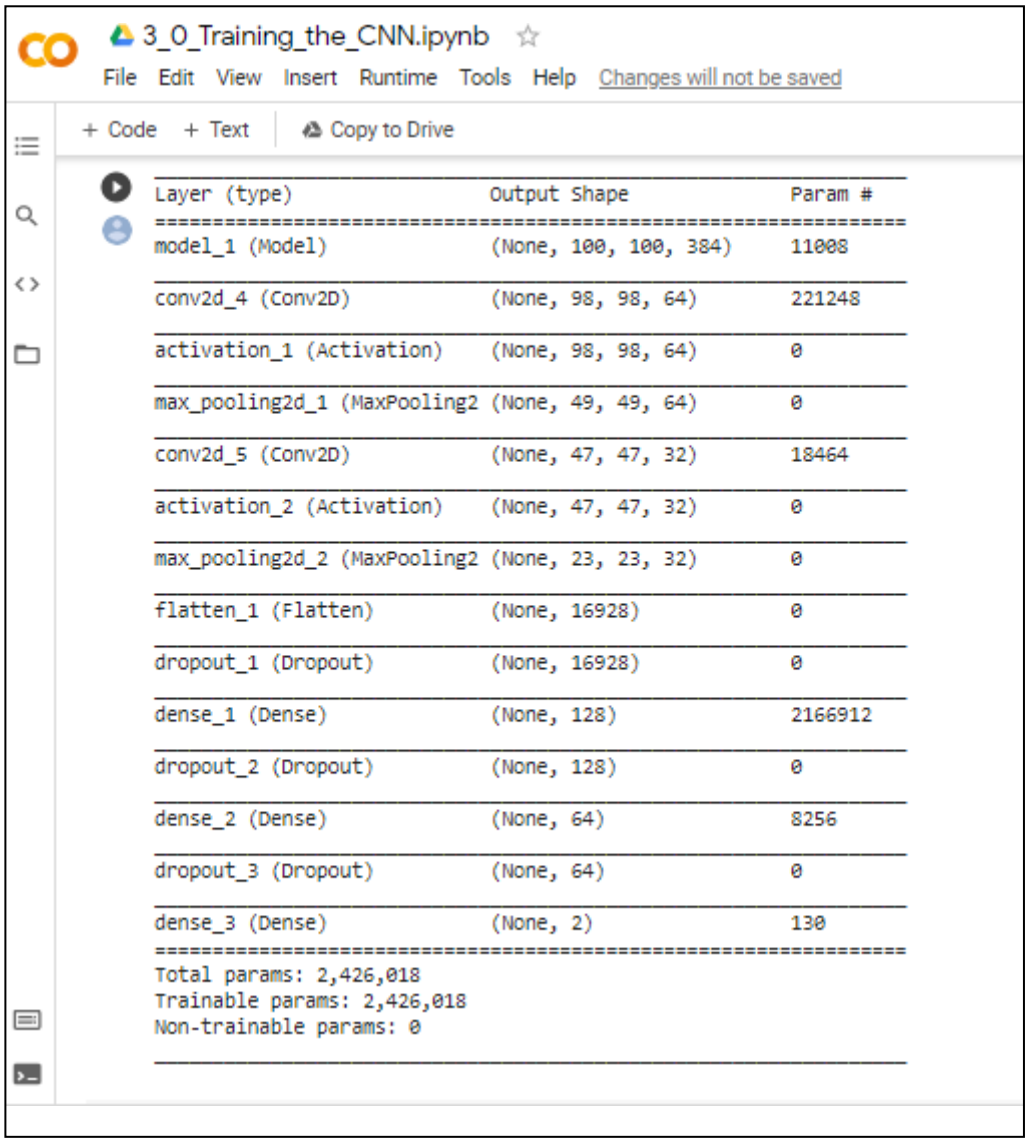
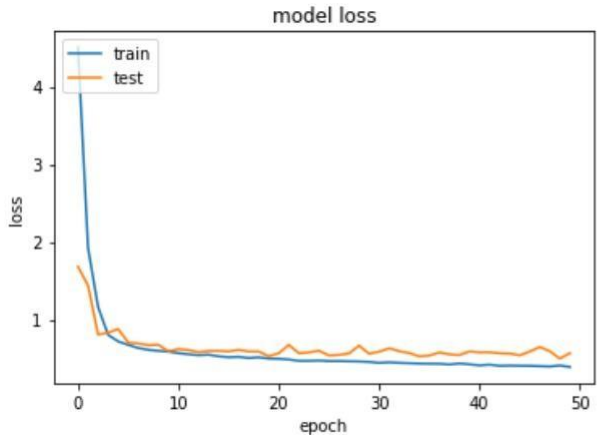
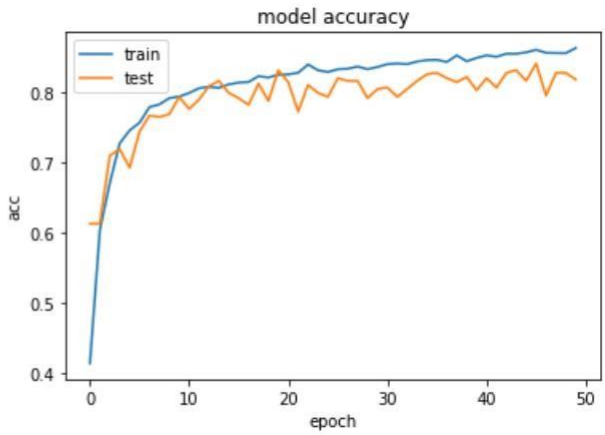


Figure 15: Model Summary

10.4 Results

Training Metrics:

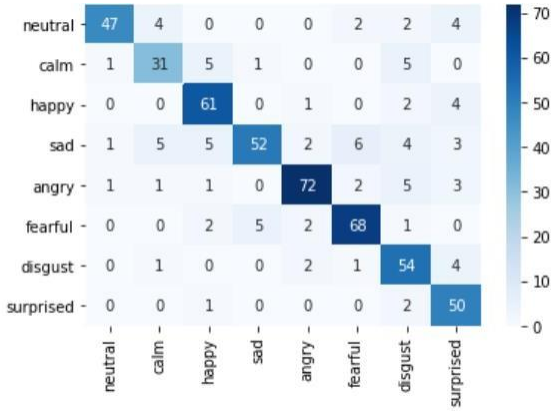




Model loss: Model accuracy

	precision	recall	f1-score	support
neutral	0.91	0.88	0.90	59
calm	0.81	0.70	0.75	43
happy	0.69	0.90	0.78	68
sad	0.88	0.74	0.81	78
angry	0.88	0.85	0.86	85
fearful	0.93	0.81	0.86	78
disgust	0.72	0.84	0.78	62
surprised	0.84	0.89	0.86	53
accuracy			0.83	526
macro avg	0.83	0.82	0.82	526
weighted avg	0.84	0.83	0.83	526

Summary at the end



Computation

10.5 GUI development with FLASK

A GUI is provided using a web interface developed with Flask framework running on Gunicorn server and Postgres database in Linux environment (see figure 20).

The front end provides the user with 3 functionalities:

1. Add an image file to the server (see figure 21)
2. Delete an image file from the server (see figure 22)
3. Predict whether patient is covid-19 or not which is on the server (see figure 23)

When a file is uploaded to the server, it's stored in the file storage of the server and its address is added in the Postgres database.

When the file is deleted from the server, the database deletes the entry of that files' location.

When the user chooses to predict covid-19 from an image file, the file is tested on the pre created model which is stored in the file storage of the server. The model returns the predicted data as a string which is then displayed on the frontend (see figure 24).

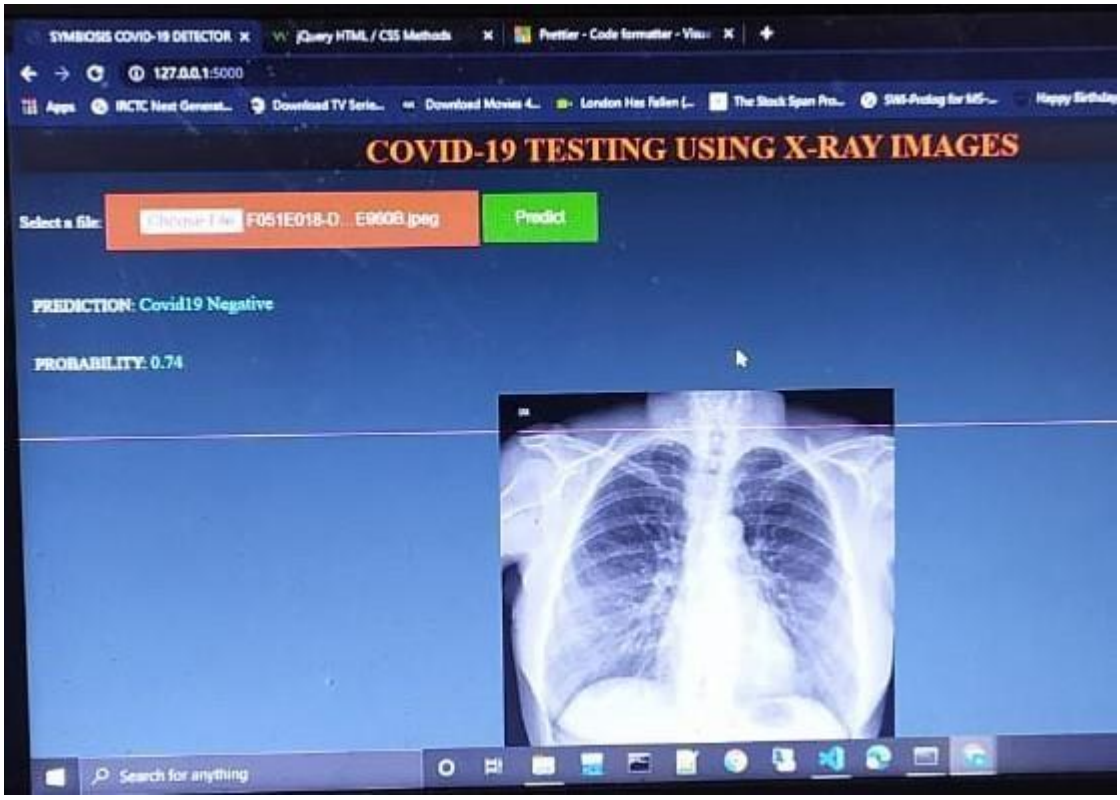


Figure 20: Homepage of the GUI for Covid-19 detection

11. Conclusion

The idea behind creating this project is to build a machine learning model that could recognize chest x-ray images in field of radiology.

We explored various datasets including,

1. Kaggle Repositories
2. Github Repositories
3. Collected from actual medical institutes

And came to conclusion that images collected from actual hospital are best to work on. Using

Kaggle dataset only, we obtained validation accuracy of 90%.

Using Kaggle and GitHub datasets combined, we tried 3 different methods and compared their results.

1. Decision Tree Classifier gave 80% accuracy
2. Random Forest gave 85% accuracy
3. Convolutional Neural Network (CNN) gave 96% accuracy

The model works as expected and is successful at predicting infection in patients for an accuracy 96.5%.

An interactive GUI is also developed in Flask to provide an interface to interact with the model.

The model can be enhanced more in future and used in real world applications. It can be used in medical institutes and hospital for better and efficient detection.

12. Challenges

The challenges that we faced were

1. Finding well labelled datasets having patient who are both covid-19 and normal. The image file should sound authentic and not unnatural.
2. Finding a powerful machine capable of training big datasets to build the models.
3. Experimenting with tweaking the neural network to find the best possible accuracy.

13. References

Research Papers

1.Ooi GC, Khong PL, Müller NL, YiuWC, Zhou LJ, Ho JC, Lam B, Nicolaou S, Tsang KW (2004) Severe acute respiratory syndrome: temporal lung changes at thin-section CT in 30 patients. Radiology 230(3):836–844

2.Wong KT, Antonio GE, Hui DS, Lee N, Yuen EH, Wu A, Leung CB, Rainer TH, Cameron P, Chung SS, Sung JJ (2003) Severe acute respiratory syndrome: radiographic appearances and pattern of progression in 138 patients. Radiology 228(2):401– 406

3.Xie X, Li X, Wan S, Gong Y (2006) Mining x-ray images of SARS patients. In Data Mining. Springer, Berlin, pp 282–294

4.Huang C, Wang, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497– 506

5.Wang Y, Hum, Li Q, Zhang XP, Zhai G, Yao N (2020) Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 accurately and unobtrusively. arXiv preprint arXiv:2002.05534

6.Koo HJ, Lim S, Choe J, Choi SH, Sung H, Do KH (2018) Radiographic and CT features of viral pneumonia. Radiographics 38(3):719–739

7.Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K (2020) Artificial intelligence distinguishes covid-19 from community-acquired pneumonia on chest ct. Radiology: 200905

8.Hansell DM, Banker AA, MacMahon H, McLoud TC, Muller NL, Remy J (2008) Fleischner Society: glossary of terms for thoracic imaging. Radiology 246(3):697–722

9.Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849

10.Apostolopoulos ID, Persiana TA (2020) Covid- 19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med:1

11.Basile C, Coombe C, Pizzarelli F, Covic A, Davenport A, Kanbay M, Kirmizis D, Schneditz D, van der Sande F, Mitra S (2020) Recommendations for the prevention, mitigation, and containment of the emerging SARS-CoV-2 (COVID-19) pandemic in hemodialysis centers. Nephrol Dialysis Transplantation 35:737– 741

12.Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G (2020) Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Modell 5:256–263

13.Yan L, Zhang HT, Xiao Y, Wang M, Sun C, Liang J, Li S, Zhang M, Guo Y, Xiao Y, Tang X (2020) Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv

14.Xu B, Meng X A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)

15.Ajlan AM, Ayad RA, Jamjoom LG, Alharthy A, Madani TA (2014) Middle East respiratory

syndrome coronavirus (MERSCoV) infection: chest CT findings. Am J Roentgenol 203(4):782– 787

16.Kanne JP (2020) Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist

17.Afshar P, Heid Arian S, Aberjhani F, Oikonomou A, Plataniotis KN, Mohammadi A (2020) Covid-caps: a capsule network-based framework for identification of covid-19 cases from x-ray images. arXiv preprint arXiv:2004.02696

18.Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, ..., Reaz MBI (2020) Can AI help in screening viral and COVID- 19 pneumonia? arXiv preprint arXiv:2003.13145

19.Stoecklin SB, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, Mechain M, Meurice L, Nguyen M, Bassi C, Yamani E (2020) First cases of coronavirus disease 2019 Deep learning-based detection and analysis of COVID-19 on chest X-ray images (COVID-19) in France: surveillance, investigations and control measures. Eurosurveillance 25(6):2000094

20.Alqudah AM, Qazaq S, Alqudah A (2020) Automated Systems for Detection of COVID-19 using chest X-ray images and lightweight convolutional neural networks

21.ChoiWJ, Lee KN, Kang EJ, Lee H (2016) Middle East respiratory syndrome-coronavirus infection: a case report of serial computed tomographic findings in a young male patient. Korean J Radiol 17(1):166–170

22.Ayan E, Unver HM (2019) Diagnosis of pneumonia from chest Xray images using deep learning. In 2019 Scientific Meeting on Electrical- Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-5). IEEE

23.Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU Khan K (2020) Pneumonia of unknown etiology in Wuhan, China: Potential for International Spread Via Commercial Air Travel. J Travel Med

24.Chan YH, Zeng YZ, Wu HC, Wu MC, Sun HM (2018) Effective pneumothorax detection for chest

X-ray images using local binary pattern and support vector machine. J Healthcare Eng 2018:1–11

25.ChungM, Bergheim A, Mei X, Zhang N, HuangM, Zeng X, Cui J, Xu W, Yang Y, Fayad ZA, Jacobi A (2020) CT imaging features of 2019 novel coronavirus (2019-nCoV). Radiology 295(1):202– 207

26.Cohen JP, Morrison P, Dao L (2020) COVID-19 image data collection. arXiv preprint arXiv:2003.11597

27.DadárioAMV, Paiva JPQ, Chute RC, Machado BS, Szarf G (2020) Regarding artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology:201178

28.Diao B, Feng Z, Wang C, Wang H, Liu L, Wang C, Wang R, LiuY, Liu Y, Wang G, Yuan Z (2020) Human kidney is a target for novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. medRxiv

29.Fu S, Fu X, Song Y, LiM, Pan PH, Tang T, Zhang C, Jiang T, Tan D, Fan X, Sha X (2020) Virologic and clinical characteristics for prognosis of severe COVID-19: a retrospective observational study in Wuhan, China. medRxiv

30.Hassani RTJ, Sandal O (2020) Le nouveau coronavirus Covid-19: quels risques ophtalmiques? J Francais D'Ophtalmologie 43(4):291

31.Hassanien AE, Mahdi LN, Ezzat KA, Elmousalami HH, Ella HA (2020) Automatic X-ray COVID-19 lung image classification system based onmulti-level Thresholding and support vector machine. medRxiv

32.Hemdan EED, Showman MA, Karar ME (2020) Covidx-net: a framework of deep learning classifiers to diagnose covid-19 in Xray images. arXivpreprint arXiv:2003.11055

33.Ilyas M, Rehman H, Nait-ali A (2020) Detection of Covid-19 from chest X-ray images using artificial intelligence: an early review. arXiv preprint arXiv:2004.05436

34.Jacobi A, Chung M, Bernheim A, Eber C (2020) Portable chest Xray in coronavirus disease-19

(COVID-19): a pictorial review. Clin Imaging 64:35–42

35.Maghdid HS, Asaad AT, Ghafoor KZ, Sadiq AS, KhanMK (2020) Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. arXiv preprint arXiv:2004.00038

36.Ozkaya U, Ozturk S, Barstugan M (2020) Coronavirus (COVID- 19) classification using deep features fusion and ranking technique. arXiv preprint arXiv:2004.03698

37.Ozturk T, Talon M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med:103792

38.Qiu G, Gai Z, Tao Y, Schmitt J, Kullak-Ublick GA, Wang J (2020) Dual-Functional Plasmonic Photothermal Biosensors for Highly Accurate Severe Acute Respiratory Syndrome Coronavirus 2 Detection. ACS nano

39.Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C (2020) Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 20:425–434

40.Thejeshwar SS, Checkered C, Eswaran K, Precise Prediction of COVID-19 in Chest X-Ray Images Using KE Sieve Algorithm

41.Durrani M, Haqq I u, Kalsoom U, Yousaf A (2020) Chest X-rays findings in COVID 19 patients at a University Teaching Hospital-A descriptive study. Pakistan J Med Sci 36(COVID19-S4): S22

42.Yadav SS, Jadhav SM (2019) Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data 6(1):113

43.Zhang J, Xie Y, Li Y, Shen C, Xia Y 2020 Covid-19 screening onchest x-ray images using deep learning-based anomaly detection. arXiv preprint arXiv:2003.12338

44.Zhao W, Zhong Zine X, Yu Q, Liu J (2020) Relational between chest CT findings and clinical conditions of coronavirus disease (Covid-19) pneumonia: a multicenter study. Am J Roentgenol 1-6