Analysis and Detection of Coronavirus(COVID-19) using Lung X-ray

A project report submitted in partial fulfillment of the requirements for Mini Project

by

P Madhu Chandra
(2018IMT-049)

Under the Supervision of

Dr. Ajay Kumar

ABV-INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT
GWALIOR-474 015
2021

ABSTRACT

The outbreak of Coronavirus disease 2019 (COVID-19), caused by SEVERE ACUTE RESPIRATORY SYNDROME (SARS); in response to the rapidly increasing number of cases of the emerging disease, this Analysis attempts to provide a comprehensive review of the COVID-19 Virus. I hope that this Analysis helps in understanding and taking some precautions of the disease as the Analysis provides the significant symptoms of COVID-19 and what age groups are mostly getting affected and Analysis on the number of cases through a period for each state. So, in this particular situation, one primary thing that needs to be done is manual testing, so that the actual situation can be understood and appropriate decisions can be taken. But the drawbacks of manual testing include the availability of testing kits, which are costly and inefficient blood tests; a blood test takes hours to generate the result. So, the idea is to overcome these limitations using the Deep Learning technique for efficient treatment. The faster we produce the results, the fewer cases in the city, that's why we can use CNN to get our job done.

Keywords: Coronavirus, pneumonia, outbreak, SARS-CoV-2, COVID-19

ACKNOWLEDGEMENT

I am extremely thankful to **Dr. Ajay Kumar** for giving me the opportunity to carry out my research work under his guidance. He have supported me constantly throughout the research work, not only academic support but he have motivated me whenever needed and also increased my confidence which was very helpful for me to trust myself that I will be able to complete my research successfully. He have always answered all my doubts with patience and appreciated my thoughts and also explained me about the improvements which can be done to achieve better results. The research work could be completed successfully without any problem only because of his encouragement and support.

I would also like to thank my peer students and also the Institute for encouraging me at each step and helping me in all possible ways till I could successfully complete my research work.

Signature:

Name: P Madhu Chandra
Roll. No: 2018IMT-049
Date: Thursday 14 May, 2021

TABLE OF CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

ANN	Artificial Neural Network
BP	Back Propagation
CNN	Convolutional Neural Network
MERS	Middle East Respiratory Syndrome
SARS	Severe Acute Respiratory Syndrome

Chapter 1

INTRODUCTION

On December 31 2019, a cluster of cases of pneumonia of unknown cause, in the city of Wuhan, Hubei province in China, was reported to the World Health Organisation. In January 2020, an anonymous new virus was identified and named the 2019 novel coronavirus.

Coronavirus is a family of viruses that affect the respiratory system of a person. Respiratory diseases can be the common cold to more severe diseases as SARS and MERS

- Middle East Respiratory Syndrome (MERS-CoV)
- Severe Acute Respiratory Syndrome (SARS-CoV)

Coronaviruses got its name because of the way they look under a microscope. The virus consists of a genetic material surrounded by an envelope with protein spikes,which appears like a crown. The word Corona means "crown" in Latin.

Convolutional Neural Network (ConvNet or CNN) is a special type of Neural Network used effectively for image recognition and classification. The reason why Convolutional Neural Networks (CNNs) do so much better than classic neural networks on images and videos. After learning a certain pattern in a picture, a convolutional network can recognize it anywhere, CNN takes advantage of local spatial coherence of images.

Can we distinguish between both X-rays if they haven't been labelled? No, we can't,but a CNN can.

We can distinguish the x-ray image using a deep-learning method called Convolution-Neural-Networks, which stores the features of the different label images.

The problem is a classification problem where we classify Normal vs COVID-19 cases. Limitation:

1. More time saving; less expensive; easy to operate
2. Practically we need more accuracy as we can't wrongly predict the x-ray image as it might lead to further spread of disease which is not highly encouraged.

Still, this model can return good accuracy and can be further improved.

1.1 Project Objectives and deliverables

The objective of the project is to do an analysis and With the Chest X-Ray dataset, Develop a Machine Learning Model to classify the X Rays of Healthy vs Pneumonia (Corona) affected patients this model powers the AI application to test the Corona Virus in Faster Phase Predict with decent accuracy so that it will helpful for anyone who uses it. And the datasets are available on Kaggle. Dataset 1: A dataset containing information regarding COVID 19 in India Dataset 2: And the second dataset Collection Chest X-Ray of Healthy vs Pneumonia (Corona) affected patients, infected patients, along with few other categories such as SARS (Severe Acute Respiratory Syndrome), Streptococcus ARDS (Acute Respiratory Distress Syndrome)

1.2 Background Information /Motivation/ Literature Survey

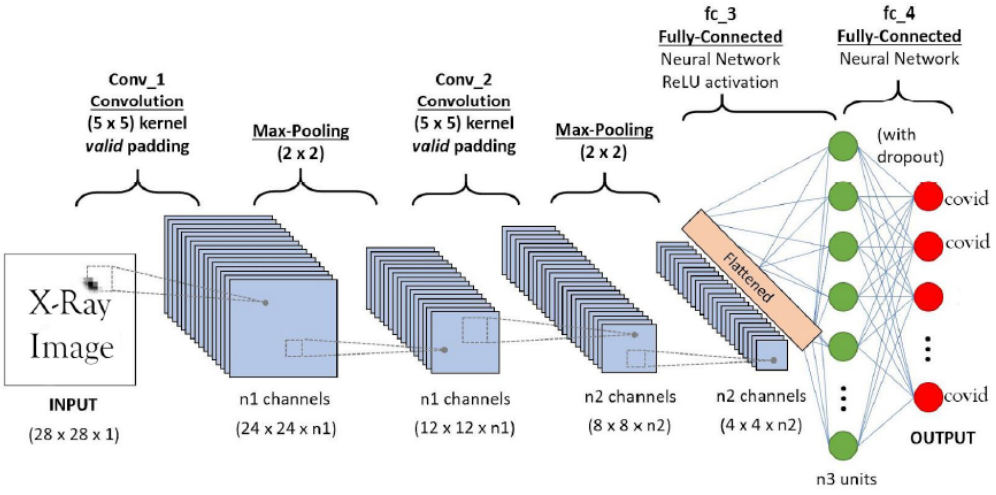
The Motivation for using CNN for the classifying the X-ray images is the powerful architecture of the CNN. That is to identify features from the images. The basic Idea of CNN is to classify a given image as either a cat or a dog. It's not about what we classify here. It's about the method we use here (CNN). The CNN learns the features from the given image here a dog or cat and can classify the test-image either into a dog if it finds the features of a dog in the image or cat if it finds the features of a cat. Same idea can be used here to classify a x-ray image; it learns the features of both the x-ray images(Normal x-ray image and covid x-ray image) and later on is used to classify the test-images.

Author	Title	Year	Publisher	Work
François Cholle	Deep Learning with python	2017	Google	Understanding neural networks,CNN and Back-propagation.
C. Szegedy etal	Going deeper with convolutions,"inProceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition	2015	IEEE	Inception-V2 is a combined architecture proposed by that uses the idea of inception blocks and residual layers together.
Taban Majeed, Rasber Rashid,Dashti Ali, and Aras Asaad	Covid-19 detection using CNN transfer learning from X-ray Images	2020	Doi.org	Quantitative analysis to evaluate 12 o-the-shelf convolutional neural networks (CNNs) for the purpose of COVID-19 X-ray image analysis.
T. Ai et al	Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease (COVID-19)	2020	pubs.rnsa.org	Review studies devoted to the use of radiography images to aid and complement PCR in China diagnosing COVID-19 case.
H. S. Maghdid, A.T. Asaad,K. Z. Ghafoor,A. S. Sadiq, and M. K. Khan	"Diagnosing COVID-19 Pneumonia from X-Ray and CT Images using Deep Learning and Transfer Learning Algorithms	2020	arxiv.org	A model of 16layers that can detect covid-19 using ct scans and xrays on small datasets.
K. Simonyan and A. Zisserman.	Very Deep Convolutional Networks for Large- ScaleImage Recognition	2015		Brief knowledge about vgg16.

Table 1.1: LITERATURE SURVEY

Chapter 2

Methodology



The Methodology includes 4 steps

- Convolution operation
- Pooling
- Flattening
- Full Connection

2.1 SYSTEM ARCHITECTURE

As we are dealing with images in computer terms, an image is just a 2d matrix with pixel values between 0 and 255 including them where 0 represents brightness, and 255 represents black, so a black and white image is nothing but a matrix with pixel values 0's and 255's

As we can see the image above the rightmost matrix represents the Leftmost image here the pixel value 1 represents black pixel and white as 0 .

We've got an input image as we discussed that's how we're going to look at images just ones and zeros as we can see the input image it's the smile image we looked before.

Feature Detector or filter is like storing a pattern here it's 3*3 it can be 5*5 or 7*7 so now we slide this filter on the input image to get the feature map.

So what is this feature map and why do we need it?
We need this Feature map for each input image because we need to find the unique features of the image.

2.1.1 STEP 1- CONVOLUTION

So what have we created here?
This is called Feature map and sometimes called convold feature Points to be noted here

And that's a very important function of the feature detector of this whole convolution step is to make the image smaller because that'll be it'll be easier to process it and it'll be just faster.

The very first question here is are we losing information?
Yes, we are losing some part of the information but at the same time we are storing the features of the image. That's what a feature detector does for us; it detects the important features for us.

We get the max value when the pattern matches in the image. So how come there are many feature maps?
We don't just have a single feature detector we need to detect as many as features as possible from the input image to classify it so we have different feature detectors for a input image for the feature map we used a feature detector similar to the one we just saw for the next one we use a different feature detector.

So get a clear idea about max pooling let's take a example let we want to identify a cat in an image so in an image the cheetah can look in one direction in an another

image it can look in another direction it can be in one part in an image it can be in anotherposition so at the end we need to identify it as a cheetah no matter what direction its looking or where it is in the image.

2.1.2 STEP 2- MAX POOLING

One of the distinct feature of the a cheetah is the tears that are going from its eyes to side of its nose so if the neural net is trying to find the feature it should find it no matter where it is in the image and what direction the cheetah is looking.

That's what Pooling is all about
There's several different types of pooling methods like mean pooling Max pooling But for now we're just applying Max pooling so we take a box of two by two pixels.

We just flatten the max pooling matrix into a vector so that we can use an input for the neural network later on for classification as shown in the image below.

2.1.3 STEP 3- FLATTENING

The flatten max pooled matrix is the input for artificial neural network initially the neurons in the neural network have random weights and the weights are changed accordingly to maximize the output accuracy.

2.1.4 STEP 4- FULL CONNECTIONS

The flatten max pooled matrix is the input for artificial neural networks.Initially the neurons in the neural network have random weights and the weights are changed accordingly to maximize the output accuracy.

2.2 DATA PREPROCESSING

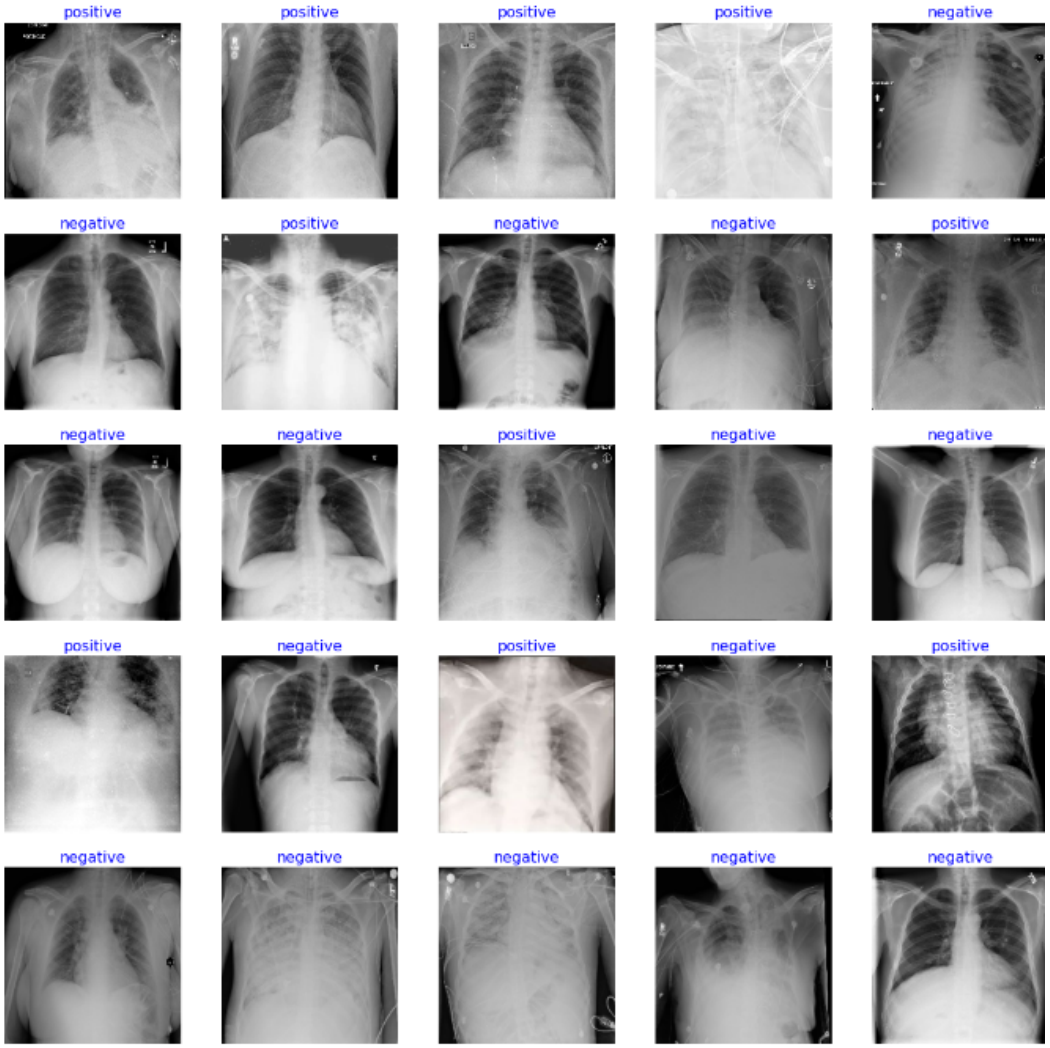


Figure 2.1: Covid and Non-Covid Lung X-ray

These image paths are stored in a data-set path images in a column and labels of the image and then we iterate a loop over the data-set and classify these images into their respective directories.

Epoch	Loss	Accuracy	V_loss	V_acc	LR	Next LR	Monitor	Duration
1 /10	5.907	90.628	5.56295	64.352	0.00100	0.00100	val_loss	312.06
2 /10	3.932	97.477	3.73533	81.944	0.00100	0.00100	val_loss	144.48
3 /10	2.966	98.816	2.85666	86.574	0.00100	0.00100	val_loss	136.14
4 /10	2.311	99.125	2.22851	93.056	0.00100	0.00100	val_loss	135.43
5 /10	1.831	99.382	1.78517	93.287	0.00100	0.00100	val_loss	134.28
6 /10	1.472	99.382	1.51221	93.981	0.00100	0.00100	val_loss	136.19
7 /10	1.196	99.408	1.25889	93.056	0.00100	0.00100	val_loss	139.00
8 /10	0.976	99.459	1.16883	94.907	0.00100	0.00100	val_loss	137.54
9 /10	0.801	99.537	1.07930	95.833	0.00100	0.00100	val_loss	136.69
10 /10	0.671	99.331	1.35431	96.065	0.00100	0.00050	val_loss	136.82

Figure 2.2: Epoch Table

The above table tells about the no.of times it runs over the dataset taken into consideration ,so there are 10 epochs and the accuracy and loss values.

2.2.1 Scaling

Scaling the pixel values are between 0 and 255 we divide each pixel value by 255 so that we scale them down to 0 and 1.

2.2.2 Splitting of Data into Train, Test, validation, and Data Augmentation

Total data in the data-sets are divided into Test, Train data accordingly.

- Train Data: 80 percentage or 0.75.
- Validation Data: 10 percentage or 0.1.
- Test Data: 10 percentage or 0.1

	Training	Test
Non-Covid	1072	269
Covid	113	28

Table 2.1: Ratio of Covid and Non-Covid images

As we have few images data-augmentation plays a key role here.Data-augmentation is a technique to increase the diversity of your training set by applying random transformations such as image rotation and flip images in the vertical and horizontal direction.

2.3 IMPLEMENTATION DETAILS

2.3.1 Dataset

- A data-set containing information regarding COVID 19 in India.
- The dataset is a collection of X Rays of Chest with
  X- Rays of both Healthy lungs and Pneumonia Lungs i.e Lungs affected by Corona.

2.3.2 Tools and Hardware

- Python
- Numpy
- Pandas
- Keras or Tensor flow
- Seaborn
- Sklearn
- CPU or GPU(recommended)

2.3.3 Hardware Setup

- If your computer does not have GPU(GPU is recommended for faster Computation) we can use GPUs that are provided by Google or Kaggle.
- Kaggle is recommended as the datasets are available on kaggle. Datasets can be loaded with ease if you are using Kaggle kernel. If you want to do it on Google Colab you need to upload the datasets to your Google Drive first.
- In a kaggle kernel we can use the GPU thats provided for us by Select the Settings tab. Then select the checkbox for Enable GPU. Verify the GPU is attached to your kernel in the console bar, where it should show GPU ON next to your resource usage metrics. And turn on the internet to download the relvent libraries.

Chapter 3

Results and Discussion

3.1 Analysis

3.1.1 Indian Analysis

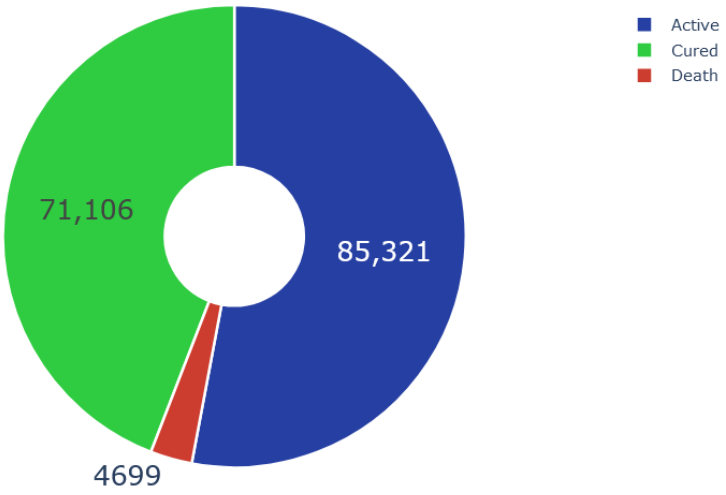


Figure 3.1: Current situation in Indian as of May 2020

By analysing the above graph portraits that :

- 49.7 percentage of the Indian population is still suffering from COVID-19
- 47.4 percentage of the Indian population have recovered from COVID-19

- And 2.88 percentage have been deceased .

3.1.2 Gender Analysis

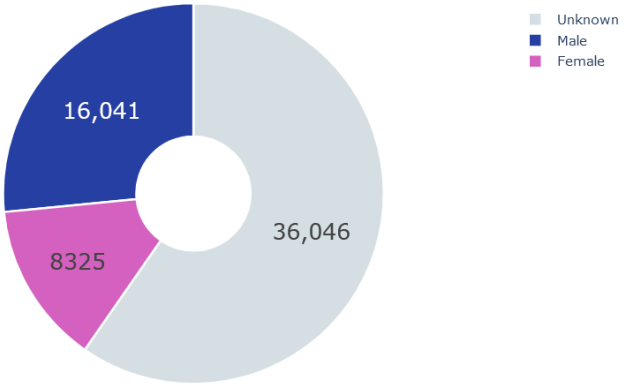


Figure 3.2: Gender wise active cases

From the above pie cart it portraits that there are almost 81 percentage of missing cases ,and from the data which is available we can say that 13 percentage of males and where as 7 percentage of females have been affected from COVID-19.

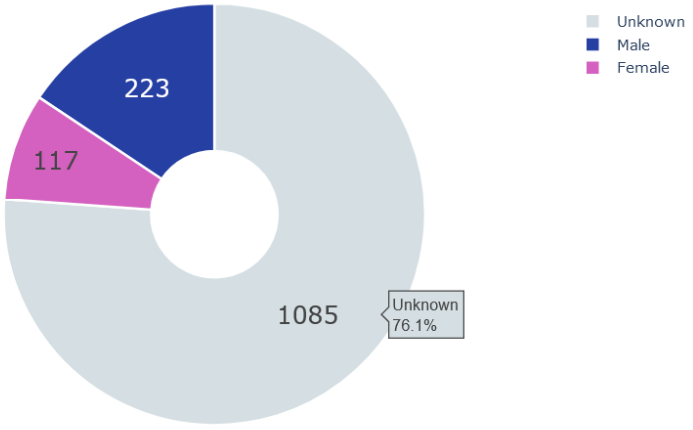


Figure 3.3: Gender wise Deaths

From this pie chart we can tell that neglecting missing data a percentage of 15.6 male have deceased where as a percentage of 8.2 female have been deceased.

3.1.3 Age Group Analysis

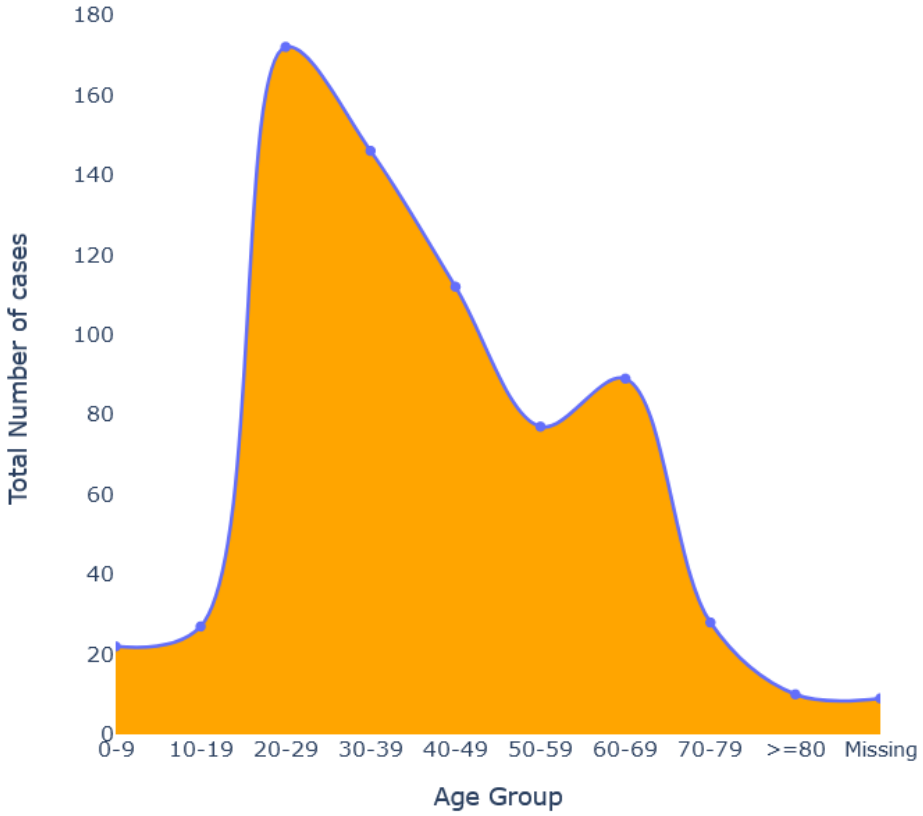


Figure 3.4: Age wise analysis

The most effected from Covid-19 is from the age from of 20-60,where as from the age group 0-20 the cases are very less and the age group ¿60 is also not that affected when compared to the (20-60)group.20-60. And the least effected are ages between 0-19 and ¿60 which makes upto 17

3.1.4 State wise Analysis

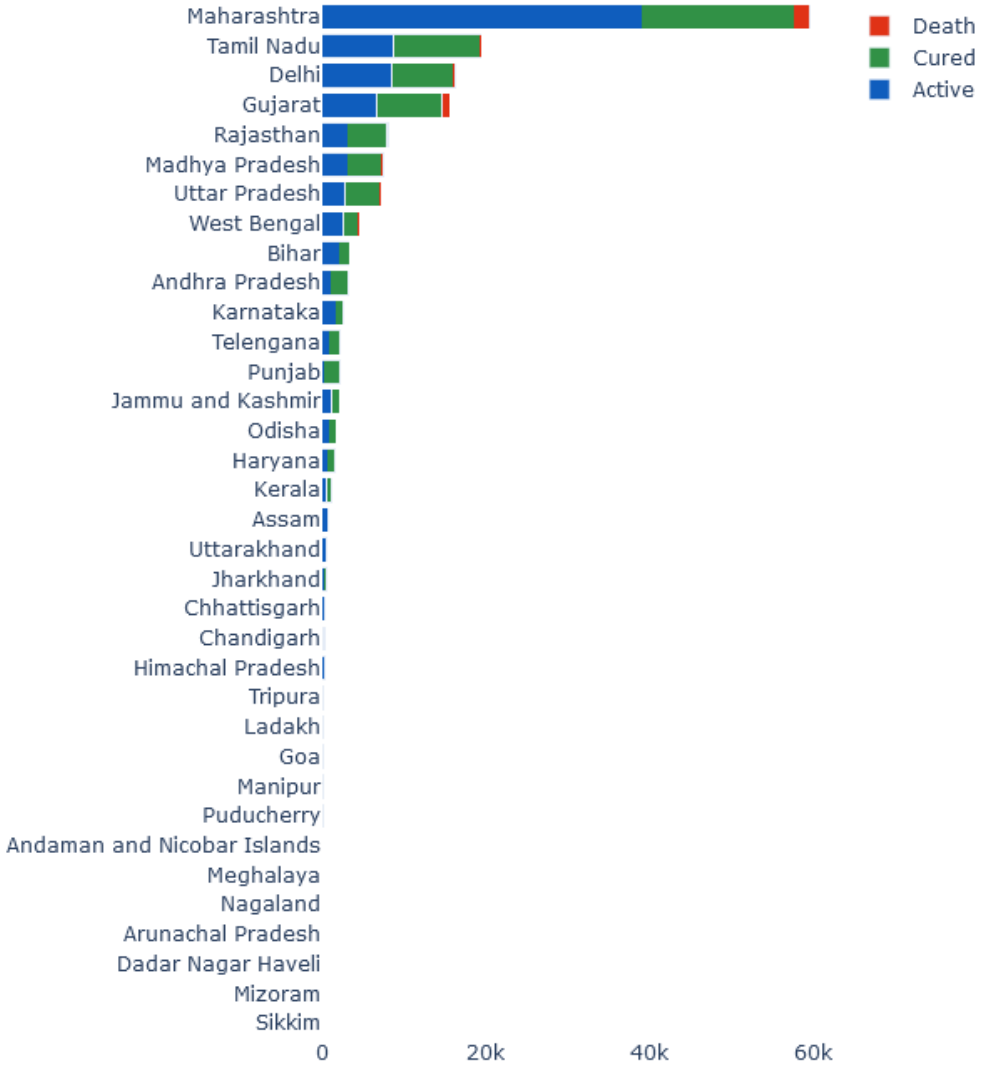


Figure 3.5: Age wise analysis

Of all the states and Union Territories, Only Lakshadweep has not reported any case so far the remaining 35 have reported COVID-19 cases so far.the total number of cases reached 19,690,596 in which 11,68,023have recovered and 38,135 have died so far. The most affected states according the analysis are in the order

- Maharashtra
- TamilNadu

- Delhi
- Gujarat

These states mentioned above also have the most number of death rates. Makes upto 60 percent of the country's COVID-19 toll as you can see from above DataFrame.

Cure-rate in Delhi, Haryana, Tamilnadu are high when compared to otherstates.

3.1.5 Symptoms of COVID-19

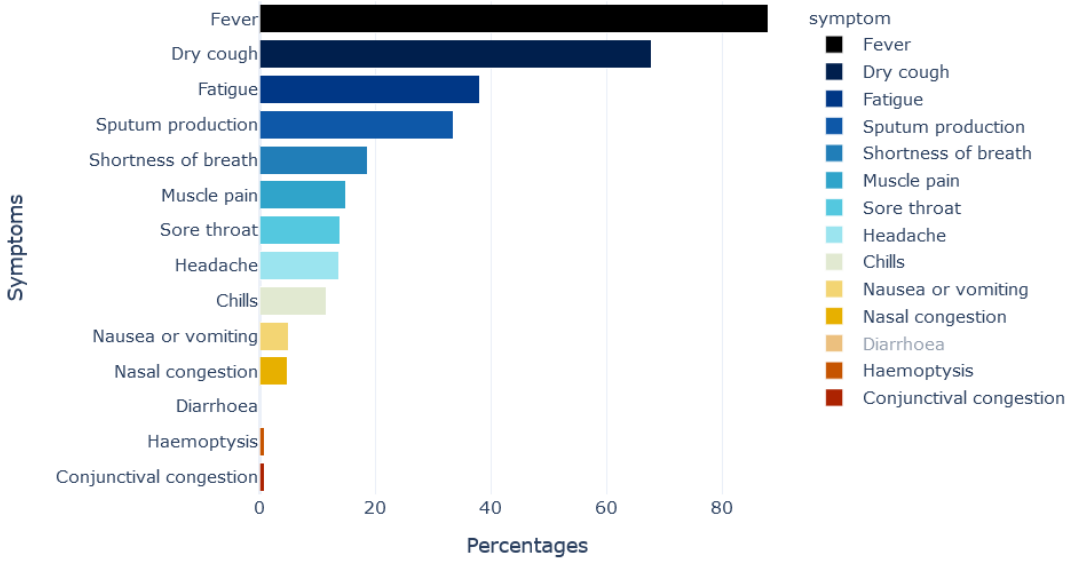


Figure 3.6: Symptoms of COVID-19

From the graph we can conclude the major symptoms of COVID-19 are Fever,Dry-cough,fatigue,Difficulty of Breath,Sore throat,Headache.

- A percentage of 87.9 people who have affected with COVID-19 have a symptom of fever.
- Where as 67.7 percentage of people have Dry-cough.
- 38.1 percentage of people have fatigue.
- 18.6 percentage of people have Difficulty in breath.
- 13.9 percentage have Sore throat.
- 13.3 percentage of people have Headache.

3.1.6 CNN-MODEL

The performance metrics used to evaluate the model are the accuracy and the loss. our main aim was to increase the accuracy at the same time decrease the loss.

1. The figure shows the comparison between accuracy and epochs, and also between loss and epochs. From the figure the accuracies i.e., both the training and validation accuracies are increasing from the first epoch itself which tells that the model is learning.

2. TensorFlow Framework was used. The dataset is taken from the Kaggle.

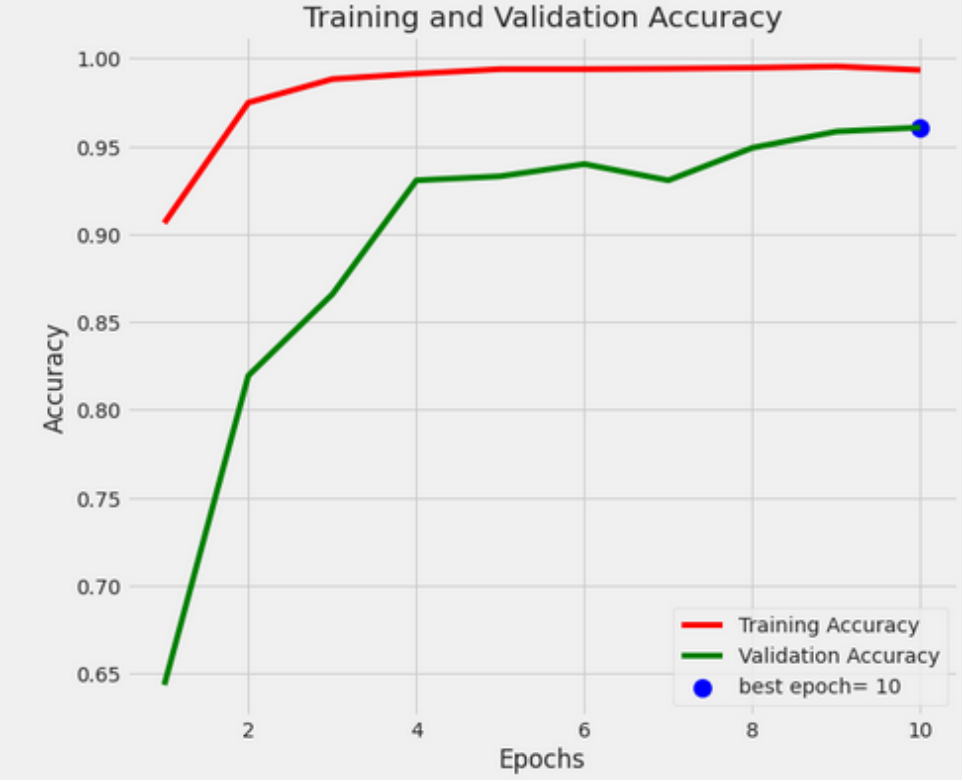


Figure 3.7: Accuracy metrics for Inception-V3

From the Figure 3.7 it can be observed that the model's training accuracy and validation accuracy are increasing for each epoch thus the model is neither over fitting nor underfitting.

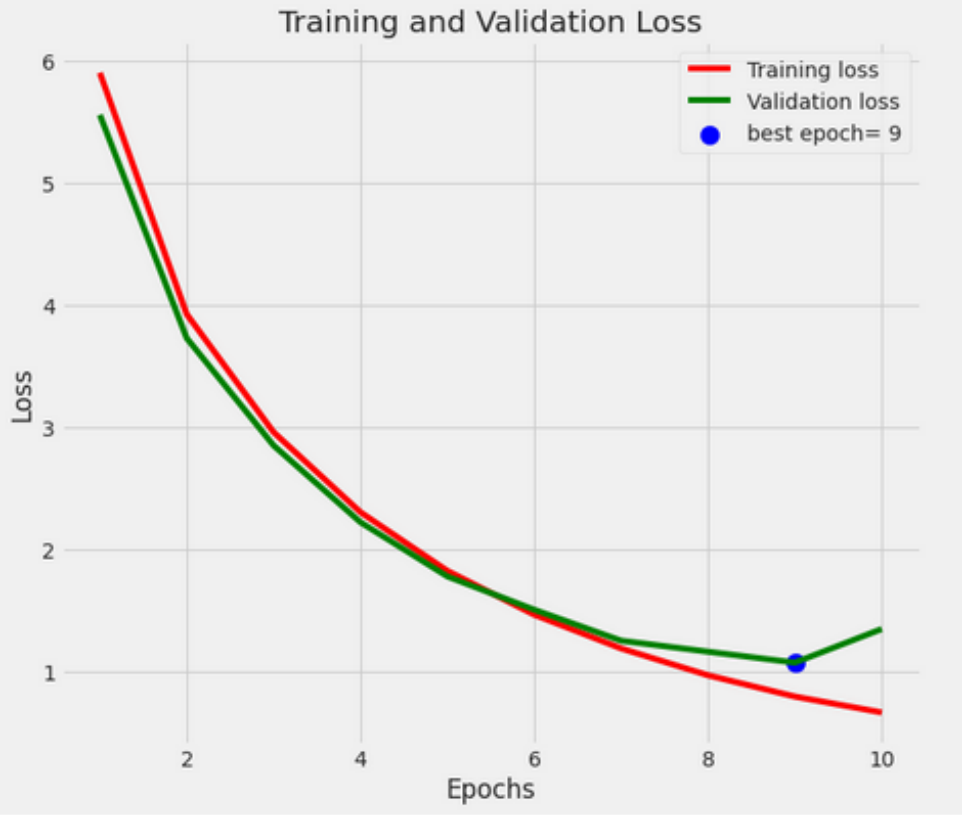


Figure 3.8: Loss metrics for Inception-V3

These metrics are for 10 epochs and a learning rate of 0.001. If the no.of are increased overfitting is observed we used just 10 epochs as the available data is less. **Note:**

Model	Train Accuracy	Test Accuracy
Inception-V3	99.53	97.5
CNN from Scratch	70.98	65.39

Table 3.1: Accuracy Table

These results are for normal x-ray images and 141 COVID-images these metrics may change if the no.of input images change.

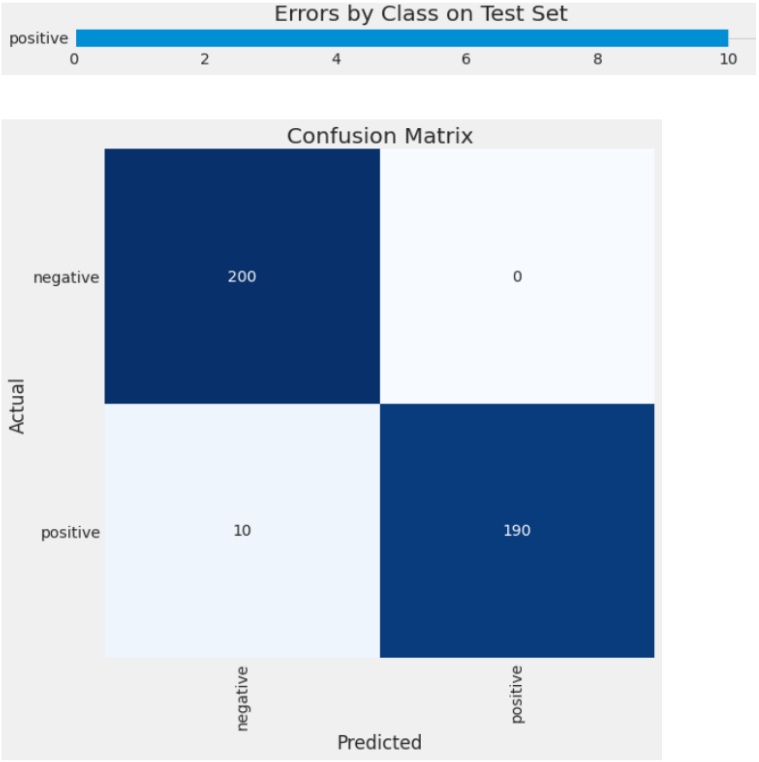


Figure 3.9: Confusion Matrix

```
Classification Report:
----------------------
              precision    recall  f1-score   support

    negative       0.95      1.00      0.98       200
    positive       1.00      0.95      0.97       200

    accuracy                           0.97       400
   macro avg       0.98      0.97      0.97       400
weighted avg       0.98      0.97      0.97       400
```

Figure 3.10: Final Report

In the above matrix we have predicted on the x-axis and actual on the y-axis ,from this matrix we can tell the no.of negative and positive cases predicted and the no.of negative and positive cases in actual.

Chapter 4

CONCLUSION

This report provides a critical analysis for 3 CNN-Architectures, proposed originally for image analysis. These CNN-Architectures are used to differentiate COVID-19 disease vs healthy based on chest X-ray images. We also proposed a simple CNN architecture but gave us low accuracy when compared to other standard CNN architectures it was observed that a learning rate of 0.01 achieved good accuracy the number of epochs is restricted to 3 as any other value greater than that over fitting is observed. And the analysis on the COVID-19 covering what age groups are mostly getting affected, Major symptoms, state-wise analysis with cure rates, and death rates. The result of the CNN that we built from scratch and compared with other deep learning CNN models such as VGG16, and Inception-V3. The pre-trained models came up with higher accuracy when compared to the model that we have built from scratch. VGG-16 and Inception-V3showed the same metrics.

4.1 ADVANTAGES

1. The amount of time required for the result and the cost is less when compared to manual testing.
2. And the limited availability of test kits.
3. And can be used for dierent x-ray models for classification.

4.1.1 LIMITATIONS

1. Number of x-rays we have are really less. For better accuracy we need more images.
2. And one more important limitation is that not every persons lungs get affected by Covid-19.

4.1.2 FUTURE SCOPE

The future work may be:

1. Removing other unnecessary noise in the image such as text writing and medical devices marked on chest X-rays for a better vision and understanding.
2. And the collection of more COVID Images will improve the models accuracy.

Chapter 5

REFERENCES

1. Deep learning with François Chollet
2. Taban Majeed, Rasber Rashid, Dashti Ali, and Aras Asaad, "Problems of Deploying CNN Transfer Learning to Detect COVID-19 from Chest X-rays", May 19, 2020.
3. C. Szegedy et al."Going deeper with convolutions", in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, vol.07-12-June-2015, pp. 1–9
4. T. Ai et al,"Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease(COVID-19) in China: A Report of 1014 Cases", Radiology, p. 200642, Feb. 2020 . Simonyan and A. Zisserman,"Very Deep Convolutional Networks for Large- ScaleImage Recognition", 2015.
5. H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. Khan,"Diagnosing COVID-19 Pneumonia from X-Ray and CT Images using Deep Learning and Transfer Learning Algorithms", arXiv, Mar. 2020.