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

by

P Madhu Chandra

2018IMT-049

ABV–INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT GWALIOR (M.P.)-(474015)

2021

Outline

- Introduction
- Explanation of:
  - Project Objectives and deliverables
  - Limitations of Manual testing
  - System Architecture/ Methodology
  - Convolution operation
  - Pooling
  - Flattening
  - Full Connection
- Salient Features of CNN
- Data Preprocessing
  - Splitting of Data into Train, Test, validation, and Data Augmentation
- Results
- Conclusion
- References

Introduction

- Covid-19 It was first identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic. 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.
- 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 to be like a crown. The word Corona means "crown" in Latin.

Project Objectives

- The objective of the project is to do an analysis on COVID-19 this Analysis helps in understanding and taking some precautions of the disease as the Analysis provides the significant symptoms of COVID-19
- Later develop and train a deep learning model on the chest X-Ray dataset to classify the given x-ray images of affected patients vs healthy people. And train some pre-trained models like Inception[3] and VGG-16[5] and compare the accuracies of the different model.

Project Objective Cont...

- Limitations of Manual testing
  1. The amount of time required for the result and the cost for manual testing.
  2. And the limited availability of test kits.

Background Information/Motivation/Literature Survey

1. 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.
2. 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).
3. 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.

Literature Survey

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.

Figure: Literature Survey

Methodology

The Methodology includes 4 steps [1]

- Convolution operation
- Pooling
- Flattening
- Full Connection

Methodology Cont...
Convolution

- 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[1].
- Feature Detector or filter is 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[1].

Methodology Cont...
Convolution

- This is called Feature map and sometimes called convold feature
- 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.
- 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[1].

Methodology Cont...
Max Pooling

- we take the feature map from the Convolution step. And apply max pooling on it.
- we take a 2*2 box and slide it and take the max value because we already know that the max value in the feature map represents a feature

Methodology Cont...
Max Pooling

- so we just take the max from the stride so we take the prominent features at the same time we reduced size
- There's several different types of pooling methods like mean pooling,Max pooling,Global pooling[1].

Methodology Cont...
Flattening

- We just flatten the max pooling matrix into a vector so that we can use an input for the neural network latter on for classification[1].

Methodology Cont...
Full Connection

- 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.
- Back propagation plays a key role in the changing the weights accordingly to reduce the loss[1].

Salient Features

- After learning a certain pattern in a picture, a convnet can recognize it anywhere in the image.
- A densely connected neural network would have to learn the pattern anew if it appeared at a new location.
- This makes convnets data efficient when processing images they need fewer training samples to learn representations that have generalization power[1].
- The first convolution layer will learn small local patterns,a second convolution layer with max pool and non linearity will learn larger patterns made of the features of the first layers, and so on. This allows convnets to efficiently learn all the features in the image[1].

System Architecture

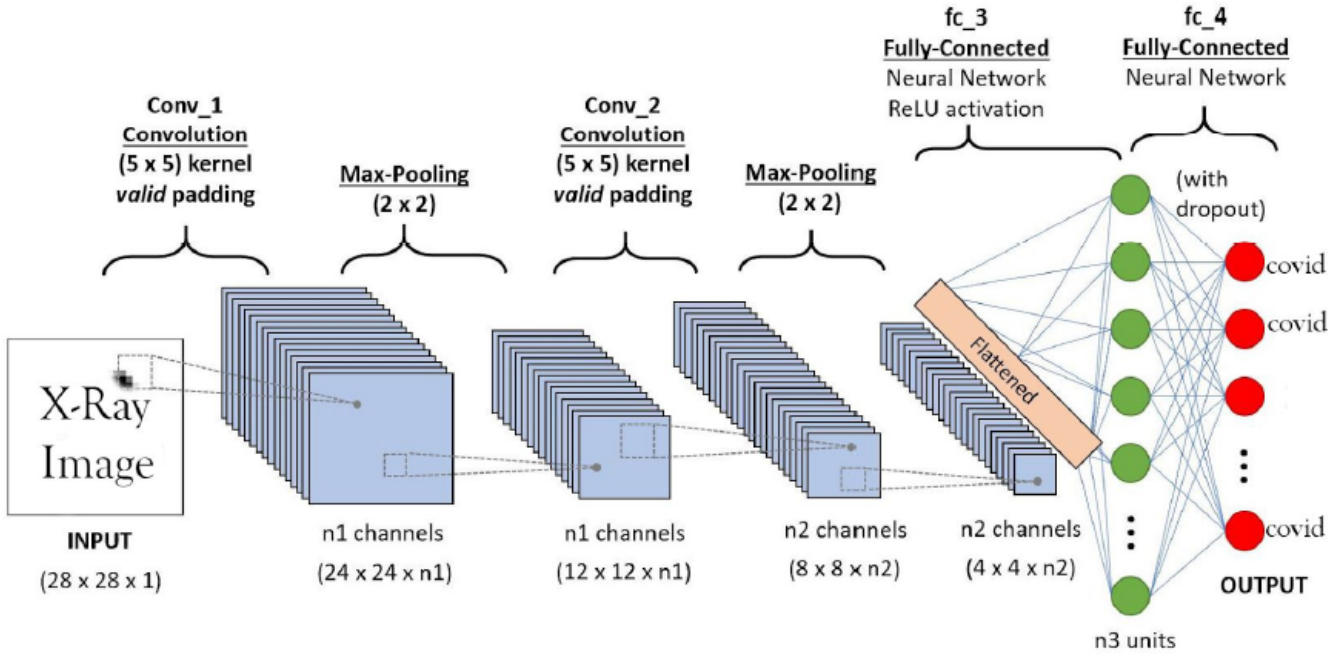

Figure: Image from towards data science

Data Preprocessing

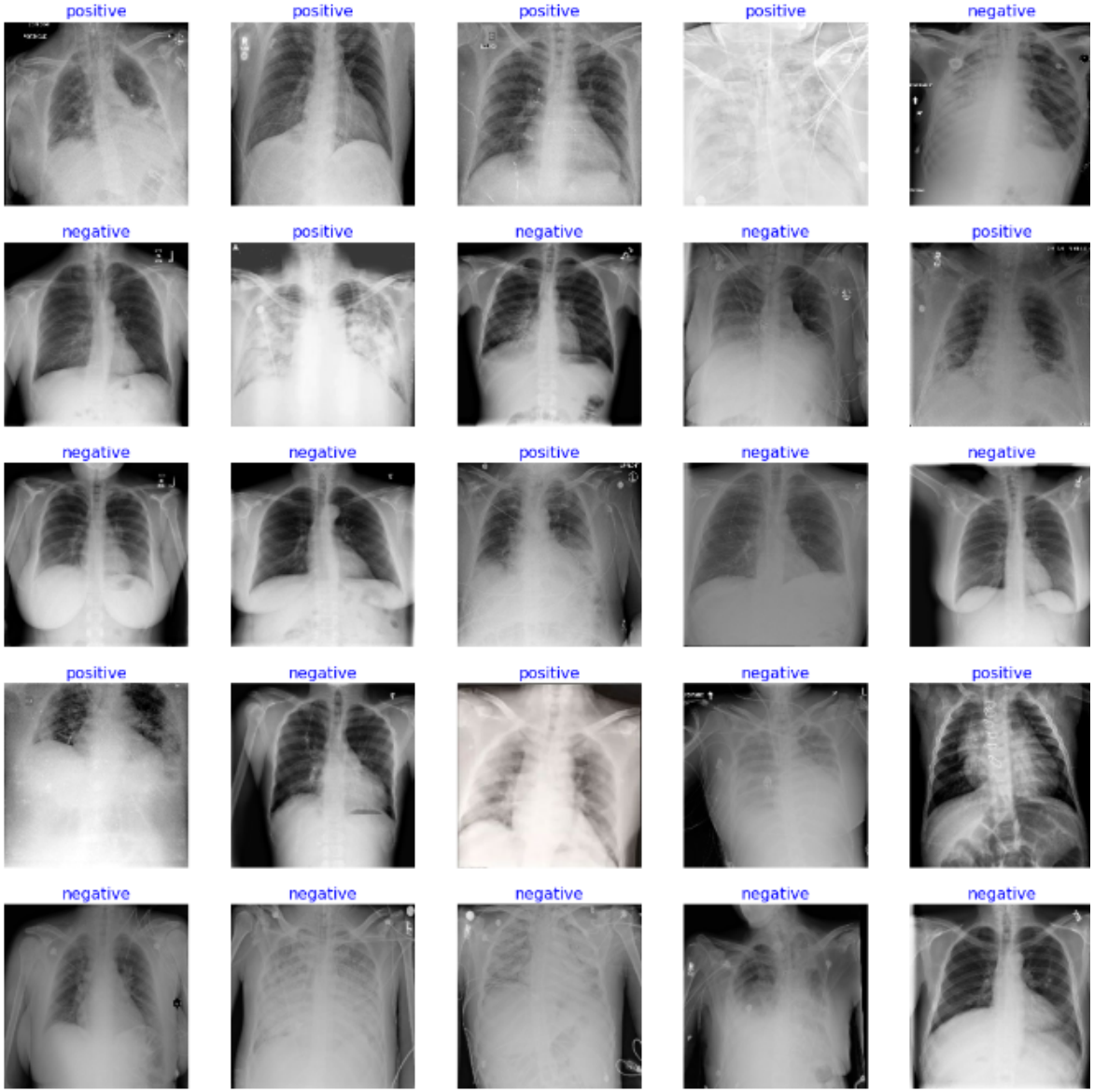

Figure: Covid and Non-Covid Lung X-ray

Data Preprocessing cont...

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: 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.

Data Preprocessing Cont...
Splitting of Data into Train, Test, validation, and Data Augmentation

- Total data in the datasets are divided into Test, Train and validation data accordingly.
- Train Data: 80 percent
- Validation Data: 10 percent
- Test Data: 10 percent
- 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

Experimental Analysis and Results

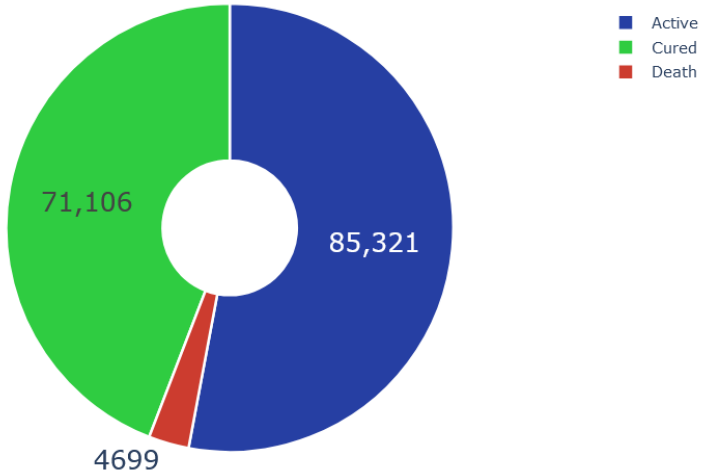

Figure: Current situation in India as of May 2020

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 .

Experimental Analysis and Results cont...

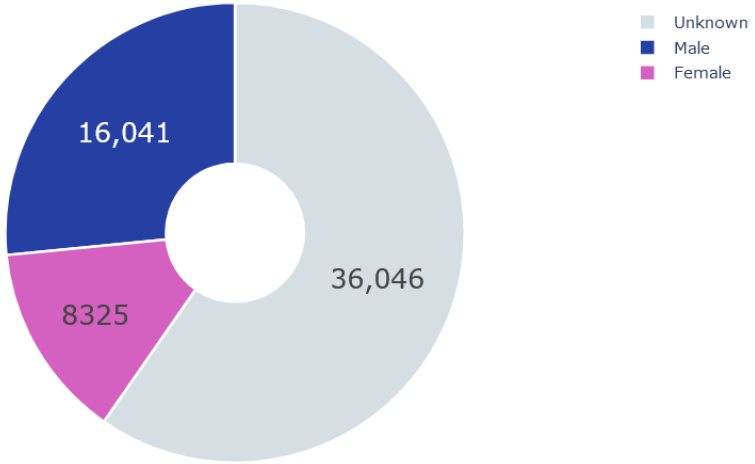

Figure: Gender wise active cases

- 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.

Experimental Analysis and Results cont...

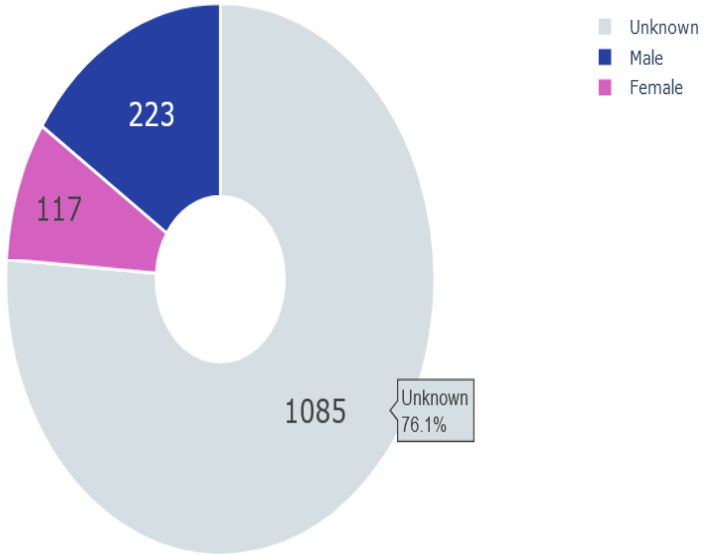

Figure: Gender wise Deaths

- 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.
- Neglecting the missing data Men are the most affected compared to women.

Experimental Analysis and Results cont...

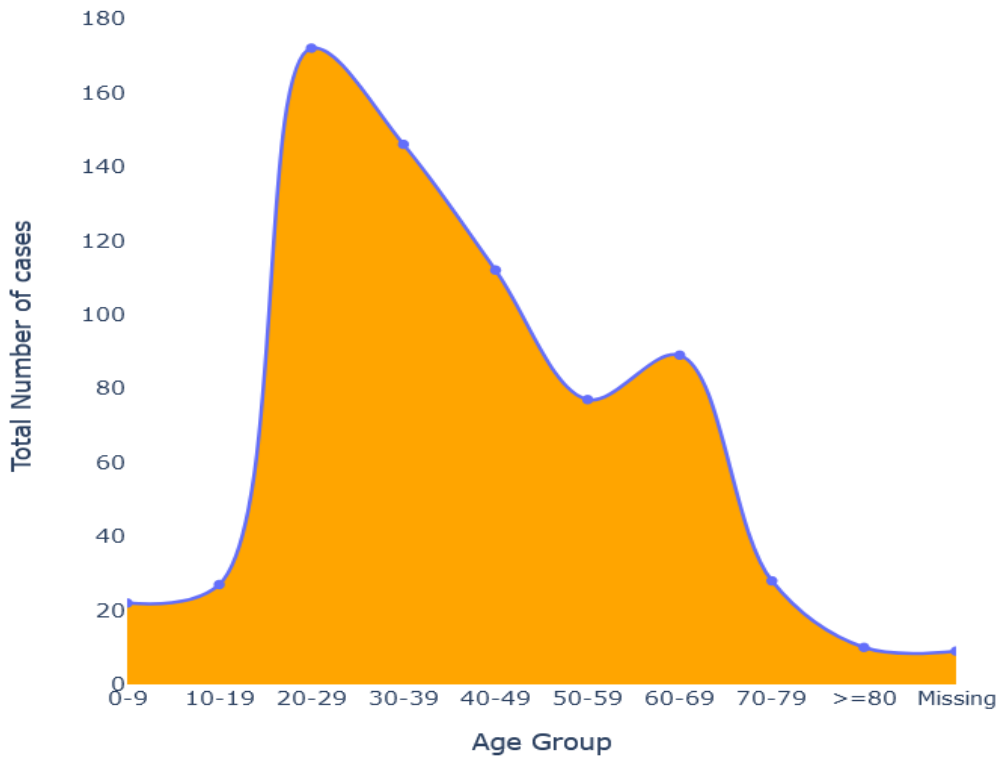

Figure: Age wise Analysis

- 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

Experimental Analysis and Results Cont...

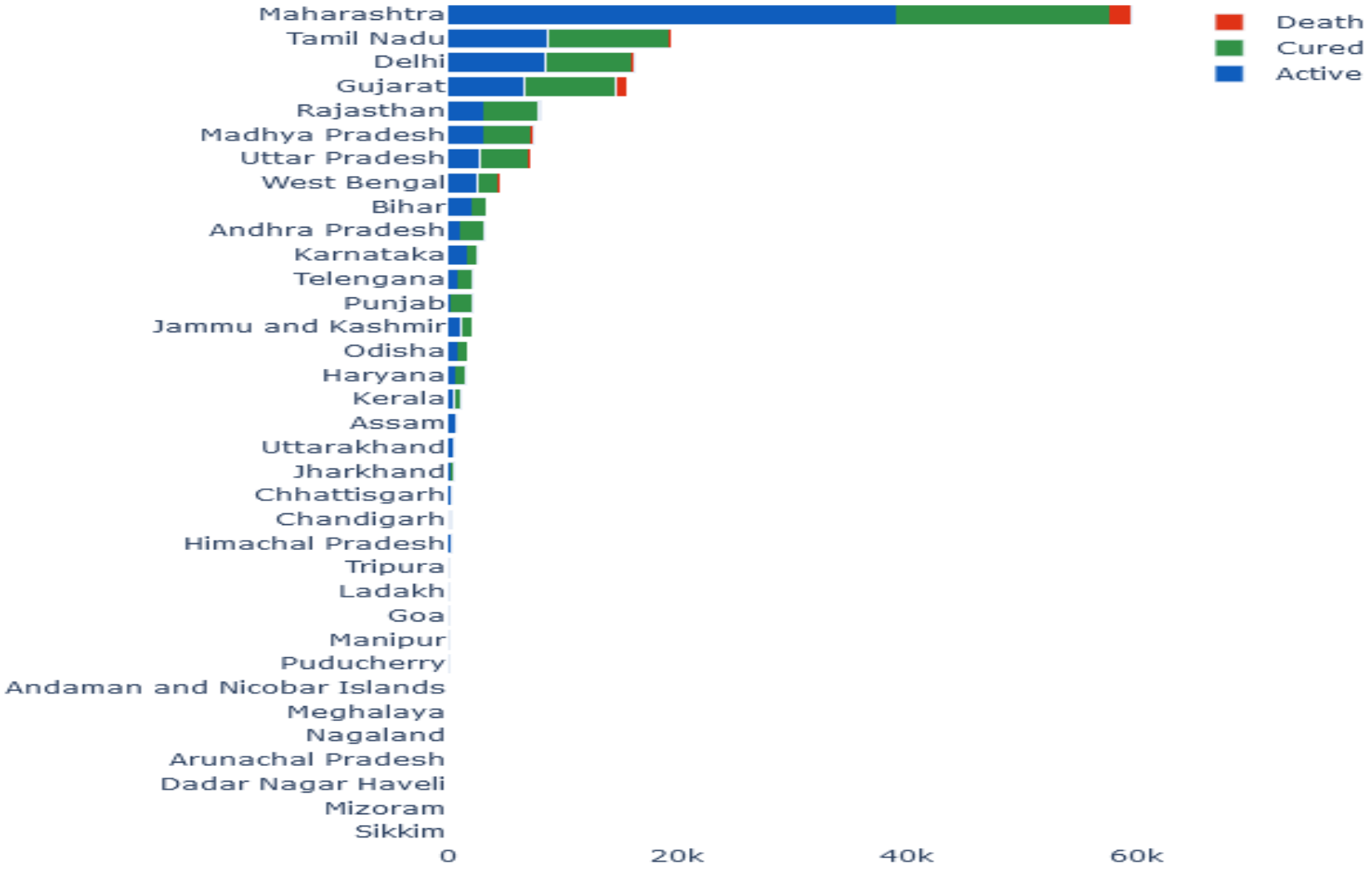

Figure: State wise Analysis

Experimental Analysis and Results Cont...

- 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 11,68,023 have recovered and 38,135 have died so far
- The most affected states according the analysis are in the order
  1. Maharashtra
  2. Tamil Nadu
  3. Delhi
  4. Gujarat

- These states mentioned above also have the most number of death rates. Makes up to 60 percent of the country's Covid-19 toll as you can see from above Data frame.

Experimental Analysis and Results Cont...

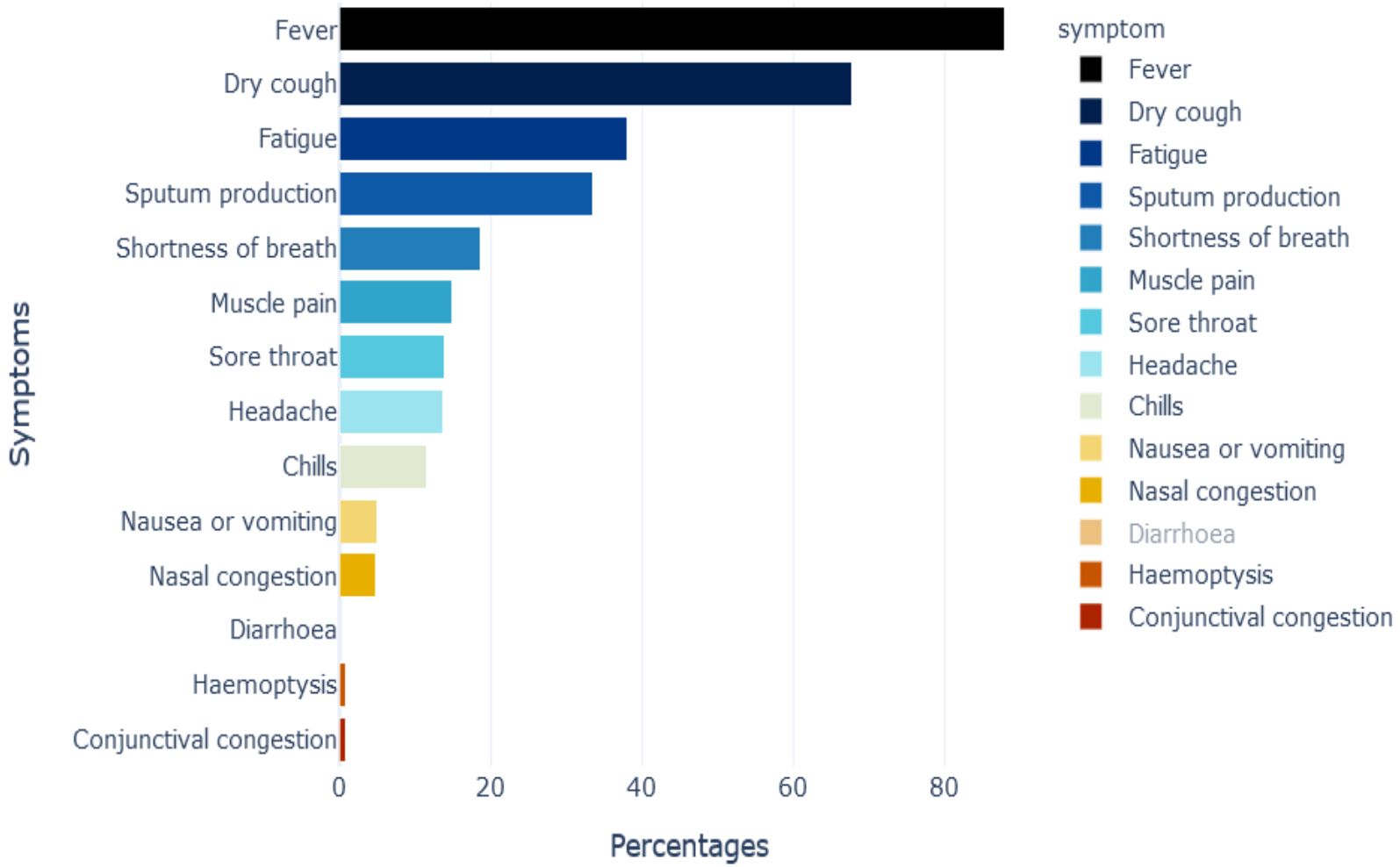

Figure: Major Symptoms

Experimental Analysis and Results Cont...

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.

Experimental Analysis and Results Cont...

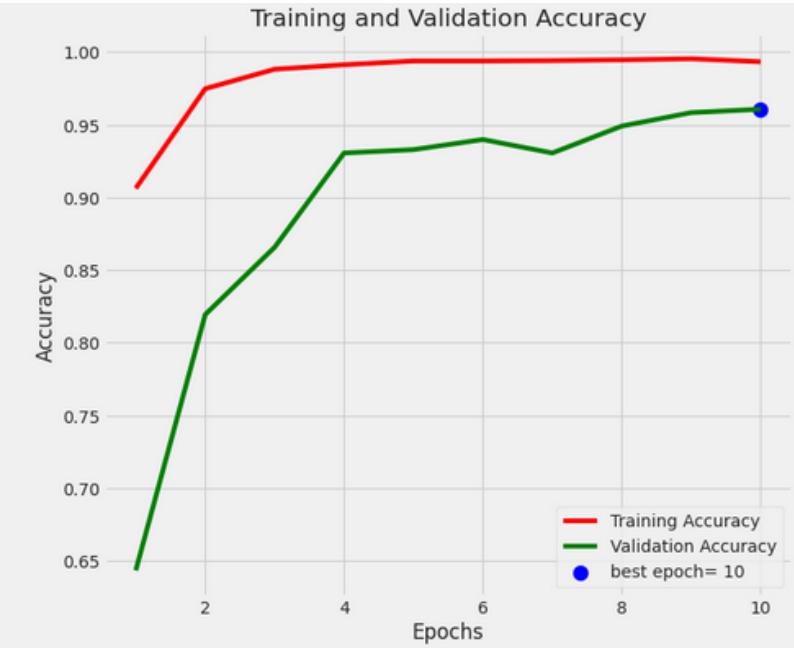

Figure: a)Train Accuracy vs Test Accuracy for VGG16 b)Train Loss vs Test Loss for VGG16

- These metrics are for 10 epochs and a learning rate of 0.001. If the number of epochs are increased overfitting is observed we used just 10 epochs as the available data is less.

Experimental Analysis and Results Cont...

- 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.
- 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.

Experimental Analysis and Results Cont...

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

Figure: Final Accuracies of Models

Note: These results are for 1000 normal x-ray images and 141 COVID-images these metrics may change if the number input images change.

Experimental Analysis and Results Cont...

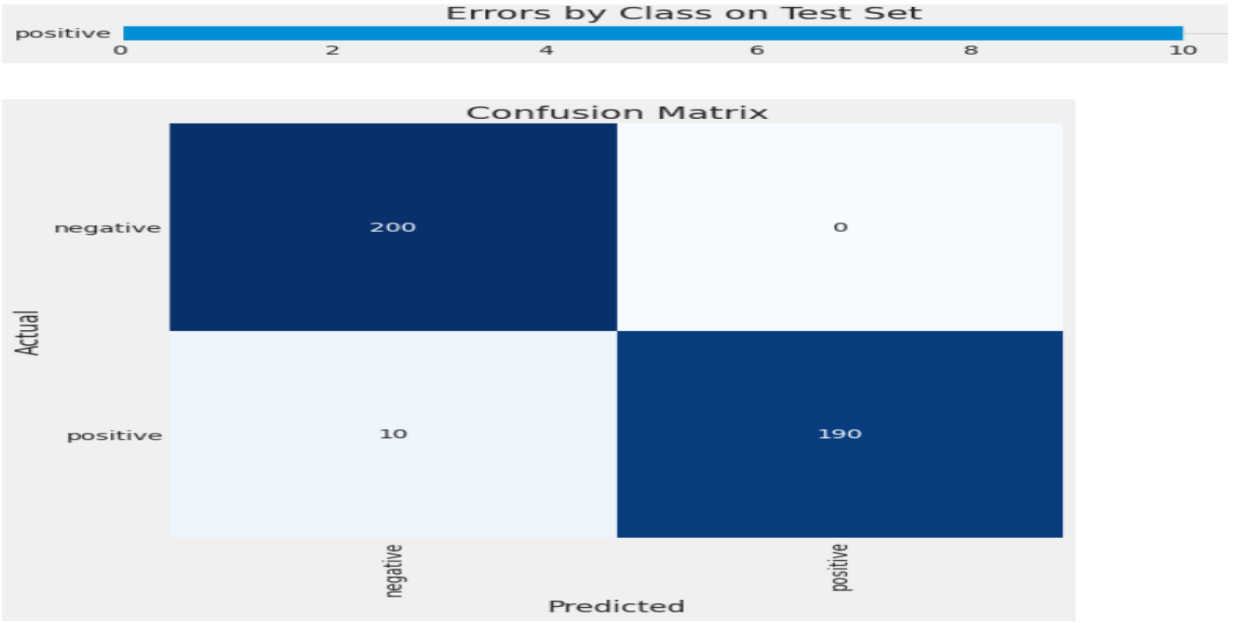

Figure: Confusion Matrix

```
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: Final Report

Conclusion

- This report provides a critical analysis for 3 CNN-Architectures, proposed originally for image analysis.
- We Proposed a simple CNN architecture[6] 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 overfitting is observed.
- The pre-trained models came up with higher accuracy when compared to the model that we have built from scratch. VGG-16[5] and InceptionV3[3] showed the same metrics.
- 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.

Future Scope

- Works in the future may include 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.
- And the collection of more COVID Images will improve the models accuracy.

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
5. K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015
6. 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.

Thank You