

# PANDEMIC DATA SCIENCE

G B Harsha Vardhan T Rohith Kumar M Rithik  
19110083 19110103 19110098

**Abstract** – The study of Pandemic data is very vital to manage the world at the time of crisis. In the following sections of the paper we will be discussing about statistical information and present some of the visualizations of the available data. It is important to understand and apply modelling to the data. It will help in making predictions for the future and understand how each parameter is affecting the situation. We will also discuss the predictive models we have built. We will also see the analysis on vaccination and how it is making the situation better.

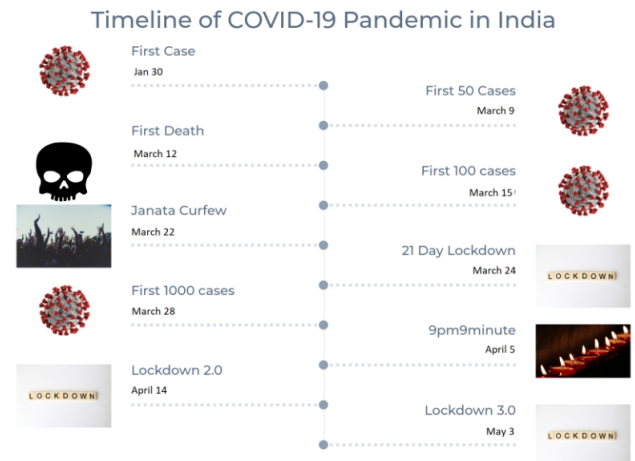


Fig. 1. Timeline of covid 19 pandemic in India

First lockdown was on march 22 and then followed by a 21 day lockdown from April 14. Now let us analyse and visualize the Covid 19 data that has been collected over the past 2 years.

## I. Introduction

Covid 19 impacted the functioning of the world dynamics. It changed the world upside down, and going back to pre covid 19 times will mostly probably cannot be seen in the near future. Covid 19 pandemic impacted the world Socially, Economically, Regionally, and Financially in a very catastrophic way by taking an enormous toll on humanity disturbing lives and livelihoods in a great deal. Covid 19 affected almost every country with some precedence. The first case of novel corona virus was reported by Wuhan municipal health centre on 31 December. On 12 January 2020 China officially shared the genetic sequence of Covid 19. World has seen 3 waves of Covid 19 and it has left its trail on everyone. First Case in India was reported on 30 January 2020 whereas a total of 7818 cases have been reported till 30 January 2020. Let us see the timeline of series of events during First wave of Covid 19 in India.<sup>1</sup>

<sup>1</sup><https://www.kaggle.com/code/nitishabharathi/the-story-of-covid-19-in-india-eda-and-prediction/notebook>

## II. Dataset

We are working the covid data set which consists of different forms of data involving daily cases, cumulative cases, deaths, state wise cases information, health facilities, resources. There are data sets on daily vaccinations containing about the doses administered, statewide vaccination doses etc.

## III. Data Exploration and Health Metrics

We have collected data from the official website of **GOVERNMENT OF INDIA**. Humanitarian and development organisations are encouraged to close data gaps by sharing data that is critical but often missing. We interpreted and analysed the data with various metrics so that these might be useful to prepare ourselves better next time while battling several other deadly viruses. COVID-19 has not left any part of the world untouched and India is one of the worst affected countries in the world. The Second wave was more dangerous and severe than the first one both economically and socially. The cases in the April 2021 were at all time high in number. Over the past couple of years, millions of people affected by Covid 19, thousands of people died and created a havoc throughout the country. Let us now

discuss some the trends and interesting plots to understand what happened in the past 2 years and how it started in the beginning.

#### A. Total Number of Cases v/s Total Deaths in India

Below is the plot for Active cases and Deaths occurred during both first and second wave.

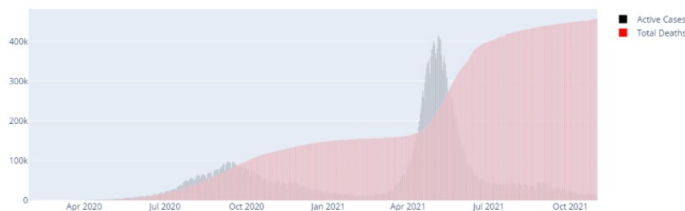


Fig. 2. Plot to compare cases versus deaths

First case was registered in india on January 30 2020. As we can infer from the above plot, the cases kept on increasing linearly from June and reached its peak at mid september and then the cases decreased gradually. One important point here to notice is the total number of deaths is postively correlated with the total number of cases. **As the total number of cases increase the slope of the increase in total deaths also increased too. This is quite obvious yet an important trend to study and understand.**

#### B. COVID Statistics

In this section Let us analyse the total number of tests,confirmed cases,deceased people,recovered people during the first and second wave of covid.

**1) Total number of tests performed :** Below is the plot for the top ten Covid testing states in India.

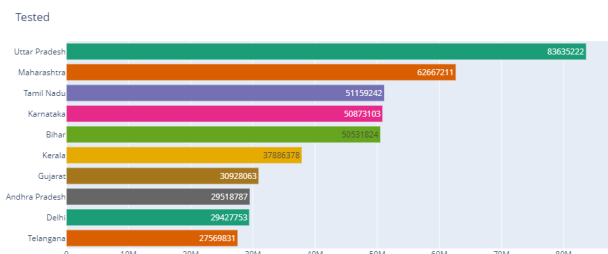


Fig. 3. Total number of Tests performed of top 10 states in India

Uttar Pradesh tops the list.A total of 83 million tests were conducted in Uttar Pradesh. Uttar Pradesh also holds the record for testing 3.7 lakh people in one day. Then followed by

Maharashtra which has 62 million number of total tests.<sup>2</sup> We can infer more information regarding these statistics through the graphs plotted in upcoming subsections.

**2) Total number of Confirmed cases:** Below is the plot for the top ten states of confirmed cases in India.

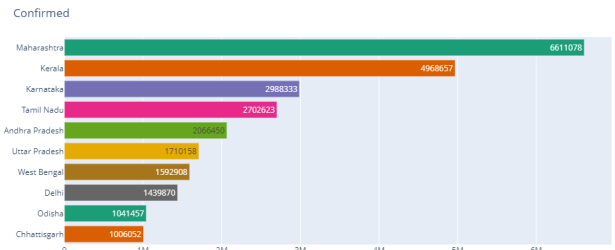


Fig. 4. Total number of confirmed cases of top 10 states in India

Maharashtra tops the list. Initially during the first wave kerala recorded the highest single day cases in India but eventually Maharashtra ended being on top. Maharashtra recorded 6.6 million cases. Even though the cases were highest in Maharashtra it's not a bad sign because it's second most populous state in India .The number of active Covid-19 cases per million in the state were 2087 whereas Karnataka has 4381 ,Tamil Nadu 2365 , Kerala 5062.

**3) Total Deaths:** Below is the plot for the top ten states of highest number of deaths in India.

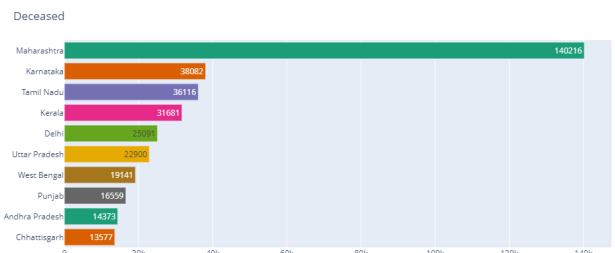


Fig. 5. Total number of deaths of top 10 states in India

Maharashtra has highest number of deaths and it scales up to 140 thousand.Karantaka and Tamil Nadu surpassed Kerala in total deaths even though kerala recorded more number of cases than them. That is where mortality rate comes into the picture and the following subtopic dwells into it more thoroughly.

**4) Mortality Rate:** Below is the plot for the top ten states of mortality rate in India. In our case, **we measured mortality rate as the number of deaths per total number of cases.**

<sup>2</sup><https://www.kaggle.com/code/nitishabharathi/the-story-of-covid-19-in-india-eda-and-prediction/notebook>

This would help us analyse and interpret why the death percentage was high and where it was lacking. The Mortality rate is interpreted below.

Comparison of Top 10 Mortality Rate of States in India

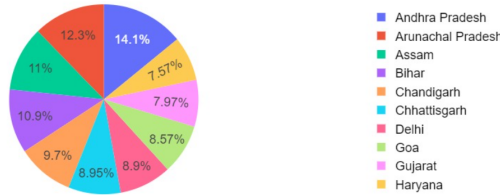


Fig. 6. Top 10 states mortality rate in India

From the above chart, we can infer that Andhra Pradesh has highest mortality rate and it is quite unusual because it is no where near top 3 statistics either in total number of cases or total number of deaths. The reason for this result is mortality rate depends how efficient the health care system is, how healthy the people of the state are and many more factors come into play. Further scope of information regarding Health Care system in India will be analysed in following sections of the paper.

### C. Health care system in India

COVID 19 pandemic put a significant stress on our already limited healthcare resources. The covid cases increased exponentially over the time period and we were not at all prepared for it during the first phase. Currently, 60 percent of the total of India's population lives in rural India.

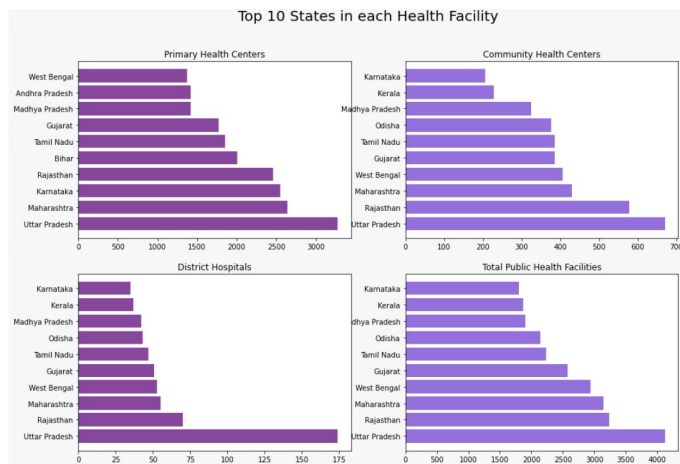


Fig. 7. Different Health care facilities in India

To provide healthcare facilities to the people living in rural India, the government has established 25743 Primary Health Centers, 158417 Sub Centers, and 5624 Community Health Centers. Currently, 713986 beds are available in government hospitals in India which amounts to 0.55 beds per 1000

population. From fig7, we can interpret and correlate the following points.

1. Karnataka has more community Health centres, District Hospitals and also total public Health centres. This explains why karnataka has low mortality rate. Kerala is also an another good example for the above analysis. **Thus, Healthcare facilities in karnataka and kerala helped to reduce the mortality rate in the both states.**
2. Similarly, Andhra Pradesh has highest mortality rate and this fact alligns with our health care plot too. Andhra pradesh only has second highest number of P.H.C's but it is no where near the top 10 other health care facilities like District Hospitals, Community Health centres, and total public health centres. **Thus we can infer from above analyses that A.P has lot of scope in development of fundamental Health care system.**

### IV. Predictive Modelling for Covid Detection

We have developed many models that help identify the trends and patterns in the data to help us predict further happenings and the results of some tests. This modelling is also applied in a wide range for making predictions regarding the Covid. In this paper, we are addressing two such successful models we were able to build on the available data.

#### A. Analysis of Chest X-Ray

Monitoring the severity of infection is so important once someone is tested positive. One such way to monitor the seriousness is to analyse the chest X-Ray. The viral load is often associated with the chest volume that appears to be infected in the X-Ray. Also, a Chest X-Ray is used to confirm Covid after one tests Positive.



Fig. 8. X-ray of covid positive report



Fig. 9. X-ray of covid negative report

Here we can see the Chest X-rays of both positive and negative patients. The target of our model is to classify the image into positive or negative.

A Convolutional neural network is a Machine Learning technique to extract features out of an image and classify it based on the calculations it makes.

**We have implemented a CNN for this image processing task.**

The Data Set we used is the Covid Radiography dataset available on Kaggle. Here is the architecture and Parameters of our neural network.

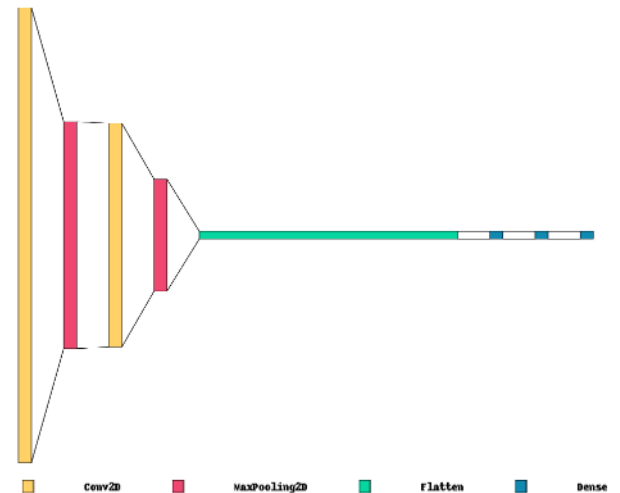


Fig. 10. CNN Architecture

\*Here are the details of our model. Here you can also understand the number of parameters for each layer.\*

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(1, 297, 297, 64)	640
max_pooling2d (MaxPooling2D)	(1, 148, 148, 64)	0
conv2d_1 (Conv2D)	(1, 146, 146, 32)	18464
max_pooling2d_1 (MaxPooling2D)	(1, 73, 73, 32)	0
flatten (Flatten)	(1, 170528)	0
dense (Dense)	(1, 64)	10913856
dense_1 (Dense)	(1, 32)	2080
dense_2 (Dense)	(1, 2)	66

Total params: 10,935,106  
Trainable params: 10,935,106  
Non-trainable params: 0

Fig. 11. Total Parameters

Here are the training epochs of our model and the accuracy in each trail

Here are the training epochs of our model and the accuracy in each trail:-

```
[Epoch 1/3] - 147s 163ms/step - loss: 0.6071 - accuracy: 0.7711 - val_loss: 0.5343 - val_accuracy: 0.8251
[Epoch 2/3] - 153s 170ms/step - loss: 0.4316 - accuracy: 0.8922 - val_loss: 0.4524 - val_accuracy: 0.9400
[Epoch 3/3] - 162s 181ms/step - loss: 0.3487 - accuracy: 0.9800 - val_loss: 0.4057 - val_accuracy: 0.9513
```

Fig. 12. Training Epochs

**We have achieved 90 percent accuracy with 900 train images and 400 test images. This Accuracy is great given the size of the data set considered.**

## V. Forecasting the Growth of Covid Cases

Linear regression is a technique to try and fit the train data into a linear plot using the loss function as mean square error. There are many variations of Regression. **We have used Linear Regression on the features corresponding to a polynomial of degree Four of the existing features.** We have tried forecasting and projecting the expected cases for different periods. Here are the results and the metrics of this prediction

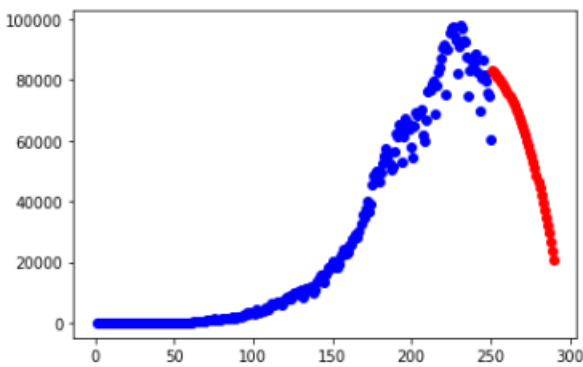


Fig. 13. Forecasting in Situation-I

### Results:-

1. RMSE= 14505.82
2. The average of true values in the predicted range= 52283.73
3. The Percentage of error taking rmse and avg of true values as error and true value= 27.75

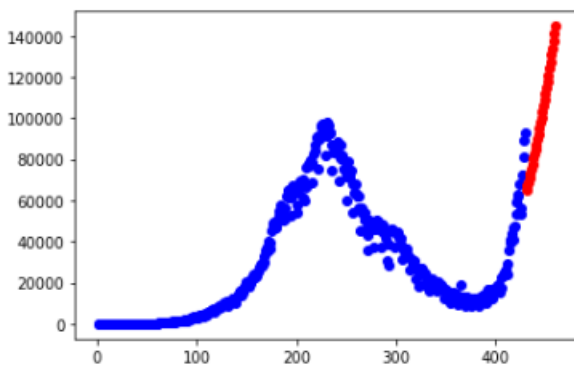


Fig. 14. Forecasting in Situation-II

### Results

1. RMSE= 184019.47
2. The average of true values in the predicted range= 269013.93
3. The Percentage of error taking rmse and avg of true values as error and true value= 68.41

**As we can see, we are getting high RMSE values compared to the average of true values.** This is not a completely desired

characteristic of a model. But we have to also to balance the degree to make sure we don't overfit the data and not very high computational power is used.

The positive side of this model is that it is working effectively to predict the further peaks while the start and the settling of the peak as they start to weaken. We can see this in the images above.

## VI. Vaccination Drive in India

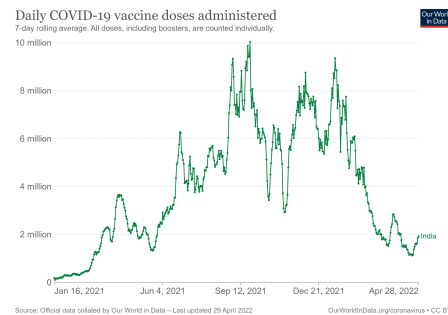


Fig. 15. Daily covid vaccination doses in in India

India consists of a huge 17.7 percent(1.39 billion) of the world's population, and vaccination production has not been easy with this population. India started the first phase of the vaccination drive on January 21, and all the people above 18 were given permission for the vaccination on July 21. The main task of the Indian government is to manage the vaccine distribution to all the regions in the country. India is currently administering two locally-manufactured vaccines, Covishield and Covaxin, and the Russian vaccine, Sputnik. There were three waves of covid, and coordinating the vaccine drive, and all the covid protocols and cases is a very tedious and complicated process.

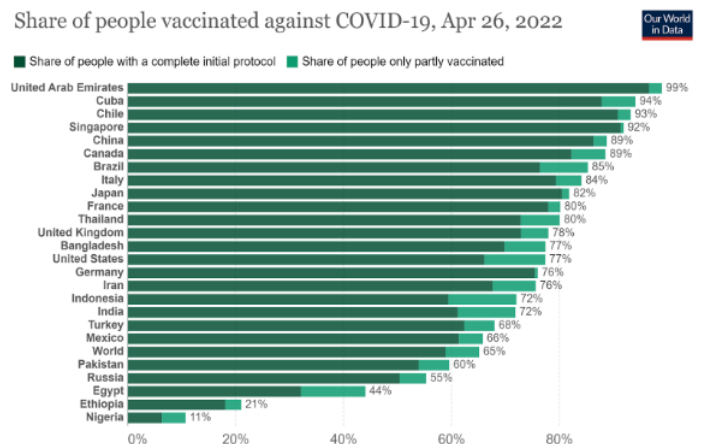


Fig. 16. Percentage of people vaccinated across the world

In the above plot fig[16], we can see India's current situation



India among all other countries in the vaccination against the corona virus. It is in the 18th position with around 62 percent population fully dosed and 72 percent of the population with atleast one dose.

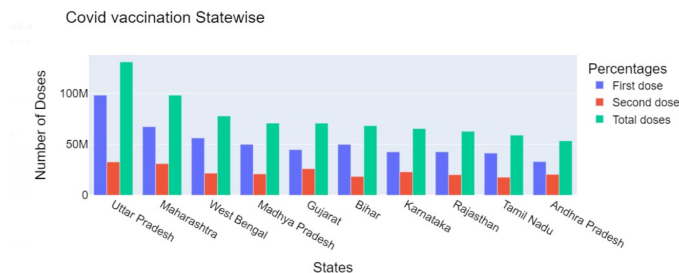


Fig. 17. State wise covid vaccination drive

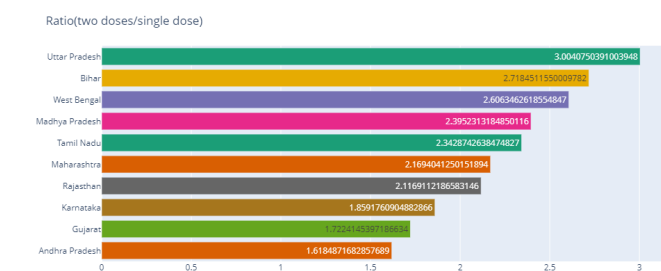


Fig. 18. State wise covid vaccination dose ratio(top10)

Among all the states, Over 152 million first doses and over 116 million second doses were administered in Uttar Pradesh, crossing all other states in the Vaccination race, and Maharashtra and West Bengal are following it(fig[17]).The ratio of people vaccinated with one dose compared to the fully vaccinated ins very high.

Demographic plot of India for the vaccination drive can be observed in fig[19]. We can see that the state Himachal Pradesh ,Jammu Kashmir and Andhra Pradesh are vaccinate above 80 percent of their population.The distribution rates in these states is very high compared to the other states.Most numbers of doses are used by Uttar Pradesh and there are very less vaccines administered in North eastern states.

**One of the reason for this is the vaccination hesitancy among the tribal population of the state and lack of proper awareness to the common people.**

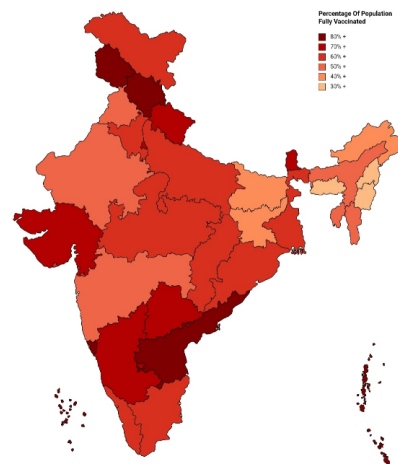


Fig. 19. Demographic plot of fully vaccinated people

**Vaccine hesitancy is a widespread challenge in India,** ignited by misinformation and lack of trust, particularly in rural areas where **65.5 percent** of the population resides has affected many regions in India during the second wave. Many deaths happened during the second wave, and one of the important reasons is the various obstacles in the vaccine distribution process and the high virality of the delta Variant. There were many restrictions on travel and lockdowns, but they still couldn't control the spread, and it shattered the entire country.

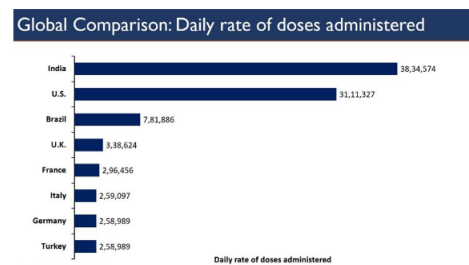


Fig. 20. Daily rates of Dose administered

We could not avoid the pandemic even with the highest numbers of vaccinations administered per day in the world during the second wave. Due to high population,even small movements are causing super spread of the virus and gathering in the vaccine centres is also major reason for the spread during the delta wave.The approach of the vaccination drive through the states with a lack of other preventive protocols has caused the situation in the last summer around April. **It is estimated that with the current rate of the vaccination drive and upcoming waves, herd immunity can be appeared in 2024.**

## VII. Conclusion

Health care facilities are directly correlated with mortality rate. Uttar Pradesh and Rajasthan Health care facilities tried to curb the mortality rate in their states. We can conclude that Health care systems need to be improved significantly in many states like Andhra Pradesh, Arunachal Pradesh, Assam and many others. We have built two models for predictive analysis. The accuracy we have achieved in the CNN we built was around 90 Percent. Such accuracies would make us feel more confident about such models. The polynomial regressor model we have built can help in detecting the future peaks and can help in predicting when a wave may settle. Applying such modelling to all kinds of data would help in better assessing the conditions and make better decisions. The Indian government must develop efficient public health measures for mass immunisation and avoid crowding people into healthcare facilities for vaccination. To minimise superspreading and to aid in the nation's mass vaccination efforts, many states have implemented door-to-door immunisation. In light of the new SARS-CoV-2 variations, the government should concentrate on maintaining a high vaccination rate and coverage of double dose of COVID-19 vaccine and shorter intervals between doses in order to attain high efficacy rates across India's whole population.

## VIII. REFERENCES

- [1] World Health Organization. 2020. "Coronavirus Disease (COVID-19) Pandemic." World Health Organization. 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- [2] Rajan Gupta, Saibal K. Pal, Gaurav Pandey. "A Comprehensive Analysis of COVID-19 Outbreak situation in India", <https://doi.org/10.1101/2020.04.08.20058347>
- [3] John Hopkins University (2020). Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE. Accessed from <https://github.com/CSSEGISandData/COVID-19> on 6th April 2020. Sharma, N. (2020).
- [4] Sharma, Niharika. n.d. "India's Swiftness in Dealing with Covid-19 Will Decide the World's Future, Says WHO." Quartz. Accessed April 30, 2022. <https://qz.com/india/1824041/who-says-indias-action-on-coronavirus-critical-for-the-world/>.
- [5] S V Subramanian, India faces a challenge with its mass vaccination efforts, The Lancet Global Health, Volume 9, Issue 9, 2021, Pages e1201-e1202, 2214-109X, [https://doi.org/10.1016/S2214-109X\(21\)00260-6](https://doi.org/10.1016/S2214-109X(21)00260-6).
- [6] Roser, Max, and Hannah Ritchie. 2020. "Coronavirus Disease (COVID-19)." Our World in Data, March. <https://ourworldindata.org/coronavirus>.
- [7] Dr. Abhishek, Singh, Deedwania Dr. Preeti, K Dr. Vinay, Chowdhury Dr. Apala Roy, and Khanna Dr. Puneet. 2020. "Is India's Health Care Infrastructure Sufficient for Handling COVID 19 Pandemic?" International Archives of Public Health and Community Medicine 4 (2). <https://doi.org/10.23937/2643-4512/1710041>.
- [8] "The Story of COVID-19 in India EDA and Prediction." n.d. Kaggle.com. <https://www.kaggle.com/code/nitishabharathi/the-story-of-covid-19-in-india-eda-and-prediction/notebook>.
- [9] "COVID-19 Radiography Database." n.d. Wwww.kaggle.com. <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>.

**Contributions:-** Equal Contribution by all the members.