 doi:10.20944/preprints202006.0214.v1

COVID-19 India prediction model based on the trend from other countries

Satyaprakash* and Pinakana Sai Deepak
Department of Civil Engineering, School of Engineering and Technology,
Sharda University, Greater Noida, UP. INDIA
satya.prakash@sharda.ac.in

Disclaimer: The study is an academic exercise to understand how the COVID-19 pandemic may evolve in the coming months in India. The study may or may not hold as there are several variables, known and unknown, which affect the spread of the disease.

Abstract
This paper is an attempt to present a COVID-19 prediction model for India. Lockdown plays an important role in the arrest of community spread of the disease. This was evident from the study of other countries such as Russia, Belgium and Germany, where peak cases were recorded within a month of the imposition of lockdown, that it showed an immediate positive effect. However, in India, even after 65 days of lockdown, there is no decrease in the number of daily new cases reported. There were many models prepared for India and almost all of them were proven wrong by the increase in the number of cases. The model in this paper is prepared using the COVID-19 trend in other countries, population density and the pandemic bell curve. Based on the available data until 24th May 2020, two scenarios have been presented. In one, the peak shall be obtained when the number of daily new cases per million reaches 190 and in the second when the daily new cases per million reach 724. One model predicts the number of cases to reach 1 million by mid-July 2020. The other model predicts the number of cases to peak by mid-July with the total cases reaching 20 million. The predicted cases were compared with the actual cases recorded for the period 25th May to 11th June 2020. It was observed that the actual values matched quite reasonably with the predicted values.

Keywords: COVID-19, Prediction model, Pandemic bell curve, India, Different scenarios

1. Introduction

Since the first case of COVID-19 was reported from Wuhan, China, there have been many attempts to predict the spread of the disease through modelling the previous data and other factors. In India, the first case was reported on 30th Jan 2020 in Kerala and since then the cases have reached more than 0.2 million as of now on 3rd June 2020, after 5 months. Some attempts have been made to predict the number of cases, but most of the models could not correctly predict the total number of infected people. A study by Singapore University of Technology (Luo, 2020) predicted that in India around 97% cases shall end by 21st May. IIT, Guwahati in association with Duke-NUS, Singapore (Ghosh at al., 2020), analysed state-wise COVID-19 situations in states like Andhra Pradesh, Delhi, Gujarat, Madhya Pradesh, Maharashtra, Uttar Pradesh and West Bengal. The researchers used the data available until 1st May for analysis. However, they did not predict the overall cases in the country.

1.1. Prediction models for India

Several other authors who have also tried to model the outbreak of COVID-19 in India. Ranjan, 2020, predicted in March that based on the SIR model, it is estimated that India will enter equilibrium by the end of May with the final number of cases being approximately 13,000. However, he reported that this estimation will be invalid if India enters the stage of community

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

transmission. Although no community transmission has been reported in India yet, the predicted numbers have been greatly exceeded. There have been a considerable number of studies published in April which also predicted the number of cases and the time when the curve would flatten in India. Bhatnagar, 2020, and Arti, 2020 do not predict the number of people infected in India. They conclude that the lockdown and social distancing plays an important role in restricting the spread of the disease. However, Bhatnagar's model predicts the cases of Italy and France quite accurately. Tiwari et al., 2020 used the time series forecasting method and predicts that the total number of confirmed cases of COVID-19 might reach around 68978, and the numbers of deaths due to 1557 around 25th Apr 2020, in India. Bhattacharjee et al., 2020 used the cumulative NCV confirmed cases; recovery cases and deaths for estimating recovery rate, caseload rate and death rate till 24th April 2020. They predicted that by 20th May 2020, the caseload rate would be lesser than the recovery rate and thereafter the number of COVID-19 patients would start reducing.

Rajesh et al., 2020 used the mathematical model SIR(D) to predict the future of the epidemic in India by using existing data. The model shows that the epidemic will be at its peak around the end of June or the first week of July with almost 100 million Indians infected if the lock-down is relaxed after 3rd May 2020. However, the total size of the infected population will become one-third of this predicted number here, if people only in the red zones (approximately one-third of India's population) are susceptible to the infection. However, they expect that the numbers of infected people will at least be of the order of 10 million. Tomar and Gupta, 2020 used a data-driven model based on LSTM techniques. Their prediction of the number of positive and recovered cases (until April 2020) obtained by the model is accurate within a certain range.

However, these models could predict neither the number of cases accurately nor the date by which they would start decreasing. This is because the study of the spread is still in very early stages and there is no mathematical model that can predict how the spread would progress in India.

2. Materials and methods

2.1. The Proposed Model

Since all the proven mathematical and statistical models were not able to predict the cases accurately due to the limitations of the models and the uncertainty in the behaviour of the spread of the disease, the authors tried a different approach. We relied on studying the spread of the disease in other countries, the trend, number of cases at the peak, total cases at the peak and population density as the factor responsible for the spread of the disease. We also took into consideration that humans tend to behave in an almost similar manner in all places. The data for the study was compiled on 24th May 2020 and all the calculations were done taking this date as the baseline.

Table-1 shows the data from those countries where the pandemic has peaked and shown considerable reduction in cases as of now. We have not taken the data from China since it was the first country where the pandemic started and hence, the people and government were caught unawares which led to the escalation of the number of cases.

Table-1: COVID-19 SPREAD IN DIFFERENT COUNTRIES AND THEIR PEAKS

S No	Country	Total Cases	Daily New Cases	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

		(per million)		(per million)		Population Density (as in 2020) #
		At Peak	At present*	At Peak	At present*	
1	Italy	777	3806	99	5	206
2	Germany	607	2159	83	6	240
3	Belgium	788	4950	160	21	383

* as on May 24th, 2020
# source: worldometers.info

The authors have taken two scenarios: one with the daily new cases per million at the peak and the other with the total number of cases per million at the peak.

2.2. Scenario 1: (Daily new cases):

It is observed from the above table that there is a correlation between the population density and the daily new cases per million. The ratio of population density to the daily new cases per million at the peak was found to be 2.08, 2.89 and 2.4 for Italy, Germany and Brazil respectively (Table-2). Taking the average of these values as the ratio for India, the daily new cases would be around 190 per million, which is 5.1 at present.

Table-2: Relationship between new cases at peak and population density

S No	Country	Daily New Cases (per million)		Population Density (as in 2020) #
		At the peak	The ratio of population density and cases at the peak	
1	Italy	99	2.08	206
2	Germany	83	2.89	240
3	Belgium	160	2.4	383
4	India	190@	2.4&	464

&Average of all the countries studied and @Calculated at the average ratio

To arrive at the peak, the pattern of new cases per week in India was studied. Table-3 shows the average number of cases per million for every week since 3rd March 2020.

Table-3: COVID-19 cases in India during lockdown stages

Wk#	Date	Average new no. of cases in the week	Stages for study	Lockdown
1	4th MAR – 10TH MAR	7.28	***Ignored*** (Average case per week was 25)	NO
2	11th MAR – 17TH MAR	12.28		
3	18TH MAR – 24TH MAR	56.28		
4	25TH MAR – 31ST MAR	123	The average increase per week was 400 cases	#1 (25th Mar – 14th April)
5	1ST APR – 7TH APR	560		
6	8TH APR – 14TH APR	882		
7	15TH APR – 21ST APR	1227.5		#2 (15th Apr – 3rd May)
8	22ND APR – 28TH APR	1606.2		
9	29TH APR – 5TH MAY	2582		
10	6TH MAY – 12TH MAY	3556		

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

11	13TH MAY – 19TH MAY	4597	The average increase per week was 1000 cases	#3 (4th May – 17th May)
12	20TH MAY – 26TH MAY	6300		#4 (18th May - 31st May)

Although the first case in India was recorded in January, a major rise in new cases was noticed on March 4th when 23 cases were recorded. However, the number of cases per week till March 24th was negligible and has not been taken into account. Figure-1 shows the daily new cases for the period from 25th March till 11th June (the date of writing this paper) and the lockdown stages (different coloured bars). Since the rate of increase in cases is increasing every week (black vertical lines show the week), even during the lockdown, it was prudent to get the future weekly increase in the number of cases from the past weekly increase trend, for the study.

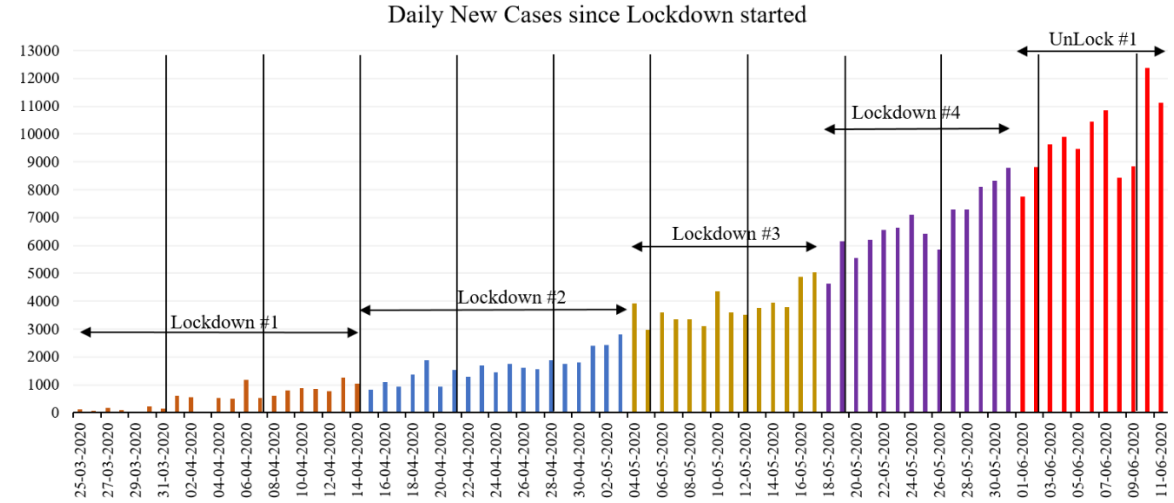

Fig-1: Daily new COVID-19 cases in India since the lockdown started

Following the above trend, and given that the restrictions imposed earlier have been eased now, (starting from Lockdown 3 and now after June 1, almost all activities have been permitted), we have assumed that the number of cases per week shall increase at a rate of almost 1000 cases till it reaches the peak of 100 cases per million (best scenario) to 190 cases per million (worst scenario).

Under the best-case scenario, Fig-2 shows the predicted average number of cases in the coming weeks/months. The daily cases shall peak in the 2nd week of October 2020 at around 130k daily cases, with total cases reaching approximately 8 million. The daily cases shall decrease to around 5k by last week of Feb 2021 and shall cease to exist by mid-May 2021. By this time, the cumulative cases shall reach to approximately 15million.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

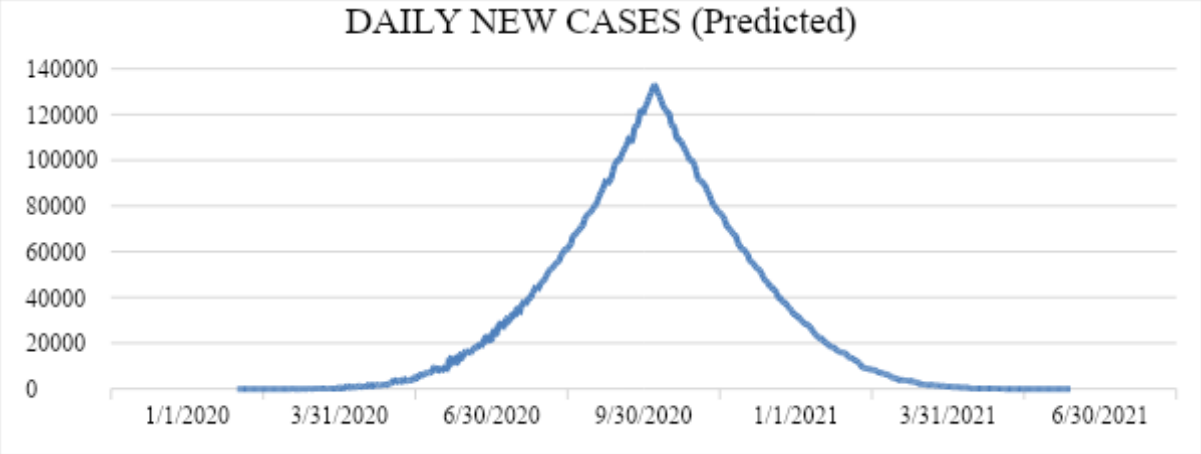

Figure-2: Predicted daily new cases for the peak at 100 cases per million

Extrapolating the same for 190 cases per million in the worst-case scenario, the peak in India shall be reached in mid-Dec 2020 and the total number of cases shall be 20 million (Fig-3). The increase in the daily average number of cases shall be around 260k and the cumulative cases will reach 40 million. By July 2021, the daily cases would reach around 5000 (Figure-4) and shall cease to exist by the end of September 2021.

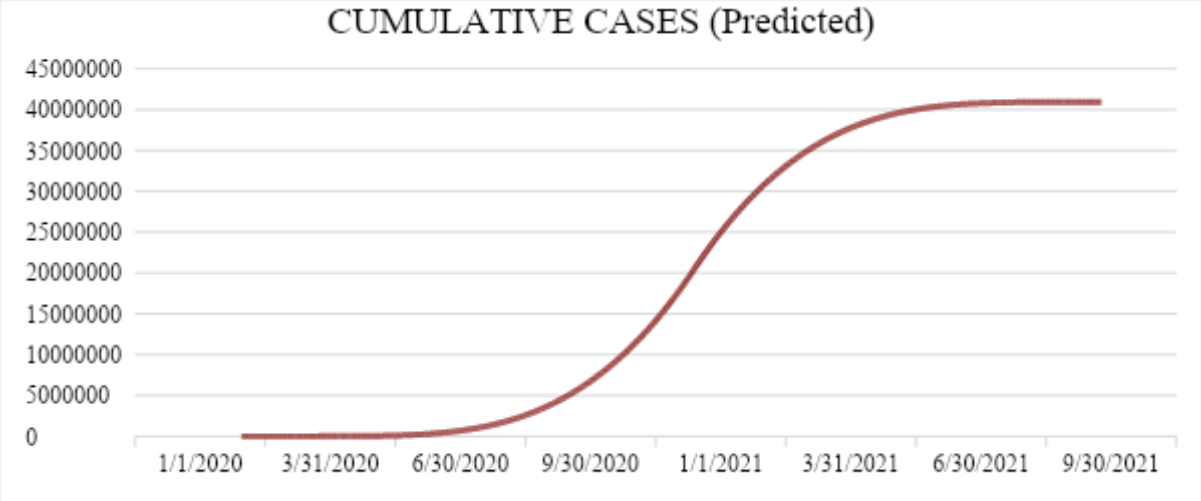

Figure-3: Cumulative cases for the peak at 190 cases per million

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

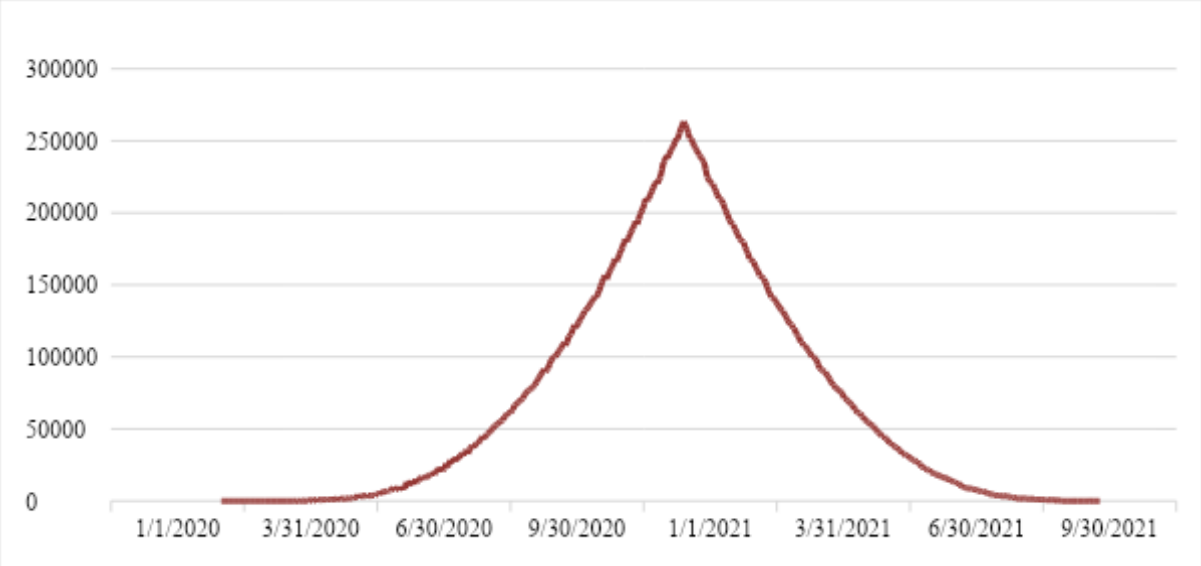

Fig-4: Daily new cases for the peak at 190 cases per million

2.3. Scenario 2 (Total cases):

In this scenario, we have taken the total number of cases per million at the peak, after which the pandemic spread started decreasing for Italy, Germany and Belgium. Table 4 shows the total cases per million for Italy, Germany and Belgium which varies between 600 and around 800. For India, we have taken the average of these three countries and assumed that the number of total cases per million shall reach 724 at the peak. At the same rate of 1000 new cases per week as assumed in scenario 1, the total cases shall reach around 1 million at the peak by mid-July 2020 when the daily new cases would be in the range of 34k and new cases per million would be around 25.

Table-4: Comparison of total cases per million

S No	Country	Total Cases (per million)	
		At Peak	At present*
1	Italy	777	3806
2	Germany	607	2159
3	Belgium	788	4950
4	India	724&	100

Beyond mid-July 2020, the cases shall start decreasing following the pandemic bell curve and by mid-September, the daily new cases shall reduce to 5000 with cumulative cases reaching 2 million. This model predicts that the cases shall cease to exist by the end of November 2020 year and from December, no new cases shall be reported. The bell curve (Fig-5) and S-curve (Fig-6) for daily new cases and cumulative cases are shown below, respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

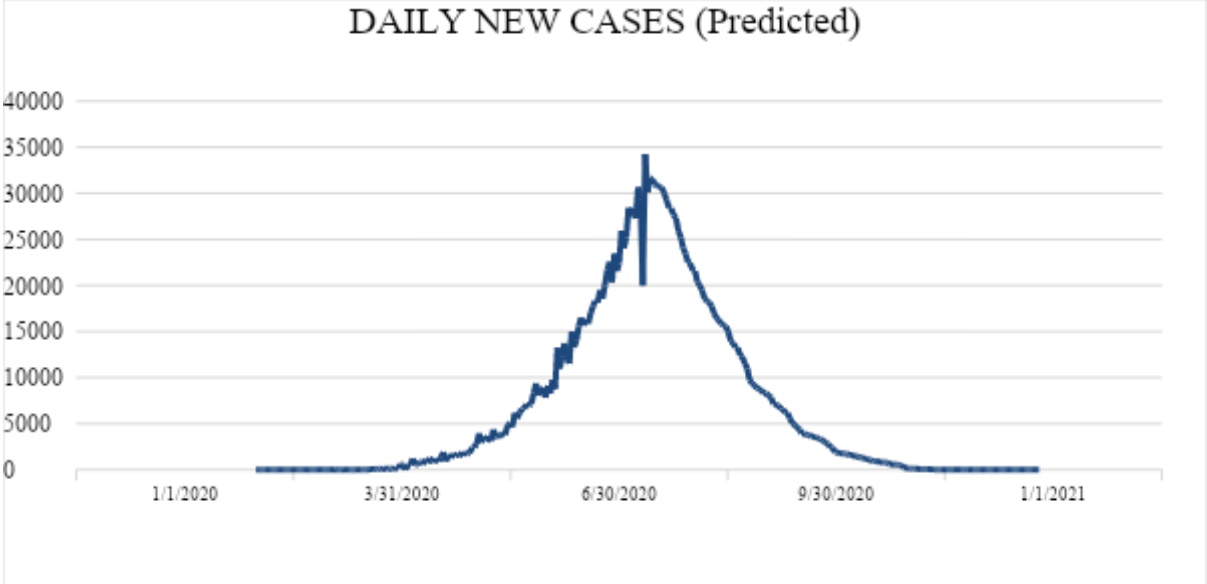

Fig-5: Daily new cases predicted for the peak at 724 cases per million

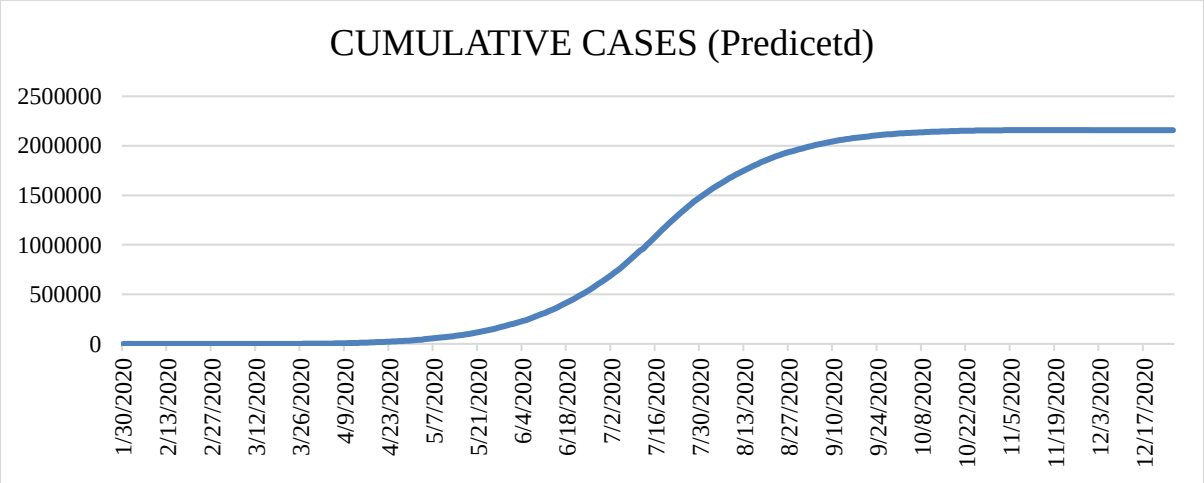

Figure-6: Cumulative cases predicted for the peak at 724 cases per million

However, this seems unlikely as the number of cases has started increasing significantly in the last one week after restrictions have been eased and almost all establishments have started functioning with domestic travel also allowed, by both rail and air.

3. Results

3.1. Model Test:

The model is based upon the data available until 24th May 2020. The validity of the model can only be assessed if it is verified with actual data. At the time of writing this article (on June 11th), the actual data obtained was matched with the predicted values. Table - 5 shows the comparison of the predicted cases vs the actual cases reported for the period 25th May till 11th June for scenario 1 (new cases reaching 190 per million at the peak).

Table-5: COMPARISON OF REAL RECORDED CASES AND PREDICTED CASES

DATE	CUMULATIVE CASES	DAILY NEW CASES

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

	Predicted	Actual	Difference (Predicted - Actual)	Error %	Predicted	Actual cases	Difference (Predicted - Actual)	Error %
25-05-2020	145516	144950	566	0.390479	6980	6414	566	8.82
26-05-2020	152675	150793	1882	1.248069	7159	5843	1316	22.52
27-05-2020	160978	158086	2892	1.829384	8303	7293	1010	13.85
28-05-2020	170347	165386	4961	2.999649	9369	7300	2069	28.34
29-05-2020	178423	173491	4932	2.842799	8076	8105	-29	-0.36
30-05-2020	187023	181827	5196	2.857661	8600	8336	264	3.17
31-05-2020	195920	190609	5311	2.786332	8897	8782	115	1.31
01-06-2020	203751	198370	5381	2.712608	7831	7761	70	0.90
02-06-2020	212875	207191	5684	2.743362	9124	8821	303	3.43
03-06-2020	221151	216824	4327	1.995628	8276	9633	-1357	-14.09
04-06-2020	230904	226713	4191	1.848593	9753	9889	-136	-1.38
05-06-2020	239673	236184	3489	1.477238	8769	9471	-702	-7.41
06-06-2020	252904	246622	6282	2.547218	13231	10438	2793	26.76
07-06-2020	263904	257486	6418	2.492563	11000	10864	136	1.25
08-06-2020	276151	265928	10223	3.844274	12247	8442	3805	45.07
09-06-2020	289875	274780	15095	5.493486	13724	8852	4872	55.04
10-06-2020	302300	287155	15145	5.274155	12425	12375	50	0.40
11-06-2020	313816	298283	15533	5.207471	11516	11128	388	3.49

The daily increase in the cumulative actual and predicted cases were plotted for dates 25th May till 11th June (Fig-7) and it was found that the maximum difference was of the order of 5% with an average of around 2.8%, which was quite reasonable. For daily new cases (Figure 8), there are only 2 days when the error % is more than 40% and three cases where the error % is more than 20%. Other than these 5 values out of 18 predicted values, the average error % for all 13 values is around 1% which translates into 73% predicted values matching with the official reported values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

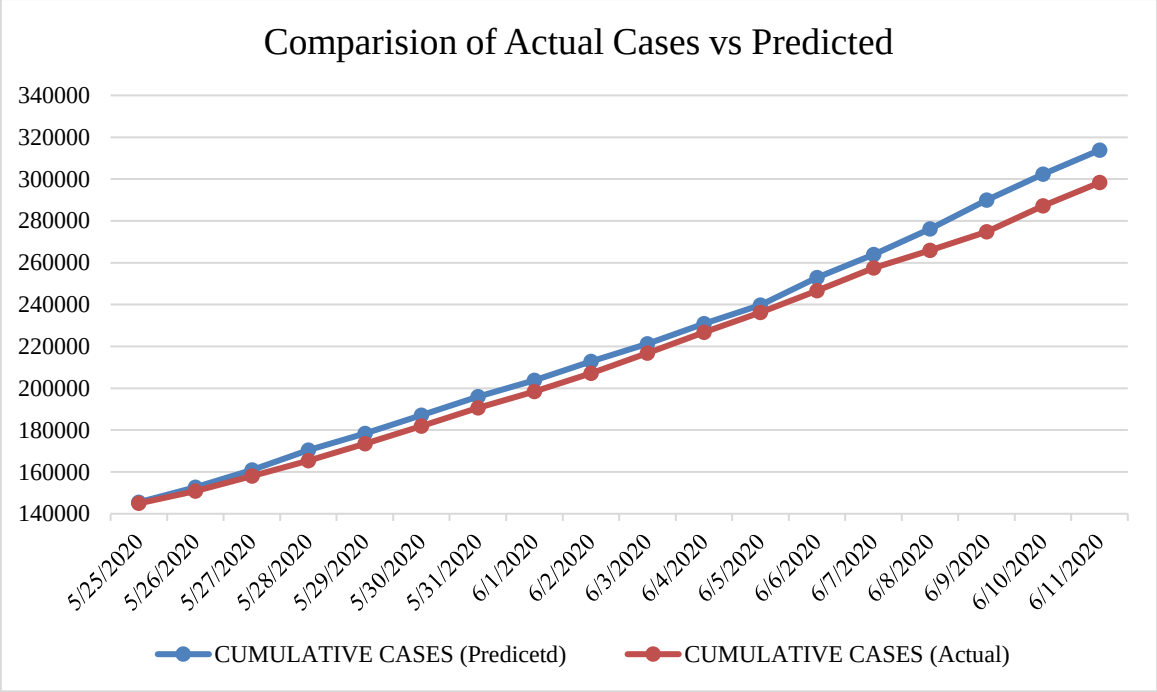

Figure-07: Comparison of Cumulative Cases (Actual Vs Predicted) 25th May – 11th June

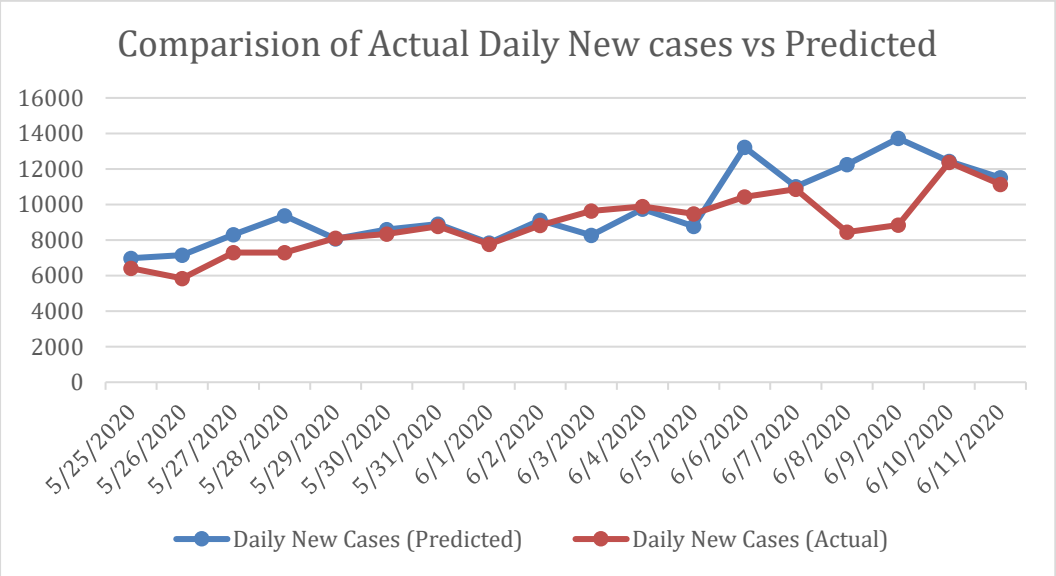

Figure-08: Comparison of Daily New Cases (Actual Vs Predicted) 25th May – 11th June

4. Discussion

A COVID-19 prediction model for India has been presented in the paper by analysing the case trends in other countries and drawing out the similarities from them. The authors used the daily new cases per million and total cases per million as two parameters to predict the number of cases along with its relationship with population density. Population density and daily new cases are co-related to some extent and can be used as a model to predict the number of cases. The disease is contagious and can spread when an infected person comes in contact with one or more persons and is liable to infect some or all of them. It may be said that population density plays an important role in the spread of the disease.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2020 doi:10.20944/preprints202006.0214.v1

India has a similar population density to the other countries considered under the present study, where the number of cases has reduced significantly. The authors used the peak arrival stage for these countries and then incorporated the average values for India to prepare the model. Based upon population density and the trend in the number of daily cases in India, it was found that the total number of cases shall reach 20 million by the end of this year and shall take another 6 months to reduce significantly and 9 months before there are no new cases reported. However, by that time, the total number of affected people in India is predicted to reach 40 million. This model fits quite reasonably for the week when the paper was written with both the cumulative cases and also, with the rise in the daily number of cases.

There is no early decrease in the peak as is also evident from the testing data for India. Fig-9 shows the daily tests done and the number of positive cases reported. It is evident from the figure that the number of tests has almost stagnated at around 150k per day, but the daily reported cases are increasing resulting in narrowing the gap between samples tested and reported cases. There is a need for increasing the number of tests per day and when the gap starts widening then it can be concluded that the disease is at its decline which does not seem likely to happen shortly, at least. This also indicates that the model prediction holds valid.

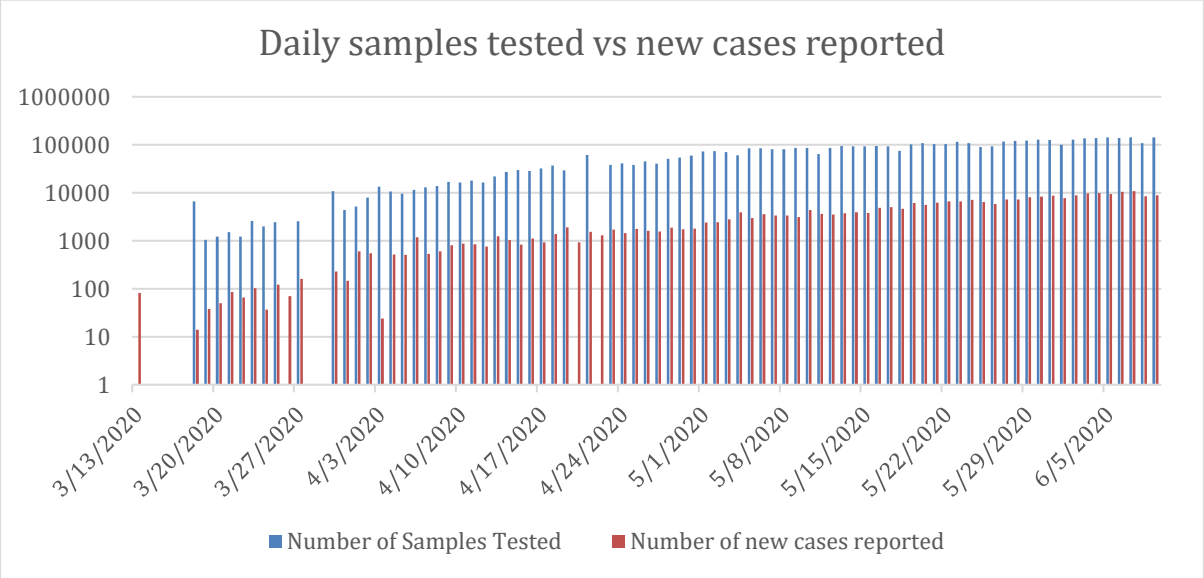

Fig-9: Samples tested vs the number of new cases reported (till 9th June 2020)

5. Conclusion

There are several uncertainties in the spread of the disease because there is no set pattern of propagation among people. At times, contact tracing results in locating the source of infection but in many cases, this is difficult or impossible. However, models can be prepared by studying the trend in other countries in terms of when the case numbers peak and reduce significantly. Two scenarios have been presented by the authors and the predicted values have been compared with the actual data for 18 days. During the 18 days, it was found that the model predicts the actual number of daily new cases for 73% of the days and there was found a mere 2% average variation from the actual cumulative cases of COVID-19 in India.

Funding

The research did not receive any funding from any agency.

Credit authorship contribution statement

Satyaprakash: Conceptualisation, Data curation, Data validation, writing the original draft, and finalisation. **Pinakana Sai Deepak:** Methodology, Formal analysis, Data visualisation, and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to acknowledge the help they received from the Department and the School of Sharda University. We also acknowledge the ICMR website (www.icmr.gov.in) for the India COVID-19 data and several contributors at www.kaggle.com for taking the pain in compiling and making the data available for the public.

References

Arti, M.K., 2020. Modeling and Predictions for COVID 19 Spread in India. DOI: 10.13140/RG.2.2.11427.81444.

Bhatnagar, Manav R., 2020. COVID-19: Mathematical Modeling and Predictions. Published on 1st April 2020. http://web.iitd.ac.in/~manav/COVID.pdf

Bhattacharjee, Atanu; Kumar, Mukesh and Patel, Kamalesh., 2020. When COVID-19 will Decline in India? Prediction by Combination of Recovery and Case Load Rate. https://www.researchgate.net/publication/341117809_When_COVID-19_will_Decline_in_India_Prediction_by_Combination_of_Recovery_and_Case_Load_Rate

Ghosh, P; Ghosh, R and Chakraborty, B., 2020. COVID-19 in India: State-wise Analysis and Prediction. medRxiv. Doi: 10.1101/2020.04.24.20077792

Luo, 2020. https://people.sutd.edu.sg/jianxi_luo/public_html/COVID19PredictionPaper.pdf

Rajesh, Aditya; Pai, Haridas; Roy, Victor; Samanta, Subhasis and Ghosh, Sabyasachi, 2020. CoVID-19 prediction for India from the existing data and SIR(D) model study. DOI: 10.1101/2020.05.05.20085902.

Ranjan, Rajesh, 2020. Predictions for COVID-19 outbreak in India using Epidemiological models. medRxiv. DOI: 10.1101/2020.04.02.20051466

Tiwari, Sunita; Kumar, Sushil and Guleria, Kalpna, 2020. Outbreak Trends of Coronavirus Disease-2019 in India: A Prediction. Disaster Med Public Health Prep, Apr 22;1-6.doi: 10.1017/dmp.2020.115.

Tomar, Anuradha and Gupta, Neeraj, 2020. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Science of The Total Environment, vol 728, #1, Aug 2020. 138762. https://doi.org/10.1016/j.scitotenv.2020.138762

 doi:10.20944/preprints202006.0214.v1

Data sources:

ICMR website: https://www.icmr.nic.in

Kaggle https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

COVID-19 daily info: https://www.worldometers.info/coronavirus/