Dynamic Model to Predict the Flattening of the COVID-19 Outbreak Curve in India

Chappa Sri Vinay*, Rishika Bhatia, Shobha Bagai

Cluster Innovation Centre, University of Delhi, New Delhi, 110007, India

Abstract

This paper aims to predict and visualize the the confirmed cases, deaths and recoveries of COVID-19 in India and also predict the end of growth of COVID-19 cases in India. The methods used for prediction of future COVID-19 cases are machine learning techniques, improved logistic growth equation with dynamic rate of infection and automation of the calculations using Python programming language. The average accuracy percentage of predictions of total confirmed cases is 85.6%, deaths is 84.5%, and recoveries is 83.8%. According to the predictions, the curve will start to flatten from October and the curve will completely flatten in the 2nd week of January. This confirms to the current situation prevailing in India.

Keywords: COVID-19 in India, Prediction, Flattening of the curve, Logistic growth model

1. Introduction

The onset of the year 2020 saw many cases of pneumonia being reported in Wuhan, China. Detailed study of the cluster of cases showed out that it was caused by an unknown novel coronavirus now known as COVID-19. The medical fraternity know that coronaviruses are a large group of viruses that contain a core of genetic material surrounded by an envelope with protein spikes giving the virus a crown like appearance. The word crown in Latin is

*Corresponding author.

Email addresses: vinay.chappa@gmail.com (Chappa Sri Vinay), rishikabhatia1703@gmail.com (Rishika Bhatia), shobhabagai@gmail.com (Shobha Bagai)

called Corona and hence these viruses are named as coronaviruses. Although different types of coronaviruses were in existence with symptoms of respiratory or gastrointestinal in nature, the effects of the new virus COVID-19 were completely unknown. This new virus was connected with the live animal and seafood market in the city of Wuhan. The disease first spread from the people of the local market of Wuhan to their family members, health care staff and eventually the disease spread from person to person through human contact. The virus has affected the human population globally. [1]

1.1. The Spread of COVID-19 in India

The first COVID-19 case in India was reported on 30th January, 2020 in Kerala and the cases further rose to three on 3rd February, 2020. On 2nd March, 2020 three more cases were discovered which included two Indians from Delhi and Hyderabad who had travel history to the affected countries. One of these three infected person was an Italian who had travelled to Jaipur. 21 new COVID-19 cases were detected on 4th March, 2020 and from then the cases started increasing rapidly.

In order to curb the spread of the infection, the government of India called for a fourteen hour nationwide curfew on 22nd March, 2020. The curfew was turned into a 21-days lockdown in 75 districts and metropolitan cities where COVID-19 cases were reported. The Government of India further extended the nationwide lockdown till 31st May, 2020 by categorising all districts into three zones according to the rate of outbreak of virus - red, orange, and green. On 30th May, 2020 the government declared to extend the lockdown from 1st June, 2020 to 30th June, 2020 in containment zones and decided to resume certain services from 8th June, 2020 in lesser contaminated zones. This phase was termed as "Unlock 1"[2]. "Unlock 2" phase was from 1st July, 2020 to 31st July, 2020 where in the lockdown was imposed only in containment zones and most of the activities were allowed in the remaining areas. Table 1 shows the brief timeline of the COVID-19 cases in India.

As of 15 August, 2020 the total number of confirmed cases in India were 2,526,192. As we can see from Table 1, the situations have been changing because of the decision taken by the government and the growth rate has been increasing rapidly and uncertainly because of various reasons. India being a huge country with most population having many factors affecting the growth of the confirmed cases, this makes the situation of COVID-19 in India uncertain. The number of COVID-19 cases may be higher than those

Date	Day	No. of Days from recent milestone	Total No. of Cases
30-01-2020	1	0	1
Phase 1: 25 March 2020 to 14 April 2020			
Phase 2: 15 April 2020 to 3 May 2020			
Phase 3: 4 May 2020 to 17 May 2020			
Phase 4: 18 May 2020 to 31 May 2020			
19-05-2020	110	110	101,139
Unlock 1.0 / Phase 5: 1 June 2020 to 30 June 2020			
03-06-2020	125	15	207,615
13-06-2020	135	10	308,993
21-06-2020	143	8	410,461
27-06-2020	149	6	508,953
Unlock 2.0 / Phase 6: 1 July 2020 to 31 July 2020			
02-07-2020	154	5	604,641
07-07-2020	159	5	719,665
11-07-2020	163	4	820,916
14-07-2020	166	3	906,752
17-07-2020	169	3	1,003,832
20-07-2020	172	3	1,118,043
23-07-2020	175	3	1,238,635
29-07-2020	181	6	1,531,669
Unlock 3.0 / Phase 7: 1 August 2020 to 31 August 2020			
07-08-2020	190	9	2,027,074
15-08-2020	198	8	2,526,192

Table 1: Uncertainty in the increase of COVID-19 cases

reported because the rate of testing the suspected patients has been less than required. Because of this, it is difficult for any machine learning model or mathematical equation to take all the factors responsible for increase in the number of COVID-19 cases into consideration and calculate the predictions of the future COVID-19 cases. All these uncertain and inconsistent situations in India makes it difficult to predict the future cases according to the present situation and it is also difficult to predict when will COVID-19 come to an end in India. The only way to predict the future is by taking the most recent present situation into consideration and calculating the predictions and considering a model that is dynamic and not static, that is, the model should take into account the data regarding the most recent situation and should update the predictions to fit the changing situations in India. The problem statement in this paper is to predict the future COVID-19 cases in India and to find the flattening of the cases in India. [3]

1.2. Existing Models for Prediction

With the spread of the infection, number of researchers initially deployed the modified SEIR (Susceptible – Exposed – Infected – Recovered) compartmental modeling to mathematically model the spread of the coronavirus and predict the end of the pandemic. A Zeb [4] proposed a mathematical model to present the dynamical behavior of COVID-19 infection by incorporating isolation class. They divided the total population into five compartments: susceptible (S), exposed (E), infected (I), isolated (I), and recovered (R) from the disease. A. Anirudh [5] discussed the outcome and the challenges of SIR, SEIR, SEIRU, SIRD, SLIAR, ARIMA, SIDARTHE models used in prediction of spread, peak, and reduction of Covid-19 cases. Although the predictions obtained from ARIMA and SEIRQ was within the range of $\pm 15\%$, the other models had a very wide difference. Baek. Et. Al [6] deployed the SEIR model to study the transmission in a tertiary hospital and assess the effects of different intervention strategies. They constructed an SEIR (susceptible-exposed-infectious-recovered) model with a compartment of doctor, nurse, patient, and caregiver to evaluate the effects of different intervention strategies such as front door screening, quarantine unit for newly admitted patients, early testing of suspected infected people, and personal protective equipment for both medical staff and visitors.

The modified SEIR model was also used for country specific spread of the disease. F. Ndairou [7] used it for Wuhan, China; M. Serhani and H. Labbardi [8] and O. Zarkary et. al [9] for Morocco; N.R. Sasmita et. al [10] for Indonesia; J.Rojas-Vallejos [11] for Chile. K. Sarkar [12] propose the S (susceptible) A (aymptotic) R (recovered) I (infected) Iq (isolated infected) Sq (susceptible quarantined) mathematical model to predict the dynamics of COVID-19 in 17 provinces of India and the overall India. Susceptible (S), exposed (E), asymptotic infected (A), asymptotic infected but not quarantined (I), symptomatic and quarantined infected (Q), hospitalised and isolated infected (H) and recovered (R) model was deployed by SK Biswas et. al. [13] to predict the spread of disease in India. Prediction of COVID-19 epidemic dissemination under the impact of non-pharmaceutical intervention in India have been studied by [14, 15, 16, 17]. More references on the initial work done using the modified SEIR model can be found in J. Wang [18]. The paper also gives the application, limitation and potential of these mathematical models.

One of the limitation of these models is that the transmission rate is generally considered as constant where as in practical these generally vary due to epidemiology, geographical and socioeconomic status. These models can be enhanced by incorporating data driven techniques specially involving machine learning. The work done in this paper is the prediction of future cases using logistic growth model and the prediction of flattening of the COVID-19 curve in India. After the COVID-19 turned into a pandemic there were many research papers about the prediction of the cases around the world and few more papers about the prediction in India using various machine learning models or mathematical models.

Prediction of COVID-19 cases around the world in different countries were done using prediction models like long short-term memory (LSTM) machine learning model[19], linear regression, Multilayer perceptron and Vector autoregression machine learning methods[20], fb prophet model[21], Logistic growth model[22], Exponential growth model[23], SIR model[24] and it's advanced models like SEIRD model, tree-based mathematical modelling[25] and supervised and unsupervised machine learning models and few more mathematical and statistical models.

All the above papers have done the prediction of the COVID-19 cases but they haven't predicted when the COVID-19 will end in India, that is, the flattening of the COVID-19 curve in India. There are very limited papers related to the flattening of the COVID-19 curve in India. Kumar and Roy[26] of the Ministry of Health and Family Welfare have used Bailey's method and regression analysis and stated that COVID-19 would end in September and it has been criticised by looking at the present situation of COVID-19 in India. Bhattacharjee et al[27] used the method based on the cases load rate and recovery rate and their difference stated that the COVID-19 patients would start reducing from 20 May, 2020 which has not happened as we see and this shows the inefficiency of the paper. Ghosh et al.[23] used three different models to predict the growth and flattening of the COVID-19 curve in India and those three models were exponential model, logistic model and susceptible infectious susceptible (SIS) model where from the graphs it is observable that they predicted the end of the pandemic to be in around June 2020 which has not happened. Ranjan[28] using exponential model and susceptible infected recovered (SIR) model stated that the equilibrium of the curve in India would be achieved by end of May 2020 which did not happen. Mahendra K. Verma et al.[29] used the best fitting of the curves and analysed the data of different countries including India and stated that the transition of the curve from an exponential regime to a power law regime may be the sign for the flattening of the curve but They haven't predicted any date for the flattening of the curve in India.

In this study, the prediction of the flattening of the curve is done using a simple but a more efficient and a realistic model by considering a dynamic infection rate. The infected model is hence being updated continuously by discretizing the modified logistic equation.

2. Methodology

Let N be the cumulative infected population, then it is assumed that the cumulative infected population will follow a Logistic Growth Model. Let N_r be the infected cases on Day r $(t = t_r)$, N_{r+1} be the cases on the next day $(t = t_{r+1})$ and a be the infection rate. The Logistic Growth Model is given by

$$\frac{dN}{dt} = a\left(1 - \frac{N}{N_{\max}}\right)N \tag{1}$$

where a is the infection rate, that is generally assumed to be constant and may be calculated from a given data set. N_{max} is the total population that can be infected. But in order that the model is more realistic the infection rate a would be a function of time t. The solution of equation (1) is given by

$$N(t) = \frac{cN_{\max}}{c + N_{\max}e^{-\int a(t)dt}} \tag{2}$$

As $t \to \infty$, $N \to N_{critical}$, where $N_{critical} = N_{\max}$ if $a(t)$ is an increasing function of t and $N_{critical} < N_{\max}$ if $a(t)$ is a decreasing function of t. Discretisation of equation (1) yields

$$\frac{N_{r+1} - N_r}{t_{r+1} - t_r} = a_r\left(1 - \frac{N_r}{N_{\max}}\right)N_r \tag{3}$$

The infection rate is recursively calculated as

$$a(t = t_r) = \frac{N_{r+1} - N_r}{N_r} \tag{4}$$

On updating the infection rate recursively, the predictions are more dynamic and realistic in contrast to a constant infection rate.

3. Computation of the Equation and Prediction

Python programming language is used to automate the process of calculation of the predictions and also to plot the graphs of the predictions. The program is divided into the following steps:

Step 1: Importing the necessary libraries for calculations and visualizations: Numpy python library is imported for the scientific and mathematical calculations for calculating prediction values. Pandas python library is used for data manipulation and analysis and is also used for CSV file input/output. Plotly graphing library is used for plotting the graphs of the data in CSV file. It is used to visualize the past data of COVID-19 cases and also to plot the predicted cases and visualize it. CSV library is imported for saving the calculated values to the CSV file easily.

Step 2: Importing the Data in CSV file format and cleaning the data:
The data is collected from the JHU CSSE COVID-19 Data[30] where the data can be downloaded in CSV file format and the data consists of the Total cases, deaths and recoveries and the growth rates of the total cases, deaths and recoveries of each day since 30 January, 2020 are calculated and used. The data is imported into the program using pandas functions and the null spaces are filled with zero and days column is added along with date.

Step 4: Creating a loop to automate the process of calculation of predictions for fixed number of days.
The training data of the infected cases is considered till 31st August 2020. The data required for predictions according to equation 3 is N_0 = Total confirmed cases on 31st Aug, 2020 and growth rate a on 31st Aug, 2020 which is calculated using the confirmed cases on 30th Aug and 31st Aug 2020. Once we have the initial values of the variables, the number of infected population $N(t)$ and the infected rate $a(t)$ the predicted infected population, the updated infected population and the infected rate are calculated from equations 3 and 4. The main part of the program is the loop and the code of the loop is given in Appendix A

Step 5: Visualization of the predicted data along with the past data
The prediction values are calculated for the fixed number of days and saved as CSV file. Using the plotly functions, the predicted values are plotted along with the past values where the line of predicted values is in blue colour (if

the initial stage is taken as 31st August 2020); red colour (if the initial stage is taken as 31st July 2020); green colour (if the initial stage is taken as 30th June 2020) and actual values is in dotted blue colour. The graphs are plotted for total confirmed cases, deaths and recoveries.

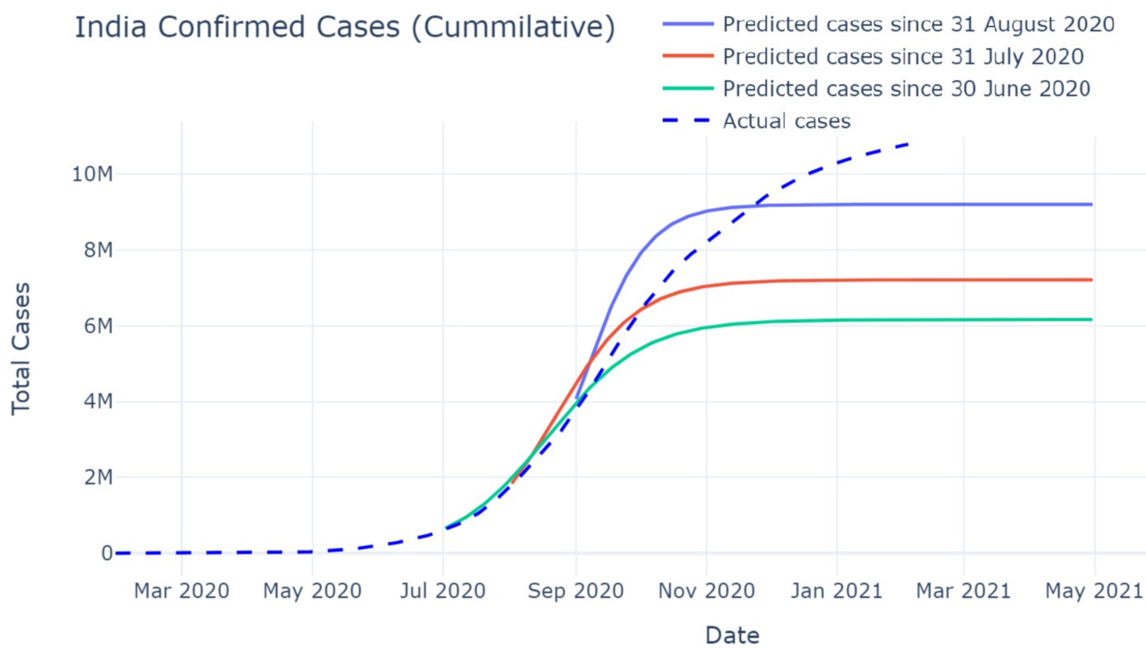

Figure 1: Graphical Representation of predicted total confirmed cases

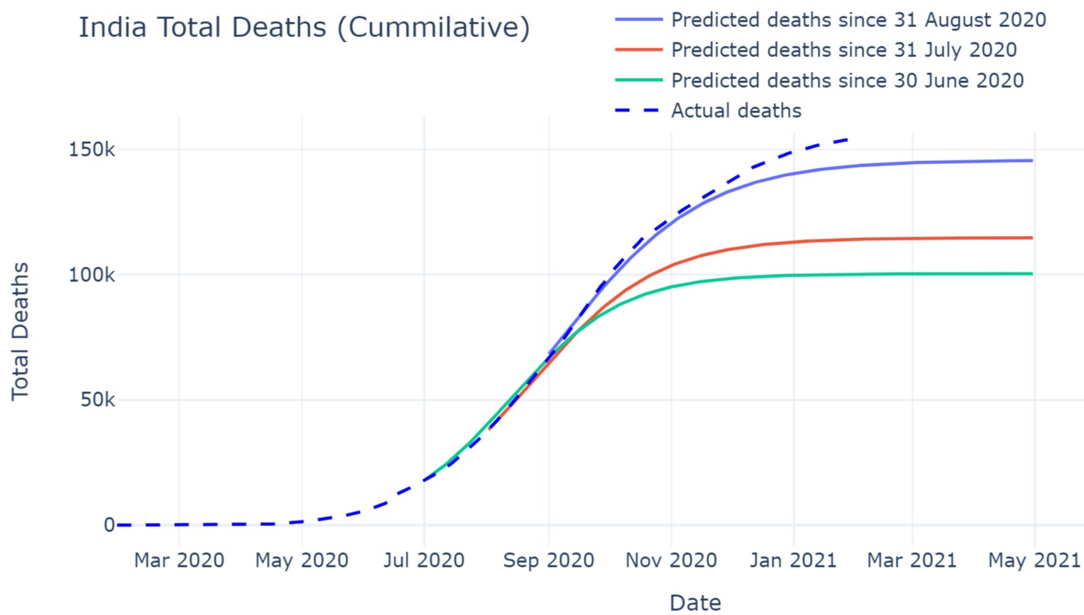

Figure 2: Graphical Representation of predicted total deaths

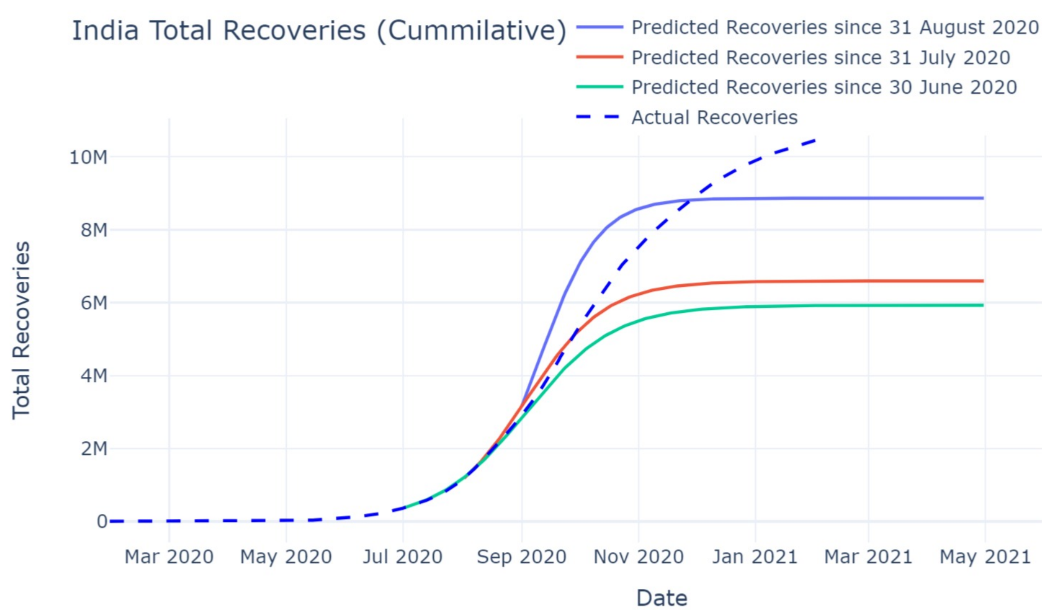

Figure 3: Graphical Representation of predicted total recoveries

4. Results and Conclusion

Figures (1 - 3) illustrate the predicted values of the cumulative infected cases, deaths and recovered cases respectively. As expected, the accuracy of the predicted values increases as we consider the initial point of N_0 to be more recent. For any initial point, there is a conformity in the predicted value and the actual values till September 2020. The deviations starts to take place from September 2020 on wards. In fact all the leading news agency had reported a sudden spike in the cases in September 2020. Unlock 4 was introduced in September in India which allowed commuting through public transport, inter state travel and even restricted foreign travels that could have attributed to higher rate of the transmission of infection. The improved logistic growth model used in this study predicts that the curve would start to flatten from October and will completely flatten by the 2nd week of January, 2021. To great extent this is correct. On 31 January, 2021 the actual values are compared with the predicted values and the percentage accuracy is given in Table 2. As the rate of infection is decreasing, $N_{critical} < N_{\max}$.

Case Type	Predicted Number	Actual Number	Accuracy
Confirmed Cases	9203709	10757610	85.6%
Deaths	130429	154392	84.5%
Recoveries	8744389	10434983	83.8%

Table 2: Prediction on 31 January, 2020

Although, a number of factors are involved to correctly predict the values when using any Mathematical model, the aim here was to give a simple yet elegant way of predicting the infected cases by using a dynamic infection rate rather than a constant rate or fitting a function for the rate of infection.

Appendix A. Code

```
# the number of days is from
# Day 129 (6 June 2020) to Day 577 (28 August, 2021)
for i in range(129, 577):
    # (Logistic Growth Model).
    # a is divided by 100 because growth rate is in percentage
    Nnext=N+(a/100)*(1-N/Nmax)*N*dt
    a = ((Nnext-N)/N)*100 #(Updating growth rate using predicted value)
```

```
N = Nnext #(Updating N for further calculations)
#saving Nnext and 'a' value to csv file
f = open('/kaggle/working/covid19previousdata.csv', 'r')
reader = csv.reader(f)
mylist = list(reader)
f.close()
mylist[i][3] = Nnext
mylist[i][4] = a
my_new_list = open('/kaggle/working/covid19previousdata.csv', 'w')
csv_writer = csv.writer(my_new_list)
csv_writer.writerows(mylist)
my_new_list.close()
```

References

[1] Coronavirus disease (covid-19), http://www.who.int/emergencies/diseases/novel-coronavirus-2019, 2020.

[2] Covid-19 pandemic lockdown in india, https://en.wikipedia.org/wiki/covid-19_pandemic_lockdown_in_india, 2020.

[3] Covid-19 pandemic in india, https://en.wikipedia.org/wiki/covid-19_pandemic_in_india, 2020.

[4] A. Zeb, E. Alzahrani, V. S. Erturk, G. Zaman, Mathematical model for coronavirus disease 2019 (covid-19) containing isolation class, BioMed Research International 2020 (2020) 1–7.

[5] A. Anirudh, Mathematical modeling and the transmission dynamics in predicting the covid-19 - what next in combating the pandemic, Infectious Disease Modelling 5 (2020) 366–374.

[6] Y. J. Baek, T. Lee, Y. Cho, J. H. Hyun, M. H. Kim, Y. Sohn, J. H. Kim, J. Y. Ahn, S. J. Jeong, N. S. Ku, J.-S. Yeom, J. Lee, J. Y. Choi, A mathematical model of covid-19 transmission in a tertiary hospital and assessment of the effects of different intervention strategies, PLoS ONE 15 (2020).

[7] F. Ndaïrou, I. Area, J. J. Nieto, D. F. Torres, Mathematical modeling of covid-19 transmission dynamics with a case study of wuhan, Chaos, Solitons, and Fractals (2020).

[8] M. Serhani, H. Labbardi, Mathematical modeling of covid-19 spreading with asymptomatic infected and interacting peoples, Journal of Applied Mathematics and Computing (2020).

[9] O. Zakary, S. Bidah, M. Rachik, H. Ferjouchia, Mathematical model to estimate and predict the covid-19 infections in morocco: Optimal control strategy, Journal of Applied Mathematics 2020 (2020) 1–13.

[10] N. R. Sasmita, M. Ikhwan, S. Suyanto, V. Chongsuvivatwong, Optimal control on a mathematical model to pattern the progression of coronavirus disease 2019 (covid-19) in indonesia, Global Health Research and Policy 5 (2020).

[11] J. Rojas-Vallejos, Strengths and limitations of mathematical models in pandemics—the case of covid-19 in chile, Medwave 20 (2020) e7874–e7874.

[12] K. Sarkar, S. Khajanchi, J. J. Nieto, Modeling and forecasting the covid-19 pandemic in india, Chaos, Solitons and Fractals 139 (2020) 110049.

[13] S. K. Biswas, J. K. Ghosh, S. Sarkar, U. Ghosh, Covid-19 pandemic in india: a mathematical model study, Nonlinear Dynamics (2020).

[14] V. Tiwari, N. Deyal, N. S. Bisht, Mathematical modeling based study and prediction of covid-19 epidemic dissemination under the impact of lockdown in india, Frontiers in Physics 8 (2020).

[15] V. P. Bajiya, S. Bugalia, J. P. Tripathi, Mathematical modeling of covid-19: Impact of non-pharmaceutical interventions in india, Chaos: An Interdisciplinary Journal of Nonlinear Science 30 (2020) 113143.

[16] P. Vijayaraghavan, S. R. Sriramkumar, Non-pharmaceutical interventions are measures to control coronavirus disease-2019 (covid-19) transmission in india, Coronaviruses 01 (2020).

[17] A. Mahajan, N. A. Sivadas, R. Solanki, An epidemic model sipherd and its application for prediction of the spread of covid-19 infection in india, Chaos, Solitons, and Fractals 140 (2020) 110156.

[18] J. Wang, Mathematical models for covid-19: applications, limitations, and potentials, Journal of Public Health and Emergency 4 (2020) 9–9.

[19] A. Tomar, N. Gupta, Prediction for the spread of covid-19 in india and effectiveness of preventive measures, Science of The Total Environment (2020) 138762.

[20] R. Sujath, J. M. Chatterjee, A. E. Hassanien, A machine learning forecasting model for covid-19 pandemic in india, Stochastic Environmental Research and Risk Assessment (2020).

[21] T. Upendra, K. Rizwan, Role of machine learning to predict the outbreak of covid-19 in india, Xi'an Jianzhu Keji Daxue Xuebao/Journal of Xi'an University of Architecture and Technology 12 (2020) 2663–2669.

[22] B. Malhotra, V. Kashyap, Progression of covid-19 in indian states - forecasting endpoints using sir and logistic growth models, COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv (2020).

[23] P. Ghosh, R. Ghosh, B. Chakraborty, Covid-19 in india: State-wise analysis and prediction, medRxiv preprint (2020).

[24] G. Anirban, Covid-19 in india: Policy suggestions using epidemiological modeling, IIMV, COVID-19 Research (2020).

[25] M. Arti, Modeling and predictions for covid 19 spread in india, ResearchGate (2020).

[26] R. Roy, A. Kumar, Application of mathematical modeling in public health decision making pertaining to control of covid-19 pandemic in india, Epidemiology International 05 (2020) 23–26.

[27] A. Bhattacharjee, M. Kumar, K. K. Patel, When covid-19 will decline in india? prediction by combination of recovery and case load rate, Clinical Epidemiology and Global Health (2020).

[28] R. Ranjan, Predictions for covid-19 outbreak in india using epidemiological models, COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv (2020).

[29] M. K. Verma, A. Asad, S. Chatterjee, Covid-19 epidemic: Power law spread and flattening of the curve, COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv (2020).

[30] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track covid-19 in real time, The Lancet Infectious Diseases (2020).