RESEARCH PAPER

A statistical and deep learning-based daily infected count prediction system for the coronavirus pandemic

Vruddhi Shah[1] · Ankita Shelke[1] · Mamata Parab[1] · Jainam Shah[1] · Ninad Mehendale[1]

Abstract

We present new data analytics-based predictions results that can help governments to plan their future actions and also help medical services to be better prepared for the future. Our system can predict new corona cases with 99.82% accuracy using susceptible infected recovered (SIR) model. We have predicted the results of new COVID cases per day for dense and highly populated country i.e. India. We found that traditional statistical methods will not work efficiently as they do not consider the limited population in a particular country. Using the data analytics-based curve we predicted four most likely possibilities for the number of new cases in India. Hence, we expect that the results mentioned in the manuscript help people to better understand the progress of this disease.

Keywords Covid-19 simulations · Coronavirus · Daily count

1 Introduction

The Chinese authorities alerted the world in December 2019, that a virus outspread had occurred in their community. Eventually, it spread to the other countries in the coming months, with cases multiplying within a span of a few days. This virus was Severe Acute Respiratory Syndrome related Coronavirus that causes COVID-19. A virus is a small collection of genetic material bordered by a protein coat. It cannot replicate on its own. Also how long, a virus can remain on the surfaces is not certain. Coronavirus mostly spreads through contact with different surfaces. It mainly spreads by the droplets, when infected people sneeze or cough.

There are two possibilities for a pandemic like a corona: rapid and slow. A rapid pandemic costs many lives which would be devastating. The worst-case for a rapid pandemic begins with a very high rate of infection as there are no countermeasures taken to slow it down beforehand. In a rapid pandemic a large number of people would get sick and all at the same time. And if these numbers become too large, then the healthcare facilities would be unable to grip it. Not enough resources, like medical staff or medical equipment like ventilators, would be left to help those in need of them. People will die untreated in such a case. If the health care workers get sick themselves, the capacity of the health care systems to operate properly will fall drastically. And if so happens, then decisions will have to be made about who gets to live and who doesn't as seen in Italy and New York. The number of deaths will tend to rise significantly in such a scenario.

This entire manuscript provides the information about likely next day count and when the end of phase 1 (EOP1), end of phase 2 (EOP2) and end of phase 3 (EOP3) (shown in Fig. 1) will arrive in India using 4 different possibilities so that people can find out in how many days India will be free from this pandemic crisis. Also, the maximum number of people getting infected during phase 2, and the end of cycle one has been predicted with the help of the data analysis-based curve generated from the analysis of different multiple countries. Our prediction can help decision-makers like government and medical supervisors to do the right thing. We expect more and more lives are saved and though this curve is unavoidable we can reduce its peak and slopes.

The COVID-19 outbreak has significantly impacted various sectors of the economy. The proposed work uses data analytics-based prediction as well as statistical analysis model (Figs. 2, 3). The data analytics-based prediction (Fig. 4) provides comparatively better results in comparison to the prediction carried out by using statistical models.

✉ Ninad Mehendale
ninad@somaiya.edu

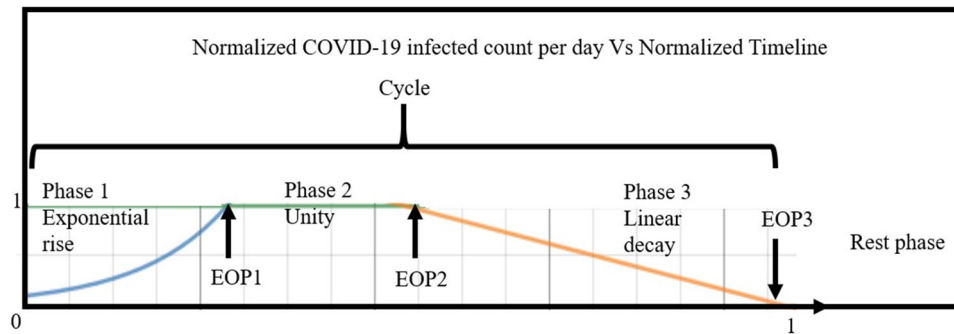1   K. J. Somaiya College of Engineering, Mumbai 400077, Maharashtra, India

Fig. 1 The normalized curve of COVID-19 infected new patient's daily count versus normalized timeline. The curve consists of a cycle having 3 phases followed by a rest phase. Phase 1 is the exponential rise where the cases keep on increasing up to a point End of Phase 1 (EOP1). After EOP1, the number of cases increases at a constant level, and a flat curve can be seen indicating the phase 2. After End of phase 2 (EOP2), the cases start to decline gradually following a skewed curve and the phase 3 begins up to a point End of phase 3 (EOP3). These 3 phases together constitute a cycle for the coronavirus cases. After EOP3 a rest phase begins

This prediction will be useful to the public and health care sector for better decision-making. As compared to existing literature, the accuracy of our predictions was 99.82%. The introduction section explains coronavirus, its impacts on the world, why it is necessary to predict the cases and its significance. The working of coronavirus is explained in the next section post the introduction. Other methodologies reported recently are covered in the literature review. The methodology section explains all the different statistical and deep learning models used for the prediction in detail. This section ends with data analysis (Figs. 5, 6, 7) and its importance in forecasting (Fig. 8). The results and discussion section explains the RMSE values obtained from the different algorithms used and also covers general observations found from the analysis of the data and compares different methods used. The conclusion summarizes the entire work and gives a future path for more development.

2 Working of corona virus

The virus begins to spread very deep into the person's body via the nasal cavity. The targets of the virus are the lungs, the spleen, or the intestines, out of which, the most noticeable effect of this virus can be seen in the lungs. A large number of epithelial cells are present at the border of the lungs. These epithelial cells, also line the organs of our body and mucosa. Coronavirus injects its genetic material inside the epithelial cell by connecting to a particular receptor on the membrane of the victim's cell. The cell executes the new commands that are given by this genetic material, which are to replicate and reconstruct new coronaviruses. This adds up further by generating more and more duplicates of the original virus till a critical point is reached where the cell receives a final call which is to self-destruct. And now the newly released coronaviruses starts attacking the other cells.

An exponential increase in the number of infected cells is seen and within 14 days. A huge number of cells, about millions, are infected and the lungs would be housing more than a billion of these viruses. This virus has so far not caused any damage to the body, but will now show a tremendous impact on the immune system. We all know that the immune system is meant for our protection, but it requires dense regulation as immune cells can also be harmful. Corona infects some of the immune cells and creates a disturbance by confusing them when these cells are poured into the lungs to fight the virus. These immune cells communicate through tiny proteins carrying information called as the cytokines. These cytokines are responsible for carrying out important immune reactions. Coronavirus causes the infected immune cells to exaggerate; it puts the immune system into a state of confusion and sends more antibodies than it should and thereby wasting a lot of its resources and causes great damage. Two types of cells in general create the disturbance. The first type of cells are Neutrophils, which are capable of killing the outside material, are sometimes even capable of killing host body cells. They arrive in large numbers and kill as many friends as enemies. The other being, the killer T-cells that ask the infected cells to commit a controlled suicide. And in this situation of conflict, they tend to ask the healthy cells to destroy themselves too. And as a huge number of these cells arrive, more damage is caused to the healthy lung tissues as well. This leads to lifelong disabilities and causes permanent irreversible damage. But in the majority of the cases, the immune system slowly regains power. The recovery begins and they start killing the infected cells and also recognize the viruses that try to infect the other new cells. In most of the cases, the patients can be mildly symptomatic. But some cases tend to become critical, and in this case, millions of epithelial cells die, the lungs lining are also damage, and the alveoli gets infected with the virus damaging the respiration. Such patients face difficulties in

breathing and need ventilators to survive. For weeks, the immune system has fought with the virus to its full capacity and made antiviral weapons to fight the coronavirus. But if the virus rapidly multiplies and enters the bloodstream, then death is more likely.

3 Literature review

Certain infectious diseases or events may or may not occur in a cyclic or rhythmic pattern. By studying and understanding their pattern we can determine when a future outbreak may occur. Due to this, it is possible to take preventative steps to minimize its impact. In year 2014, Kane et al. [1] worked in same field. He predicted the outbreak of avian influenza H5N1 in the nearest future. Kane et al.

used ARIMA and Random Forest time series models for predicting this outbreak. In the study, Kane et al. found that Random Forest time-series predictive analysis provided better results over the ARIMA model. Mean Square Error (MSE) for the ARIMA model found to be 28.7412 whereas it was 24.8101 for the Random Forest model. In the current situation, studying and analyzing the spread of the Coronavirus is an essential task. Multiple factors lead to an increase in the count of the patient. In most of the cases, it is found that people already having some respiratory disease are at the verge of having COVID-19 disease. There is a direct relationship between the increased count of Coronavirus-infected people and pre-existing respiratory problems like pneumonia. Ji et al. [2] worked on same topic. He used Multivariate Cox regression to identify risk factors associated with the progression of the disease. Ji
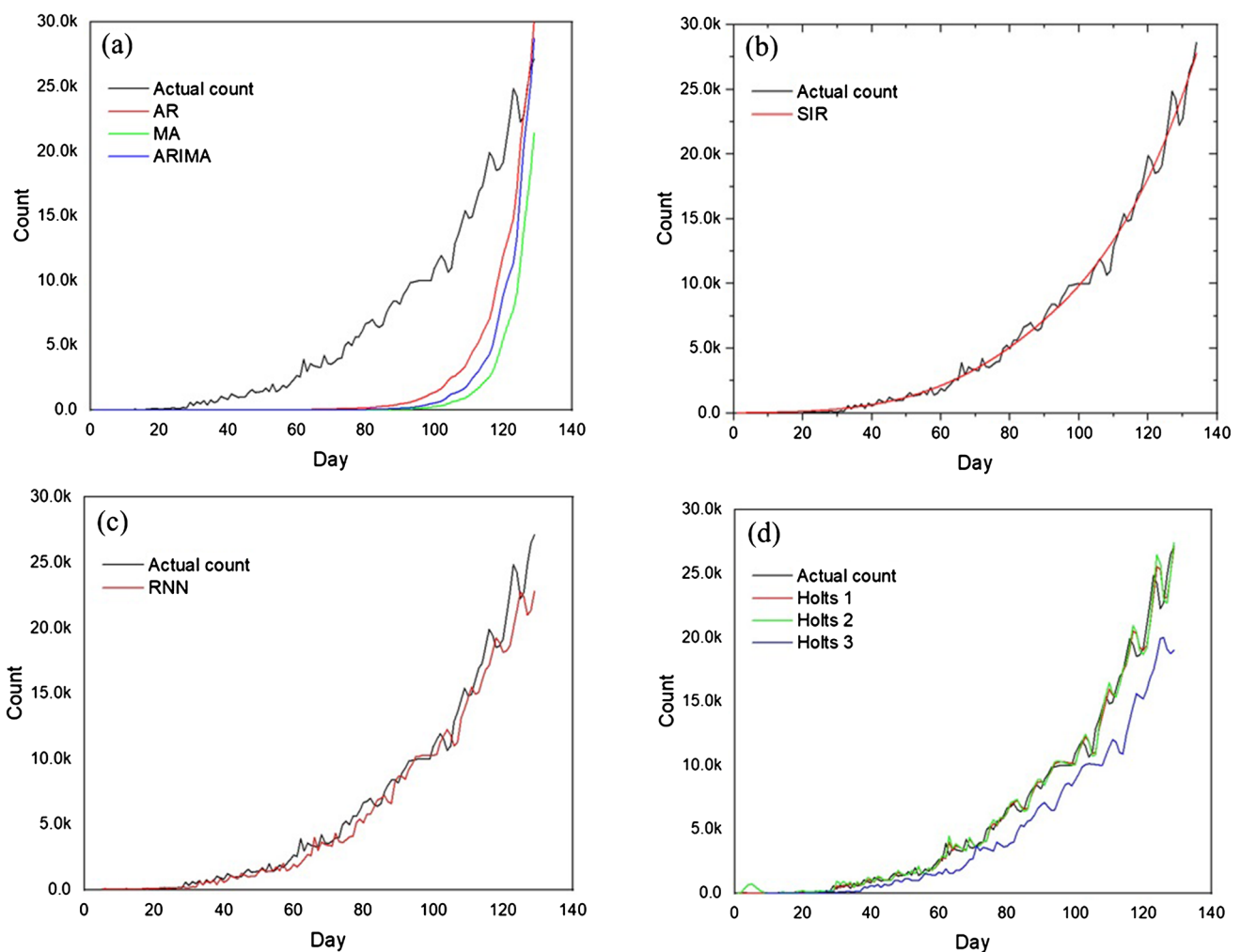

Fig. 2 Different statistical models fitting on actual data to train and forecasting the count. The black line indicates the actual count. **a** AR, MA, and ARIMA model for the fitting. They are far away from actual values most of the time. **b** Result of SIR model for fitting. The SIR model gives a good fit for the actual data. **c** Result of the RNN model for fitting on the actual data **d** 3 different Holts algorithm for fitting the data
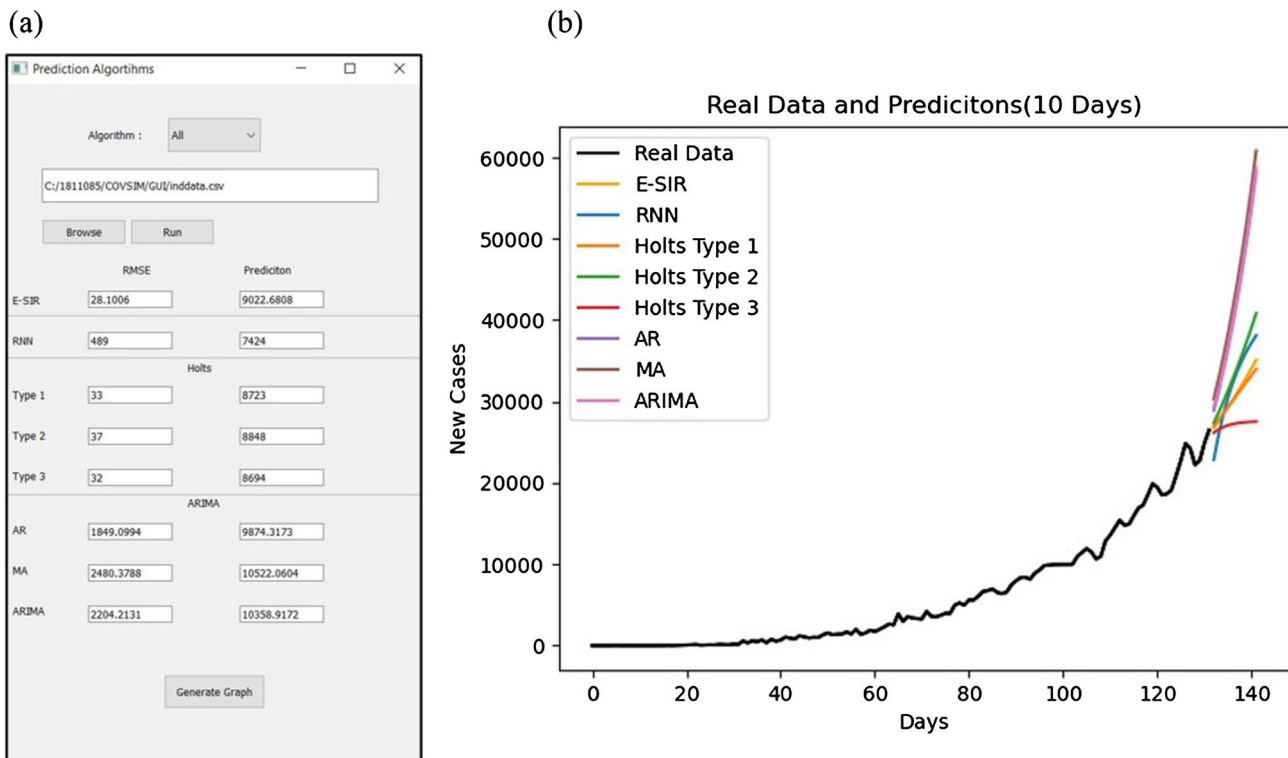
(a)

(b)

Fig. 3 **a** A graphic user interface (GUI) for the prediction of the cases. 6 different algorithms were implemented to predict the cases likely to arise in the coming days. The input given was a .csv file which consists of two rows. One row consists of the date and the second consists of the corresponding count on that date. The Run button generates the predicted next day count and the RMSE values for each algorithm (E-SIR, RNN, Holts, AR, MA, ARIMA). **b** The output of the GUI application which gives the prediction for the next 10 days in the form of graphs by just clicking the generate graphs button

et al. obtained 0.91 value for Receiver Operating Characteristics. As Ji et al. and Qasim et al. has also studied the trend of the Corona pandemic. As mentioned in, Qasim et al. [3], qasim*et al.* has been determine the trend of corona spread based on Successive Approximation Method. In this work, he calculated worldwide expected lower bound value for confirmed cases 247007 cases with a maximum limit of 1667719 cases and minimum deaths count 8660 with an upper limit of 117397 deaths in the next 30 days. The mean ratio of expected cases was found to be 0.485 whereas the mean ratio of deaths was found to be 0.49. It is an essential task to prevent the spread of Coronavirus further. To stop or reduce the spread of Coronavirus, across the world various steps have been taken like announcing the lockdown, isolation, etc. Arti *et al.* [4], used a mathematical model to determine the effect of prevention technique on the spread of Corona. Arti et al. used a tree-based model to predict the spread of the coronavirus. Dhanwant et al. [5] worked in the same field for predicting the spread of Coronavirus in different countries by considering various features of Coronavirus growth. They found that the parameter η governed by various

factors like the social-contact structure. The dependency of this parameter eta on the transmission level in society gives a sense of the effectiveness of the measures taken for social distancing. An algorithm is developed in python using Scipy that understands and learns the social-contact structure. This algorithm gives a suitable value for η. In a study by Tandon et al. [6] has compared multiple models and then employed them for forecasting future COVID-19 cases in India. They tried various models and found the ARIMA(Auto-Regressive Integrated Moving Average) model to be the most accurate based on error. Chen group [7] conducted the mathematical and numerical analysis for COVID-19 and proposed a time-dependent susceptible-infected-recovered (SIR) model that tracks 2 time series namely the transmission rate at time t and the recovering rate at time t. In a study by Jun Zhang [8], a method was proposed to reconstruct a phase space based on the chaotic time series using RNN. This method was simple, practical, and effective in chaotic time series prediction. In a study by Chris Chatfield [9], the Holts-Winters forecasting procedure was tested for practical purposes and an important distinction between automatic and non-automatic

Fig. 4 Proposed system flow diagram. The input consisted of the actual count taken from the open-source dataset. The data was initially filtered, and the different time series prediction models were applied to the processed data. Six statistical models, viz. Holts forecasting, ETS, AR, MA and ARIMA, SIR, E-SIR, and neural networks forecasting and a data analytics based prediction model was used for predicting the output. The obtained results from each model were compared to find the best-suited model for the prediction
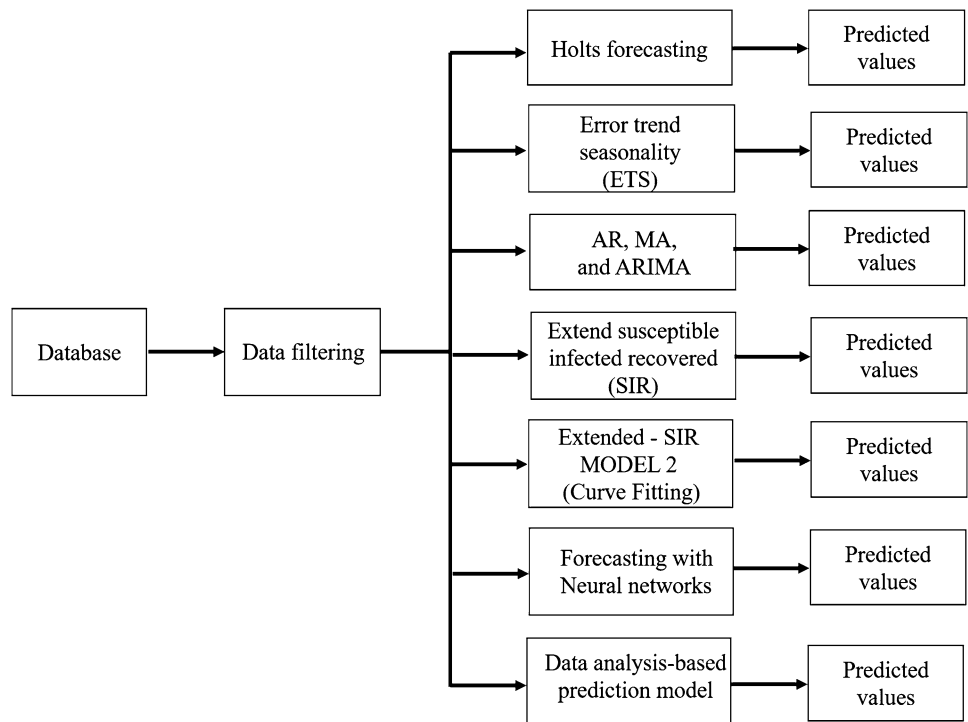

forecasting approach was done to implement the Holts-Winters algorithm.

4 Methodology

The statistics on the coronavirus pandemic for a number of countries in the world has been presented as an open source dataset for further analysis. Using this dataset [10], we carried out the prediction and the results obtained were tested using the data mentioned in the dataset. A basic flow of the methodology used has been shown in Fig. 4. We have used six time series prediction codes and

Fig. 5 Data analysis of the countries namely the UK, Italy, Germany, Ireland, New Zealand, Canada, and France where the spread of coronavirus has stopped has been done. Their normalized graphs have been plotted. A curve has been fitted on the normalized graphs of these countries shown by the blue dotted line, indicating the 3 phase nature of the simulation curve
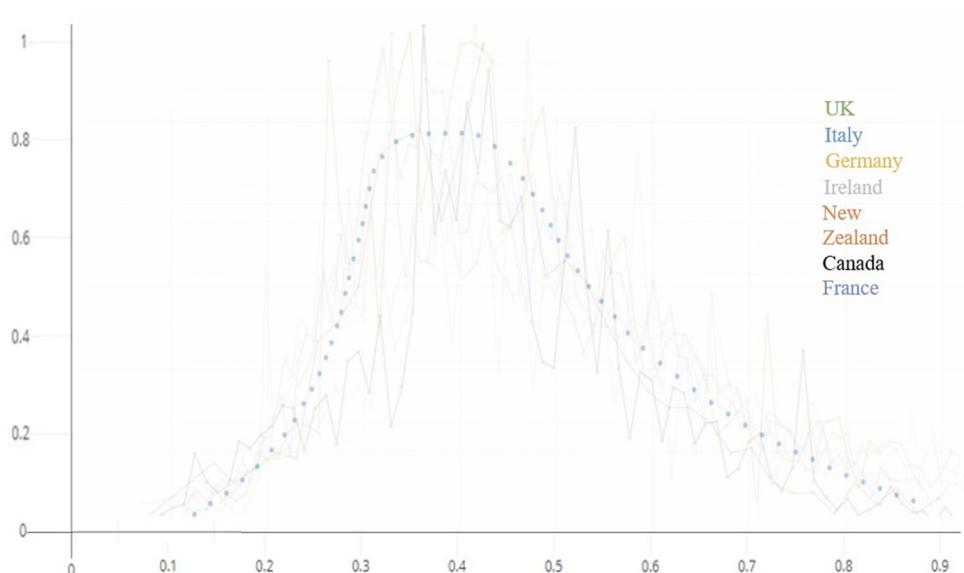
a data-analysis based prediction model has been presented. A comparison of the results obtained with each prediction algorithm was done to find a suitable model for prediction. A description on the algorithms used for prediction of next COVID-19 count daily rise value are presented in the following sub-sections. For people's benefit we are giving away all the codes with this manuscript in supplementary section.

4.1 Holts forecasting model

The simple exponential smoothing to allow the forecasting of the data with no trends was extended by Holt [11]. A forecast equation and two smoothing equations are involved in Holt's method [12](one for the level and one for the trend) where the level smoothing parameter is $0 \leq \alpha \leq 1$ is, and the trend smoothing parameter is $0 \leq \beta \leq 1$. For the long-term forecast, the forecasting with Holt's method will increase or decrease indefinitely and can extend into the future. In that case, we use the Damped trend method having a damping parameter $0 \leq \phi \leq 1$ to prevent the forecast from giving extremely large value.

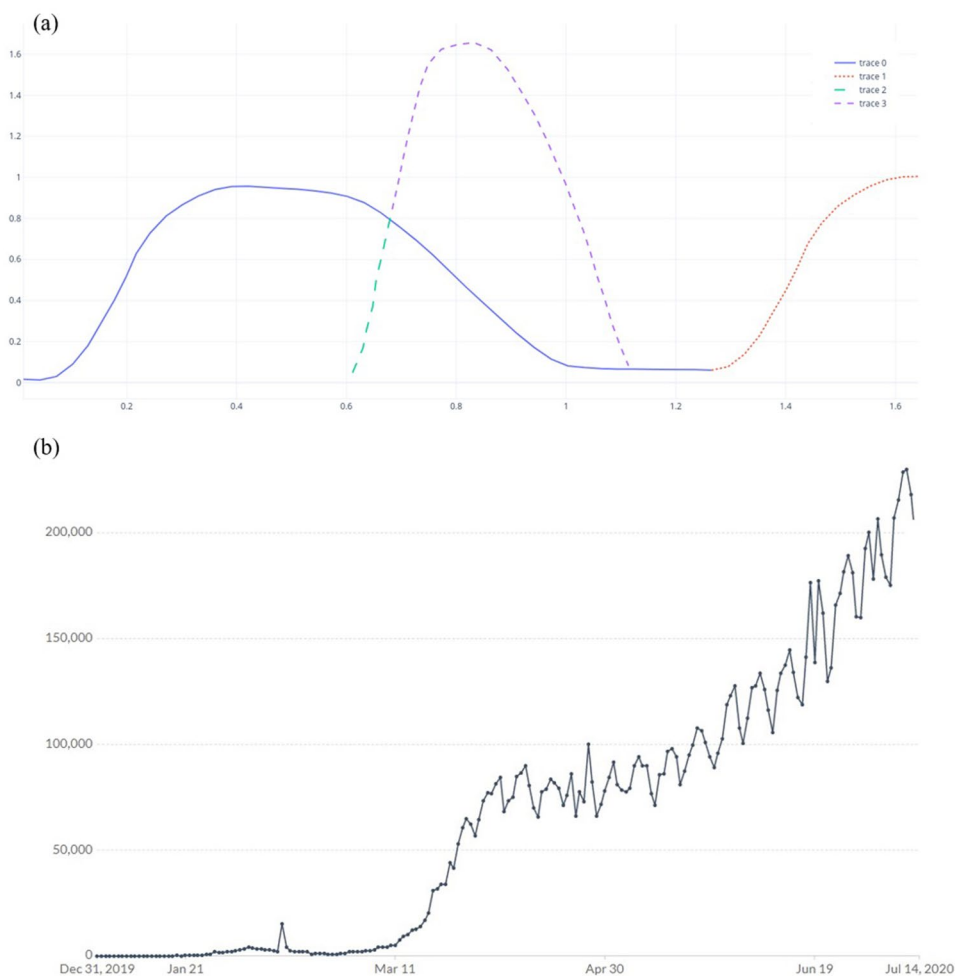Three different approaches were used here:

– Holt's linear trend
– Exponential trend
– Additive damped trend

4.2 Error trend seasonality (ETS) model

ETS stands for error trend seasonality [13], based on the nature of data, we can choose the model to forecast values. Other parameters such as trend, damped, seasonal, seasonal periods, box-cox transform, remove bias are adjusted and RMS values are calculated for all to get the top three best working models for our data.

4.3 AR, MA, and ARIMA Model

Auto-Regressive (AR) is used to describe certain time-varying processes and is a representation of a type of random process in statistics, econometrics, and signal processing [14]. The moving average (MA) model is used in time-series analysis and is a common approach for modeling univariate

Fig. 6 **a** Trace 0 indicates the regular cycle of the coronavirus if the lockdown gets implemented throughout the pandemic. Trace 1 indicates a second cycle after the rest phase which may occur in case the entire population does not infected. Trace 2 occurs when the government policies change. Trace 3 shows the actual effect of the changes made in the government policies which can be seen after 14 days of the policy changes.The red dotted line indicates the beginning of the second cycle after the rest phase was completed.The purple dotted line indicates that even though the first cycle was not completed, a second cycle was overlaid on the first cycle indicating a higher peak than the first peak. **b** The normalized graph of the world indicates a second cycle that has been overlaid to the first cycle during the phase 2 of the first cycle. The number of cases rise exponentially and the peak is higher than first one

time series [14]. A class of models that explains a given time series based on its own past values is 'Auto-Regressive Integrated Moving Average' (ARIMA) [15]. It uses its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

### 4.4 Extend susceptible infected recovered (SIR) model

SIR model [16] is used to simulate disease and hence help us find trends as well as possible future predictions for the spread of disease. In the Extended model we can include more parameters like Exposed, that is, people who have been exposed but not yet showing symptoms. Further, we add cases that have been isolated and deaths, to get closer

to the real world. Also, parameters like incubation period for the virus, mortality rate, number of people an infected infects per day, etc. are taken from real-world data as well as manual fitting of data.

4.5 Extended—SIR MODEL 2 (Curve Fitting)

Similar to previous, but we use limit module for python for best-fit curve fitting and also add more parameters like age group, critical conditions, number of beds leading to triage, etc.

Fig. 7 A simulation curve for the count of COVID-19 for different countries. **a** India and **b** Brazil have still not reached EOP1 and he numbers of cases are rising exponentially. **c** Russia is currently in phase 2 showing a flat curve indicating the number of cases increased at the same rate. **d** In the case of Bangladesh, Phase 2 has just begun. **e** Germany and **f** Italy show similar normalized nature of the curves and are in the rest phase. **g** USA shows an overlaid cycle after the phase 3 because of the relaxation of the lockdown, whereas **h** Australia shows a second cycle after the rest phase
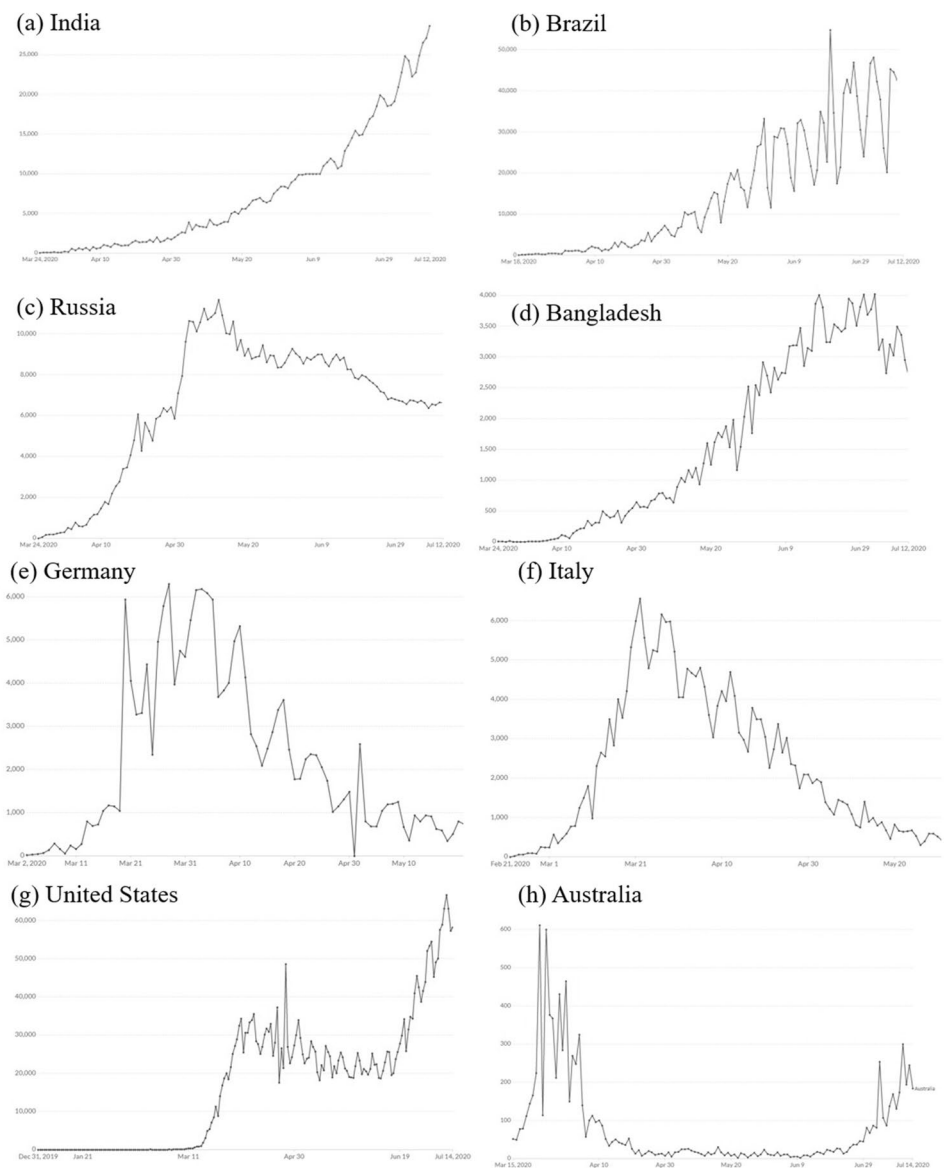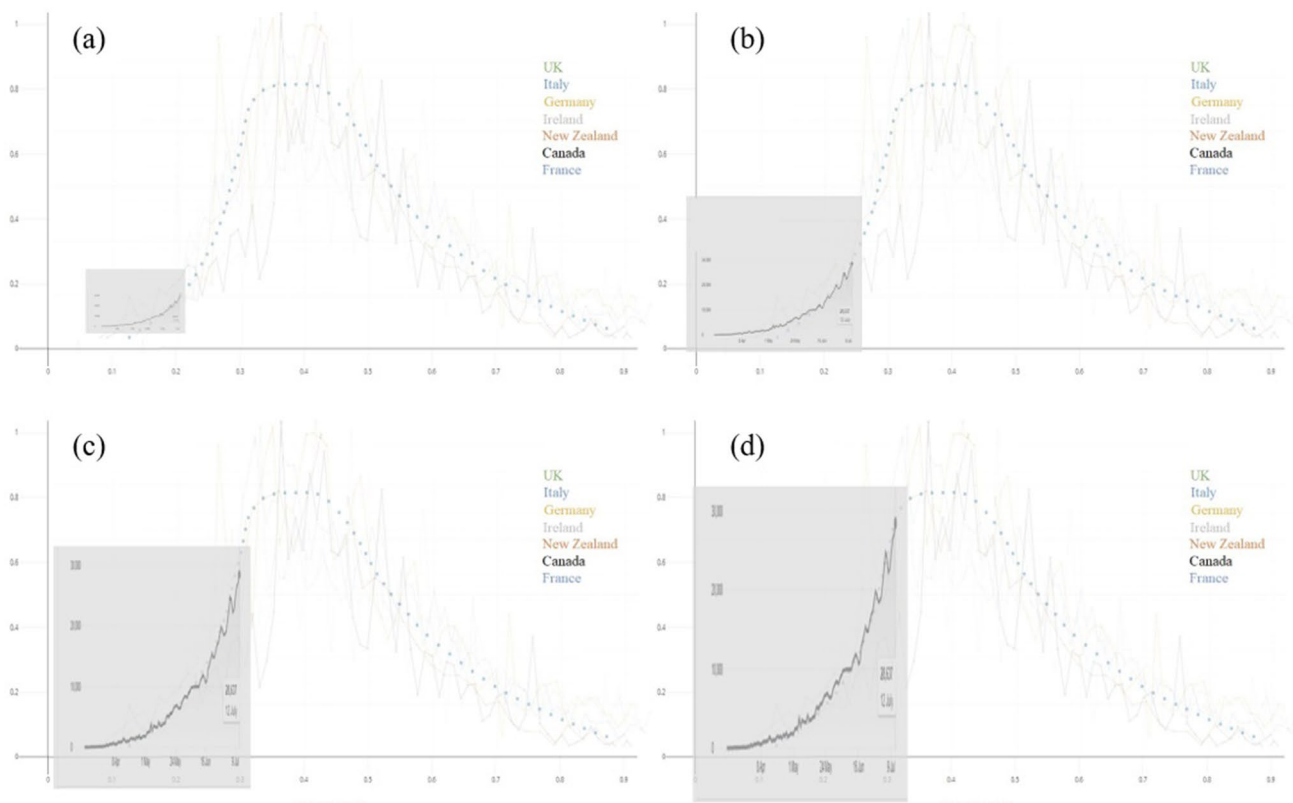
Fig. 8 From the deep learning curve fitted on the normalized graphs of different countries, the curve for India has been fitted in all possible ways to the general trend observed. 4 different possibilties were found. The gray slab in all the figures indicates the current status of India from 15 March to 12 July,2020. **a** indicates the worst case scenario with count going up to 238,743. **b** and **c** indicate moderate scenarios with count reaching up to 102,138 and 41,000 respectively. **d** shows the best case scenario with the maximum cases reaching up to 31,900

4.6 Forecasting with neural networks

An artificial recurrent neural network used in deep learning is Long short-term memory (LSTM) having feedback connections. It processes single data points and also entire sequences of data. It overcomes the vanishing gradient problem and is a recurrent neural network trained in backpropagation through time [17].

4.7 Data analysis-based prediction model

As it is not possible to predict the count of the cases using the time-series prediction algorithm, we used a data analysis-based prediction model. As seen in Fig. 5, the normalized graphs of the countries namely, UK, Italy, Germany, Ireland, New Zealand, Canada, and France where the spread of coronavirus has stopped have been plotted. After analyzing the data for those countries, we fitted a curve shown by the blue colored dotted line with the help of the deep learning model on the normalized graphs indicating the 3 phase nature of the curve for people testing positive for COVID-19. An

exponential rising phase followed by a flat curve and finally a decay in the curve constituting a cycle for the normalized curve of the COVID-19 infected new patient's daily count versus normalized timeline can be seen. This proves the 3-phase nature of the curve followed by a rest phase shown in Fig. 1.

As seen in Fig. 1, we can observe a cycle in the normalized curve. This cycle consists of 3 phases. Phase 1 is the exponential rise where the cases keep on increasing up to a point End of Phase 1 (EOP1). After EOP1, the number of cases increases at a constant level, and a flat curve can be seen indicating the phase 2. After End of phase 2 (EOP2), the cases start to decline gradually following a skewed curve and the phase 3 begins up to a point End of phase 3 (EOP3). These 3 phases together constitute a cycle for the coronavirus cases. After EOP3 a rest phase begins. A second cycle may begin after the rest phase in some cases or a second cycle may overlay on the first cycle itself, depending on the relaxation of lockdown and the government policies.

From the general trend obtained from the data analysis, we tried to fit the curve of India from 15 March to 12

July 2020 in all the possible ways to the general trend. The coronavirus is spreading very quickly in India and the number of cases is increasing drastically. We could get only 4 different possibilities for fitting the curve obtained by the deep learning model as shown in Fig. 8 and the values of EOP1, EOP2, EOP3, the maximum count during the phase 2 were calculated for all the 4 possibilities.

5 Results

All the predictions in the mentioned curve are with the assumption, that there will not be any vaccine or medicine developed during the normalized timeline period. In the case of a vaccine gets developed and distributed the nature of the curve will be affected. As seen in Fig. 2a AR, MA, and ARIMA models do not give a good fit on the data. Whereas in Fig. 2b SIR model shows a close fit according to the actual data. Fig. 2c uses the RNN model for the fitting which also does not give a good fit as required. Figure 2d uses the Holts algorithm for the fitting of the data. Holts 1 and Holts 2 give a better fit on the data as compared to the Holts 3 algorithm. But it is practically not feasible to determine the count for an infinite number of days using the above-mentioned models. A front end was developed for the prediction software as shown in Fig. 3a. The input given was a .csv file which consists of two rows. One row consists of the date and the second consists of the corresponding count on that date. The Run button generates the predicted next day count and the RMSE values for each algorithm (E-SIR, RNN, Holts, AR, MA, ARIMA). The 6 algorithms that we used for the prediction were merged into a single application which would generate the prediction graph as shown in Fig. 3b for the next 10 days by just clicking on the "Generate Graph" button.

Table 1 indicates the Actual count, predicted count of COVID-19 patients, the RMSE values, and the accuracy for different time series models. The actual count patient was given as 28,637 to model. Among all the models the nearest

count predicted by the SIR model which was 28,486. The Root Mean Square Error (RMSE) fount to be for this model was 50. The most inefficient model found among all models was MA. MA estimated a count of 31,197. The RMSE value found in this model was 888 and the accuracy was 88.94 %. Holts models 1, 2, and 3 respectively detected estimated count as 28,018, 28,455, 27,377 with RMSE values of 342.2, 451.8, and 1097.9. The models named ETS 1 and ETS 2 estimated count of patients 27,648 and 26,982 with an RMSE value of 1062 and 1073 respectively. According to the table, we found that the SIR model was nearly able to predict the count of COVID-19 patients correctly with an accuracy of 99.82 %. Table 2 gives the values of the EOP1, EOP2, EOP3, and the maximum cases predicted for India. Figure 8a shows the worst-case scenario with maximum cases reaching 238,743. The pandemic is likely to end on 04-04-2022. Figure 8b and c indicate a moderate case scenario with maximum cases reaching up to 102138 and 41000 respectively. Figure 8d indicates the best case scenario with cases reaching up to a maximum of 31,900 and the pandemic would end in 24-03-2021.

5.1 Comparison with existing literature

A comparison of our proposed models with the one's reported in the literature has been shown in Table 3. Wong [18] did a comparison of different pandemic search models

Table 2 Different possible outcomes for newly reported daily cases in India

	EOP1	EOP2	EOP3	Maximum cases in a day
Figure 8a	16-10-2020	02-04-2021	04-04-2022	238743
Figure 8b	31-08-2020	05-11-2020	25-06-2021	102138
Figure 8c	24-07-2020	20-09-2020	24-04-2021	41000
Figure 8d	13-07-2020	15-09-2020	24-03-2021	31900

Table 1 Comparison of different COVID-19 count prediction methods

Method	Actual value	Predicted value	RMSE	Accuracy
HOLT1	28637	28018	342.2	98.80504243
HOLT2	28637	28455	451.8	98.42232077
HOLT3	28637	27377	1097.9	96.16614869
ETS1	28637	27648	1062.3	96.29046339
ETS2	28637	26982	1073.1	96.25274994
AR	28637	30112	917.7	92.05399841
MA	28637	31197	888	88.94765522
ARIMA	28637	30112	917.1	92.05599097
RNN	28637	28532	224.7	99.21535077
SIR	28637	28486	50.5	99.82365471
Data analysis-based	28637	28693	56	99.60966089

Table 3 Comparison with different methods reported in literature

Method	Prediction accuracy (%)
Wong et al. [18]	99
Roda et al. [19]	95
Roosa et al. [20]	98
Salgotra et al. [19]	99.99
Proposed Method (SIR)	99.82
Proposed Method (Data analysis-based)	99.60

in the time domain and found out that the highest reported accuracy can go up to 99 %, Roda [26] on the other end generated genetic programming for predictions on India and achieved 95 % accuracy with fractals. Roosa [20] used the forecasting model in China and could reach only 98 % accuracy. Compared to the above-mentioned works, our accuracy with SIR and data analysis based prediction is much higher i.e, 99.82 % and 99.6 %. The highest accuracy to date was achieved by Salgotra [19] reported 99.99 % but only for a region of Delhi and their accuracy in other regions has gone down below 98.8 %. We predicted the curve to be a combination of an exponential curve (phase 1) then a flat curve (phase 2), and finally linear decay (phase 3), unlike the bell-shaped curve as predicted by most researchers [18, 21–23] and simulations [24, 25].

5.2 Discussions

As shown in Fig. 6a Trace 0 indicates the regular cycle of the coronavirus if the lockdown gets implemented throughout the pandemic. Trace 1 indicates a second cycle after the rest phase which may occur in case the entire population does not infect. Trace 2 occurs when government policies change. Trace 3 shows the actual effect of the changes made in the government policies which can be seen after 14 days of the policy changes. The red dotted line indicates the beginning of the second cycle after the rest phase was completed. The purple dotted line indicates that even though the first cycle was not completed, a second cycle was overlaid on the first cycle indicating a higher peak than the first peak. Figure 6b indicates the normalized curve of the world. In the case of the world, a second cycle was overlaid on the first one during phase 2 and a rising curve peak can be seen. As seen in Fig. 7a India and Fig. 7b Brazil are still in the exponential rising phase and have not yet reached EOP1. From Fig. 7c, Russia is in Phase 2 of the cycle where the cases are increasing at a constant level. As seen in Fig. 7d Bangladesh has also reached EOP1 and is in phase 2 of the cycle. Germany and Italy as shown in Fig. 7e and f respectively, are in the rest phase. Both the countries show similar normalized

nature of the curve, the reason being they both lie in the southern hemispheres. Their nature of curves is similar and not the number of cases. In the case of the USA as seen in Fig. 7g, a second cycle has been overlaid onto the first cycle during phase 2, which resulted in a rise in the curve. In the case of Australia, as seen in Fig. 7h, a second cycle has started after the completion of the first cycle when Australia was in the rest phase.

6 Conclusions

We have found that the coronavirus trend is a cycle having 3 phases. The majority of the countries worldwide follow the 3-phase curve and there are many countries like India which are still in the rising phase. The effect of lockdown and government policies may affect the nature of the graph which may then consist of an overlaid cycle or a second cycle altogether. Data analysis-based prediction model helped in analyzing the general trend of the simulation curve and predict valuable information for India. With this manuscript, we would like to alert the population about the trend that we have analyzed with multiple models. And also, would like to give away the codes we developed for future researchers to develop further.

Declarations

References

1. Kane MJ, Price N, Scotch M, Rabinowitz P (2014) Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks. BMC Bioinform 15(1):276
2. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, Chen G et al (2020) Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. Clin Infect Dis 71(6):1393–1399

3. Qasim M, Ahmad W, Yoshida M, Gould M, Yasir M (2020) Analysis of the worldwide corona virus (covid-19) pandemic trend: a modelling study to predict its spread, medRxiv

4. Arti M, Bhatnagar K Modeling and predictions for covid 19 spread in India, ResearchGate. https://doi.org/10.13140/RG.2.2.11427.81444

5. Dhanwant, JN, Ramanathan V (2020) Forecasting covid 19 growth in India using susceptible-infected-recovered (sir) model, arXiv preprint arXiv:2004.00696

6. Tandon H, Ranjan P, Chakraborty T, Suhag V (2020) Coronavirus (covid-19): arima based time-series analysis to forecast near future. arXiv preprint arXiv:2004.07859

7. Chen YC, Lu PE, Chang CS, Liu TH (2020) A time-dependent sir model for covid-19 with undetectable infected persons. arXiv preprint arXiv:2003.00122

8. Zhang J, Man K (1998) In: SMC'98 conference proceedings. 1998 IEEE international conference on systems, man, and cybernetics (Cat. No. 98CH36218), vol. 2, pp. 1868–1873

9. Chatfield C, Yar M (1988) Holt-winters forecasting: some practical issues. J Royal Stat Soc: Series D (TheStatistician) 37(2):129–140

10. Roser Max, Ritchie Hannah, Ortiz-Ospina Esteban, Hasell Joe (2020) Coronavirus pandemic (COVID-19), our world in data. https://ourworldindata.org/coronavirus

11. Kalekar PS (2004) Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi School Inform Technol 4329008(13)

12. Taylor JW (2003) Exponential smoothing with a damped multiplicative trend. Int J Forecast 19(4):715

13. Jain G, Mallick B (2017) A study of time series models arima and ets. Available at SSRN 2898968

14. Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the arma, arima, and the autoregressive artificial neural network models in forecasting the monthly inflow of dez dam reservoir. J Hydrol 476:433

15. Harvey AC (1984) A unified view of statistical forecasting procedures. J Forecast 3(3):245

16. Wang D, Zhao X (2003) Empirical analysis and forecasting for SARS epidemic situation. Beijing da xue xue bao Yi xue ban= J Peking Univ Health Sci 35:72–74

17. Vermaak J, Botha E (1998) Recurrent neural networks for short-term load forecasting. IEEE Trans Power Sys 13(1):126

18. Wong JB (2020) Pandemic surge models in the time of severe acute respiratory syndrome coronavirus-2: Wrong or useful? Ann Int Med 173(5):396–398

19. Salgotra R, Gandomi M, Gandomi AH (2020) Time series analysis and forecast of the covid-19 pandemic in india using genetic programming. Chaos, Solitons & Fractals p, p 109945

20. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, Yan P, Chowell G (2020) Real-time forecasts of the covid-19 epidemic in china from February 5th to February 24th, 2020. Infect Dis Modell 5:256

21. Kolozsvari LR, Berczes T, Hajdu A, Gesztelyi R, TIba A, Varga I, Szollosi GJ, Harsanyi S, Garboczy S, Zsuga J (2020) Predicting the epidemic curve of the coronavirus (sars-cov-2) disease (covid-19) using artificial intelligence, medRxiv

22. Fuller J (2020) Models V. Evidence, Boston Review. Serial on the Internet

23. Utsunomiya YT, Utsunomiya ATH, Torrecilha RBP, Paulan SDC, Milanesi M, Garcia JF (2020) Growth rate and acceleration analysis of the covid-19 pandemic reveals the effect of public health measures in real time. Front Med 7:247

24. Tian L, Li X, Qi F, Tang QY, Tang V, Liu J, Cheng X, Li X, Shi Y, Liu H, et al., (2020) Quantifying the infected population for calibrated intervention and containment of the covid-19 pandemic. https://arxiv.org/pdf/2003.07353v3.pdf

25. Schüttler J, Schlickeiser R, Schlickeiser F, Kröger M (2020) Covid-19 predictions using a gauss model, based on data from April 2. Physics 2(2):197

26. Roda WC, Varughese MB, Han D, Li MY (2020) Why is it difficult to accurately predict the covid-19 epidemic? Infect Dis Modell