Noname manuscript No.
(will be inserted by the editor)

Review of Forecasting Models for Coronavirus (COVID-19) Pandemic in India during Country-wise Lockdown

Abhinav Gola · Ravi Kumar Arya* · Animesh · Ravi Dugh

Received: date / Accepted: date

Abstract COVID-19 is spreading widely across the globe right now. Majority of the countries are relying on models and studies such as stochastic simulations, AceMod model, neural networks-based models, exponential growth model, Weibull distribution model, and so on to forecast the number of COVID-19 cases in the coming months. The objective on utilizing these models is to ensure that strict measures can be enacted to contain the virus spread and also predict the resources required to deal with the pandemic as the disease spreads. In the past few months, several models were used to predict the infection rate for COVID-19. These models predicted the infection rates, recovery rate or death rates for the COVID-19 patients. All these different models took different approaches and different scenarios to predict the future rates. Now, that we know the real cases, we can check how accurate these models were. Some of these models were able to predict the near future quite close to the reality but most of them went astray. In this study, we review major forecasting models that were used in the context of India during

Abhinav Gola
Department of Electrical and Electronics Engineering
National Institute of Technology, Delhi
New Delhi, India
E-mail: 171230003@nitdelhi.ac.in

Ravi Kumar Arya *
Department of Electronics & Communication Engineering
National Institute of Technology, Delhi
New Delhi, India
E-mail: raviarya@nitdelhi.ac.in

Animesh
Department of Electrical and Electronics Engineering
National Institute of Technology, Delhi
New Delhi, India
E-mail: 171230013@nitdelhi.ac.in

Ravi Dugh
Electrical and Electronics Engineering Department
The University of Rochester
Rochester, NY, USA
E-mail: rdugh@ur.rochester.edu

country-wise lockdown and compare them. From these comparisons, we can see that while the advanced warning can be helpful in mitigating and preparing for an impending or ongoing epidemic, effects can be catastrophic if poorly fitting models are used for predictions.

Keywords COVID-19 · SARS-CoV-2 · forecasting models · epidemics · pandemics · computational epidemiology · machine learning method · prediction · COVID-19 prediction · linear regression (LR) · multilayer perceptron (MLP) · vector autoregression (VAR) · ARIMA · India

1 Introduction

COVID-19 pandemic, also known as the coronavirus pandemic. is currently wreaking havoc in more than 200 countries [1] worldwide. It is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Till now, at least 17 million people are diagnosed with this virus and this extremely infectious disease has left more than 668 thousand people dead. These numbers are increasing every day and no respite from this disease can be seen until vaccines are invented.

The virus spreads through contact, mainly through coughing, sneezing or just being in close proximity to an infected person. The droplets released by the coughing or sneezing person can travel several feet and can infect the nearby people. If these droplets fall on nearby surfaces, they can contaminate such surfaces and lead to infecting others in contact with such surfaces. Currently there are actually no medications to prevent or treat this epidemic [2] but many biotech companies have stepped in the race of developing vaccine with crores of rupees at stake.

On 30 January 2020, the first COVID-19 occurrence was registered in India. According to the Ministry of Health and Family Welfare (MoHFW) [3], more than 1.6 million people have been infected by this disease. Coronavirus is also responsible for more than 36,000 deaths till now. India has the highest confirmed case in Asia since 16 June, 2020 [1].

Different nations imposed different methods to combat the spread of this disease. These varied for quarantine, mask use, closing of markets/malls, restricting social gatherings, travel restrictions, frequently washing hands, sanitizing frequently used areas etc.. Some countries imposed lockdowns in different states to restrict outbreak of the disease further. In India, to counter the spread of this deadly disease, several lockdowns were imposed country-wise. First nationwide lockdown was imposed on March 24, 2020 for 21 days. The lockdown was further extended on April 14, 2020 till May 3, 2020. There were further extensions starting May 3 and May 17, 2020. With a few exceptions, the nation began unlocking on 1 June 2020 [3]. We use the prediction models during these lockdown times. During this time, as the lockdwon was at national level, there were less variables in the supposed model. After these lockdowns, India went into partial lockdowns where some states of India were under lockdowns while others were not.

Several prediction models have been used over the past several months to predict the COVID-19 infection rate. These models forecast the rate of infection, recovery, death or some combination of these 3 parameters for the COVID-19 patients. The primary evaluation variables can be split into two groups in order to test a forecast model - virus intrinsic: time of incubation, element of virulence, etc. and outside: size, quarantine, and so on of the infected population.

While no model can accurately forecast the rates of infection and mortality, attempts have been made to consider and analyze the strengths and shortcomings of many studies and models presented regarding the coronavirus. Whereas the forecast models used by the health department or Government of India were not disclosed, we can definitely continue with existing models in separate publications. Each of these models took different approaches and techniques to predict the future rates. In this study, we review major forecasting models that were used in the context of India and compare them with real/true data. In this way, the difference of confirmed cases of COVID-19 between forcasted and true cases gives us the error by which these models made bad predictions. For our study, we take published models before national unlocking period. All the prediction models were chosen so that they could predict the cases during lockdowns.

This study is organized into five main sections. The paper starts with the general information about history and information of the disease. Section 2 provides the survey of the multiple forecasting techniques employed to predict the confirmed cases in Indian context. Sections 3 and 4 discuss our findings and results. We conclude this research work in section 5 with the best performing forecasting model along with some interesting observations.

2 Technical Background

Forecasting techniques can be inculcated thereby assisting the government in designing better strategies and in taking productive decisions. These techniques assess the situations of the past, enabling better predictions about the situation to occur in the future. Such predictions will help governments all over the world to prepare for the forthcoming situations. These type of forecasting techniques can play a very important role in yielding nearly accurate predictions.

We provide here a basic introduction to some of these techniques and concepts that were used in the research studies surveyed by us. For further details about different techniques, readers are advised to read research works pertaining to particular used technique.

2.1 Regression Model

Regression analysis is a form of predictive modeling technique which investigates the relationship between a dependent and independent variables. In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

Given a data set of n statistical units, a linear regression model [4] assumes that the relationship between the dependent variable y and the p-vector of regressors x is linear. This relationship is modeled through a disturbance term or error variable ϵ - an unobserved random variable that adds "noise" to the linear relationship between the dependent variable and regressors. Thus, the model takes the general form as shown by Eq. 1.

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where ^{T} denotes the transpose, so that $\mathbf{x}_i^\top \boldsymbol{\beta}$ is the inner product between vectors x_i and β.

Whereas, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as an nth degree polynomial in x. Research work [6] used the polynomial regression for COVIE-19 predictions.

2.2 A Multilayer Perceptron

A multilayer perceptron (MLP) represents any type of feed-forward neural network composed of multiple hidden layers. MLPs [5] are being increasingly used in complex predicting tasks which can't be solved using traditional machine learning methods. An MLP with a single hidden layer can be represented graphically as shown in Fig. 1.

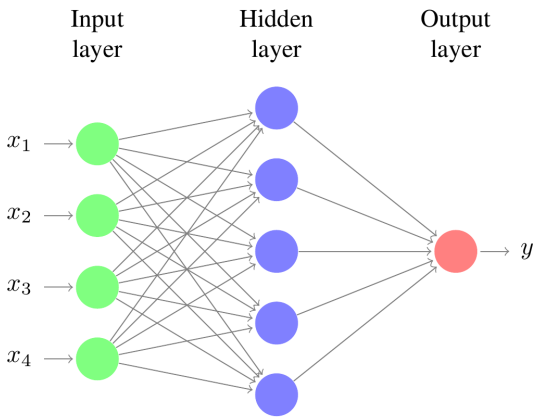


Fig. 1: Multilayer perceptron with one hidden layer

A one hidden-layered MLP can be represented in matrix notation by the following equation:

$$f(x) = G\left(b^{(2)} + W^{(2)}\left(s\left(b^{(1)} + W^{(1)}x\right)\right)\right) \tag{2}$$

where $b^{(1)}, b^{(2)}$ are bias vectors, $W^{(1)}, W^{(2)}$ are weight matrices and G and s are activation functions.

The vector $h(x) = \Phi(x) = s\left(b^{(1)} + W^{(1)}x\right)$ constitutes the hidden layer.

2.3 Vector Autoregression

Vector Autoregression (VAR) [7] is a regression technique capturing the relationship of 2 time series which interact with each other. It is employed where the

interdependence between the time dispositions included is bi-directional. The general equation for VAR model is:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t \tag{3}$$

where α is the intercept, $\beta_1, \beta_2, ...\beta_p$ are the coefficients of the lags of Y and ϵ_t is the error considered as white noise.

2.4 Susceptible Exposed Infected Recovered (SEIR) Model

The SEIR model is constituted of four major factors which are - Susceptible (S) depicting the individuals which can get the disease, Exposed (E) referring to those people which have been already exposed to the disease, Infected (I) depicting the number of people which have been infected and can infect others and Recovered (R) which refers to those people who have became immune to the disease after contracting it once. Fig. 2 gives a pictorial representation of the SEIR Model [6].

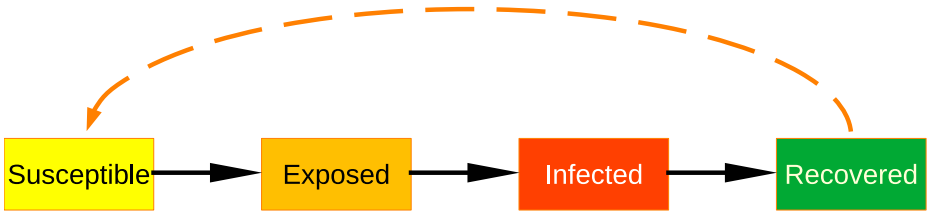


Fig. 2: Graphical representation of SEIR model

2.5 Time Series Forecasting using Waikato Environment for Knowledge Analysis (Weka) Software

Time series analysis is the method of analysing and describing a time-based sequence of data points using statistical techniques. Its forecasting involves the use of a formula to produce future event forecasts based on established events from the past. Time series data have a normal sequential order-different from traditional data mining and computer analysis applications where each data point offers an individual indication of the principle to be studied. Its topics include: capacity preparation, inventory refilling, demand projections and projected staffing ratios [8].

Weka was used to develop the forecasting model [8]. It is basically a series of data mining machine learning algorithms which can be implemented or named from the own Java code directly into a dataset. It provides resources to pre-process, identify, rectify, cluster, associate rules and visualize results.

2.6 Recurrent Neural Networks (RNNs)

Recurrent neural networks (RNNs) fall in the domain of deep learning methods where the output from the previous step is fed as input to the current step. In traditional neural networks, its inputs and output data are independent of each other. But for tasks requiring the use of sequential information, we would require the information from previous computations done by the neural network [9].

Thus, RNNs came into existence, which possessed a 'memory; like feature to remember the information from previously computed calculations. Theoretically, they can make use of this information in randomly long sequences but in practical cases they run into the vanishing gradient problem which inspired researchers to develop LSTMs.

2.7 Long Short Term Memory Networks

Long Short Term Memory (LSTM) was developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. The equations for the gates in a LSTM [9] are as follows:

$$\begin{aligned}
f_t &= \sigma_g\left(W_f x_t + U_f h_{t-1} + b_f\right) && (a)\\
i_t &= \sigma_g\left(W_i x_t + U_i h_{t-1} + b_i\right) && (b)\\
o_t &= \sigma_g\left(W_o x_t + U_o h_{t-1} + b_o\right) && (c)\\
\tilde{c}_t &= \sigma_h\left(W_c x_t + U_c h_{t-1} + b_c\right) && (d)\\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t && (e)\\
h_t &= o_t \circ \sigma_h\left(c_t\right) && (f)
\end{aligned} \tag{4}$$

where i_t represents input gate, f_t represents forget gate, o_t represents output gate, h_{t-1} is the output of the previous lstm block (at timestamp t-1), x_t is the input at current timestamp, b_x represents biases for the respective gates (x), w_s refers to the weight for the respective gate (x) neurons and σ: represents sigmoid function.

Equation 4(a) is for the input gate which tells us what new information we are going to store in the cell state (that we will see below). Second equation 4(b) is for the forget gate which tells the information to throw away from the cell state. Third one 4(c) is for the output gate which is used to provide the activation to the final output of the LSTM block at timestamp 't'. A block of LSTM at any point t is shown in Fig. 3.

2.8 Susceptible Infected Recovered (SIR) Model

SIR model is a disease modelling system like SEIR which segregates the whole community into 3 sections - Susceptible, Infected and Recovered. The definitions of S, I and R sections are same as those of SEIR model. The three categories

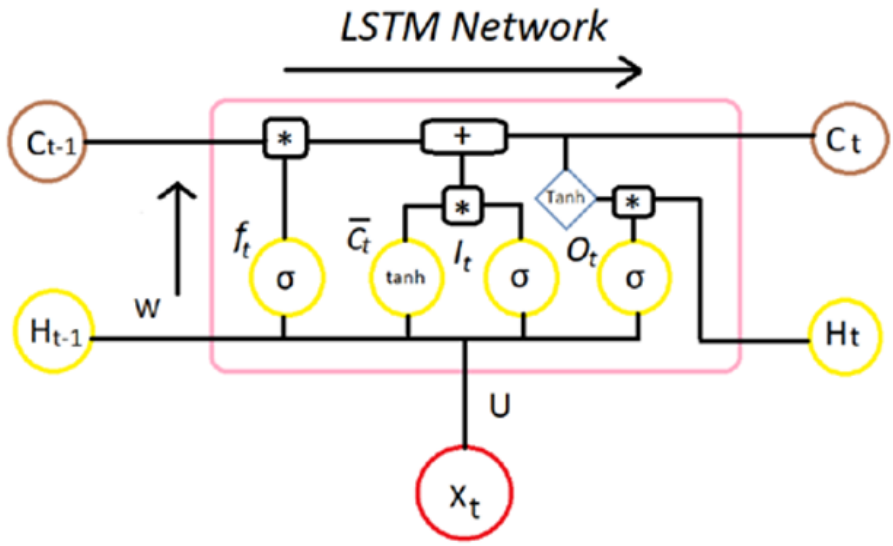


Fig. 3: Block of LSTM network at some point of time 't'

are interrelated with each other and with parameters according to the following equations [10]:

Rate of change of Susceptible Population is given by:

$$\frac{dS}{dt} = -\beta * I * S \tag{5}$$

Rate of change of Infected Population is given by:

$$\frac{dt}{dt} = \beta * I * S - \gamma * I \tag{6}$$

Rate of change of Recovered Population is given by:

$$\frac{dR}{dt} = \gamma * I \tag{7}$$

2.9 Curve Fitting

In this technique, we fit a curve or mathematical function that fits best to a sequence of input data values, conditioned upon some constraints. It can entail use of interpolation, where an accurate fit to the data is a necessity. Otherwise a smoothing function is adopted where only an approximate fit to the points is required [11]. Previous research [12] fitted the following polynomial function to their input data with least squares as their loss function:

$$y = -1\text{E}{-}06\text{x}^6 + 0.318\text{x}^5 - 34942\text{x}^4 + 2\text{E}{+}09\text{x}^5 - 7\text{E}{+}13\text{x}^2 + 1\text{E}{+}18\text{x} - 9\text{E}{+}21 \tag{8}$$

One [13] of the other researchers used an exponential fitting to predict the values. The equation employed is:

$$N(t) = \exp(12.52372 + 0.040t) \tag{9}$$

2.10 ARIMA Model

ARIMA, short for 'Auto-Regressive Integrated Moving Average,' is a family of equations that describe a certain time series dependent upon their prior values, i.e. its own lags and lagged prediction errors, such that the equation can be used to estimate future values. The ARIMA process is also known as the Box-Jenkins method. The Box-Jenkins approach is for a merged ARIMA configuration to be fitted to a particular data set [**?**].

3 Analysis of Past Forecasting Models

3.1 Methodology Used

The analysis of each study is conducted by comparing their forecasted values for a certain duration with the actual (also called true) number of cases for that duration. This was implemented by extracting values from tables and graphs presented in these research papers and then juxtaposing them with the values obtained from Ministry of Health and Family Welfare website of Government of India [3]. We have used Mean Absolute Percentage Error (MAPE) metric for contrasting between predicted and true values. MAPE was chosen because of its scale-independence property which would remove the size of test data bias from the models.

MAPE is a measure of how accurate a forecasting model is and gives its output as a percentage value. It is calculated by normalising the average error at each point. This is done by dividing the errors by actual values and then averaging their sum using the total size of data using the following mathemaical equation:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{10}$$

3.2 Reviewed Past Research Papers

For the purpose of exhaustive research, we considered 10 research studies which focused on forecasting the number of COVID cases in India. These were chosen keeping in mind to include a variety of forecasting techniques being used. These studies are shown in Table 1 along with some other information such as the models used by them, source of their data, duration of their forecasting and the MAPE score they achieved.

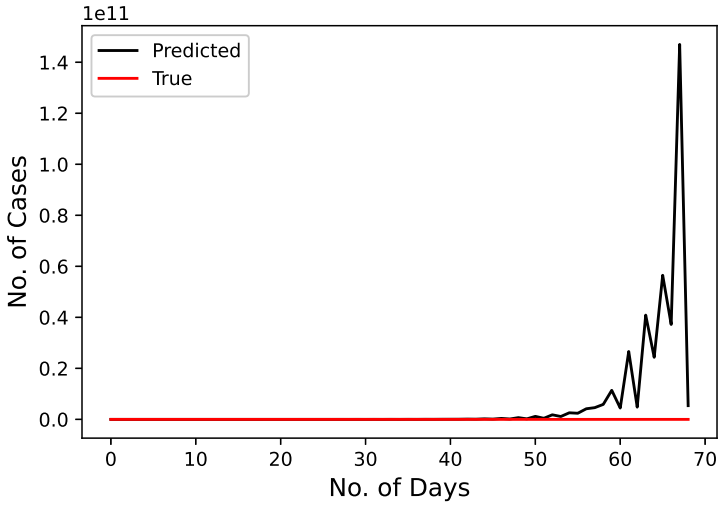
Table 1: Research Papers Reviewed

References	Forecasting Techniques	Duration of Forecasting	Data Source	MAPE score
[5]	LR, MLP and VAR	69 days	Kaggle [17]	LR: 1745454.432, MLP: 80.057, VAR: 43289.290
[6]	SEIR and Regression models	20 days	JHU [18]	SEIR: 25.533, Regression: 21.889
[8]	Time Series Forecasting using Weka	24 days	Kaggle [20]	55.489
[9]	LSTM	25 days	Govt. Of India [21]	63.357
[14]	Genetic programming-based model (GP) [GEP model]	10 days	Mendley Time series datasets [24]	7.827
[15]	Exponential Growth Model and ML based LR model	33 days	Humanitarian website for COVID data	LR: 662.441, Exponential: 2096.000
[13]	Short term forecasting using exponential curve fitting	15 days	[22]	6.4807
[10]	SIR Model	36 days	WHO [23]	66.819
[12]	Least square fitted model	35 days	[22]	39.816

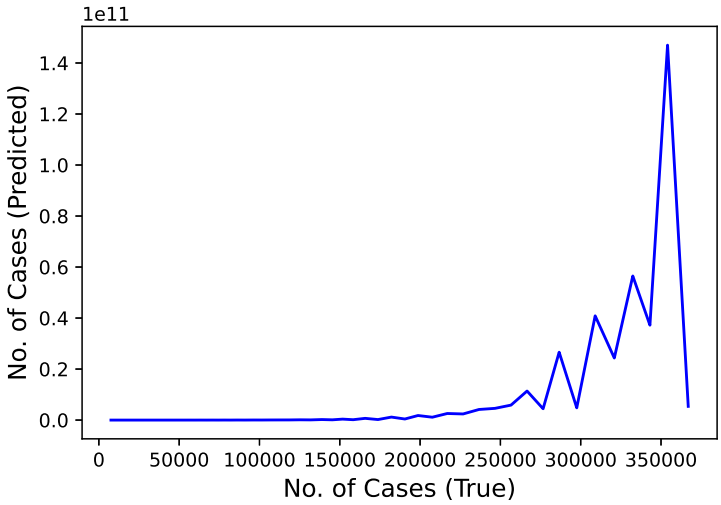
4 Results

In this section we demonstrate how well these models fared against the true number of cases in India. This information is shown by 2 different curves for each study. First plot represents both the predicted and true curve against the predicted number of days while the second plot represents the relationship between the forecasted and true data. Ideally, this plot should tend to achieve linearity.

Figure 4 represents the results of Linear Regression technique proposed by the authors of [5]. This research used the official number of cases from January 22, 2020 to April 10, 2020 as their input data and predicted values for a duration of 69 days starting from April 11, 2020 to June 18, 2020. This model scores really



(a) Comparison of True and Predicted Data

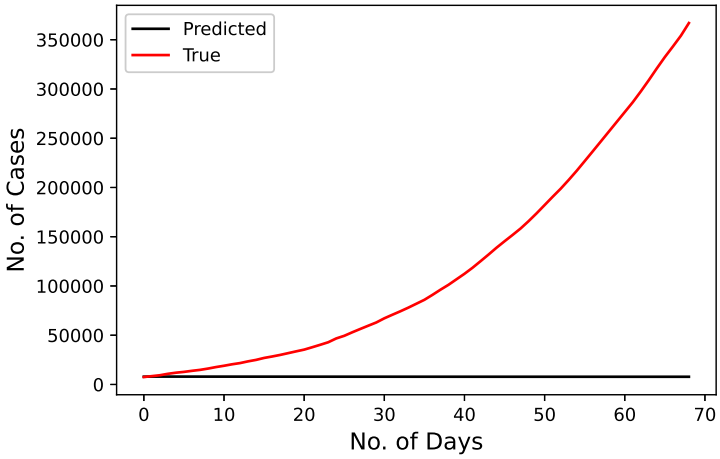


(b) Forecasted vs True datae

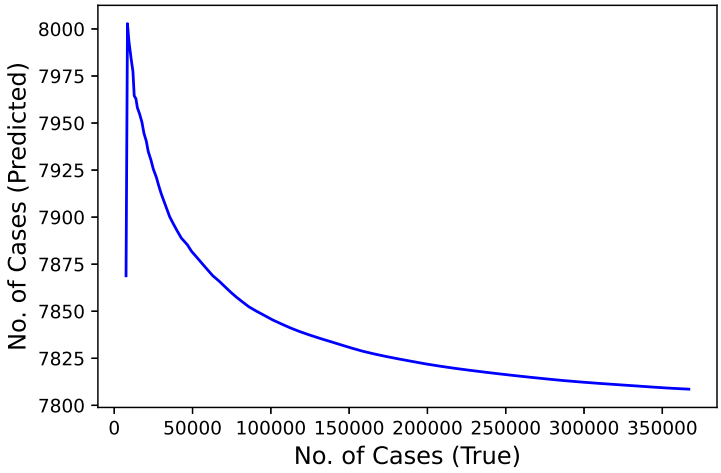
Fig. 4: "A machine learning forecasting model for COVID-19 pandemic in India" [Linear Regression model]

low on our metric with a score of 1745454.432. The error for some points reaches upto the magnitude of 10^{11}. It means that this research was very far off from the true value of cases.

Figure 5 represents the results of Multilayer Perceptron model proposed by the authors of [5]. Duration of both the training and testing data are same as that of the Linear Regression model. This model performed much better than the LR technique with a MAPE of 80.057.



(a) Comparison of True and Predicted Data



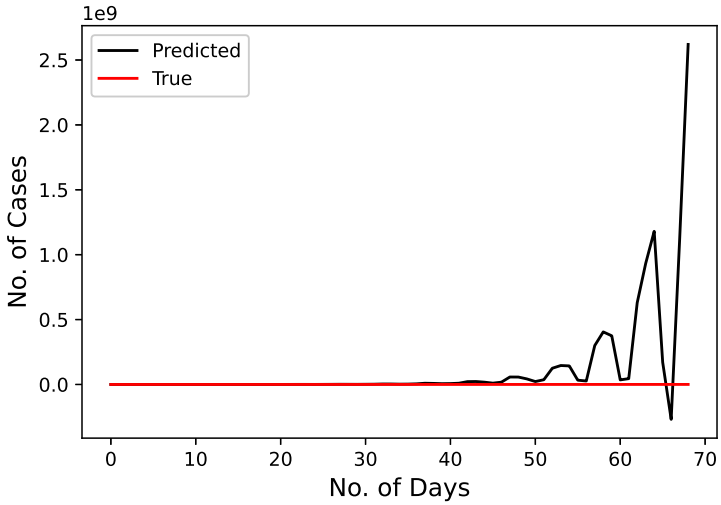
(b) Forecasted vs True datae

Fig. 5: "A machine learning forecasting model for COVID-19 pandemic in India" [Multilayer Perceptron model]

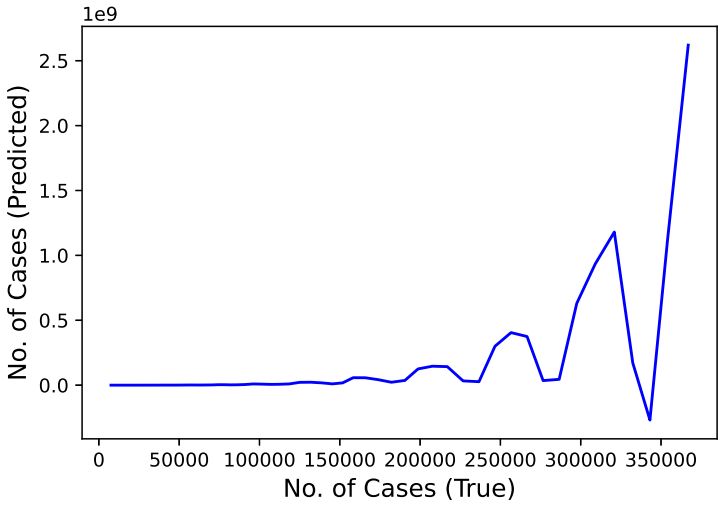
Figure 6 shows the results of Vector Autoregression technique proposed by the authors of [1] on the same input data as previous two techniques. This model again gives a poor performance with MAPE of 43289.29.

Figure 7 displays the results of time series forecasting conducted by the authors of [8]. They used the official no. of cases from January 22, 2020 to April 3, 2020 to train their model. The predicted duration was of 24 days starting from April 4, 2020 to April 27, 2020. This study scores a MAPE score of 55.48984343.

Figure 8 represents the results of the LSTM model used by the authors of [9] to predict for a duration of 25 days from April 5, 2020 to April 29, 2020. They



(a) Comparison of True and Predicted Data

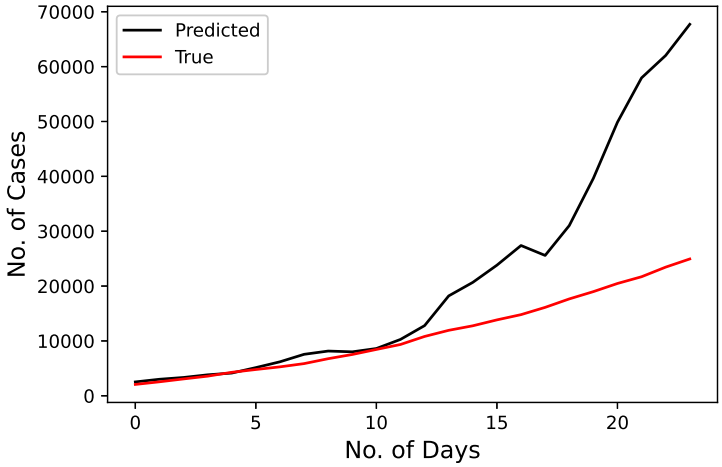


(b) Forecasted vs True datae

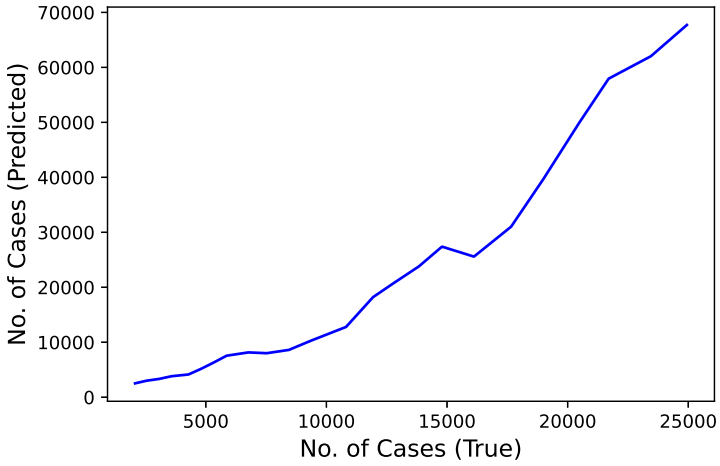
Fig. 6: "A machine learning forecasting model for COVID-19 pandemic in India" [Vector Autoregression model]

have taken data starting from January 30, 2020 to April 4, 2020 for their model. Their system scores good on our metric with a value of 63.357.

Figure 9 represents the results of GEP model proposed by the authors of [14]. They predicted the number of infected cases from May 14, 2020 to May 23, 2020 for a duration of 10 days. Input for their model were the number of official cases from March 24, 2020 to May 13, 2020. This model got a very good score of 7.827.



(a) Comparison of True and Predicted Data



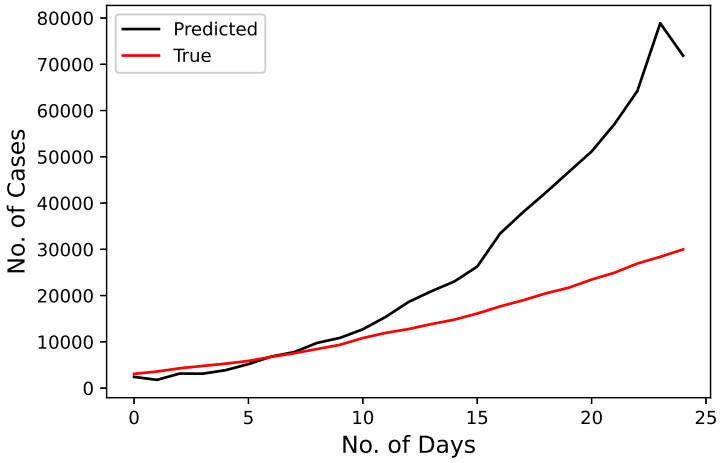
(b) Forecasted vs True data

Fig. 7: "Outbreak trends of Coronavirus (COVID-19) in India: A Prediction. Disaster Medicine and Public Health Preparedness"

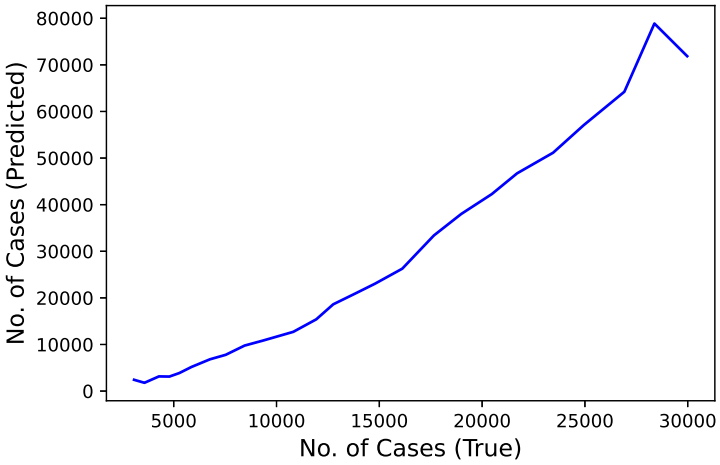
Figure 10 demonstrates the results of Linear Regression technique proposed by the authors of [15]. They used the official number of cases from February 11, 2020 to May 11, 2020 as their input data to train their model. The predicted duration was of 33 days starting from May 12, 2020 to June 30, 2020. This model scores quite low on our metric with a score of 662.441.

Figure 11 displays the results of the exponential model used by the authors of [15]. Size of both training and test data remains analogous to the linear regression technique. This model gives metric result of 2096.000.

Figure 12 demonstrates the results of SEIR model proposed by the authors of [6]. They have used the official number of cases from January 30, 2020 to March



(a) Comparison of True and Predicted Data



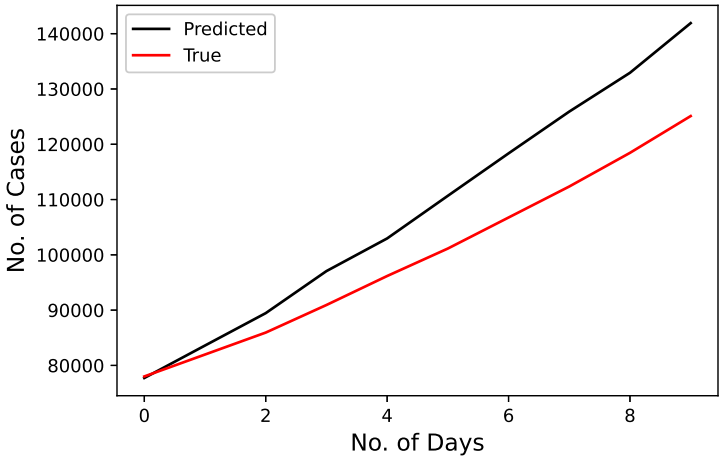
(b) Forecasted vs True data

Fig. 8: "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures"

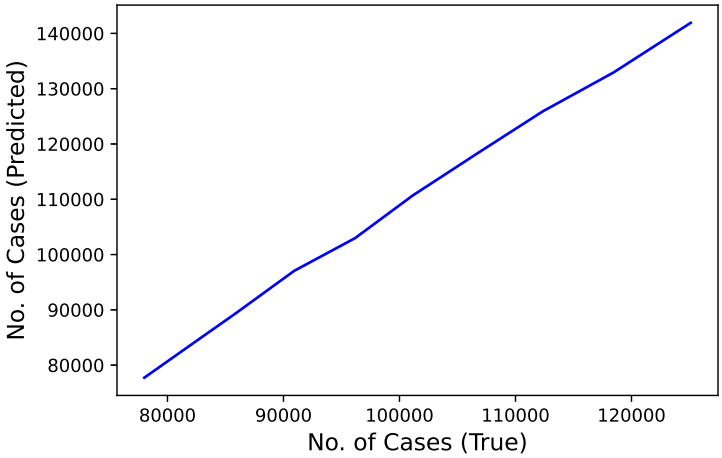
30, 2020 as their input data to train their model. The predicted duration was of 20 days staring from March 31, 2020 to April 13, 2020. This model scores good on the MAPE metric with a score of 25.533.

Figure 13 represents the results of Regression model proposed by the authors of [14]. Duration of both the training and testing data are same as that of the SEIR model. Also, this model performed better than the SEIR technique with a MAPE of 21.889.

Figure 14 represents the results of Least Square Fitted model proposed by the authors of [12]. They have used the official no. of cases from January 30, 2020 to April 26, 2020 as their input data to train the model. The predicted duration was



(a) Comparison of True and Predicted Data



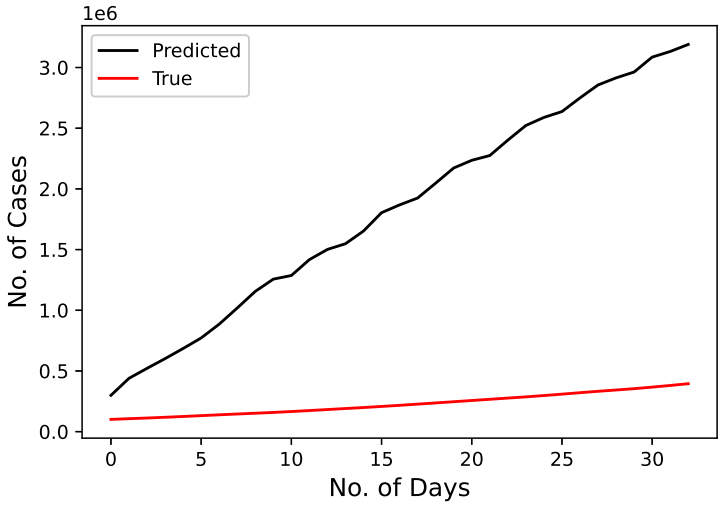
(b) Forecasted vs True data

Fig. 9: "Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming"

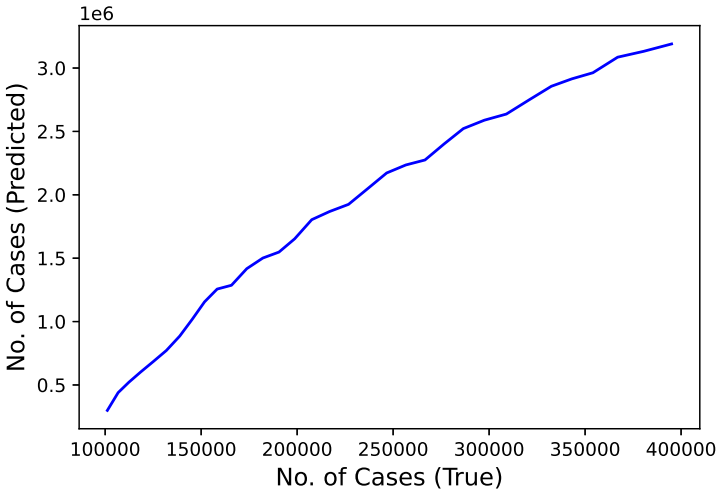
of 35 days starting from April 27, 2020 to May 31,2020. This model gives a MAPE score of 39.816.

Figure 15 represents the results of the SIR model used by the authors of [10]. They predicted the number of cases for a duration of 36 days starting from March 23, 2020 to April 27, 2020 using the data from March 3, 2020 to March 22, 2020 . This model scores a modicum MAPE of 66.819.

Figure 16 represents the results of Linear Regression technique proposed by the authors of [13]. They have used the official number of cases from 1st June,2020 to 10th June, 2020 as their input data/to train their model. The predicted duration



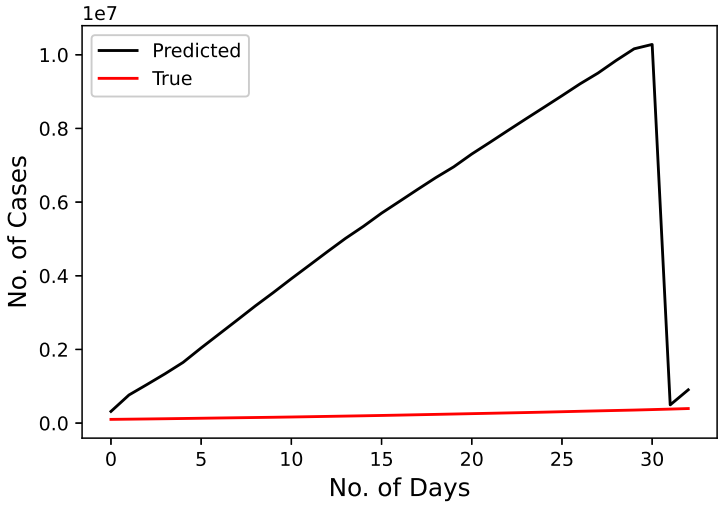
(a) Comparison of True and Predicted Data



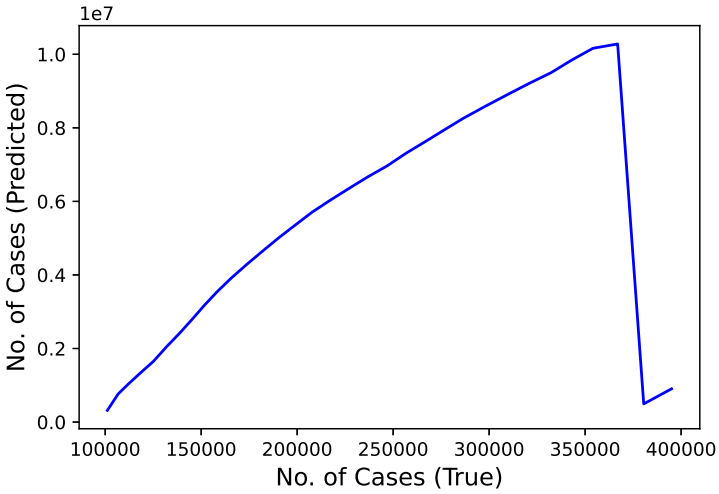
(b) Forecasted vs True data

Fig. 10: "Predicting the COVID-19 positive cases in India with concern to Lockdown by using Mathematical and Machine Learning based Models" [Linear Regression model]

was of 15 days starting from 10th June,2020 to 24th June,2020. This model scores best on the MAPE metric with a score of 6.4807.



(a) Comparison of True and Predicted Data

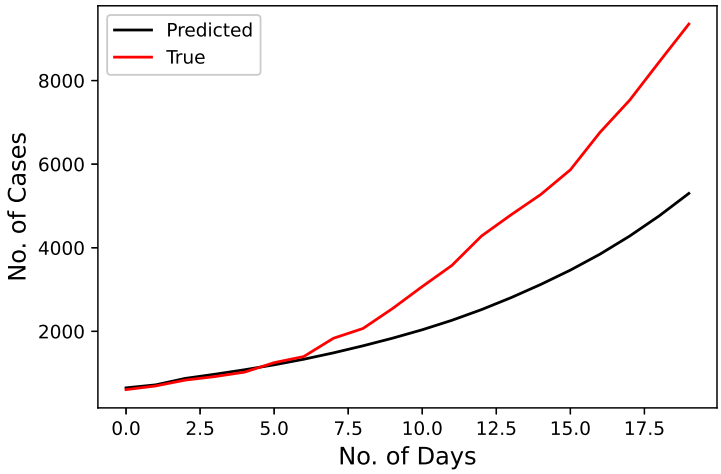


(b) Forecasted vs True data

Fig. 11: "Predicting the COVID-19 positive cases in India with concern to Lockdown by using Mathematical and Machine Learning based Models" [Exponential model]

5 Conclusion

This research work considered data related to daily confirmed cases of COVID-19 in India during lockdowns on national level. From the results shown in the previous section, we observe that even after using a size independent metric, models which predicted for lesser durations (less than 20 days) [14], [13] scored best on our



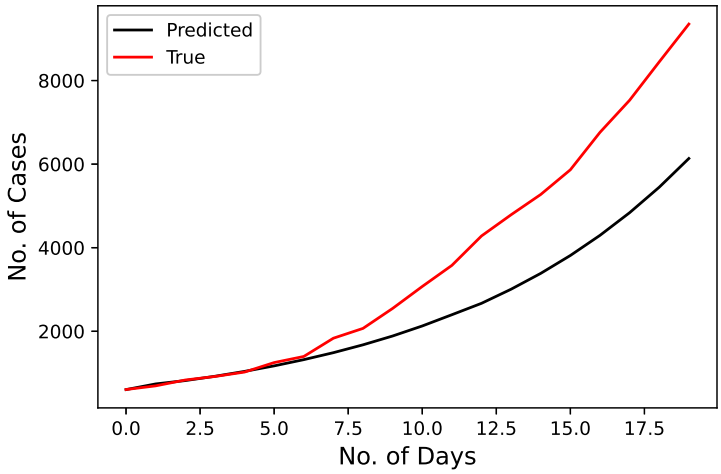
(a) Comparison of True and Predicted Data

(b) Forecasted vs True data

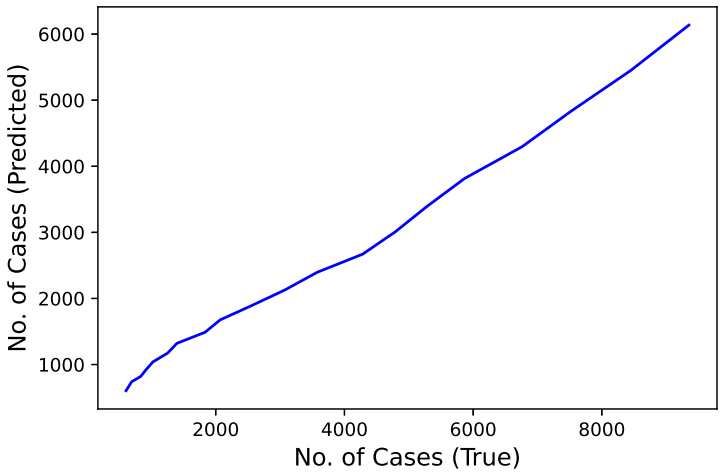
Fig. 12: "SEIR and Regression Model based COVID-19 outbreak predictions in India" [SEIR model]

MAPE metric. Overall, exponential curve fitting [13] on recent data produced the lowest value score of 6.4807. But its use should be limited to only short-term forecasting as predicting values for a longer duration will lead to the fitted curve moving away from the true values.

Discarding these outlier models which conducted forecasting for smaller durations, we conclude that the least square fitted model [12] outperforms all the other techniques and thus, should be used for future purposes of predicting the infection rate under similar conditions. Also, disease modeling systems such as SEIR and



(a) Comparison of True and Predicted Data

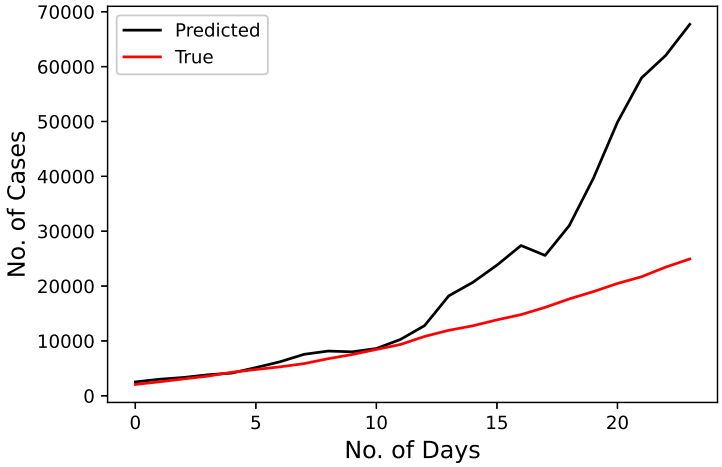


(b) Forecasted vs True data

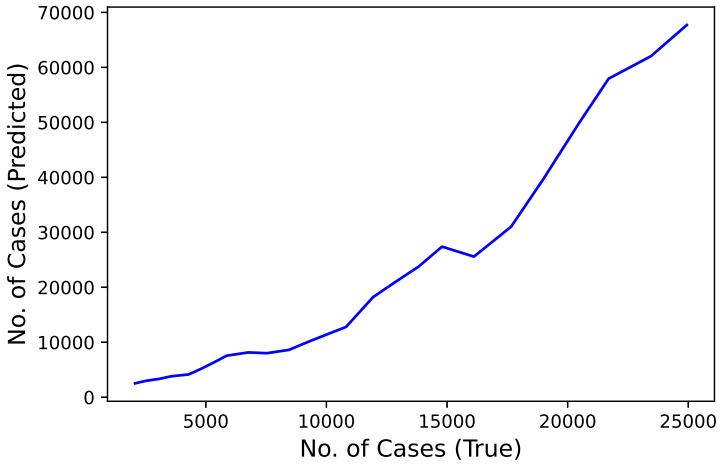
Fig. 13: "SEIR and Regression Model based COVID-19 outbreak predictions in India" [Regression model]

SIR [6], [10] fared much better than other regression models and we believe that if applied with more precision, can achieve good results.

For future purposes, this study can be expanded to include forecasting models which were developed to predict the infection rate for some of the major countries affected by the COVID-19 pandemic. That research would serve as a pivotal guide to determine the apropos forecasting techniques for such tasks. As countries are going to lockdown again, this work can give the guidelines for the future prediction model that can be applied and have least error.



(a) Comparison of True and Predicted Data

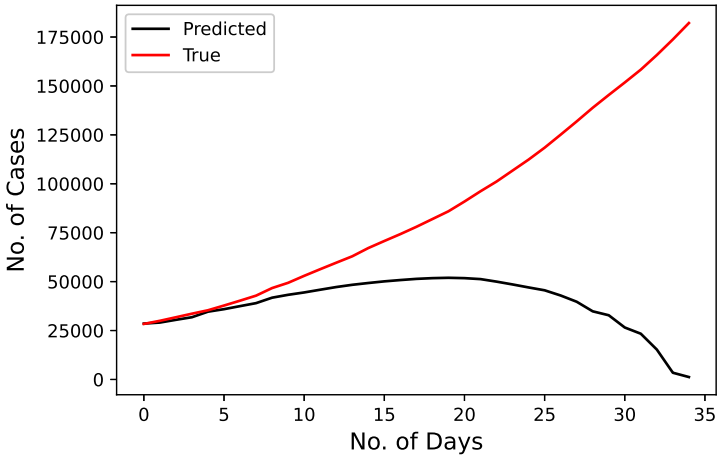


(b) Forecasted vs True data

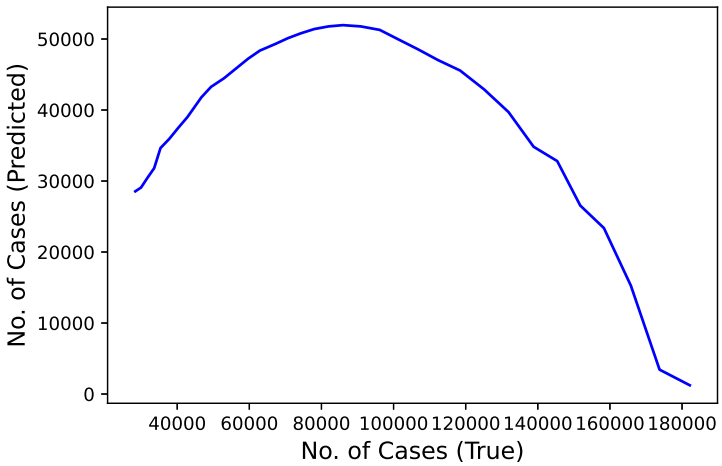
Fig. 14: "OVERVIEW: SITUATION ASSESSMENT AND PREDICTION OF CORONA VIRUS IN INDIA"

References

1. `https://covid19.who.int` accessed on July 31, 2020
2. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses` accessed on July 31, 2020
3. Ministry of Health and Family Welfare,, `https://www.mohfw.gov.in` accessed on July 31, 2020
4. `https://en.wikipedia.org/wiki/Linear_regression`
5. R. Sujath, Jyotir Moy Chatterjee & Aboul Ella Hassanien, "A machine learning forecasting model for COVID-19 pandemic in India" *Stochastic Environmental Research and Risk Assessment*, volume 34, pages 959–972(2020), doi: 10.1007/s00477-020-01827-8



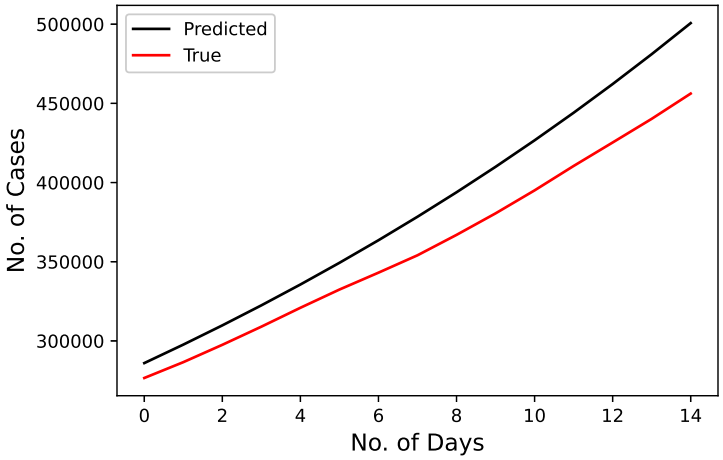
(a) Comparison of True and Predicted Data



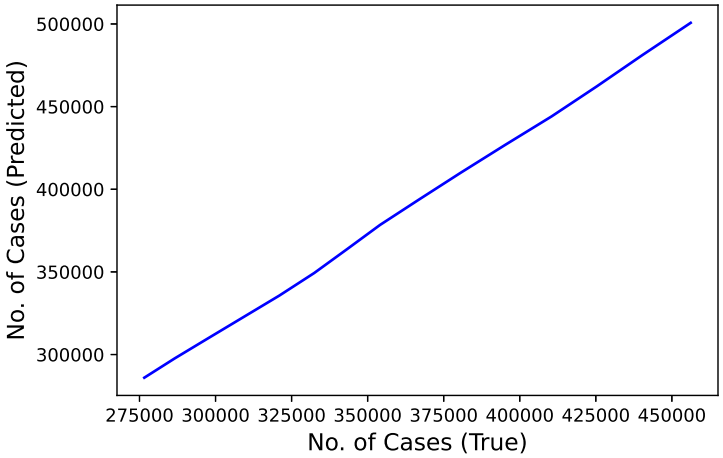
(b) Forecasted vs True data

Fig. 15: Forecasting COVID 19 growth in India using Susceptible-Infected-Recovered (S.I.R) model

6. Gaurav Pandey, Poonam Chaudhary, Rajan Gupta, Saibal Pal, "SEIR and Regression Model based COVID-19 outbreak predictions in India", doi: 10.1101/2020.04.01.20049825
7. https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/
8. Sunita Tiwari, Sushil Kumar, Kalpna Guleria, "Outbreak trends of CoronaVirus (COVID-19) in India: A Prediction" Disaster Med Public Health Prep. 2020 Apr 22: 1–6., doi: 10.1017/dmp.2020.115
9. Anuradha Tomar, Neeraj Gupta. "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures" Science of The Total Environment Volume 728, 1 August 2020, 138762, doi: 10.1016/j.scitotenv.2020.138762
10. Jay Naresh Dhanwant, V. Ramanathan, "Forecasting COVID 19 growth in India using Susceptible-Infected-Recovered (S.I.R) model", arXiv:2004.00696 [q-bio.PE]
11. https://en.wikipedia.org/wiki/Curve_fitting



(a) Comparison of True and Predicted Data



(b) Forecasted vs True data

Fig. 16: "Nearly Perfect Forecasting of the Total COVID-19 Cases in India: A Numerical Approach"

12. Mr. Sudip Ghosh, "An Overview: Situation Assessment and Prediction of Corona Virus in India" Mukt Shabd Journal Volume IX Issue V, MAY/2020 Issn No: 2347-3150
13. Hemanta Kumar Baruah, "Nearly Perfect Forecasting of the Total COVID-19 Cases in India: A Numerical Approach", doi: 10.1101/2020.06.13.20130096
14. Rohit Salgotra, Mostafa Gandomi, Amir H Gandomi, "Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming" Chaos, Solitons & Fractals Volume 138, September 2020, 109945, doi: 10.1016/j.chaos.2020.109945
15. Ajit Kumar Pasayat, Satya Narayan Pati, Aashirbad Maharana, "Predicting the COVID-19 positive cases in India with concern to Lockdown by using Mathematical and Machine Learning based Models" doi: 10.1101/2020.05.16.20104133
16. Shinde, G.R., Kalamkar, A.B., Mahalle, P.N. et al. Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. SN COMPUT. SCI. 1, 197 (2020). https://doi.org/10.1007/s42979-020-00209-9

17. https://www.kaggle.com/imdevskp/corona-virus-report/data
18. https://coronavirus.jhu.edu/
19. WHO. Situation report–1. Novel coronavirus (2019-nCoV). https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
20. Kaggle. Novel corona virus 2019 dataset, https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset
21. https://www.mygov.in/covid-19/?cbps=1
22. https://www.worldometers.info/
23. https://www.who.int
24. https://data.mendeley.com/datasets/tmrs92j7pv/1