Comprehensive Review on Statistical Modeling Approach to Predict the COVID-19 Transmission

Vallaippan Raman(✉), Navin Aravinth, Preetha Merlin Joy, and Kowsalya

Department of Artificial Intelligence and Data Science, Coimbatore Institute of Technology, Coimbatore, India
valliappan@cit.edu.in

Abstract. This study aims to focus on the statistical model for forecasting the transmission of covid-19. The dynamics of the spreading nature can be determined by prediction models. Various prediction models are devised and/or used to know the disease dynamics and the existing ones based on statistical models are being developed for single or multiple countries. Many review articles commonly address the statistical models adopted, whereas the studies indicate effective models that address disease dynamics and forecast potential contagion scenarios viz. Data-driven techniques were created on different parameters. This work aims at collating the basic working philosophies of most cited COVID-19 dynamic prediction model reports by a systematic literature study. The review highlights the dynamic models strength and their weakness in predicting of SARS Covid-19. words.

Keywords: Forecasting · COVID-19 · Statistical Models · Machine Learning Methods

1 Introduction

Wuhan Hospital noted an uncommon case of severe acute respiratory illness caused by a novel virus in December of 2019 and it spread quickly. Because of how similar it is to the original SARS-CoV from 2002, they later refer to it as SARS-CoV-2. The World Health Organization (WHO) refers to this virus as COVID-19, a new coronavirus (nCOV-19). This virus can remain in an individual for up to 14 days without causing any symptoms, which can cause it to evolve from a Wuhan outbreak into a global pandemic that affects the whole world. Machine learning (ML), a part of artificial intelligence (AI), is used by researchers and governments to estimate the number of COVID-19 cases in a timely way and to stop the virus's spread and incubation. Despite the drawbacks of using a prediction model, data collection is complex. Any infectious disease report has been filled with difficulties and problems. The foundation of our approach, like that of many others, depends on the testing and reporting of infections. The suggested model will be a prototype that can forecast the covid-19's short-term predictions. The study's objective is to create a real-time interactive dashboard and chatbot using a statistical model that

S. Jayasingh et al. (Eds.): ICETBM 2023, AEBMR 242, pp. 112–129, 2023.
https://doi.org/10.2991/978-94-6463-162-3_11

predicts the number of incident cases in the future based on past data and offers a clear understanding of dynamic transmission and disease control. The following is how the paper is set up: The literature study is described in Sect. 3 and contains various machine-learning techniques. Section 4 discusses the methodology's importance and Sect. 5 concludes with recommendations for further research.

2 Literature Review

Artificial intelligence, which is widely defined as a machine's ability to mimic intelligent human behavior, includes the subfield of machine learning. Artificial intelligence (AI) systems are used to carry out complicated tasks in a manner akin to how people solve issues. Due to their ability to adapt to the unique characteristics of the data without being constrained by presumptions like decision-making functions that are functional or the variables' probability distribution, machine learning algorithms have emerged as a developed as a common paradigm in modern scientific research [1]. According to Kavadi, Patan, Ramachandran, and Gandomi [2], managing a wide range of data and easily identifying trends and patterns of an unknown nature are the main benefits of using ML approaches. that discusses the epidemic is covered initially. Therefore, we may estimate the future pattern of the disease by using methodologies to look at the increased spread analysis of infectious disease. First proof that ML methods might be used to identify COVID-19-related distress in HCWs early on, opening the door to potential targeted therapies. Machine learning (ML) may be essential for uncovering hidden patterns from the mining of enormous raw datasets and for creating superior predictive models. An effective Gaussian process regression based model for predicting the cured and positive of COVID-19 cases in two significantly afflicted nations was developed using Bayesian optimization to modify the Gaussian process regression hyperparameters [3].

2.1 Deep Learning

Deep learning is a component of machine learning techniques that use representation learning and artificial neural networks. Unsupervised, semi-supervised, and supervised learning are all possible. Machine learning, which is simply a neural network with three or more layers, is a subset of deep learning. These neural networks make an effort to mimic the behavior of massive data sets. For COVID-19 forecasting, deep learning models have been examined. The results reveal that the Variation Auto Encoder model performs well [4]. Because these networks are capable of learning long-term connections between time steps of input, LSTMs are frequently used to learn, process, and classify sequential data. It is a type of artificial neural network used in deep learning and artificial intelligence.

LSTM has feedback connections in contrast to traditional feed-forward neural networks. Such a recurrent neural network may process complete data sequences in addition to single data points. According to [5], an LSTM-based forecasting model with a rolling update mechanism is suggested for the long-term pandemic trend of COVID-19. Multi-layer perceptron's and LSTM-RNN can be used for data modeling and prediction based on multivariate time series. The use of a smoothed the LSTM approach is anticipated to lessen the impact of outliers and increase forecasting precision. The algorithm performs best when predicting the total number of infections for the upcoming week and month. In these provinces, each case's model will be assessed using the root mean squared error (RMSE). For short-term prediction (1–3 days), bi-LSTM produces highly accurate results (error less than 3%) when LSTM variations including deep LSTM, convolutional LSTM, and bi-directional LSTM models are tested on 32 states and union territories [6]. The Pearson Correlation Heat Map shows the linear correlations between various characteristics, and the inferences from the Bi-LSTM model outperform in terms of approved indices. Bi-LSTM can be used for pandemic prediction for improved planning and management due to its demonstrated robustness and increased prediction accuracy [7]. LSTM-Markov model, which employs the Markov model to lessen the LSTM model's prediction error. We estimated the LSTM training errors and built the probability transfer matrix of the Markov model by the errors using confirmed case data from the US, UK, Brazil, and Russia [8].

An explanation of the relationship between one or more independent variables and a response, dependent, or target variable is provided by a regression model. For instance, a linear regression model might be used to explain the link between height and weight. The one can predict how many new COVID-19 occurrences there will be in the near future using a regression-based ensemble learning model that combines Linear regression, Ridge, Lasso, ARIMA, and SVR. This model considers the data from the previous 14 days. The ensemble method offers superior prediction performance for the vast majority of these countries, with less than 10% error for 5 countries and less than 40% error for 27 countries [9]. The most consistently accurate probabilistic projections of incident deaths caused by COVID-19 at the state and national level from April 2020 through October 2021 were produced via a multimodel ensemble forecast that incorporated predictions from numerous organizations each week [10]. To find the protective and risk factors connected to mental health parameters, quantitative analyses, one-way analysis of variance and linear regression were utilized. According to logistic regression analysis, the attitude was found to be the best predictor of successful practice outcomes, whereas knowledge was the strongest predictor of a positive attitude [11].

By combining homogenous convolutional neural network (CNN) classifiers, a hybrid deep neural network model is created. By adjusting the initialization of the neural network's weights and the input features' diversity, the ensemble of classifiers are created. For both time series, hybrid models considerably outperformed the corresponding single models in terms of capturing the linear, nonlinear, and seasonal pandemic trends [12]. A hybrid model divides complicated, non-stationary data into various intrinsic mode functions (IMFs), ranging in frequency from low to high. This is based on a particle swarm optimization algorithm-optimized rolling mechanism and grey prediction (PSO). [13] state that the EAMA hybrid model is well suited for predictions based on historical and current data.

2.2 Predictive Analytics

A subset of advanced analytics called predictive analytics uses historical data along with statistical modelling, data mining, and machine learning to forecast future results. One important tool for assisting the government and others in planning and constructing health services during a pandemic is prediction. In the process-relational method, which enhances processes, data quality, and model quality, prediction and machine learning are related. The Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to independently evaluate studies for bias risk and applicability [14]. Short-term forecasts were found to have a decent correlation in the study, highlighting the necessity for predictive model corrections as more and more data become available [15]. The claimed predictive capabilities of the majority of published prediction model research are likely to be optimistic due to inadequate reporting and a significant risk of bias [16]. The best prediction model for Covid-19 illness outcome has a Receiver Operating Characteristic AUC of 0.92, Sensitivity of 0.88, and Specificity of 0.82. The PSO-optimized Grey Rolling the Modelling technique outperforms the traditional prominent method for building predictive models with a short dataset is grey modelling [17].

2.3 Forecasting

The goal of ML forecasting methods is the same as that of conventional approaches: to enhance forecast accuracy while minimising a loss function. ML forecasting algorithms frequently use techniques involving more complex characteristics and predictive methodologies. Stochastic theory mathematical models and data science/machine learning approaches are two categories of forecasting techniques. Additionally, the use of data from numerous platforms is essential for predicting [18]. Once these approaches were able to learn the nonlinearities present in the assessed epidemiological time series, it is reasonable to conclude that SVR and stacking-ensemble learning model are effective tools to anticipate COVID-19 instances for the majority of the adopted states [19]. We outline three regional-scale models for predicting and analysing the pandemic's trajectory. The goal of this research is to show how useful sparse models are for comprehending the pandemic and to create a usable framework for developing insights into its trajectory that can be used to anticipate COVID-19. We produce predictions using the Richards growth model, the generalised logistic growth model, and the sub-epidemic wave model; the predictions from the sub-epidemic model additionally takes into account the possibility of further sustained transmission [20]. Depending on how often it performs better than the competing models, the SSA approach is determined to be a credible alternative for predicting the number of daily confirmed cases, fatalities, and recoveries brought on by COVID-19 [21]. A prognostic yet deterministic model is created using system modelling and identification techniques to predict the spread of COVID-19 in India [22]. Fitting multiple phenomenological models to non-linear development curves (Richards, 3 and 4 parameters logistic, Weibull and Gompertzto generate short-term projections of COVID-19 incidence and fatalities both nationally and provincially [23].

Autoregressive Integrated Moving Average (ARIMA). The standard abbreviation for ARIMA models is ARIMA (p,d,q), where p denotes the order of the moving-average model, d the degree of differencing and q the order of the autoregressive model. ARIMA models transform a non-stationary time series into a stationary one via differencing and then extrapolate future values from the past. Since our relief problem is nonlinear and the ARIMA model has a linear model structure, accuracy should be degraded if the model has strong nonlinearity [24]. The auto-ARIMA model was used to pick the initial combinations of the model parameters, and then the optimal model parameters were found based on the best match between the predictions and test data. To determine the optimal model parameter combinations, the modified ARIMA model was utilised to determine the best match between the test and forecast data [25]. ARIMA offers a straightforward yet effective way of producing accurate time series forecasts because it expressly caters to a some of common time series data types. It incorporates the concept of integration and is a generalisation of the easier autoregressive moving average [26]. The SutteARIMA approach is preferable to ARIMA for calculating daily forecasts of Covid-19 and IBEX-confirmed cases in Spain. For confirmed cases of Covid-19 in Spain, the MAPE value was 0.1905 (less than 0.04 compared to the MAPE value of ARIMA), while it was 0.02 for IBEX stock [27]. The Levenberg-Marquardt optimization training technique (LM) was used to optimise the ten-neuron NAR model architecture, which has an overall R2 value of 0.97 [28].

2.4 Heuristic Model

In machine learning (ML) and artificial intelligence (AI), heuristics are employed when a specific problem cannot be effectively solved by a step-by-step method. Heuristic methods prioritise speed above accuracy, hence they are frequently paired with optimization techniques to provide better outcomes. The Heuristic model aids in determining whether these intervention techniques are the most effective ones for illness control and how they might impact the dynamics of the disease [29]. Based on stringent epidemic prevention methods and the known COVID-19 spreading characteristics, a heuristic spread model is suggested. The basic regeneration number and the model's equilibria (disease-free equilibrium and endemic equilibrium) are examined [30]. A stochastic individual compartment model called the individual contact SEIQHRF model predicts the prevalence of disease in the susceptible, infected, recovered, and deadly compartments [31]. By using two strategies—quarantining the sick and treating them—this research aims to lower the size of vulnerable, infected, exposed, and asymptomatic groups and, as a result, eradicate the infection. Based on our spread forecast, we suggest a straightforward heuristic to estimate the COVID-19 fatality. The suggested multi-period curve model yields a respectably high level of accuracy in the prediction of the confirmed cases and fatality, according to numerical trials [32]. Artificial neural network-based curve fitting techniques in prediction and forecasting of the Covid-19 number of rising cases and death cases [33]. Gompertz models for predicting the effects of the COVID-19 outbreak and helping with decision-making, both in terms of health care needs and public health outcomes. These models depend only on three parameters (the initial number of infected individuals, the maximum number of infected people, and the infection growth rate), which can be identified by fitting the historical data [34].

2.5 Statistical Model

Statistical modelling is a complex technique for producing sample data and making predictions about the real world using a variety of statistical models and explicit assumptions. In this procedure, there is a mathematical connection between random and non-random variables. Modeling the observed incidence cases by using a Poisson distribution for the daily incidence number and a gamma distribution for the series interval; estimating the effective reproduction number under the assumption that its value remains constant over a brief time interval; and deriving next occurrences cases from their posterior distributions under the assumption that the present transmission rate will not change or just marginally alter. To predict future COVID-19 cases, specifically the positive and recovery number, an autoregression model using the Poisson distribution was used. It was discovered that Poisson autoregression could produce an accurate prediction with a MAPE below 20% and tended to follow the actual data for the next 8 to 14 days. An ensemble of Kalman filter (EnKf) method is created to estimate unknown parameters and immeasurable state variables in a COVID-19 model [35].

2.6 Statistical Model

A tool called data mining is used to find fresh, precise, and practice patterns in data or important pertinent information. Machine learning is the process of identifying algorithms that have improved thanks to experience-derived data. With an overall accuracy of 99.85%, the data mining method is more effective at predicting the likelihood that infected patients from the COVID-19 pandemic will recover [36]. The algorithms can be used to forecast outbreak trends. This forecast may aid in planning and resource allocation for policymakers and healthcare administrators [37]. Utilizing epidemiological data of COVID-19 patients from South Korea, data mining models were created to predict the recovery from COVID-19 infected patients [38].

2.7 Covid-19 Data Analysis Approach

The entire data analysis procedure is automated by machine learning to deliver richer, quicker, and more thorough insights. AI's subset of machine learning uses algorithms to examine enormous volumes of data. The purpose of this study is to evaluate how SDMs (such as isolation and quarantine) can stop the spread of COVID-19. Research on COVID-19 and the WHO database on the disease will be used for the first studies evaluating the effectiveness of SDMs (such as isolation and quarantine) in reducing COVID-19 transmission and will be published by accordance with the PRISMA statement. This protocol will be created using the PRISMA-P checklist [39]. The current study makes an effort to produce a more precise estimate of the number of Coronavirus cases by taking into account data from the past as well as other potent elements related to the virus. Data analysis, along with the creation of a network-based neural algorithm [i.e., nonlinear autonomous exogenous input (NARX)], can be utilised for this purpose [40]. C-reactive protein, platelets, and D-dimers were shown to be the variables most closely connected with COVID-19 severity prediction when data analysis techniques were employed to identify patterns and key characteristics in the data [41].

2.8 Evolution Model

Natural phenomena serve as the inspiration for EC techniques, which employ stochastic methods to tackle optimization issues. They can provide a dependable and efficient method to handle complicated issues in practical applications. Recently, EC algorithms have been applied to enhance the effectiveness of Machine Learning (ML) models and the calibre of their output. An evolutionary algorithm (EA) is a procedure that uses mechanisms drawn from nature to solve issues by simulating the actions of living things. EA is a component of both evolutionary computing and computing influenced by biology. The ideas of Darwinian evolution serve as inspiration for EAs. The CC and DCs of COVID-19 were developed using a robust variant of GEP; the proposed gene expression programming models are extremely trustworthy and may be regarded as the industry standard for time series prediction for COVID-19 in Australia [42]. By assisting in the processing of enormous data volumes, complex system modelling, and sourcing derivations from healthcare data and simulations, scientific programming is seen as a key tool and important contributor to the provision of solutions to current and future problems involved in the management of large-scale data in healthcare.

3 Comparative Analysis of Different Models to Predict COVID-19

Research addressing the comparative analysis and quality assessment of the covid-19 and type of the models. The following Table 1 gives a summary of models applied to forecast COVID-19.

4 Methodology

4.1 Susceptible-Exposed-Infectious-Removed (SEIR), Susceptible-Infected-Recovered (SIR)

A common Susceptible-Infected-Recovered model states that at a given time t, the population (of size N) can be divided into three groups: those who are susceptible (S(t), infected (I(t), and recovered (R)). The SEIR model is an extension of the classic SIR model, and both Susceptible-Exposed-Infectious-Removed and Susceptible-Infected-Recovered models serve as the basis for many epidemiological modelling techniques (t) [43]. The Worldometer data set has been exempted from the pandemic up until January 2022. The newly developed modified SEIR model's parameters were fitted. To solve the modified SEIR dynamics model examined, the Euler integral algorithm was used [44]. Hence equations are as follows.

$$\mathrm{S} + \mathrm{E} + \mathrm{I} + \mathrm{R} = 1 \tag{1}$$

$$\begin{aligned}
\mathrm{dS} &= -\mathrm{dS_temp} \\
\mathrm{dE} &= \mathrm{dS_temp} - \mathrm{dE_temp} \\
\mathrm{dR} &= (\mathrm{I/t_inf}) * \mathrm{delta_t}
\end{aligned}$$

$$dI = dE\_temp - dR$$

The analysis of the COVID-19 outbreak is suggested using certain updated SEIR models that take many parameters into account. For instance, a nonlinear incidence rate is employed; public policy is taken into consideration; and some generic control techniques, such as hospital, quarantine, and external input, are taken into consideration. a compartmental SEIR model, where a set of differential equations are used to simulate

Table 1. Models applied to forecast COVID-19.

No	Study	Objective	Type of model	Results
1	Istaiteh, O et al., 2020 [6]	To provide a tool for global predicting COVID-19 confirmed cases for the future week.	LSTM, RNN	The final refined model was tested in a real-world setting to forecast the COVID-19 cases for the following seven days using a test set to compare the four models.
2	Raushan Raj et al., 2020 [9]	Using ensemble regression models, make short-term predictions for confirmed COVID-19.	Ensemble Regression Models	With less than 10% error for 5 nations and less than 40% error for 27, our results in terms of relative percentage error demonstrate that the ensemble method offers superior prediction performance for the great majority of these countries.
3	Saleh I. Alzahrani et al., 2020 [24]	In order to predict the anticipated daily number of COVID-19 cases in Saudi Arabia over the ensuing four weeks, we used the ARIMA Model.	ARIMA	If strict preventative and control measures are not put in place to stop the spread of COVID-19, the trend in Saudi Arabia will continue expanding and may reach up to 7,668 new cases per day and over 127,129 cumulative daily cases in a matter of four weeks.

(*continued*)

Table 1. (*continued*)

No	Study	Objective	Type of model	Results
4	Ansari Saleh Ahmar et al., 2020 [27]	Our study's goal is to use the SutteARIMA approach to estimate the short-term distribution of confirmed cases of COVID-19 and IBEX in Spain.	SutteARIMA	The SutteARIMA approach is preferable to ARIMA for calculating daily forecasts of Covid-19 and IBEX confirmed cases in Spain.
5	Farhan Mohammad Khan et al., 2020 [28]	The number of COVID-19 infection cases that can be anticipated in India over the next few days has been predicted using the univariate time series model.	ARIMA, NAR	Based on information available as of April 4, 2020, the results indicated an upward trend in the actual and projected numbers of COVID-19 cases, with an average of 1500 cases per day.
6	S.K. Tamang et al., 2020 [33]	Covid-19 case forecasting based on curve fitting for artificial neural networks in predictions.	ANN curve fitting model	The outcomes have demonstrated that ANN is capable of accurately predicting future COVID 19 outbreak cases in any nation.
7	L. J. Muhammad et al., 2020 [36]	Epidemiology Dataset for COVID-19 Infection Prediction.	Decision Tree	The decision tree model has the highest accuracy, according to the results of the models' performance evaluation.

(*continued*)

Table 1. (*continued*)

No	Study	Objective	Type of model	Results
8	Seyed Mohammad Ayyoubzadeh et al., 2020 [37]	To estimate the COVID-19 occurrence in Iran by using Google Trends data.	Data mining, Deep learning	Algorithms for data mining can be used to forecast outbreak trends. This forecast might help managers of the healthcare system and lawmakers allocate resources for healthcare in a sensible way.
9	Sunidhi Shrivastava et al., 2020 [38]	Utilizing an epidemiological dataset of COVID-19 patients from South Korea, data mining models for the prediction of recovering COVID-19-infected patients were created for the study.	Data mining	The study's findings revealed the coronavirus patient's forecast. ARIMA used artificial intelligence and machine learning technology for that. The clustering was done using KNN-based clustering
10	Ma, Ruifang et al., 2021 [8]	By integrating the LSTM and Markov methods, one may forecast and analyse the COVID-19 pandemic pattern.	Prediction and Analysis	All of those indications show that the suggested LSTM-Markov model has a higher prediction accuracy than the LSTM model, leading to a more precise prediction of COVID-19.
11	Senthilkumar Mohan et al., 2021 [13]	A method for predicting the effects of COVID-19 using supervised machine learning.	Hybrid Model	The EAMA hybrid model works well for forecasts based on historical and current data.

(*continued*)

Table 1. (*continued*)

No	Study	Objective	Type of model	Results
12	ZeynepCeylan et al., 2021 [17]	In order to limit harm to human health, it is essential to foresee the spread of the coronavirus illness of 2019 (COVID-19).	PSO	When it comes to estimating the total number of positive cases of COVID-19 instances, PSO algorithm outperforms the traditional Grey Modelling, Grey Rolling Modelling, and nonlinear autoregressive artificial neural network models. The current study can serve as a crucial foundation for nations to manage health resources and create effective epidemic prevention measures.
13	Kalantari M et al., 2021 [21]	Employing the best single spectrum analysis to predict the COVID-19 pandemic.	Forecasting	Forecasting the number of daily positive cases, fatalities, and discharge brought on by COVID-19 is proven to be possible using the SSA approach.
14	Tarylee Reddy et al., 2021 [23]	To create short-term projections of COVID-19 incidence and fatalities both nationally and provincially.	Forecasting	Forecasts made using logistic growth models were more precise than those made using the Richards model 10 days in advance.

(*continued*)

Table 1. (*continued*)

No	Study	Objective	Type of model	Results
15	Kulshreshtha, V et al., 2021 [26]	To compare the general population's mental health during the pandemic in seven middle-income countries (MICs) in Asia (China, Iran, Malaysia, Pakistan, Philippines, Thailand, and Vietnam).	Regression	Identification of protective and risk factors connected to mental health parameters was done using analysis of variance and linear regression.
16	Fernández-Martínez et al., 2021 [34]	To forecast the mathematical models for the COVID-19 pandemic's short- and long-term growth.	Mathematical model	Both the Verhulst and the Gompertz models offer comparable outcomes and can be used to track and foretell future outbreaks. The Verhulst model, however, appears to be simpler to use and interpret.
17	Yasminah Alali et al., 2022 [3]	The goal of this project is to create a data-driven model that makes no assumptions and can predict COVID-19 spread with accuracy.	Random forest	Results show that employing the suggested dynamic machine learning models, significant improvement can be attained.
18	Sandeep Maan et al., 2022 [25]	For the purpose of forecasting daily confirmed cases and cumulative confirmed cases, this study developed a hybrid autoregressive integrated moving average and Prophet model.	Prediction model	This study will assist the nation and the states in developing the best public health strategies for the next COVID-19 waves.

(*continued*)

Table 1. (*continued*)

No	Study	Objective	Type of model	Results
19	Yang B et al., 2022 [30]	On the basis of stringent epidemic control measures and the known COVID-19 spreading characteristics, a mathematical spread model is constructed.	Mathematical Model	The mean relative error is used to assess the impact of simulations and fitting. The model can accurately reflect the spread dynamics of COVID-19, according to simulation data.
20	Shahpoori et al., 2022 [40]	The current study makes an effort to produce a more precise estimate of the number of COVID-19 cases by taking into account historical data as well as other potent elements related to the virus.	Data analysis	It is possible to use data analysis to create a network-based neural algorithm, or NARX (nonlinear autonomous exogenous input).
21	Mariam laatifi et al., 2022 [41]	The goal of this work is to create and evaluate models for COVID-19 severity prediction based on machine learning.	Machine Learning	Uniform Manifold Approximation and Projection (UMAP), a new feature engineering technique built on topological data analysis, has demonstrated that it produces better outcomes.
22	Estee Y. Cramer et al., 2022 [10]	The accuracy of short-term projections of reported deaths attributable to COVID-19 during the first 1.5 years of the pandemic in the United States is compared in this research.	Short-term forecasts	A stand-alone model's accuracy varies widely between and within it, whereas an ensemble model's accuracy, which aggregated forecasts from all eligible models, is more consistent.

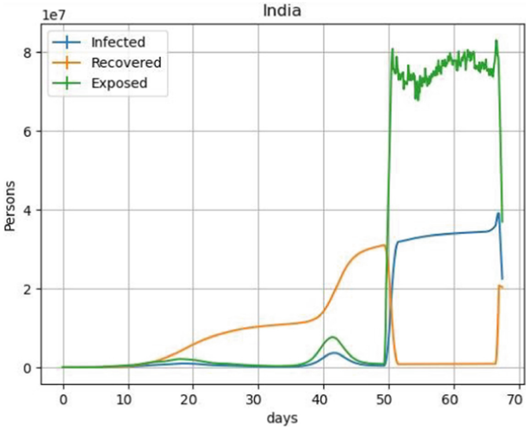

Fig. 1. SEIR model

how people move between compartments. The With the aid of MATLAB, 1000 Monte Carlo simulations of various situations were modelled. SimVoi software was used to simulate the requirements for intensive care units (ICUs) and fatalities. It was assessed how non-pharmacological treatments (NPIs), such as social isolation and lockdown, affected containing the outbreak [45]. Model parameters used in this study are shown (Fig. 1).

5 Conclusion and Future Works

The complicated behaviour of COVID-19 data could not be reliably modelled using machine learning algorithms. The explanation is that a variety of variables, such as the local environment and physical closeness, affect the number of reported instances. The study has thus far proven to be a regression problem in these conditions. It is based on some complex supervised ML regression model (LR), the lowest absolute and selective shrinking operator, vector supports, and exponential smoking models. Accuracy metrics were used to evaluate the performance of all models. According to research studies, time series models are used in models based on analysis with the goal of platforms to capturing any potential nonlinearity in the data, but only some models are capable of making predictions with high accuracy rates. Here, we use a Kalman filter to provide it to the SEIR model with the maximum accuracy rate.

In Future studies we try to bring train and test the model with a high accuracy rate and predict the cases for the next 7 to 14 days. Future works have been designed to bring the AI chatbot embedded with the dashboard for users and healthcare workers like Q&A where the chatbot answers the question along the dashboard. Based on the inference, it is presumed that similar studies may also be deployed for epidemic diseases like Swine Flu, Influenza and other variants.

Acknowledgement. We would like to thank ICMR for funding this project, ID NO-2021-6415 “INTERACTIVE COVID-19 CHATBOT DASHBOARD: Design and development of statistical model for short-term Prediction of Dynamic COVID-19 transmission”.

References

1. Peng, Y., Nagata, M.H.: An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. Chaos, Solitons & Fractals. 139, 110055 (2020). https://doi.org/10.1016/j.chaos.2020.110055.
2. Kavadi, D.P., Patan, R., Ramachandran, M., Gandomi, A.H.: Partial derivative Nonlinear Global Pandemic Machine Learning prediction of COVID 19. Chaos, Solitons & Fractals. 139, 110056 (2020). https://doi.org/10.1016/j.chaos.2020.110056.
3. Alali, Y., Harrou, F., Sun, Y.: A proficient approach to forecast COVID-19 spread via optimized dynamic machine learning models. Sci Rep. 12, 2467 (2022). https://doi.org/10.1038/s41598-022-06218-3.
4. Zeroual, A., Harrou, F., Dairi, A., Sun, Y.: Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. Chaos, Solitons & Fractals. 140, 110121 (2020). https://doi.org/10.1016/j.chaos.2020.110121.
5. Wang, P., Zheng, X., Ai, G., Liu, D., Zhu, B.: Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. Chaos, Solitons & Fractals. 140, 110214 (2020). https://doi.org/10.1016/j.chaos.2020.110214
6. Istaiteh, O., Owais, T., Al-Madi, N., Abu-Soud, S.: Machine Learning Approaches for COVID-19 Forecasting. In: 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA). pp. 50–57 (2020). https://doi.org/10.1109/IDSTA50958.2020.9264101.
7. Shahid, F., Zameer, A., Muneeb, M.: Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos, Solitons & Fractals. 140, 110212 (2020). https://doi.org/10.1016/j.chaos.2020.110212.
8. Ma, R., Zheng, X., Wang, P., Liu, H., Zhang, C.: The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method. Sci Rep. 11, 17421 (2021). https://doi.org/10.1038/s41598-021-97037-5.
9. Raj, R., Seetharam, A., Ramesh, A.: Ensemble Regression Models for Short-term Prediction of Confirmed COVID-19 Cases.
10. Cramer, E.Y., Ray, E.L., Lopez, V.K., and Reich, N.G.: Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. Proceedings of the National Academy of Sciences. 119, e2113561119 (2022). https://doi.org/10.1073/pnas.2113561119.
11. Singh, P.K., Anvikar, A., Sinha, A.: COVID-19 related knowledge, attitudes, and practices in Indian Population: An online national cross-sectional survey. PLOS ONE. 17, e0264752 (2022). https://doi.org/10.1371/journal.pone.0264752.
12. Perone, G.: Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. Eur J Health Econ. 23, 917–940 (2022). https://doi.org/10.1007/s10198-021-01347-4.
13. Mohan, S., A, J., Abugabah, A., M, A., Kumar Singh, S., kashif Bashir, A., Sanzogni, L.: An approach to forecast impact of Covid-19 using supervised machine learning model. Software: Practice and Experience. 52, 824–840 (2022). https://doi.org/10.1002/spe.2969.
14. Miller, J.L., Tada, M., Goto, M., Chen, H., Dang, E., Mohr, N.M., Lee, S.: Prediction models for severe manifestations and mortality due to COVID-19: A systematic review. Academic Emergency Medicine. 29, 206–216 (2022). https://doi.org/10.1111/acem.14447.
15. Kotwal, A., Yadav, A.K., Yadav, J., Kotwal, J., Khune, S.: Predictive models of COVID-19 in India: A rapid review. Medical Journal Armed Forces India. 76, 377–386 (2020). https://doi.org/10.1016/j.mjafi.2020.06.001.

16. Wynants, L., Calster, B.V., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Albu, E., Arshi, B., Bellou, V., Bonten, M.M.J., Dahly, D.L., Damen, J.A., Debray, T.P.A., Jong, V.M.T. de, Vos, M.D., Dhiman, P., Ensor, J., Gao, S., Haller, M.C., Harhay, M.O., Henckaerts, L., Heus, P., Hoogland, J., Hudda, M., Jenniskens, K., Kammer, M., Kreuzberger, N., Lohmann, A., Levis, B., Luijken, K., Ma, J., Martin, G.P., McLernon, D.J., Navarro, C.L.A., Reitsma, J.B., Sergeant, J.C., Shi, C., Skoetz, N., Smits, L.J.M., Snell, K.I.E., Sperrin, M., Spijker, R., Steyerberg, E.W., Takada, T., Tzoulaki, I., Kuijk, S.M.J. van, Bussel, B.C.T. van, Horst, I.C.C. van der, Reeve, K., Royen, F.S. van, Verbakel, J.Y., Wallisch, C., Wilkinson, J., Wolff, R., Hooft, L., Moons, K.G.M., Smeden, M. van: Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 369, m1328 (2020). https://doi.org/10.1136/bmj.m1328.
17. Ceylan, Z.: Short-term prediction of COVID-19 spread using grey rolling model optimized by particle swarm optimization. Applied Soft Computing. 109, 107592 (2021). https://doi.org/10.1016/j.asoc.2021.107592.
18. Shinde, G.R., Kalamkar, A.B., Mahalle, P.N., Dey, N., Chaki, J., Hassanien, A.E.: 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.
19. Ribeiro, M.H.D.M., da Silva, R.G., Mariani, V.C., Coelho, L. dos S.: Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos, Solitons & Fractals. 135, 109853 (2020). https://doi.org/10.1016/j.chaos.2020.109853.
20. Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., Yan, P., Chowell, G.: Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. Journal of Clinical Medicine. 9, 596 (2020). https://doi.org/10.3390/jcm9020596.
21. Kalantari, M.: Forecasting COVID-19 pandemic using optimal singular spectrum analysis. Chaos, Solitons & Fractals. 142, 110547 (2021). https://doi.org/10.1016/j.chaos.2020.110547.
22. Rafiq, D., Suhail, S.A., Bazaz, M.A.: Evaluation and prediction of COVID-19 in India: A case study of worst hit states. Chaos, Solitons & Fractals. 139, 110014 (2020). https://doi.org/10.1016/j.chaos.2020.110014.
23. Reddy, T., Shkedy, Z., Janse van Rensburg, C., Mwambi, H., Debba, P., Zuma, K., Manda, S.: Short-term real-time prediction of total number of reported COVID-19 cases and deaths in South Africa: a data driven approach. BMC Med Res Methodol. 21, 15 (2021). https://doi.org/10.1186/s12874-020-01165-x.
24. Alzahrani, S.I., Aljamaan, I.A., Al-Fakih, E.A.: Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. Journal of Infection and Public Health. 13, 914–919 (2020). https://doi.org/10.1016/j.jiph.2020.06.001.
25. Maan, S., Devi, G., Rizvi, S.: Prediction of Third Covid Wave in India using ARIMA Model. JSR. 66, 149–155 (2022). https://doi.org/10.37398/JSR.2022.660220.
26. Kulshreshtha, V., Garg, N.K.: Predicting the New Cases of Coronavirus [COVID-19] in India by Using Time Series Analysis as Machine Learning Model in Python. J. Inst. Eng. India Ser. B. 102, 1303–1309 (2021). https://doi.org/10.1007/s40031-021-00546-0.
27. Ahmar, A.S., del Val, E.B.: SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain. Science of The Total Environment. 729, 138883 (2020). https://doi.org/10.1016/j.scitotenv.2020.138883.
28. Khan, F.M., Gupta, R.: ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. Journal of Safety Science and Resilience. 1, 12–18 (2020). https://doi.org/10.1016/j.jnlssr.2020.06.007.

29. Mandal, M., Jana, S., Nandi, S.K., Khatua, A., Adak, S., Kar, T.K.: A model based study on the dynamics of COVID-19: Prediction and control. Chaos, Solitons & Fractals. 136, 109889 (2020). https://doi.org/10.1016/j.chaos.2020.109889.
30. Yang, B., Yu, Z., Cai, Y.: A spread model of COVID-19 with some strict anti-epidemic measures. Nonlinear Dyn. 109, 265–284 (2022). https://doi.org/10.1007/s11071-022-07244-6.
31. Sengupta, P., Ganguli, B., SenRoy, S., Chatterjee, A.: An Analysis of COVID-19 Clusters in India-Two Case Studies on Nizamuddin and Dharavi. In Review (2020). https://doi.org/10.21203/rs.3.rs-68814/v1.
32. Majeed, B., Li, A., Peng, J., Lin, Y.: A Multi-Period Curve Fitting Model for Short-Term Prediction of the COVID-19 Spread in the U.S. Metropolitans. Frontiers in Public Health. 9, (2022).
33. Tamang, S.K., Singh, P.D., Datta, B.: Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique. Global Journal of Environmental Science and Management. 6, 53–64 (2020). https://doi.org/10.22034/GJESM.2019.06.SI.06.
34. Fernández-Martínez, J.L., Fernández-Muñiz, Z., Cernea, A., Kloczkowski, A.: Predictive Mathematical Models of the Short-Term and Long-Term Growth of the COVID-19 Pandemic. Computational and Mathematical Methods in Medicine. 2021, e5556433 (2021). https://doi.org/10.1155/2021/5556433.
35. Nkwayep, C.H., Bowong, S., Tewa, J.J., Kurths, J.: Short-term forecasts of the COVID-19 pandemic: a study case of Cameroon. Chaos, Solitons & Fractals. 140, 110106 (2020). https://doi.org/10.1016/j.chaos.2020.110106.
36. Muhammad, L.J., Algehyne, E.A., Usman, S.S., Ahmad, A., Chakraborty, C., Mohammed, I.A.: Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset. SN COMPUT. SCI. 2, 11 (2020). https://doi.org/10.1007/s42979-020-00394-7.
37. Ayyoubzadeh, S.M., Ayyoubzadeh, S.M., Zahedi, H., Ahmadi, M., Kalhori, S.R.N.: Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health and Surveillance. 6, e18828 (2020). https://doi.org/10.2196/18828.
38. Shrivastava, S., Gugnani, P.: Data Mining Prediction Models for Novel Coronavirus (COVID-19) using Machine Learning, https://papers.ssrn.com/abstract=3759263 (2021). https://doi.org/10.2139/ssrn.3759263.
39. Regmi, K., Lwin, C.M.: Impact of social distancing measures for preventing coronavirus disease 2019 [COVID-19]: A systematic review and meta-analysis protocol, https://www.medrxiv.org/content/10.1101/2020.06.13.20130294v1 (2020). https://doi.org/10.1101/2020.06.13.20130294.
40. Shahpoori, P.K., Mirzaei, A.: Determining the efficiency of data analysis systems in predicting COVID-19 infected cases. Journal of Family Medicine and Primary Care. 11, 2405 (2022). https://doi.org/10.4103/jfmpc.jfmpc_1205_21.
41. Laatifi, M., Douzi, S., Bouklouz, A., Ezzine, H., Jaafari, J., Zaid, Y., El Ouahidi, B., Naciri, M.: Machine learning approaches in Covid-19 severity risk prediction in Morocco. J Big Data. 9, 5 (2022). https://doi.org/10.1186/s40537-021-00557-0.
42. Salgotra, R., Gandomi, A.H.: 21 - Time series analysis of the COVID-19 pandemic in Australia using genetic programming. In: Kose, U., Gupta, D., de Albuquerque, V.H.C., and Khanna, A. (eds.) Data Science for COVID-19. pp. 399–411. Academic Press (2021). https://doi.org/10.1016/B978-0-12-824536-1.00036-8.
43. Zhao, H., Merchant, N.N., McNulty, A., Radcliff, T.A., Cote, M.J., Fischer, R.S.B., Sang, H., Ory, M.G.: COVID-19: Short term prediction model using daily incidence data. PLOS ONE. 16, e0250110 (2021). https://doi.org/10.1371/journal.pone.0250110.

44. Cao, S., Feng, P., Shi, P.: [Study on the epidemic development of COVID-19 in Hubei province by a modified SEIR model]. Zhejiang Da Xue Xue Bao Yi Xue Ban. 49, 178–184 (2020). https://doi.org/10.3785/j.issn.1008-9292.2020.02.05.
45. Chatterjee, K., Chatterjee, K., Kumar, A., Shankar, S.: Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. Medical Journal Armed Forces India. 76, 147–155 (2020). https://doi.org/10.1016/j.mjafi.2020.03.022.