Optimized Gaussian Process Regression by Bayesian Optimization to Forecast COVID-19 Spread in India and Brazil: A Comparative Study

Optimized Gaussian Process Regression by Bayesian Optimization to Forecast COVID-19 Spread in India and Brazil: A Comparative Study

Yasminah Alali, Fouzi Harrou, Ying Sun
King Abdullah University of Science and Technology (KAUST)
Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division
Thuwal, 23955-6900, Saudi Arabia
Email: yasminah.alali@kaust.edu.sa, fouzi.harrou@kaust.edu.sa

***Abstract*—On June 29 2021, the World Health Organization (WHO) reported around 45,951 confirmed cases and 817 deaths of COVID-19 in India, and 64,903 confirmed cases and 1,839 deaths in Brazil. This virus has been determined as a global pandemic by WHO. Accurate forecast of COVID-19 cases has become a crucial task in the decision-making of hospital managers to optimally manage the available resources and staff. In this study, the Gaussian process regression (GPR) model tuned by Bayesian optimization (BO) was used to forecast the recovered and confirmed COVID-19 cases in two highly impacted countries, India and Brazil. Specifically, the BO algorithm is employed to find the optimal hyperparameters of the GPR model to improve the forecasting quality. We compared the performance of the Optimized GPR with 14 models, including Support vector regression with different kernels, GPR with different kernels, Boosted trees, and Bagged trees. We also applied the BO to the other investigated predictors to maximize their forecasting accuracy. Three statistical criteria are used for the comparison. The daily records of confirmed and recovered cases from Brazil and India are adopted in this study. Results reveal the high performance of the GPR models compared to the other models.**

***Index Terms*—COVID-19, forecast, data-driven model, Ensemble models, Bayesian optimization, time series.**

I. Introduction

At the end of 2019, the world population faced a new virus pandemic called SARS COV-2 (COVID-19). The virus starts to spread across the countries fast with a high contagion rate until it becomes a global pandemic. The number of confirmed cases reached more than 180 million and around 4 million deaths. Also, the outbreak of COVID-19 led to a socioeconomic crisis that put high pressure on the contraries [1]. The primary issue the healthcare encountered was difficulty diagnosing the virus. It can stay inpatient for 14 days with no symptoms, so healthcare finds challenging to control the spread of COVID-19. Recently, machine learning (ML) in the artificial intelligence field showed an essential role in forecasting COVID-19 spread. Importantly, ML provides a forecast of the future trends of COVID-19 cases based on the past recorded COVID-19 time series. Indeed, COVID-19 cases forecasting provides relevant information to countries and hospital managers to plan for the future COVID-19 spread by setting policies and rigorous strategies to avoid riskiness and reduce the number of contaminated cases.

Identify applicable funding agency here. If none, delete this.

Recently, there have been many studies conducted to understand and manage the COVID-19 pandemic [2]–[4]. The researchers in [5] use datasets from ten Brazilian cities, including Rio de Janeiro and Sao Paulo, to forecast COVID-19 confirmed cases based on various machine learning methods (e.g., Random Forests and support vector regression (SVR)). The result shows that SVR achieves the highest performance among all considered algorithms. Another study in [6] presents short-term forecasting models using improved adaptive network-based fuzzy inference system (ANFIS) by chaotic marine predators algorithm (CMPA) algorithm. They collect the data from reported cases of WHO between 26th of March to 1st of June 2020 in Brazil and Russia, then divided the data into a training subset and test subset. Results showed improved performance by using the CMPA approach compared to other models, including the conventional ANFIS. In [7] the authors applied seven models, such as Auto-Regressive Integrated Moving Average model (ARIMA), and Double Exponential Smoothing (DES), to forecast COVID-19 transmission in India and Brazil. They evaluate the models by focusing on Mean Absolute Error, MAE. It has been shown that the DES model achieves the lowest MAE among all other models. In [8], the authors investigate predicting the spread of COVID-19 on five states in India, particularly in Kerala, Karnataka, Maharashtra, Tamil Nadu, and Andhra Pradesh cities. Four models are compared in this study, namely Gompertz growth, exponential, logistic, and ARIMA. Results indicate that ARIMA outperforms the other models. In [9] several models are used to forecast COVID-19 applied confirmed, recover, and death cases in Indian, including Multilayer perceptron, Vector autoregression, and Linear regression. The Multilayer perceptron reaches the most reasonable forecast accuracy. In [10], ARIMA, exponential smoothing, and Holt-Winters models are applied to forecast COVID-19 based on data recorded between the 4th of March to 11th of July. Here, the ARIMA model showed the best forecasting compared to the other studied models. In [11], the forecasting performance

of ARIMA is compared to those of Prophet model [12] when forecasting COVID-19 spread in five cities in India: Andhra Pradesh, Maharashtra, Uttar Pradesh, Karnataka, and Tamil Nadu. Results show that the ARIMA model provided better forecasting results than the Prophet method.

The ability to accurately forecast the recovered and confirmed COVID-19 cases could aid slow down the transmission of COVID-19 by making appropriate decisions. This study performed data-driven methods to forecast confirmed and recovered COVID-19 cases based on machine learning models. Specifically, this study investigates the forecasting ability of the optimized GPR, a kernel-based machine learning method, in forecasting the COVID-19 time series. This choice is motivated by the desirables features of the GPR model, including its simple and flexible construction using the mean and covariance functions, its ability and superior nonlinear approximation, and the possibility to explicitly provide a probabilistic representation of forecasting outputs [13], [14]. Specifically, the BO algorithm is employed to find the optimal hyperparameters of the GPR model to improve the forecasting quality. We compared the performance of the Optimized GPR (OGPR) with 14 models, including Support vector regression (SVR) with different kernels, GPR with different kernels, Boosted trees (BT), and Bagged trees (BS). Here, A BO is employed in all forecasting approaches for improving their accuracy. Of course, we would like to examine the effectiveness of the OGPR model on the COVID-19 datasets of limited size and assess its performance compared to traditional machine learning methods. Datasets from India and Brazil where the COVID-19 spread was significantly high are used to evaluate and compare the sixteen methods' forecasting accuracy. The used COVID-19 time series are recorded from January 22, 2020, to June 12, 2021.

The remaining of this study is structured as follows. Section II presents the used COVID-19 datasets, provides a brief description of the GPR model and the BO algorithm. The results and discussions were given in section III to show model performances and comparisons. The conclusions are outlined in Section IV.

II. Methodology

A. Data description

Here, daily confirmed and recovered COVID-19 data from two highly impacted countries, India and Brazil, are utilized to evaluate the forecasting capacity of the 14 investigated data-based methods. The daily record of cumulative confirmed and recovered cases of COVID-19 from the first case, in India and Brazil on the 30th of January and 26th of February 2020, are available in (https://github.com/CSSEGISandData/COVID-19). The dataset automated update for delayed data in the website without any missing value. Figure 1(a-b) displays the confirmed and recovered COVID-19 cases dataset used in this study. We observe that India has the highest number of confirmed cases. Considering the population in each country, India is receiving the most considerable impact from COVID-19. On the other hand, India shows rapid growth in recovered cases, indicating their prompt and effective response to this public health event.

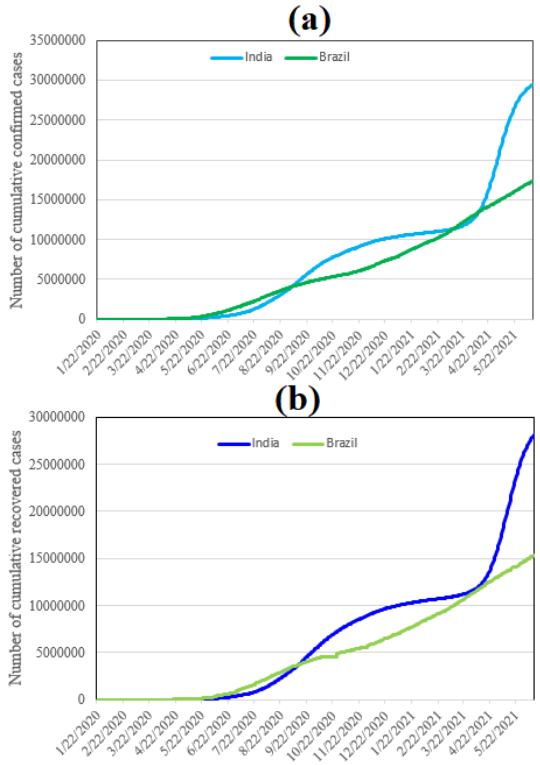

Fig. 1. The number of (a) confirmed and (b) recovered COVID-19 cases from January 22, 2020, through June 12, 2021, in Brazil and India.

B. GPR model

The GPR is a supervised nonparametric machine learning algorithm. GPR is an effective kernel-based approach to learn implicit correlations among various variables in the training set, making GPR especially suitable to deal with challenging nonlinear prediction [15]. For a prediction problem, the output y of a function f at the input x in GPR can be expressed as,

$$\mathbf{y}_i = f(\mathbf{x}_i) + \varepsilon_i. \tag{1}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. In GPR, the term, $f(x)$, is assumed to be a random variable that is distributed according to a particular distribution. Indeed, observing the output of the function at various input points could reduce the uncertainty regarding f. The observations are always tainted with a noise term ε that reflects their inherent randomness.

Assume $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ is the input-output data and $f(\cdot)$ to be approximated and supposed to follow a Gaussian process. For sake of simplicity, let assume that x_i's and y_i 's are scalar observations while ε_i 's are independent, and identically distributed random noises following the normal distribution with mean value $\bar{\varepsilon}_i = 0$ and variance σ^2.

Let's consider the measured y_i values $[y_1, y_2, \dots y_n]^\top$ are finite values of the function $f(\cdot)$ contaminated with noises. Thus, y_i 's follow a joint Gaussian distribution:

$$\mathbf{y} = [y_1, y_2, \dots y_n]^\top \sim \mathcal{N}(\mathbf{m}(\mathbf{x}), \mathbf{K} + \sigma^2\mathbf{I}), \tag{2}$$

where $\mathbf{m}(\mathbf{x}) = [m(x_1), m(x_2), \dots m(x_n)]^\top$ represents the mean vector $m(\cdot)$, $\mathbf{I}$ refers to the identity matrix, and $\mathbf{K}$ denotes the $n \times n$ covariance matrix with $(i,j)^{th}$ element $\mathbf{K}_{ij} = k(x_i, x_j)$. For a GPR model, $k(x_i, x_j)$ is usually termed a kernel function [16].

The optimal values of the kernel parameters are achieved by maximization of the following likelihood.

$$\boldsymbol{\theta}_{\text{opt}} = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) \tag{3}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots]$ denotes kernel parameters, the mean values $m(.)$ are taken to be zero, and

$$L(\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{K} + \sigma^2\mathbf{I}|}} \exp\left(-\frac{1}{2}(\mathbf{y}^\top(\mathbf{K} + \sigma^2\mathbf{I})\mathbf{y})\right). \tag{4}$$

Let x_* is a new input, then the predictive mean and variance associated with $\hat{y}_* = f(x_*) = f_*$ are respectively expressed as follows:

- the mean value

$$\hat{y}_* = \mathbf{k}_*^\top(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y} \tag{5}$$

- and variance

$$\hat{\sigma}_* = k_{**} - \mathbf{k}_*^\top(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{k}_*. \tag{6}$$

- and $\mathbf{y}_*$ follows a conditional distribution:

$$y_*|\mathbf{y} \sim \mathcal{N}(\hat{y}_*, \hat{\sigma}_*) \tag{7}$$

For more details about GPR model, see [17].

C. Bayesian Optimization of Model Parameters

Numerous machine learning models, such as SVR and GPR, comprise several hyperparameters to be chosen (e.g., kernel types in GPR and parameters). The chosen hyperparameters significantly impact the performance of any machine learning model [18]. Several optimization procedures are proposed in the literature for hyperparameter tunings, such as grid search, random search, and Bayesian optimization [19]. This study used the Bayesian optimization (BO) procedure to find the optimal hyperparameters of the investigated methods, namely SVR, GPR, boosted trees, and bagged trees. Importantly, the BO algorithm is designed based on Gaussian processes and Bayesian inference. It could be employed to optimize functions with unknown closed-form [20].

The essence of the BO algorithm is to construct a probabilistic proxy model for the cost function based on outcomes of historical experiments as training data. Essentially, the proxy model, such as the Gaussian process, is more inexpensive to compute, and it gives sufficient information on where we should assess the true objective function to obtain relevant results. Let's consider m hyperparameters $\mathbf{P} = \mathbf{p}_1, \dots, \mathbf{p}_m$ to be tuned. The aim is to determine

$$\mathbf{P}^* = \arg \min_{\mathbf{P}} \mathbf{g}(\mathbf{P}|(x_i, y_i)_{i=1}^n (8)$$

where $\mathbf{g}$ is a cost function. The whole optimization procedure is controlled via a suitable acquisition function (AF) that defines the following set of hyperparameters to be assessed. Crucially, any acquisition function requires adjusting within exploration and exploitation. Generally speaking, exploration is an area search with high uncertainty, where we expect to discover a new set of parameters that enhance the model's prediction accuracy. At the same time, exploitation refers to an area search nearby to already computed high estimated values [21].

D. Alternative models for Comparison

In this study, we investigated the performance of the OGPR and compared its forecasting accuracy with the set of machine learning-based forecasting models listed in Table 1. In short, a total of fifteen forecasting methods are applied to predict COVID-19 time-series data: 6 SVR methods [15], [22], 4 GPR methods, 2 ensemble learning techniques (i.e., BT and BS) [23]–[25] and six SVR models [26], and 3 optimized methods.

TABLE I
FORECASTING METHODS INVESTIGATED IN THIS STUDY.

Model Approach	Model Name	Model Description	Kernel Function[(1)]
Support Vector Regression (SVR)	SVR_L	SVR with the Linear kernel	${x_i}^T x_j$
	SVR_Q	SVR with the Quadratic kernel	$(1 + {x_i}^T x_j)^2$
	SVR_C	SVR with the Cubic Kernel	$(1 + {x_i}^T x_j)^3$
	SVR_FG	SVR with the Fine Gaussian kernel	$e^{(-\frac{\sqrt{p}}{4}\|x_i - x_j\|^2)}$
	SVR_MG	SVR with the Medium Gaussian kernel	$e^{(-\sqrt{p}\|x_i - x_j\|^2)}$
	SVR_CG	SVR with the Cubic Gaussian kernel	$e^{(-4\sqrt{p}\|x_i - x_j\|^2)}$
Gaussian Process Regression (GPR)	GP_RQ	GPR with the Rational Quadratic kernel	$\sigma_f^2\left(1 + \frac{r^2}{2\alpha\sigma_l^2}\right)^{-\alpha}$
	GP_SE	GPR with the Squared Exponential kernel	$\sigma_f^2 e^{\left(\frac{r^2}{2\sigma_l^2}\right)}$
	GP_M52	GPR with the Matern 5/2 kernel	$\sigma_f^2\left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right)e^{(\frac{\sqrt{5}r}{\sigma_l})}$
	GP_Exp	GPR with the Exponential kernel	$\sigma_f^2 e^{\left(\frac{r}{\sigma_l}\right)}$
Ensemble Learning (EL)	BST	Boosted Trees	
	BT	Bagged Trees	
Optimised models	OSVR	Optimized SVR	
	OGPR	Optimized GPR	
	OEL	Optimized EL	

[(1)] $r = \sqrt{(x_i - x_j)'(x_i - x_j)}$ in the GPR-based kernel function

E. Evaluation metrics

In this study, we assess the accuracy of the forecasting models using three metrics: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}, \tag{9}$$

$$MAE = \frac{\sum_{t=1}^{n}|y_t - \hat{y}_t|}{n}, \tag{10}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|\%, \tag{11}$$

where y_t is the number of COVID cases, $\hat{y}_t$ is its corresponding forecasted COVID cases, and n is the number of records. Lower RMSE, MAE, and MAPE values would imply better precision and forecasting quality.

F. Forecasting framework

The general procedure performed in this study to forecast COVID-19 cases is represented in Figure 2. Firstly the daily recovered and confirmed time-series data are split into training subsets. All models are trained using the training set and evaluated using the testing set. The best forecasting model is indicated by four statistical criteria, namely RMSE, MAE, and MAPE.

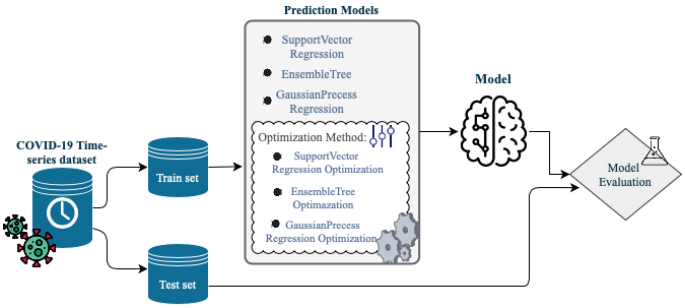

Fig. 2. Illustration of the used forecasting framework.

III. Results and Discussion

This study's training data to construct each model includes confirmed and recovered cases from January 22, 2020, to June 6, 2021. We used six days for the testing period from June 7, 2021, to June 12, 2021. For the OGPR, OSVR, and OEL models, we applied the BO procedure to get the optimal parameters maximizing the forecasting precision. The hyperparameter search ranges for each model are listed in Table II. The computed values of the hyperparameters of each model using the BO algorithm are summarized in Table III.

The fifteen models are constructed using training data and then applied for forecasting confirmed and recovered COVID-19 cases for a 6 day forecast horizon from June 7, 2021. The observed test set together with model forecasts of confirmed and recovered cases in India and Brazil are displayed in Figures 3 and 4, respectively. From Figure 3, we observe that the forecasted values of the confirmed and recovered cases in India from the considered models are closer to the actual data, indicating good forecast performance. For the confirmed and recovered cases in Brazil, Figure 4, shows broader bands around the actual cases, indicating wider variations among model predictions. In this scenario, models showed relatively better forecasts for India confirmed and recovered cases series.

Tables IV and IV quantifies the performances of each model in terms of RMSE, MAE, and MAPE, for COVID-19 data recorded in India and Brazil, respectively. In terms of all metrics calculated, the GPR models showed the best performance in terms of RMSE, and MAE. It could be attributed to its capacity to capture dynamics in time-series data.

TABLE II
HYPERPARAMETERS SEARCH RANGE.

Model	Hyperparameter search range
SVRO	• Box constraint: 0.001-1000 • Kernel scale: 0.001-1000 • Epsilon: 0.18495-18495.1816 • Kernel function: Gaussian, Linear, Quadratic, Cubic • Standardize data: true, false
GPRO	• Sigma: 0.0001-1441.9316 • Basis function: Constant, Zero, Linear • Kernel function: Nonisotropic Exponential, Nonisotropic Matern 3/2, Nonisotropic Matern 5/2, Nonisotropic Rational Quadratic, Nonisotropic Squared Exponential, Isotropic Exponential, Isotropic Matern 3/2, Isotropic Matern 5/2, Isotropic Rational Quadratic, Isotropic Squared Exponential • Kernel scale: 0.498-498 • Standardize: true, false
ELO	• Ensemble method: Bag, LSBoost • Number of learners: 10-500 • Learning rate: 0.001-1 • Minimum leaf size: 1-249 • Number of predictors to sample: 1-2

TABLE III
OPTIMIZED HYPERPARAMETERS USING THE BO ALGORITHM.

Model	Optimized Hyperparameters
SVRO	• Box constraint: 1.7128 • Kernel scale: 1 • Epsilon: 1.3156 • Kernel function: Cubic • Standardize data: true
GPRO	• Sigma: 1217.1288 • Basis function: Linear • Kernel function: Nonisotropic Matern 5/2 • Kernel scale: 493.0376 • Standardize: false
ELO	• Ensemble method: LSBoost • Number of learners: 11 • Learning rate: 0.98438 • Minimum leaf size: 2 • Number of predictors to sample: 2

Figure 5 displays the heatmap of the MAPE values achieved by the investigated model for the confirmed and recovered COVID-19 data from Indian and Brazil. We observe that GPR models achieved the best forecasting performance with the lowest MAPE values. This could be attributed to the extended capacity of the GPR models to learn dynamics in COVID-19 time-series data. Furthermore, this study shows the capability of machine learning models to forecast the future trends of COVID-19.

TABLE IV
TH OBTAINE STATISTICAL CRITERIA FOR CONFIRMED AND RECOVERED COVID-19 CASES FORECASTS IN INDIA.

Series	Model	RMSE	MAE	MAPE
Confirm India	SVRO	22337053.113	22334295.732	38.960
Confirm India	SVR_C	1012357.846	1008215.980	3.365
Confirm India	SVR_{CG}	1382637.913	1369677.347	4.967
Confirm India	SVR_{FG}	6967701.735	6507706.629	30.360
Confirm India	SVR_L	759414.577	759392.528	2.697
Confirm India	SVR_{MG}	2356188.304	2280681.541	8.581
Confirm India	SVR_Q	1024262.932	1019713.136	3.401
Confirm India	GPR_{RQ}	37398.517	32479.864	0.112
Confirm India	GPR_{SE}	36208.928	30442.130	0.105
Confirm India	GPR_{M52}	14350.001	12258.416	0.072
Confirm India	GPR_{Exp}	972208.519	905902.811	3.233
Confirm India	GPRO	111506.899	108951.780	0.374
Confirm India	BT	2005011.686	1974055.703	7.325
Confirm India	BS	2779609.635	2757363.556	10.538
Confirm India	ELO	1625388.219	1587044.713	5.806
Recoved India	SVRO	21100097.634	21087507.583	11.320
Recoved India	SVR_C	1125965.648	1107063.877	3.919
Recoved India	SVR_{CG}	1771552.327	1718529.593	6.779
Recoved India	SVR_{FG}	10306345.167	9339278.915	59.969
Recoved India	SVR_L	754472.349	754424.474	2.877
Recoved India	SVR_{MG}	3670129.936	3373505.792	14.480
Recoved India	SVR_Q	1179022.579	1155346.231	4.080
Recoved India	GPR_{RQ}	167795.963	143454.775	0.527
Recoved India	GPR_{SE}	30214.921	23379.830	0.085
Recoved India	GPR_{M52}	54374.745	48482.147	0.178
Recoved India	GPR_{Exp}	1524148.063	1336405.700	5.208
Recoved India	GPRO	58832.766	46691.520	0.052
Recoved India	BT	3078226.027	2990707.504	11.681
Recoved India	BS	3467540.137	3390087.093	14.359
Recoved India	ELO	1871290.946	1723538.715	6.127

TABLE V
TH OBTAINE STATISTICAL CRITERIA FOR CONFIRMED AND RECOVERED COVID-19 CASES FORECASTS IN BRAZIL.

Series	Model	RMSE	MAE	MAPE
Confirm Brazil	SVRO	178495.629	176859.741	1.055
Confirm Brazil	SVR_C	859664.899	846897.020	4.749
Confirm Brazil	SVR_{CG}	856493.084	849454.839	5.279
Confirm Brazil	SVR_{FG}	2681791.796	2400829.100	17.153
Confirm Brazil	SVR_L	658423.358	657587.460	4.041
Confirm Brazil	SVR_{MG}	1201574.167	1157539.469	7.350
Confirm Brazil	SVR_Q	100175.382	92240.318	0.552
Confirm Brazil	GPR_{RQ}	22347.367	20542.504	0.122
Confirm Brazil	GPR_{SE}	36548.399	29617.766	0.175
Confirm Brazil	GPR_{M52}	25452.517	22951.994	0.136
Confirm Brazil	GPR_{Exp}	499989.497	426469.985	2.585
Confirm Brazil	GPRO	22821.043	21485.641	0.127
Confirm Brazil	BT	819117.717	776873.997	4.811
Confirm Brazil	BS	1414426.255	1390388.795	8.951
Confirm Brazil	ELO	1471998.737	1448916.717	3.363
Recoved Brazil	SVRO	30627.182	30583.310	7.618
Recoved Brazil	SVR_C	167774.226	167591.124	4.806
Recoved Brazil	SVR_{CG}	151929.129	151847.006	5.351
Recoved Brazil	SVR_{FG}	259341.166	254940.374	18.745
Recoved Brazil	SVR_L	129890.658	129878.098	3.400
Recoved Brazil	SVR_{MG}	181760.427	181092.640	7.876
Recoved Brazil	SVR_Q	30362.953	30188.755	1.473
Recoved Brazil	GPR_{RQ}	1719.567	1558.378	0.241
Recoved Brazil	GPR_{SE}	1656.806	1525.704	0.247
Recoved Brazil	GPR_{M52}	1667.333	1508.890	0.188
Recoved Brazil	GPR_{Exp}	40549.196	37897.498	2.935
Recoved Brazil	GPRO	1667.329	1508.883	0.188
Recoved Brazil	BT	84573.683	83349.510	4.581
Recoved Brazil	BS	247302.171	246886.202	9.251
Recoved Brazil	ELO	36794.352	33885.945	3.489

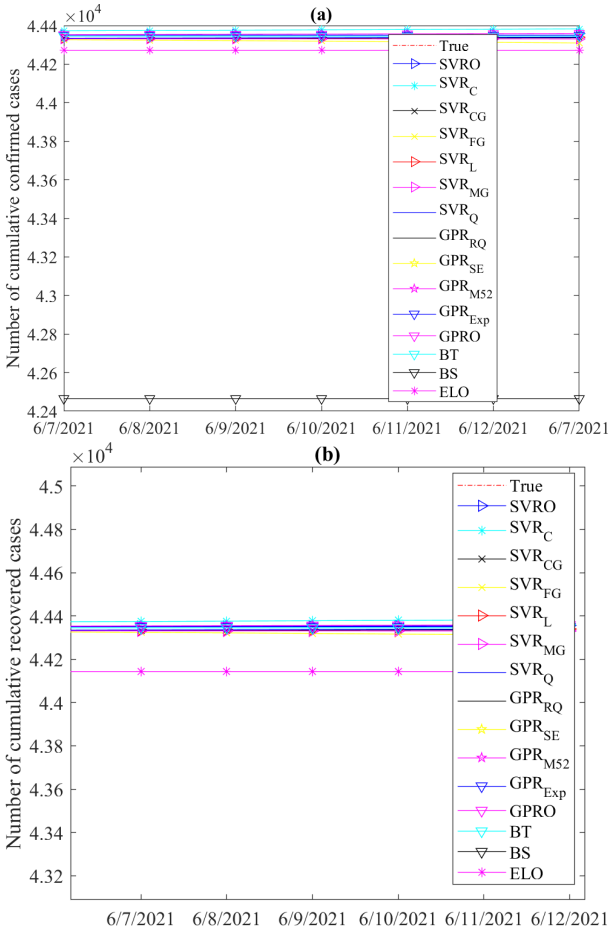

Fig. 3. Records and forecasts of (a) confirmed and (b) recovered COVID-19 cases in India for testing period, using the fifteen machine learning methods.

IV. CONCLUSION

In this work, we applied 16 machine learning models to forecasting COVID-19 cases on datasets of two countries, India and Brazil. Here, the GPR model has been optimized using Bayesian optimization and compared with fifteen machine learning models. In short, the forecasting result shows that GPR models achieved superior performance compared to the other models in terms of RMSE, MAE, and MAPE. Of course, machine learning approves its ability in the medical field to forecast COVID-19 cases. We will consider the dynamic models in future work, which consider the number of past days 1-7 instead of one day only to have more features that will help get a better result. Also, using deep learning to build models works with a small dataset.

REFERENCES

[1] WHO. The Coronavirus (COVID-19), WHO, Geneva, Swit-zerland. [Online]. Available: https://covid19.who.int/

[2] A. Dairi, F. Harrou, A. Zeroual, M. M. Hittawe, and Y. Sun, "Comparative study of machine learning methods for covid-19 transmission forecasting," *Journal of Biomedical Informatics*, vol. 118, p. 103791, 2021.

[3] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting covid-19 time-series data: A comparative study," *Chaos, Solitons & Fractals*, vol. 140, p. 110121, 2020.

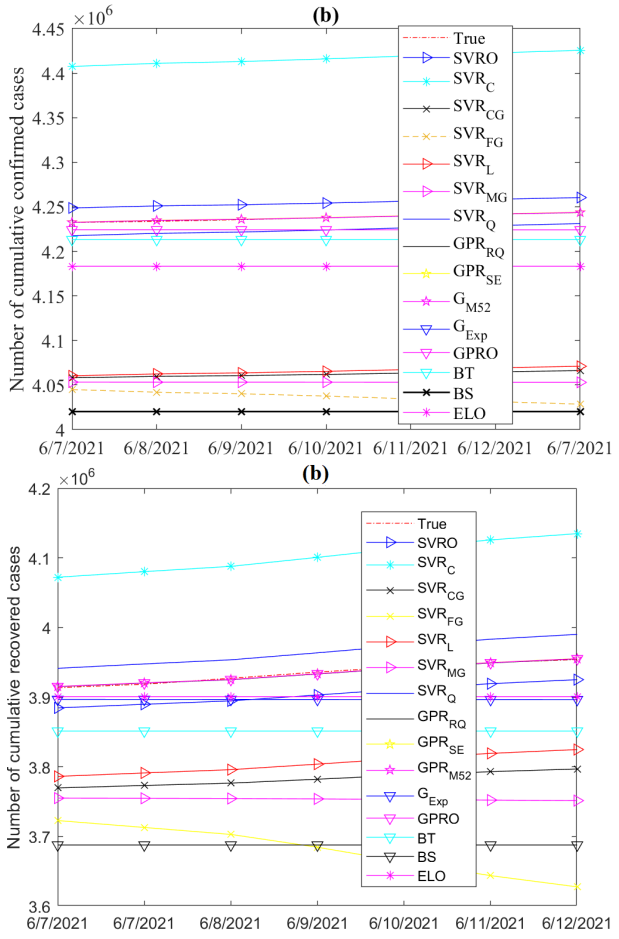

Fig. 4. Records and forecasts of (a) confirmed and (b) recovered COVID-19 cases in Brazil for testing period, using the fifteen machine learning methods.

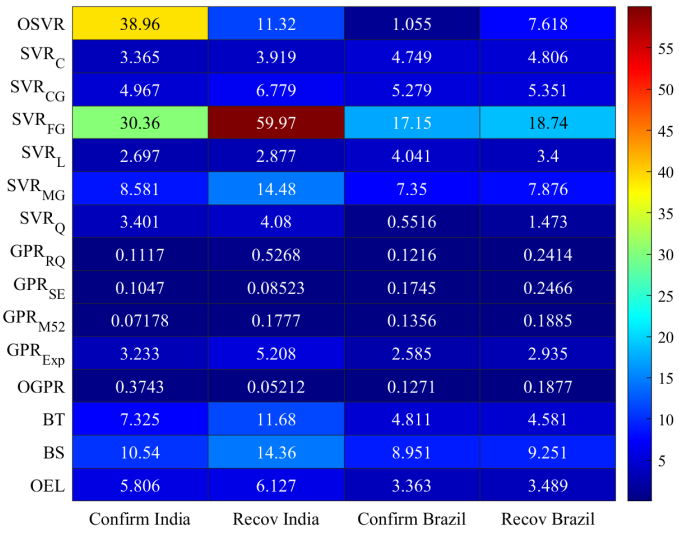

Fig. 5. HeatMap of MAPE values obtained using the fiteen models.

[4] F. Harrou, A. Dairi, F. Kadri, and Y. Sun, "Forecasting emergency department overcrowding: A deep learning framework," *Chaos, Solitons & Fractals*, vol. 139, p. 110247, 2020.

[5] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," *Chaos, Solitons & Fractals*, vol. 135, p. 109853, 2020.

[6] M. A. Al-Qaness, A. I. Saba, A. H. Elsheikh, M. Abd Elaziz, R. A. Ibrahim, S. Lu, A. A. Hemedan, S. Shanmugan, and A. A. Ewees, "Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil," *Process Safety and Environmental Protection*, vol. 149, pp. 399–409, 2021.

[7] K. Konarasinghe, "Modeling COVID-19 epidemic of india and brazil," *Journal of New Frontiers in Healthcare and Biological Sciences*, vol. 1, no. 1, pp. 15–25, 2020.

[8] S. Mangla, A. K. Pathak, M. Arshad, and U. Haque, "Short-term forecasting of the covid-19 outbreak in india," *International health*, 2021.

[9] R. Sujath, J. M. Chatterjee, and A. E. Hassanien, "A machine learning forecasting model for COVID-19 pandemic in India," *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 959–972, 2020.

[10] V. K. Sharma and U. Nigam, "Modeling and forecasting of covid-19 growth curve in india," *Transactions of the Indian National Academy of Engineering*, vol. 5, no. 4, pp. 697–710, 2020.

[11] S. Chordia and Y. Pawar, "Analyzing and Forecasting COVID-19 Outbreak in India," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, 2021, pp. 1059–1066.

[12] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[13] F. Harrou, A. Saidi, Y. Sun, and S. Khadraoui, "Monitoring of photovoltaic systems using improved kernel-based learning schemes," *IEEE Journal of Photovoltaics*, vol. 11, no. 3, pp. 806–818, 2021.

[14] Y. Xie, K. Zhao, Y. Sun, and D. Chen, "Gaussian processes for short-term traffic volume forecasting," *Transportation Research Record*, vol. 2165, no. 1, pp. 69–78, 2010.

[15] J. Lee, W. Wang, F. Harrou, and Y. Sun, "Reliable solar irradiance prediction using ensemble learning-based models: A comparative study," *Energy Conversion and Management*, vol. 208, p. 112582, 2020.

[16] C. K. Williams and C. E. Rasmussen, "Gaussian processes for regression," 1996.

[17] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, vol. 85, pp. 1–16, 2018.

[18] E. Protopapadakis, A. Voulodimos, and N. Doulamis, "An investigation on multi-objective optimization of feedforward neural network topology," in *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2017, pp. 1–6.

[19] A. D. Bull, "Convergence rates of efficient global optimization algorithms." *Journal of Machine Learning Research*, vol. 12, no. 10, 2011.

[20] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.

[21] J. T. Springenberg, A. Klein, S. Falkner, and F. Hutter, "Bayesian optimization with robust bayesian neural networks," *Advances in neural information processing systems*, vol. 29, pp. 4134–4142, 2016.

[22] V. Vapnik, S. E. Golowich, A. Smola *et al.*, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in neural information processing systems*, pp. 281–287, 1997.

[23] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 308–324, 2015.

[24] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[25] B. Khaldi, F. Harrou, S. M. Benslimane, and Y. Sun, "A data-driven soft sensor for swarm motion speed prediction using ensemble learning methods," *IEEE Sensors Journal*, 2021.

[26] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer school on machine learning*. Springer, 2003, pp. 63–71.