

Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization



Hossein Abbasimehr*, Reza Paki

Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran

ARTICLE INFO

Article history:

Received 24 September 2020

Accepted 23 November 2020

Available online 28 November 2020

Keywords:

COVID-19

Deep learning

Multi-head attention

CNN

LSTM

Bayesian optimization

ABSTRACT

COVID-19 virus has encountered people in the world with numerous problems. Given the negative impacts of COVID-19 on all aspects of people's lives, especially health and economy, accurately forecasting the number of cases infected with this virus can help governments to make accurate decisions on the interventions that must be taken. In this study, we propose three hybrid approaches for forecasting COVID-19 time series methods based on combining three deep learning models such as multi-head attention, long short-term memory (LSTM), and convolutional neural network (CNN) with the Bayesian optimization algorithm. All models are designed based on the multiple-output forecasting strategy, which allows the forecasting of the multiple time points. The Bayesian optimization method automatically selects the best hyperparameters for each model and enhances forecasting performance. Using the publicly available epidemical data acquired from Johns Hopkins University's Coronavirus Resource Center, we conducted our experiments and evaluated the proposed models against the benchmark model. The results of experiments exhibit the superiority of the deep learning models over the benchmark model both for short-term forecasting and long-horizon forecasting. In particular, the mean SMAPE of the best deep learning model is 0.25 for the short-term forecasting (10 days ahead). Also, for long-horizon forecasting, the best deep learning model obtains the mean SMAPE of 2.59.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Coronavirus 2019 (COVID-19) pandemic [1] has spread from Wuhan, China to other countries in the world. It has high viral infectivity and a rapid rate of spread compared to prior infectious diseases which makes its control hard [2]. Since its emergence, COVID-19 disease has encountered people in the world with many problems. It has more negative impacts on people's health and interrupted the economy. As a result, many countries have implemented strong interventions to control the spread of the epidemic and to reduce the negative effects of COVID-19 disease [3]. Although the interventions vary between countries, the commonly adopted interventions are social distancing, border closure, school closure, lockdown, travel banning, and public events banning [4]. The effectiveness of interventions across 11 European countries has been investigated in Flaxman, Mishra [4] concluding that the adopted interventions were effective in reducing the rate of transmission of COVID-19 epidemic.

To evaluate the success of controlling COVID-19 epidemic, it is vital to accurately monitor and reveal the data about the number

of cases infected with it [2]. Making public the data of confirmed cases of countries in the world allow academics to conduct modeling on data in order to gain useful knowledge about the trend of the disease. Johns Hopkins University's Corona Virus Resource Center [5] has collected and published the data about the COVID-19 confirmed cases which are used by scholars to model the spread of the disease and perform data analysis.

Given the negative impacts of COVID-19, accurately forecasting the number of cases infected with this virus is a vital task to reveal the trend of the disease and thereby to help governments to take preventive measures [6]. Previous researches on COVID-19 time series forecasting have adopted mathematical and computational intelligence models to forecast the number of confirmed cases. In [7] the adaptive neuro-fuzzy inference system (ANFIS) was employed to forecast the number of infected cases in China. In [3] mathematical and computational models such as Logistic, Gompertz, and ANN were applied to model the number of cases in Mexico. Castillo and Melin [8] proposed a new combined approach with fuzzy fractal and fuzzy logic to predict the number of confirmed cases of COVID-19 in 10 countries. Also, in [9], a new ensemble approach based on ANNs and fuzzy aggregation was proposed and its performance was evaluated on COVID-19 time series of Mexico and its 12 states which showed significant improvement

* Corresponding author.

E-mail address: abbasimehr@azaruniv.ac.ir (H. Abbasimehr).

Table 1

Summary of studies on COVID-19 infection forecasting.

Reference	Modeling techniques	Country	Date
[7]	ANFIS	China	21 January, 2020 to 18 February, 2020
[19]	Logistic model, Bertalanffy model and Gompertz model	China	15 January, 2020 to 4 April 2020
[20]	Gompertz and Logistic	China, South Korea, Italy, and Singapore	Until 27 March, 2020
[3]	Gompertz, Logistic Artificial Neural Networks	Mexico	February 27, 2020 to May 8, 2020
[6]	ANN, ARIMA	Iran	Trainset: 19 February, 2020 to 24 March, 2020 Test set: 25 March, 2020 to 31 March, 2020
[8]	Fuzzy Fractal	Ten countries: US, United Kingdom, Turkey, Spain, Mexico, Italy, Iran, Germany, France, and Belgium	July 22, 2020 to 7 August, 2020
[9]	An ensemble of neural network models with fuzzy aggregation	Mexico and 12 states in Mexico	Not available
[2]	ARIMA, nonlinear autoregression neural network (NARNN), and LSTM	Denmark, Belgium, Germany, France, United Kingdom, Finland, Switzerland and Turkey	Until 3 May, 2020
[10]	Bi-directional LSTM, Stacked LSTM, and Convolutional LSTM	India (32 Indian states)	March 14, 2020- May 14, 2020
[11]	ARIMA, support vector regression (SVR), LSTM, GRU, and Bi-LSTM	Ten countries: Brazil, China, Germany, India, Israel, Italy, Russia, Spain, UK, USA	Until June 27, 2020
[12]	LSTM	Russia, Peru and Iran	Until July 7, 2020

than single ANN. In recent studies [2,10–12], deep learning methods such as LSTM and bidirectional LSTM (BiLSTM) have been utilized for COVID-19 time series forecasting. The results indicated that LSTM and its variants have good performance in predicting the COVID-19 time series. In the literature review section, we will give a comprehensive review of studies related to COVID-19 time series forecasting.

Although LSTM was recently applied for COVID-19 infection forecasting, the predictive power of other deep learning methods that are suitable for sequence processing problems has not been explored in COVID-19 forecasting context. Therefore, in this paper, in addition to LSTM [13], we focus on the other deep learning models including the multi-head attention [14], and CNNs [15] to forecast the number of cases of COVID-19. Furthermore, the performance of deep learning methods mainly influenced by hyperparameter tuning [16]. There are several hyperparameters that must be specified when employing a deep learning model. The previous studies on COVID-19 forecasting using the LSTM method have not exploited an optimization method to identify the optimal hyperparameters. Most of those studies (e.g. [2,10,12]) have implemented models using hand-tuned hyperparameters. As another contribution, in this study, we utilize the Bayesian Optimization method [17] in order to optimize the hyperparameters of Multi-head attention, LSTM, and CNN. Besides, the design of proposed methods is based on the multiple output approach that allows forecasting of the number of cases for multiple next days.

Overall, the main contributions of this study are as follows:

- 1 Adopting the deep learning models to predict the number of daily infected cases with COVID-19.
- 2 Exploiting the Bayesian Optimization for optimal parameter selection.
- 3 Adopting a multiple-output modeling approach: The models are designed to be multi-output to predict the next few days. The usual approach to multi-step-ahead prediction is iterated one-step-ahead forecasting in which the forecasting of the n next steps performed as a n single step-ahead forecasting. Multi-output forecasting is an effective choice for long-horizon forecasting [18].

The deep learning models are applied on COVID-19 data of the top 10 countries with the highest number of infections. To evaluate the performance of the proposed models, we perform two sets of experiments. The first set of experiments explores the effectiveness

of the proposed models in short-term forecasting and compares their performance with the results of the fuzzy fractal model presented in [8]. The results indicated the deep models achieve better performance than the fuzzy fractal across all countries. Also, the second set of experiments are conducted to investigate the prediction power of the devised models in a wider forecasting window. The results can help governments in long-term decision making to control the pandemic.

The rest of this paper is organized as follows. In Section 2, we provide a comprehensive literature review on models and methods proposed for COVID-19 time series forecasting. Section 3 describes the structure of the proposed models. In Section 4, we describe the data and provide the detailed results of the proposed models and compare their performance to the benchmark model. Section 5 concludes the paper and outlines future work.

2. COVID-19 time series forecasting

In this section, we summarize the previous studies in the context of COVID-19 time series prediction. Since the publicly available data of COVID-19 contains daily statistics of the confirmed cases, so it is considered as a time series data and the time series forecasting techniques can be exploited to this data. Table 1 illustrates the researches on COVID-19 time series forecasting. The table highlights the modeling techniques, the countries, and the time period of the utilized data in each study. As Table 1 indicates, various types of methods including mathematical, statistical, machine and deep learning, and fuzzy logic-based techniques have been employed for COVID-19 time series forecasting. From mathematical models, the Gompertz model and logistic models have been used in several studies (i.e. [3,19,20]). Also, from statistical methods, the Auto-Regressive Integrated Moving Average (ARIMA) approach has been employed in some studies such as [2,6,11]. Besides, the machine and deep learning techniques such as ANN and LSTM have exhibited improvements in COVID-19 time series forecasting studies (e.g. [2,10,12]). Also, some methods based on fuzzy logic have been proposed in the literature (e.g. [7,8]). As the literature review indicates, the exploitation of deep learning models has led to improvements in the prediction of COVID-19 cases [2,10–12]. Since the COVID-19 time series forecasting task is a kind of sequence processing, other deep learning models can be adopted to forecast the COVID-19 time series [12]. The remarkable characteristic of the machine and deep learning methods is their ability

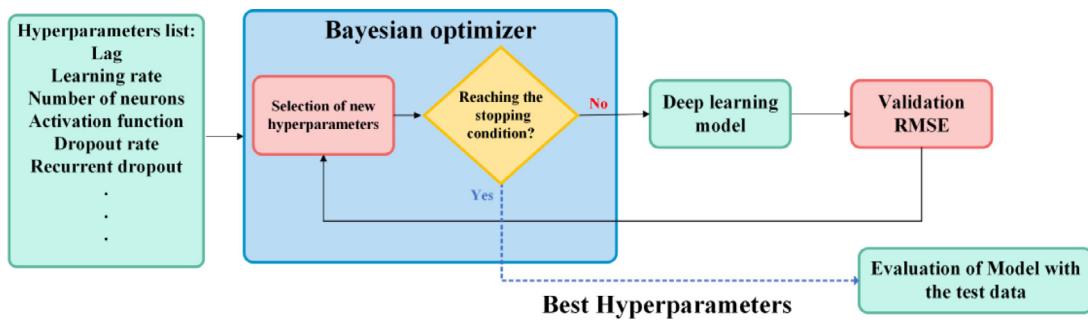


Fig. 1. The general procedure of the proposed models.

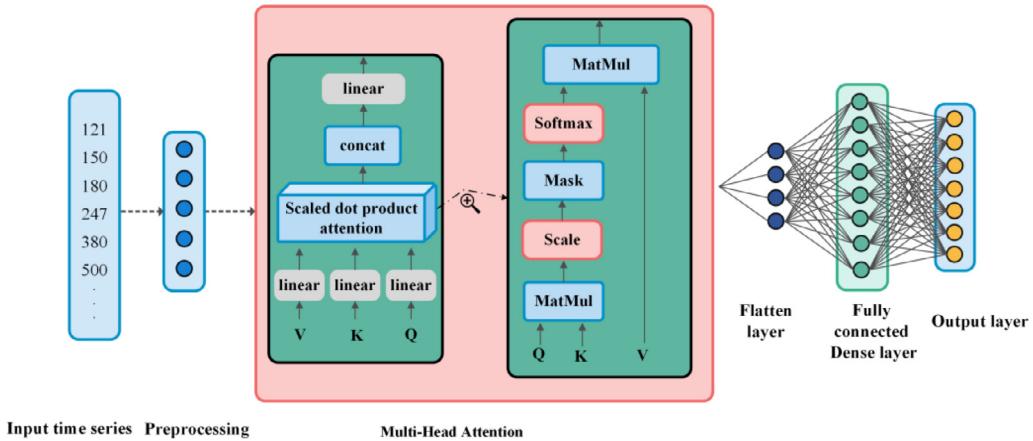


Fig. 2. The proposed attention-based model (ATT_BO).

to capture nonlinear patterns [21], which makes them suitable for modeling complex time series.

In the recent years, in addition to the LSTM model, other types of deep learning models such as methods based on the attention mechanisms and convolutional neural networks have demonstrated promising results in many areas of applications such as natural language processing (NLP) [22], stock market price forecasting [21] and so on. Investigating the literature on COVID-19 forecasting reveals that attention mechanism and the convolutional neural network have not been employed for COVID-19 prediction. Therefore, this study aims to propose deep learning models based on these methods to evaluate their effectiveness in forecasting COVID-19 infected cases.

3. The proposed models

In this study, we consider three different deep learning methods to predict the cumulative number of cases. The three proposed methods are the multi-head attention-based method (ATT_BO), CNN-based method (CNN_BO), and LSTM-based method (LSTM_BO). As illustrated in Fig. 1, all proposed methods are combined with the Bayesian optimization algorithm to select the optimal values of hyperparameters. In Fig. 1, the Bayesian optimizer [23] accomplishes the task of identifying the optimal hyperparameters. A common alternative to Bayesian optimization is the grid search which is a time-consuming method. The reason for choosing Bayesian optimization are: (1) the superiority of Bayesian optimization over grid search has been proved in previous studies [24] (2) unlike grid search, Bayesian optimization can efficiently find the optimal hyperparameters with fewer iterations [25]. In the following subsections, we describe the structure of the proposed models.

3.1. ATT_BO

Recently attention mechanisms have been employed successfully in the sequence processing tasks and especially in natural language processing applications [21,22]. The study of Vaswani, Shazeer [26] demonstrated the effectiveness of the attention mechanism for processing sequence data. In this study, we propose a multi-head attention-based model for COVID-19 forecasting using the multi-head attention mechanism developed in [26] (Fig. 2). An attention function takes a query Q and a set of keys and values $< K, V >$ to get the output O . This procedure is often called Scaled Dot-Product Attention. Multi-head attention is a set of multiple heads that jointly learn different representations at every position in the sequence [14]. The proposed attention method (ATT_BO) has three main parts including the multi-head attention layer, the flatten layer, and the fully connected layer. After preprocessing the input data and creating the instances, the multi-head attention layer computes a new representation of the input data which are more informative than the input data. The output of the multi-head attention layer is reshaped using the flatten layer and finally, the outputs are produced using the fully connected layer. The superiority of the proposed model is attributed to the multi-head attention layer which has the ability to capture the most important input features and gives higher weights to them.

3.2. LSTM_BO

Deep learning methods such as RNNs are suitable for sequence processing as they consider the temporal behavior of a given time series [21]. But, the main shortcoming of RNNs is the vanishing/exploding gradient problem that makes their training a difficult task [27]. To overcome this problem, LSTM which is a kind of gated RNNs are often employed [28]. The structure of an LSTM block is

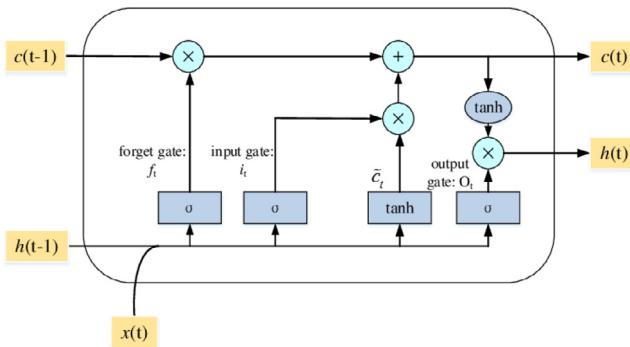


Fig. 3. The structure of the LSTM [27].

depicted in Fig. 3. Each LSTM block consists of a memory cell along with three gates including an input gate i_t , the forget gate f_t and the output gate o_t which regulate the flow of information to its cell state c_t :

Each of the three gates accomplishes a different operation [29]:

- The forget gate determines which information is discarded.
- The input gate decides which information is input to the cell state.
- The output gate regulates the outgoing information of the LSTM cell.

The architecture of the proposed LSTM-based (LSTM_BO) is articulated in Fig. 4. This method consists of three main parts, including the LSTM layer, the flatten layer, and the fully connected layer. The input time series is firstly preprocessed and then is fed into the LSTM layer, which learns a new representation of data considering the dependency among data. Afterward, the output of the LSTM layer is reshaped into a suitable format using a flatten layer and then is fed into a fully connected layer. Finally, the fully connected layer produces multiple outputs.

3.3. Convolutional model

CNNs are quite successful in processing machine vision problems [15]. In this study, we implement CNN for COVID-19 time series forecasting. The convolutional layers in CNNs take input data and apply convolution operation on data using convolution kernels to extract new features. The convolution kernel is a small window that slides over the input data and performs convolutional operations to extract new features [30]. The derived features using the convolution operation are usually more discriminative than the raw input data, therefore, improving the forecasting. The architecture of the proposed CNN-based model (CNN_BO) is described in Fig. 5. CNN_BO contains three main parts: the convolution layer, the flatten layer, and the fully connected layer. After preprocessing of the input data, features are extracted from the input time

series using the convolution layer, and then the flatten layer reshapes data into a format that can be used by the fully connected layer and the fully connected layer generates the multiple outputs.

4. Empirical study and analysis

4.1. Data

The data utilized in this study was obtained from the Humanitarian Data Exchange (HDX) [31]. In this study, we perform two sets of experiments using two different datasets, including Dataset 1 and Dataset 2 that are described in Table 2. The first set of experiments examine the usefulness of the proposed deep learning model in a shorter 10 days window. To perform the first set of experiments, we utilize Dataset 1 which contains the data used in [8]. To compare the results of the proposed methods, we choose the fuzzy fractal method proposed by Castillo and Melin [8] as the benchmark.

Also, to evaluate the performance of the three proposed models in long-horizon forecasting, we use Dataset 2 that includes the updated data of COVID-19 cases until 3 August. Similar to Dataset 1, Dataset 2 contains data for ten countries with the highest number of cases. In selecting the top ten countries of Dataset 2, we firstly aggregate the data of all cities for each country.

4.2. Evaluation measures

To evaluate the effectiveness of the proposed methods on COVID-19 time series forecasting, we employ three primary measures including symmetric mean absolute percentage error (SMAPE), mean absolute percentage error (MAPE), and root mean square error (RMSE), as well as the following aggregate measures, which are based on the primary measures including mean of SMAPEs (Mean SMAPE), mean of the SMAPE ranks (Rank SMAPE), mean of MAPEs (Mean MAPE), mean of the MAPE ranks (Rank MAPE), mean of RMSEs (Mean RMSE) and the mean of RMSE ranks (Rank RMSE).

The definitions of SMAPE, MAPE, and RMSE are given by Eqs. (1)–(3) respectively:

$$\text{SMAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{\frac{|\hat{y}_t| + |y_t|}{2}} \times 100 \quad (1)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{|y_t|} \times 100 \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (3)$$

where \hat{y}_t and y_t are the predicted and actual value at time point t .

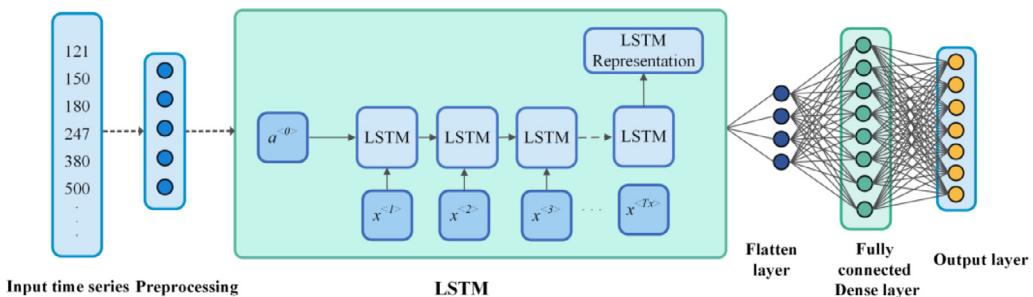


Fig. 4. The Proposed LSTM-based model.

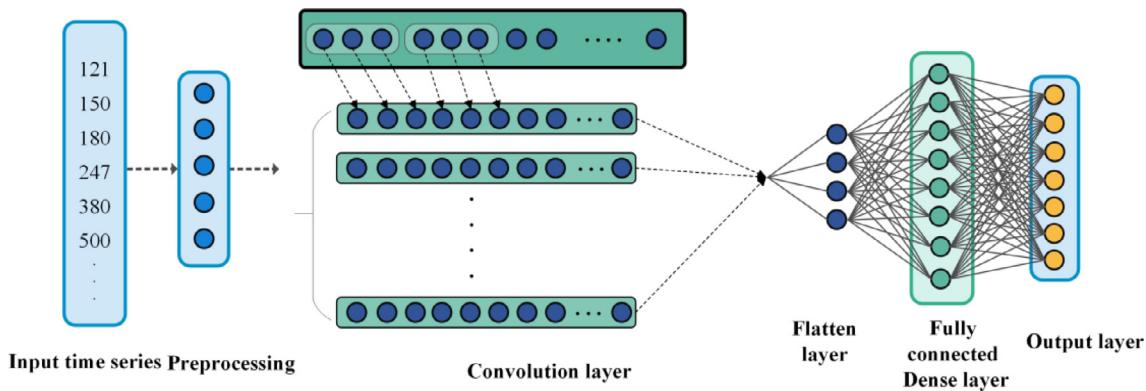


Fig. 5. The proposed CNN-based model.

Table 2
The description of data.

Dataset	Countries	Time period
Dataset 1	US, United Kingdom, Turkey, Spain, Mexico, Italy, Iran, Germany, France, Belgium	January 20, 2020–August 1, 2020
Dataset 2	US, Brazil, India, Russia, South Africa, Mexico, Peru, Chile, Colombia, Iran	January 20, 2020–August 3, 2020

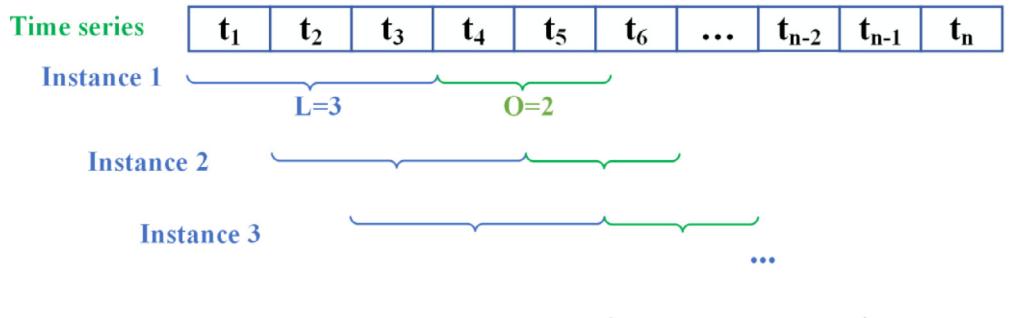


Fig. 6. The Process of instance generation.

4.3. Preprocessing of data

In this study, as the architectures of the three proposed models indicate, we design the models following the multi-output forecasting strategy, which allows forecasting of multiple time steps rather than a single time step that is applied in the single-output strategy.

The proposed models require the input to be instances (data objects) of input-output format. So, the input time series must be converted into the input-output format. Therefore, considering the input size, L (Lag), which refers to the length of the input window, and the output size, O , which denotes the length of the output window, subsequences of length $L + O$ are extracted from the series. The first L points of a sequence are considered as the input, and the last O points are considered as the output values. For example, as depicted in Fig. 6, the process of the construction of the instances iteratively generates the instances using the input= 3 ($L=3$) and the output size $O=2$.

4.4. Experiment setup

In this study, we combine the proposed methods with the Bayesian optimization algorithm to identify the optimal hyperparameter value. The proposed methods the proposed method are implemented using Keras library in python [32]. To prevent all methods from overfitting and improving their generalization to

Table 3
The range of hyperparameters used in the Bayesian optimization process.

Model	Hyperparameter range
ATT_BO	Activation function: (ReLU, Linear)
LSTM_BO	Activation function: (ReLU, Linear, Tanh) Dropout rate: (0.0, 0.1, 0.2, 0.3, 0.4, 0.50) Number of neurons: (32, 64, 128, 256)
CNN_BO	Size of kernel: (2, 3, 4, 5, 6) Stride: (1, 2) Number of neurons: (32, 64, 128, 256)

new data, we use early stopping [33]. To employ early stopping, we set the epoch limit to 500.

4.4.1. Hyperparameter selection

To utilize the Bayesian optimizer, the range of the hyperparameters should be specified. One important hyperparameter which significantly impacts time series forecasting accuracy is the size of the input window (Lag). The range of Lag is set to (10, 11, 12, 13, 14, 15) for all proposed methods. Table 3 provides the range of hyperparameters utilized throughout the experiments. As the fully connected and output layers have been incorporated after the main layer of the proposed methods; for all deep learning models, we set the range of hyperparameters corresponding to these layers identical. To limit the search space of the Bayesian optimization algorithm, for these layers, we include their activation functions

Table 4

The performance of the proposed methods in terms of SMAPE on Dataset 1.

Country	ATT_BO	LSTM_BO	CNN_BO	Fuzzy fractal
US	0.4082	0.5325	0.2776	1.0755
UK	0.0464	0.056	0.0504	1.0147
Turkey	0.0412	0.0475	0.0984	0.0085
Spain	0.6536	0.62	0.6119	0.3572
Mexico	0.5171	0.5668	0.5684	0.693
Italy	0.0438	0.1117	0.0626	1.5343
Iran	0.0685	0.1313	0.0577	1.5343
Germany	0.1562	0.2321	0.1823	0.1174
France	0.3956	0.3169	0.313	0.2894
Belgium	0.2754	0.4366	0.2519	0.4281

Table 5

The performance of the proposed methods in terms of MAPE on Dataset 1.

Country	ATT_BO	LSTM_BO	CNN_BO	Fuzzy fractal
US	0.317	0.5314	0.276	1.0691
UK	0.0402	0.0542	0.0456	1.0214
Turkey	0.0412	0.0182	0.0984	0.0085
Spain	0.4977	0.5947	0.6025	0.3581
Mexico	0.4389	0.5355	0.5187	0.6901
Italy	0.0409	0.1114	0.0624	0.0551
Iran	0.0538	0.1269	0.0428	1.5196
Germany	0.1461	0.2128	0.1804	0.1173
France	0.3208	0.3033	0.3088	0.2893
Belgium	0.2609	0.39432	0.2491	0.4287

Table 6

The performance of the proposed methods in terms of RMSE on Dataset 1.

Country	ATT_BO	LSTM_BO	CNN_BO	Fuzzy fractal
US	20023.27	26415.24	15181.05	27609.68
UK	164.8403	193.7	180.567	3494.91
Turkey	99.43	115.42	240.66	27.303
Spain	2320.7	2273.54	2269.22	1398.52
Mexico	2511.54	2825.99	2781.7	3069.18
Italy	126.71	298.89	173.7	168.08
Iran	243.015	417.944	198.11	5135.7
Germany	395.75	537.7	436.9	333.42
France	1035.42	910.24	894.56	782.001
Belgium	230.52	369.39	208.2	312.61

in the hyperparameter selection process. For both layers, “ReLU” and “Linear” activation functions [15] are utilized. Also, the range of learning rate parameter for all models is set to (0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05).

4.5. Results and analysis

In this section, we give the results of the experiments conducted based on the two datasets. In the analysis of the first set of experiments, we consider the results of the fuzzy fractal model proposed in [8]. The main reason behind choosing the fuzzy fractal method as the benchmark is that this method was comprehensively evaluated in the recent study conducted by Castillo and Melin [8] using Dataset 1. Besides, on the second set of experiments, we explore the performance of our developed models on a wider forecasting window by adopting a multi-output forecasting strategy.

4.5.1. Results of the first set of experiments on Dataset 1

To make the forecasting comparable with the results of the fuzzy fractal model [8], for Dataset 1, we consider the last 10 days as the test points .The results of the proposed models as well as the benchmark model on Dataset 1 are illustrated in Tables 4–6. As

Table 7

The performance of all methods in terms of Mean SMAPE, Mean MAPE, Mean RMSE, Rank SMAPE, Rank MAPE, rank RMSE (the best results are marked bold) on Dataset 1.

Method	ATT_BO	LSTM_BO	CNN_BO	Fuzzy fractal
Mean SMAPE	0.2606	0.3051	0.2474	0.7052
Mean MAPE	0.2157	0.2883	0.2385	0.5557
Mean RMSE	2715.12	3435.80	2256.47	4233.14
Rank SMAPE	2.1	3	2.2	2.7
Rank MAPE	2	3	2.4	2.6
Rank RMSE	2	3.3	2.1	2.6

the results indicate, in terms of SMAPE (Table 4), ATT_BO achieves better performance compared to the Fuzzy fractal in 6 countries out of 10 countries such as the US, UK, Mexico, Italy, Iran, and Belgium. Furthermore, CNN_BO obtains better performance in terms of SMAPE in comparison with the Fuzzy fractal method in 6 countries including the US, UK, Mexico, Italy, Iran, and Belgium. Also, the results of LSTM_BO indicate that it has similar performance to the Fuzzy fractal method. While LSTM_BO performs better than and fuzzy fractal for the US, UK, Mexico, Italy, and Iran, fuzzy fractal achieves a lower SMAPE than LSTM_BO for the remaining five countries. Overall, the results indicate that ATT_BO and CNN_BO achieve better results compared to the fuzzy fractal model. The Mean SMAPE and Rank SMAPE values over the ten countries are given in Table 7. The Mean SMAPEs of the three deep learning models are significantly lower than the fuzzy fractal's one (Mean SMAPE=0.7052) (as seen in Table 7). Furthermore, the ATT_BO and CNN_BO models outperform the fuzzy fractal model in terms of Rank SMAPE.

Table 5 illustrates the results of all models in terms of MAPE. The best result for each country is denoted using the boldface. Deep learning models achieve the best results for 6 countries compared to the fuzzy fractal model that obtains the best results in 4 countries. In terms of MAPE, ATT_BO model outperforms the fuzzy fractal model in 6 countries. Compared to the fuzzy fractal model, CNN and LSTM archives better results in 5 and 4 countries, respectively. Also, in terms of Mean MAPE as seen in Table 7, all deep learning methods outperform the fuzzy fractal method. Besides, ATT_BO reaches the first Rank MAPE.

Looking at the results in terms of RMSE, as illustrated in Table 5 it is seen that the ATT_BO model performs better than the fuzzy fractal method in 6 cases out of 10 countries including US, UK, Mexico, Italy, Iran, and Belgium. Also, CNN_BO model has similar performance to the fuzzy fractal as both methods give better results in 5 countries. Furthermore, LSTM_BO reaches a lower RMSE in 4 countries compared to the fuzzy fractal. The Mean RMSE and Rank RMSE measures are provided in Table 7. We see that all deep learning models outperform the fuzzy fractal model in terms of Mean RMSE. Besides, the ATT_BO obtains the first Rank RMSE.

The overall results provided in Table 7 indicate that all proposed models perform significantly better in terms of Mean SMAPE, Mean MAPE, and Mean RMSE. The results demonstrate the performance of deep learning methods for COVID-19 forecasting. The better forecasting performance of the deep learning methods mainly attributed to their inherent characteristics in handling sequence data.

To illustrate the performance of methods, in Figs. 7–16, we also visualize the forecasted and actual cases for each country with the best models achieved from the deep learning models as well as the fuzzy fractal method. In all of the following figures, the black line indicates the real values, the green line corresponds to the forecasted cases using the best deep learning model, and the red line plot the forecasted cases with the fuzzy fractal.

Fig. 7 shows the forecast of confirmed cases for US, where the difference between the deep learning model (the green line) and

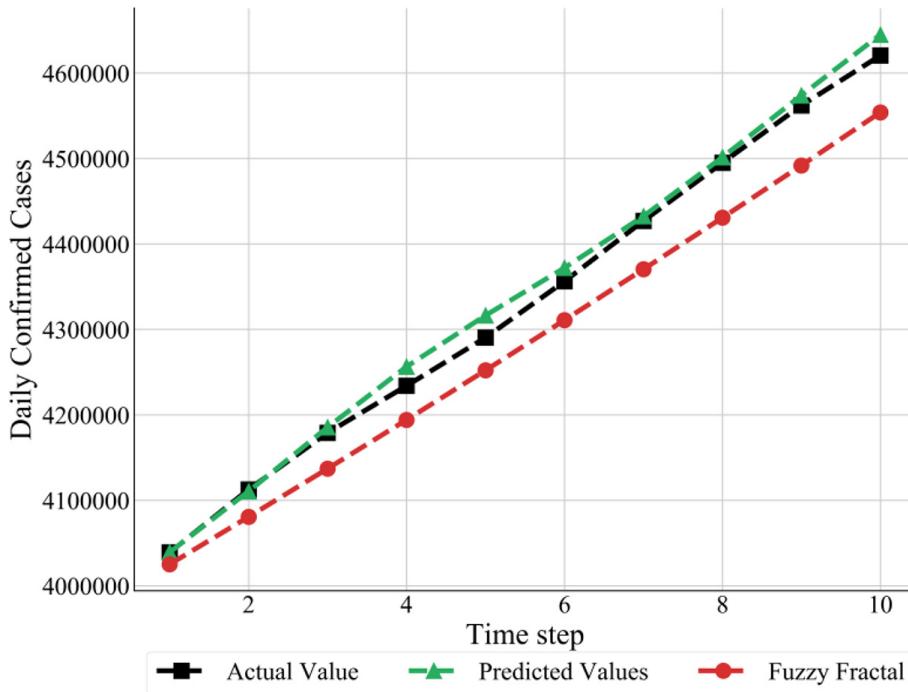


Fig. 7. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for US.

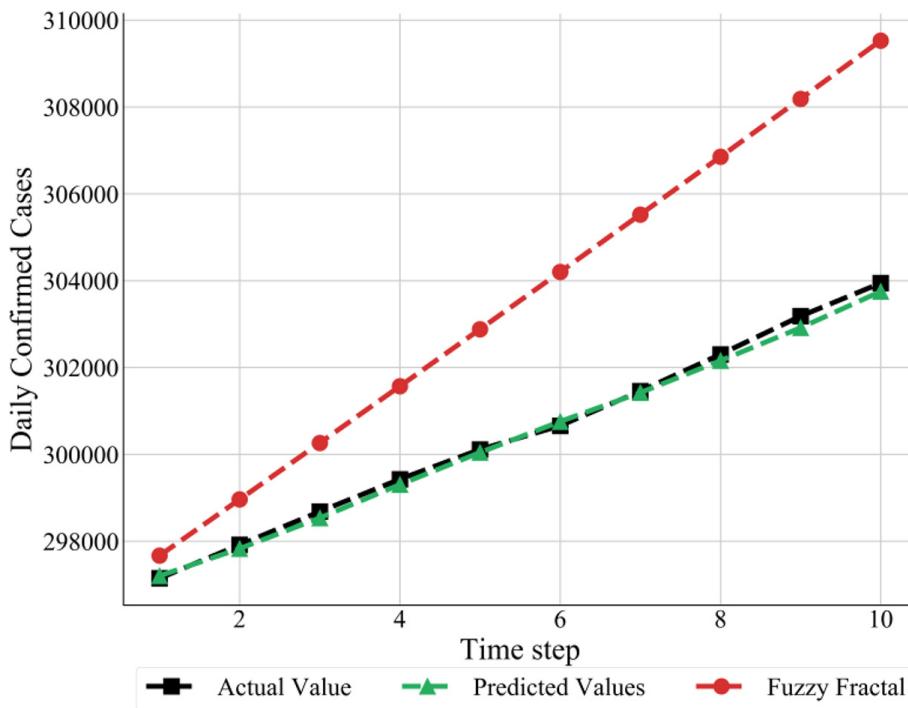


Fig. 8. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for UK.

the fuzzy fractal method (the red line) is clear. The forecasted cases with the deep learning model are very close to the real values. Fig. 8 shows the forecasted values for UK, where the difference between the deep learning model and the benchmarking model is apparent. Fig. 9 illustrates similarly the predicted values for Turkey, where the forecasted values using both the deep learning model and the benchmarking model are very close to the real values.

Fig. 10 plots the forecasted values for Spain, where the benchmark model slightly predicts better than the deep learning model. Figs. 11–13 show the predicted values for Mexico, Iran, and Italy respectively, where the forecasted values using the deep learning method are very close to the actual ones. The plots for Germany and France are illustrated in Figs. 14 and 15, respectively, which indicate the fuzzy fractal model predicted slightly better than the

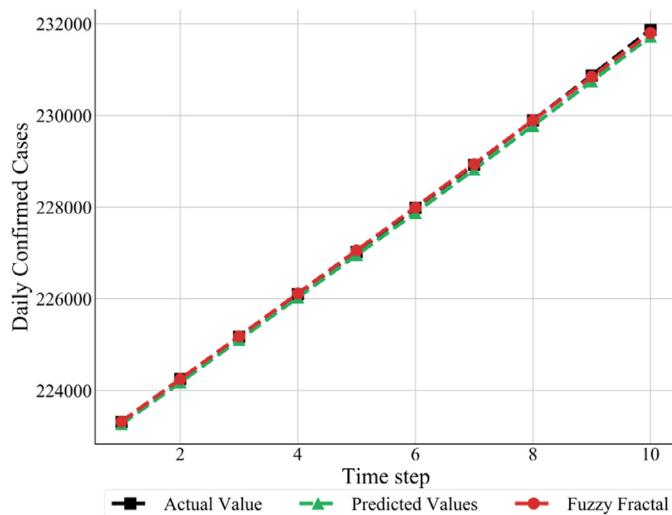


Fig. 9. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for Turkey.

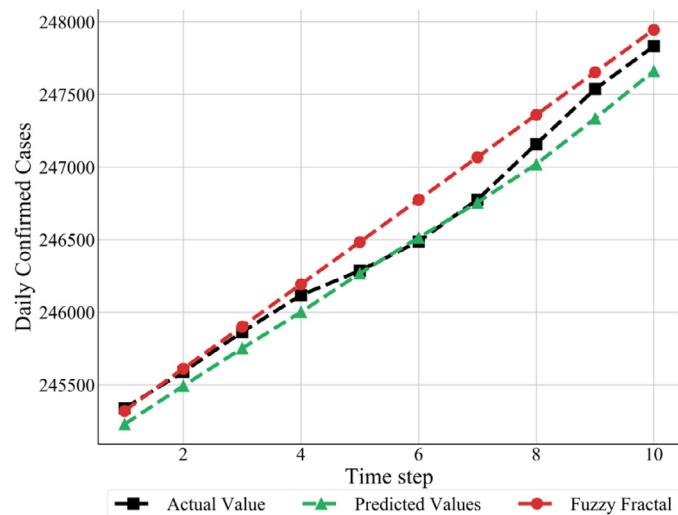


Fig. 12. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for Italy.

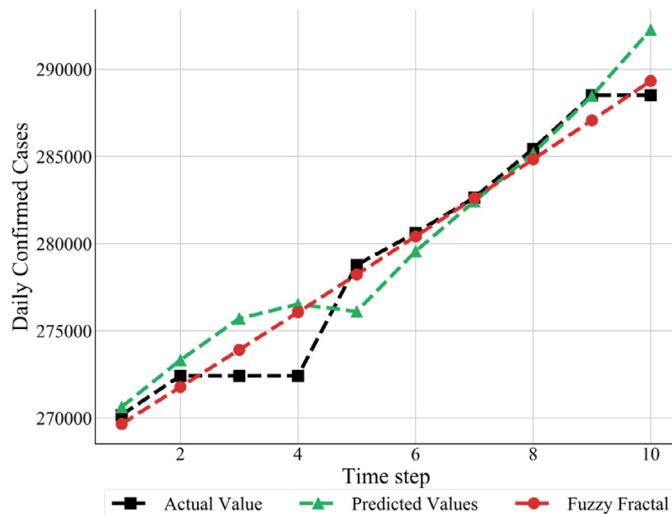


Fig. 10. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for Spain.

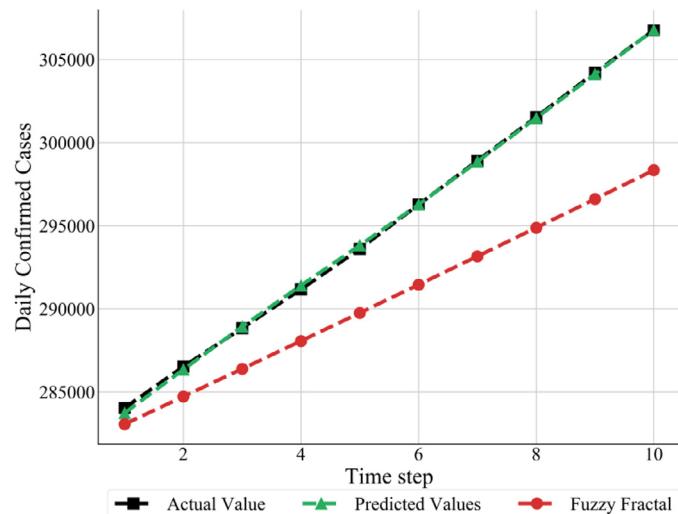


Fig. 13. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for Iran.

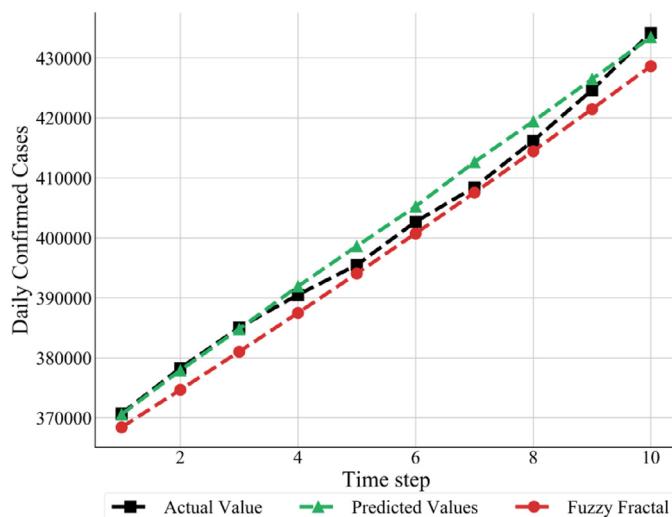


Fig. 11. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for Mexico.

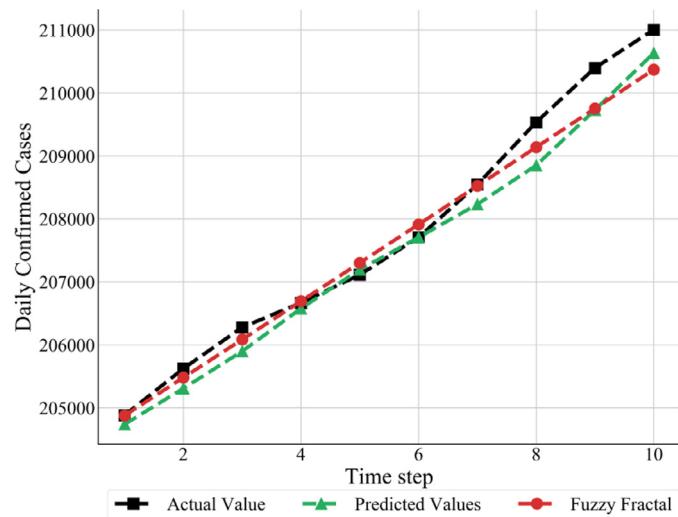


Fig. 14. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for Germany.

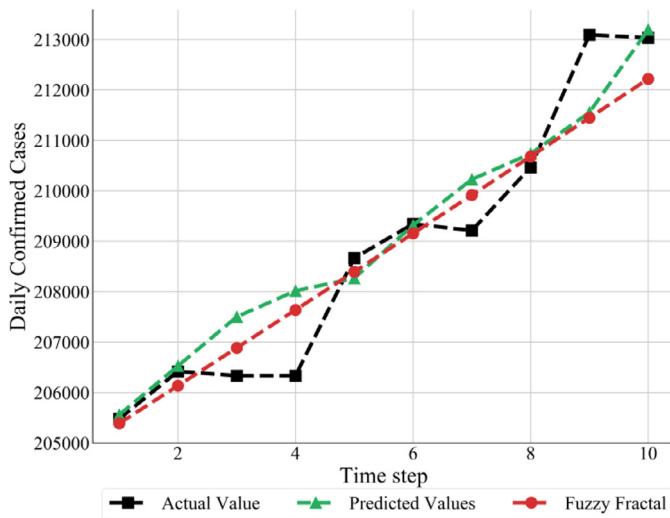


Fig. 15. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for France.

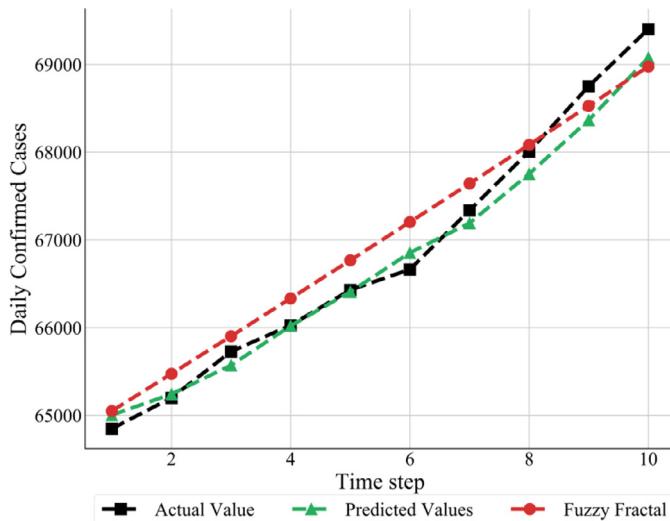


Fig. 16. The actual and predicted number of cases for 10 days (22 Jul to 1 August) for Belgium.

deep learning model. Fig. 16 illustrates the forecasted values for Belgium, where our proposed method predicts values as close as the actual value.

Analyzing the figures indicates that for the majority of countries, the best deep learning model achieves better performance than the fuzzy fractal model. For all countries, it is apparent that the fuzzy fractal model fits a linear model to predict the confirmed cases. Analyzing the figures indicates that for the majority of countries, the best deep learning model achieves better performance than the fuzzy fractal model. For all countries, it is apparent that the fuzzy fractal model fits a linear model to predict the confirmed cases. Also, as the figures display, the deep learning model was able to capture both linear and nonlinear patterns, which enhances its accuracy. The results confirm the suitability of the proposed model for COVID-19 time series forecasting.

4.5.2. Results of the second set of experiments on Dataset 2

After validating the effectiveness of the deep learning-based model on a shorter-window forecasting task, in this section, we perform the second set of experiments on Dataset 2 to examine

Table 8

The performance of the proposed methods in terms of SMAPE on Dataset 2 (The best results are marked bold).

Country	ATT_BO	LSTM_BO	CNN_BO
US	0.6914	0.8117	0.9946
Brazil	4.1811	3.4828	3.0081
India	0.8735	0.7711	1.1117
Russia	0.7723	0.4747	1.7461
South Africa	8.0334	8.1889	9.4018
Mexico	1.1996	1.1139	1.5866
Peru	4.3358	3.5406	3.5637
Chile	1.87	2.4096	3.2176
Colombia	4.8034	4.233	4.2347
Iran	1.0407	0.8953	1.4831

Table 9

The performance of the proposed methods in terms of MAPE on Dataset 2 (the best results are marked bold).

Country	ATT_BO	LSTM_BO	CNN_BO
US	0.6901	0.8105	0.9883
Brazil	4.2974	3.5924	3.0692
India	0.878	0.7748	1.08
Russia	0.7681	0.4732	1.7688
South Africa	8.6522	8.8127	10.2508
Mexico	1.1919	1.1055	1.5692
Peru	4.21	3.4325	3.4734
Chile	1.8376	2.3693	3.147
Colombia	4.9383	4.3337	4.3363
Iran	1.0485	0.9017	1.4977

the performance of the proposed models in longer-horizon forecasting. Longer-horizon forecasting reveals the trend of the pandemic in the long term and thus help governments to make appropriate decisions. To conduct experiments on Dataset 2, we adopt the hold-out method and split each COVID-19 time series into two parts: train set (80%) and test set (out-of-sample (20%)). The model building process is accomplished on the train set. The test set is used for evaluating the obtained models throughout the experiments. Also, for each time series, 20% of the train set is considered as the validation set data that is used in the hyperparameter identification process. As mentioned before, we adopt a multi-output forecasting strategy, so we set the output size=7. Therefore, the proposed model can forecast the number of cases for 7 next days.

The results of experiments in terms of SMAPE are provided in Table 8. For Dataset 2, ATT_BO achieves the best SMAPE for US, South Africa, and Chile. Also, LSTM_BO exhibits a significant performance and obtains the best SMAPE for 6 countries including India, Russia, Mexico, Peru, Colombia and Iran. CNN performs worse among these three methods and obtains the best performance only for Brazil.

Table 9 shows the results of experiments with respect to the MAPE measure. Similar to the results given in Table 8, LSTM_BO, ATT_BO, and CNN_BO achieve the best performance in 6, 3, and 1 countries, respectively.

The results of models in terms of RMSE are given in Table 10. We observe that regarding RMSE, LSTM_BO achieves the lowest RMSE in 5 cases. Also, the second-best performing method is the ATT_BO, which obtains the lowest RMSE in 3 countries. CNN_BO obtains the best forecasting only for Brazil.

To gain a more understanding of the overall performance of the proposed methods and their rank across all countries, we calculate Mean SMAPE, Mean MAPE, Mean RMSE, Rank SMAPE, Rank MAPE, Rank RMSE over all 10 countries data (as seen in Table 11).

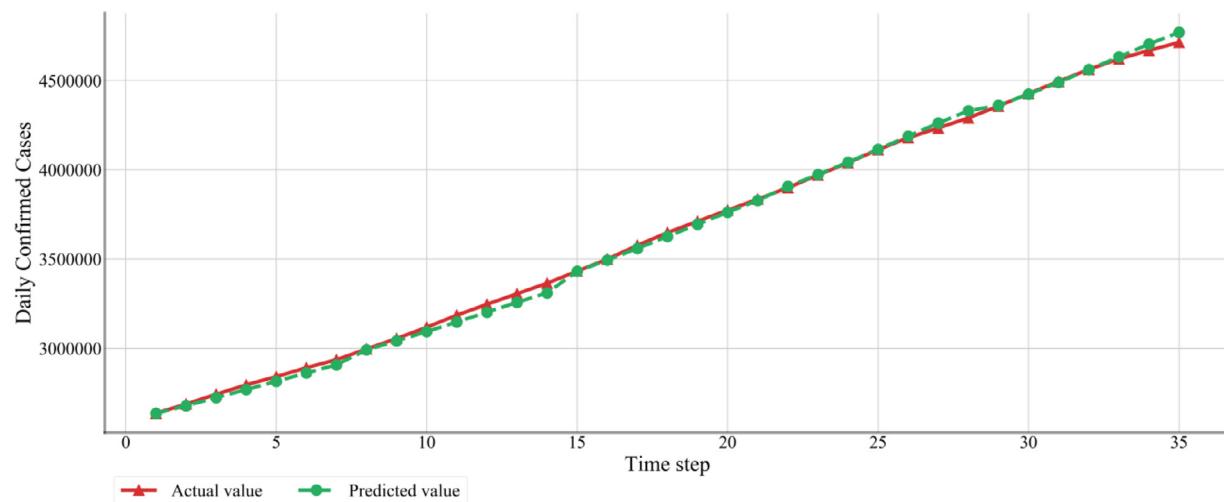


Fig. 17. The actual and predicted number of cases for test set-US.

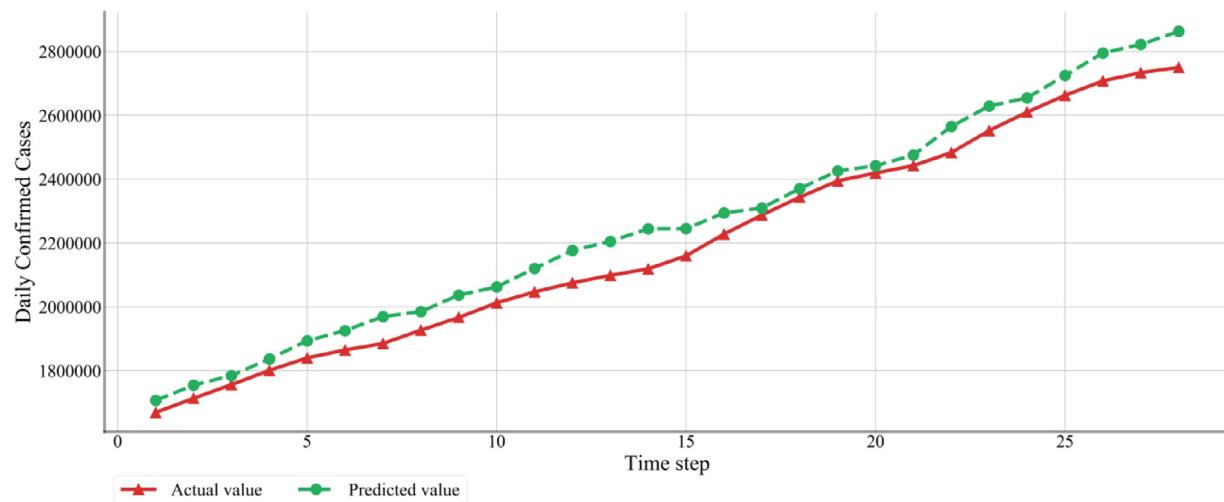


Fig. 18. The actual and predicted number of cases for test set-Brazil.

Table 10
The performance of the proposed methods in terms of RMSE on Dataset 2 (the best results are marked bold).

Country	ATT_BO	LSTM_BO	CNN_BO
US	31661.82	36223.39	43230.38
Brazil	110229.9	98053.65	75718.32
India	13834.01	13194	16290.89
Russia	7054.7	4275.18	15662.22
South Africa	55269.28	55611.48	65740.78
Mexico	5360.93	4935.58	7080.47
Peru	20003.09	17376.54	16049.06
Chile	7438.9	9388.29	13168.58
Colombia	12356.65	10721.77	10936.55
Iran	3279.91	3143.49	4802.84

The results demonstrate that the LSTM_BO method outperforms ATT_BO and CNN_BO in terms of all overall performance measures and is a suitable choice for a longer horizon forecasting task.

To further illustrate the forecasting power of the deep learning-based methods on dataset 2, in Figs. 17–26, we also visualize the actual and predicted cases for each country with the results of the best model obtained from the deep learning models. In Figs. 17–26, the red line indicates the actual values, and the green line corresponds to the forecasted cases using the best deep learning model. As Figs. 17, 19, 20, 22, and 26 show, the forecasted cases for countries including US, India, Russia, Mexico, and Iran are very close to the actual values. Besides, for these countries, in most of the time points, the forecasted values overlap the actual ones. The results confirm the power of deep learning models in COVID-19 time series forecasting. Moreover, for countries such as Brazil, South Africa, Peru, Chile, and Columbia as shown in Figs. 18, 21, 23, 24, and 25, respectively, the differences between the actual and predicted number of cases are not significant and at some points, the actual and predicted values are very close.

Table 11
The performance of all methods in terms of Mean SMAPE, Mean MAPE, Mean RMSE, Rank SMAPE, Rank MAPE, rank RMSE on Dataset 2 (the best results are marked bold).

Method	ATT_BO	LSTM_BO	CNN_BO
Mean SMAPE	2.7801	2.5922	3.0348
Mean MAPE	2.8512	2.6606	3.1181
Mean RMSE	26648.92	25292.337	26868.01
Rank SMAPE	2	1.4	2.6
Rank MAPE	2	1.4	2.6
Rank RMSE	2	1.5	2.5

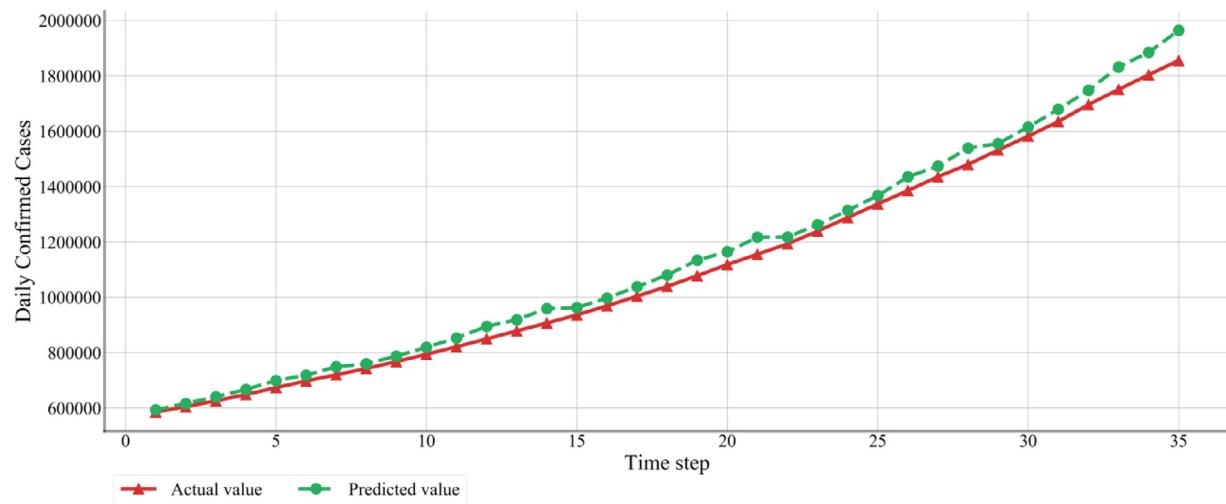


Fig. 19. The actual and predicted number of cases for test set-India.

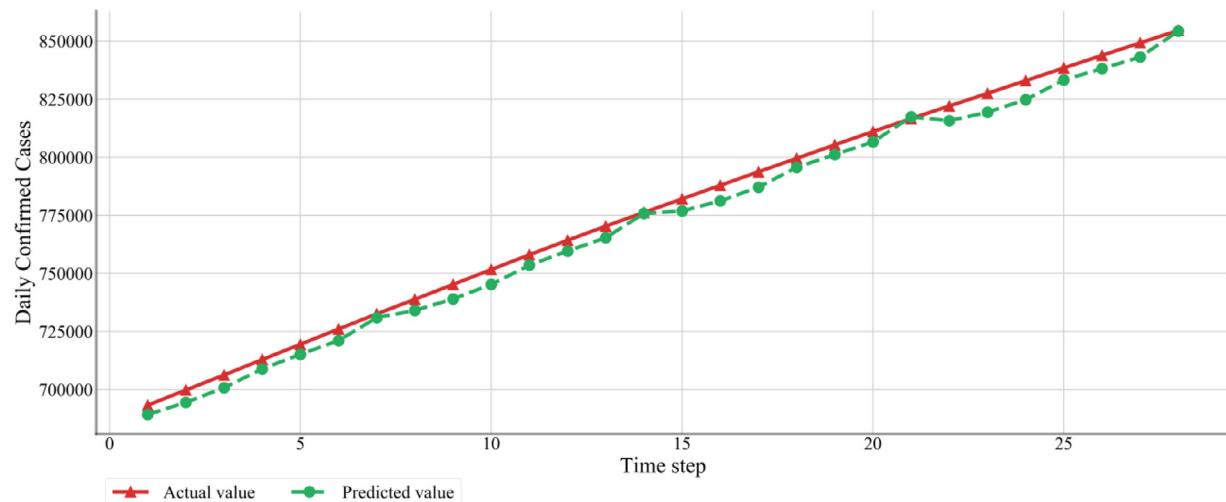


Fig. 20. The actual and predicted number of cases for test set-Russia.

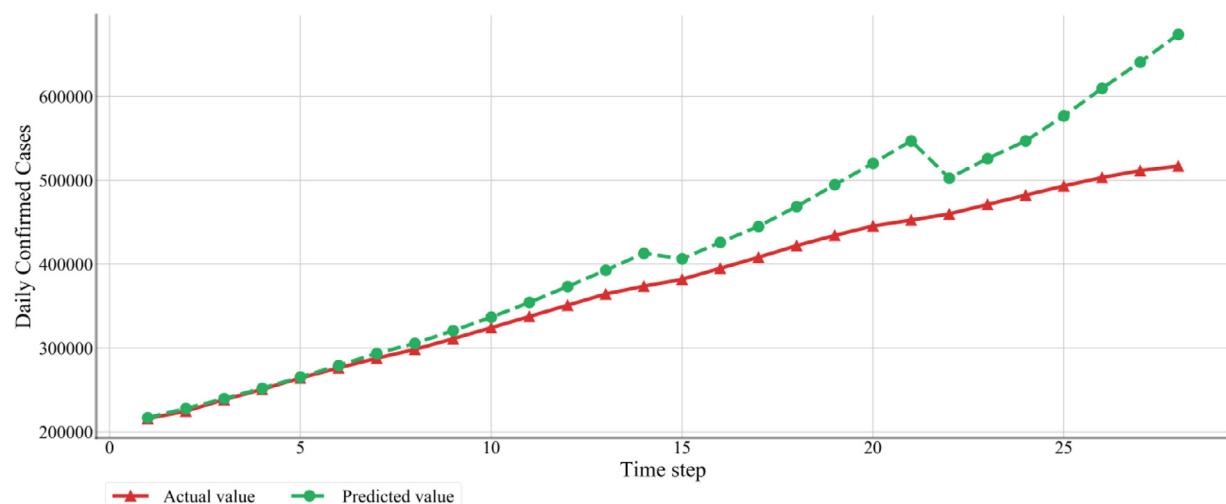


Fig. 21. The actual and predicted number of cases for test set- South Africa.

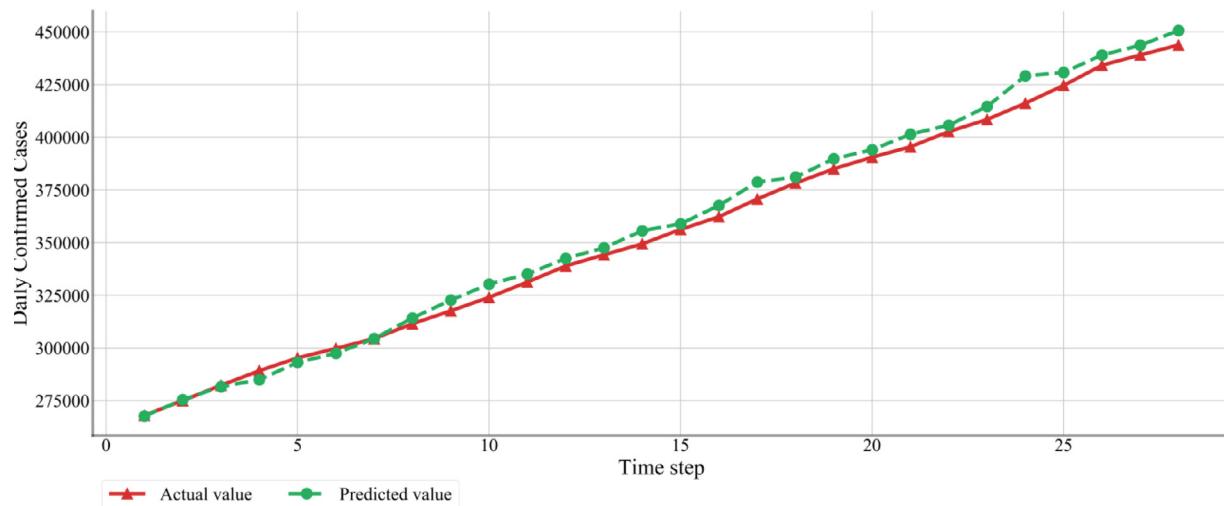


Fig. 22. The actual and predicted number of cases for test set-Mexico.

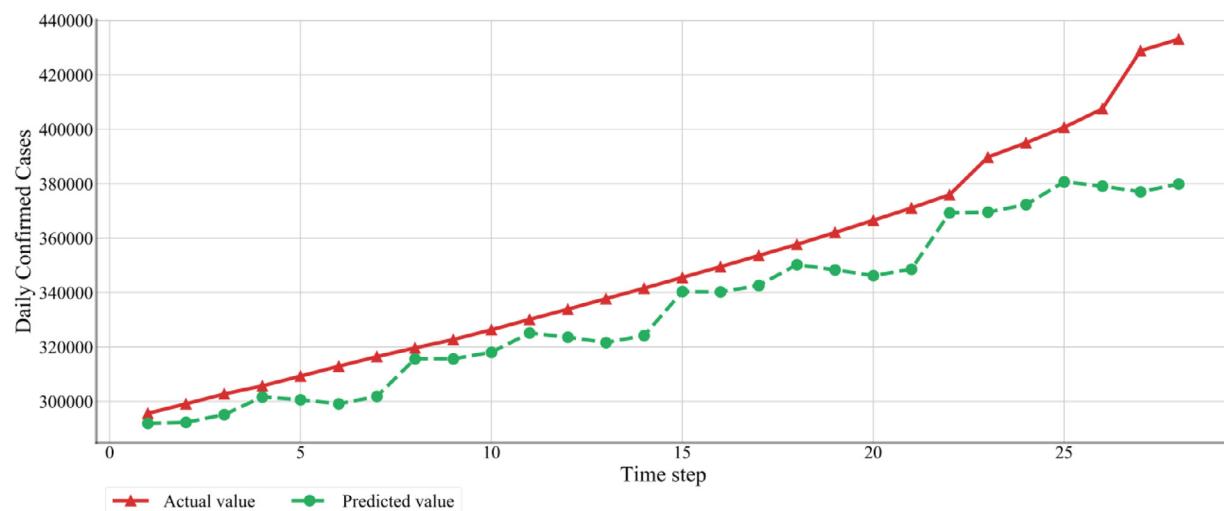


Fig. 23. The actual and predicted number of cases for test set-Peru.

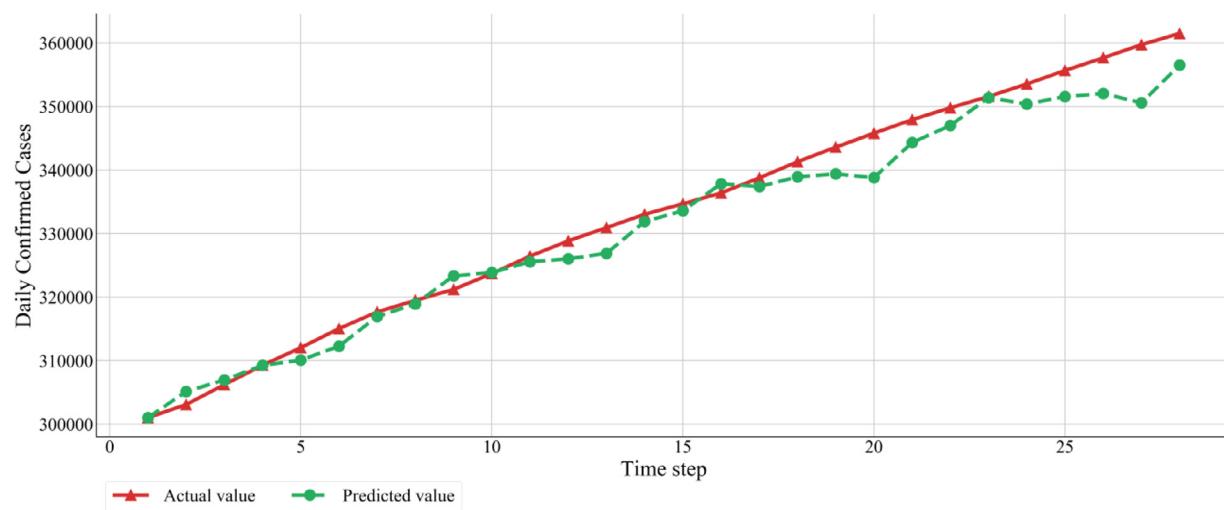


Fig. 24. The actual and predicted number of cases for test set-Chile.

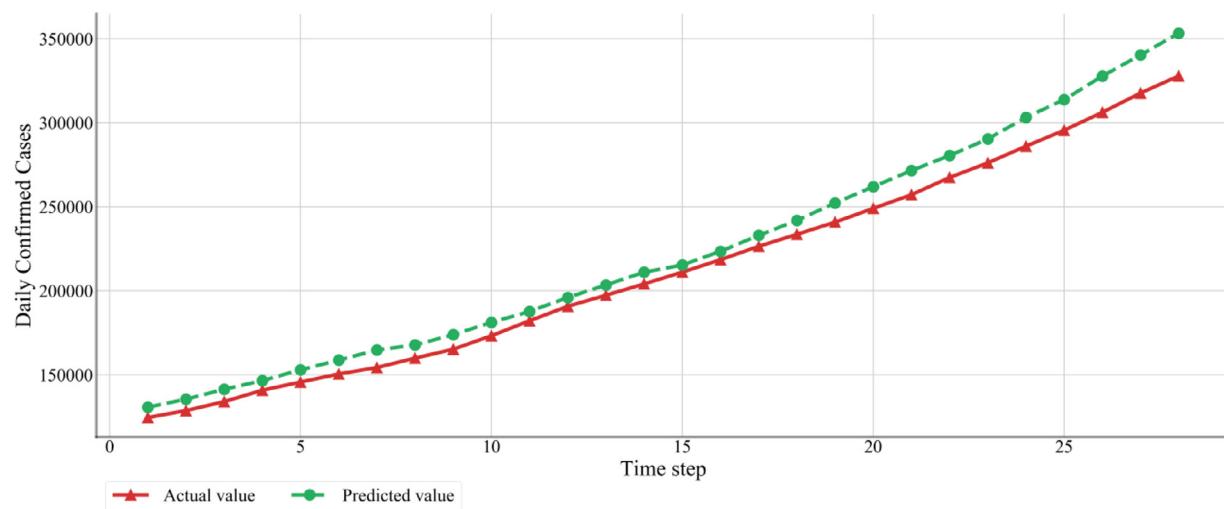


Fig. 25. The actual and predicted number of cases for test set-Colombia.

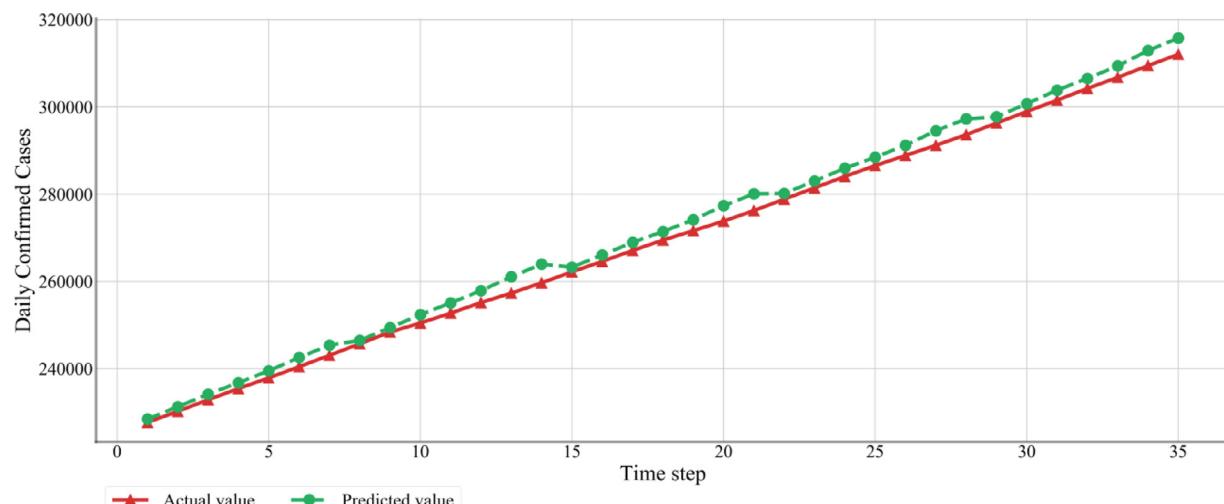


Fig. 26. The actual and predicted number of cases for test set-Iran.

5. Conclusion

In this study, three methods based on combining the deep learning models such as multi-head attention, CNN, and LSTM with the Bayesian optimization algorithm were developed to forecast COVID-19 time-series data. The main advantage of the proposed methods is their ability in processing the sequence data. Also, as another advantage, the design of the devised models is based on the multi-output forecasting strategy that allows forecasting multiple next days. The proposed methods were applied on the COVID-19 time series data considering two settings, the short-term forecasting, and the long horizon forecasting. For short-term forecasting, we adopted the fuzzy fractal method as the benchmarking model. the best deep learning model outperforms the fuzzy fractal model in 6 countries out of 10 countries. The significant result is that in terms of all overall measures such as Mean SMAPE, Rank SMAPE, Mean MAPE, Rank MAPE, Mean RMSE, and Rank RMSE, the three proposed methods perform significantly better than the benchmark model. Also, as the long-horizon forecasting is beneficial for long-term decision making on COVID-19 interventions, we explored the ability of the proposed methods on a longer horizon forecasting. The results of experiments indicated that among the three proposed models, the LSTM_BO achieves the best SMAPE in 6 countries. Besides, in terms of the performance measures computed across all countries, LSTM_BO outper-

formed ATT_BO and CNN_BO. Moreover, visualizing the actual and forecasted values demonstrated the effectiveness of the proposed methods in COVID-19 time series forecasting. As future work, we aim to extend the proposed methods by extracting the informative features from time series and incorporating them into the deep learning models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Hossein Abbasimehr: Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Reza Paki:** Software, Data curation, Visualization.

Reference

- [1] Gorbatenko AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020;5:536–44. doi:[10.1038/s41564-020-0695-z](https://doi.org/10.1038/s41564-020-0695-z).

- [2] İ Kirbaş, A Sözen, Tuncer AD, FŞ Kazancıoğlu. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos Solitons Fractals* 2020;138:110015. <https://doi.org/10.1016/j.chaos.2020.110015>.
- [3] Torrealba-Rodríguez O, Conde-Gutiérrez R, Hernández-Javier A. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. *Chaos Solitons Fractals* 2020;109946.
- [4] Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 2020;584:257–61. doi:[10.1038/s41586-020-2405-7](https://doi.org/10.1038/s41586-020-2405-7).
- [5] Critical Trends: Tracking critical data. <https://coronavirus.jhu.edu/data/>; 2020 [accessed 2020/06/01].
- [6] Leila M, Mozhgan S, Marziyeh Sadat S. exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. *Iran J Public Health* 2020;49. doi:[10.18502/ijph.v49i5.3675](https://doi.org/10.18502/ijph.v49i5.3675).
- [7] Al-Qaness MA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. *J Clin Med* 2020;9:674.
- [8] Castillo O, Melin P. Forecasting of COVID-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic. *Chaos Solitons Fractals* 2020;110242.
- [9] Melin P, Monica JC, Sanchez D, Castillo O. Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: the case of Mexico. *Healthcare (Basel, Switzerland)* 2020;8:181. doi:[10.3390/healthcare8020181](https://doi.org/10.3390/healthcare8020181).
- [10] Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos Solitons Fractals* 2020;139:110017. <https://doi.org/10.1016/j.chaos.2020.110017>.
- [11] Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* 2020;140:110212. <https://doi.org/10.1016/j.chaos.2020.110212>.
- [12] 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* 2020;140:110214. <https://doi.org/10.1016/j.chaos.2020.110214>.
- [13] Olah C. Understanding lstm networks, 2015. URL <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. 2015.
- [14] Li J, Tu Z, Yang B, Lyu MR, Zhang T. Multi-head attention with disagreement regularization. In: 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics; 2018. p. 2897–903.
- [15] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.
- [16] Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 2017;28:2222–32. doi:[10.1109/TNNLS.2016.2582924](https://doi.org/10.1109/TNNLS.2016.2582924).
- [17] Law T, Shawe-Taylor J. Practical Bayesian support vector regression for financial time series prediction and market condition change detection. *Quant Finance* 2017;17:1403–16.
- [18] Ben Taieb S, Sorjamaa A, Bontempi G. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing* 2010;73:1950–7. <https://doi.org/10.1016/j.neucom.2009.11.030>.
- [19] Jia L, Li K, Jiang Y, Guo X. Prediction and analysis of Coronavirus Disease 2019. arXiv preprint arXiv:200305447. 2020.
- [20] Castorina P, Iorio A, Lanteri D. Data analysis on Coronavirus spreading by macroscopic growth laws. *Int J Mod Phys C* 2020;2050103.
- [21] Ntakaris A, Mirone G, Kannainen J, Gabbouj M, Iosifidis A. Feature engineering for mid-price prediction with deep learning. *IEEE Access* 2019;7:82390–412.
- [22] Sangeetha K, Prabha D. Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. *J Ambient Intell Hum Comput* 2020;1–10.
- [23] Brochu E, Cora VM, De Freitas N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:10122599. 2010.
- [24] Cornejo-Bueno L, Garrido-Merchán EC, Hernández-Lobato D, Salcedo-Sanz S. Bayesian optimization of a hybrid system for robust ocean wave features prediction. *Neurocomputing* 2018;275:818–28. <https://doi.org/10.1016/j.neucom.2017.09.025>.
- [25] He F, Zhou J, Z-k Feng, Liu G, Yang Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Appl Energy* 2019;237:103–16. <https://doi.org/10.1016/j.apenergy.2019.01.055>.
- [26] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017:5998–6008.
- [27] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–80.
- [28] Abbasimehr H, Shabani M, Yousefi M. An optimized model using LSTM network for demand forecasting. *Comput Ind Eng* 2020;106435. <https://doi.org/10.1016/j.cie.2020.106435>.
- [29] Generating sequences with recurrent neural networks. <https://arxiv.org/>; 2013 [accessed January 10, 2020].
- [30] Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput* 2017;29:2352–449.
- [31] Novel Coronavirus (COVID-19) Cases Data. <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>; 2020
- [32] Keras. <https://github.com/fchollet/keras>; 2015 [accessed January 12, 2020].
- [33] Prechelt L. Early stopping – but when?. In: Montavon G, Orr GB, Müller K-R, editors. *Neural Networks: Tricks of the Trade*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 53–67.