DOI: https://doi.org/10.52756/ijerr.2024.v45spl.009 *Int. J. Exp. Res. Rev.*, Vol. 45: 106-118 (2024)

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)
© Copyright by International Academic Publishing House (IAPH)
ISSN: 2455-4855 (Online)
www.iaph.in

An Improved Parallel Heterogeneous Long Short-Term Memory Model with Bayesian Optimization for Time Series Prediction

Tirupati G[1]*, Krishna Prasad MHM[2] and Srinivasa Rao P[3]

[1]Department of Information Technology, GVP College of Engineering for Women(A), India; [2]Department of Computer Science & Engineering, College of Engineering (A), JNTUK, India; [3]Department of Computer Science & Engineering, MVGR College of Engineering (A), India

E-mail/Orcid Id:

GTR, gtr@gvpcew.ac.in, https://orcid.org/0000-0002-9160-7259; *MHM,* krishnaprasad.mhm@gmail.com, https://orcid.org/0000-0003-2150-7807; *PSR,* psr.sri@gmail.com, https://orcid.org/0000-0003-4851-744X

Article History:

Received: 15th Jun., 2024
Accepted: 27th Oct., 2024
Published: 30th Nov., 2024

Keywords:

Deep Learning, Hyperparameter, Optimization, Parallel LSTM, Prediction

How to cite this Article:
Tirupati G, Krishna Prasad MHM and Srinivasa Rao P. (2024). An Improved Parallel Heterogeneous Long Short-Term Memory Model with Bayesian Optimization for Time Series Prediction. *International Journal of Experimental Research and Review*, *45*, 106-118.
DOI:https://doi.org/10.52756/ijerr.2024.v45spl.009

Abstract: Currently, Deep Learning (DL) with the Recurrent Neural Networks (RNN) variants is being applied successfully in many domains of Engineering for prediction. In view of the demand for precise forecasting and the aid of Artificial Intelligence Tools, time series prediction reveals a vital task in decision-making and risk assessment. However, the application of novel Recurrent DL models for obtaining an accurate prediction of time series is yet to be explored. Recent trends reveal that Hybrid Neural Networks and DL models are appropriate for time series forecasts. At the same time, the model's selection and the hyperparameter's tuning can greatly impact its performance. To address this problem, a parallel long-term memory (PLSTM) model integrated with Bayesian hyperparameter optimization (PLSTM-BO) is proposed for time series prediction. The model is tuned in terms of key parameters, including the number of neurons, dropout, learning rate, and optimization technique. The model's performance is assessed using the SARS-COVID-19 cumulative cases, deaths, recovery cases, and NIFTY 50 stock closing price time series dataset. The obtained results convey that the current model exhibits remarkable performance compared to existing models.

Introduction

Time series prediction has a wide scope and challenges in research for so many years, with significant applications including disease spread prediction (Abbasimehr and Paki, 2020; Chandra et al., 2022), weather conditions (Haque et al., 2021; Dubey et al., 2024), air pollutants (Drewil and Al-Bahadili, 2022), stock price prediction (Kumar et al., 2022; Zhou et al., 2022) and many more. Ample research work has been committed in recent decades to improve and develop ML Models for prediction. The era of Machine Learning (ML) and Artificial intelligence (AI) has promoted cutting-edge technologies with low prediction errors in all domains. With the help of prediction models, an individual or Government can make decisions and forecast future needs. The objective of prediction is to gather past data that can be used to build a model that describes the relationship between the variables. In the last decade, researchers have proposed many statistical (Malki et al., 2020), conventional machine learning (Wu et al., 2019; Zivkovic et al., 2021), and hybrid models for forecasting life-threatening diseases such as Zika, hepatitis, chicken pox, COVID-19, and more. The SARS CoV-2 Virus is spread by airborne transmission and has a lengthy incubation span related to other Diseases. The respiratory system and other organs (Di Gennaro et al., 2020) are severely affected by COVID-19, which might cause death.

Time series models such as Auto Regressive Intergraded Moving Average (ARIMA) and Seasonal Auto Regressive Intergraded moving Average (SARIMA) are well-liked due to their prediction using Box-Jenkins in the model erection process. The presumptive linear form of these models will not be able

***Corresponding Author:** gtr@gvpcew.ac.in

to catch the nonlinear sequences. The estimation of linear models is not always adequate for complex problems. However, the challenging task is to develop an accurate and reliable unique model for all types of time series prediction and a set of hyperparameters for the respective model. Hyper parameter optimization is done in two ways: manual tuning and automatic tuning. Manual tuning (Tirupati et al., 2021) depends on the expertise of the developer who develops the model and requires domain knowledge and experience. To avoid the problem of manual search, several authors have proposed automatic search algorithms such as grid, random, and many more. Grid search uses a set of possible combinations of values for hyperparameters and has the limitation of the curse of dimensionality, i.e., performance declines as search space expands. On the other hand, in high dimensional space, random search is more effective than grid search. However, hyperparameter tuning using a random approach is uncertain for learning composite models. Random and Grid hyperparameter optimization is fully unaware of past estimations.

The work has been structured as follows: Part 2 designates the current work associated with the forecasting domain. In Part 3, we describe the material & methods used to obtain prediction, followed by the results in Part 4 and the conclusions of the study presented in Part 5. Finally, in part 6, some future research directions have been put forward.

Related work

Recent studies highlight the significance of Recurrent Neural Networks (RNN) in time sequence prediction due to their flexible nonlinear modeling capability and their ability to memorize past data. Recurrent Neural Networks (RNNs) are dynamic models that effectively utilize temporal information from the input time series, both for prediction (Bergstra et al., 2011; Muhammed, 2023), and classification (Huang et al., 2022). Disease control and prevention are of the at most utmost importance, a classical LSTM and Convolutional Neural Networks (CNN) based on Bayesian optimization (Abbasimehr & Paki, 2020) methodology is proposed for predicting the COVID-19 Daily cumulative cases in India, the USA and other countries. The results indicate that LSTM with BO shows low MAPE in the majority of COVID-19 cumulative Confirmation Cases. A large amount of research on diverse sets of time-series data has been conducted (Bergstra et al., 2011) using parallelized LSTM (PCLSTM) on the weather time series of Brazilian Cities. The Analysis has shown that PCLSTM achieves low MAPE and RMSE. The benefit of hyperparameter optimization has been pointed out in (Kumar et al., 2022), study demonstrates training the neural networks with global optimization approaches including Tree-structured Parzen Estimator (TPE), Gaussian Process (GP), and Random Search. The Analysis shows that the TPE approach gives better results than other hyperparameter approaches. Recent work (Wu et al., 2019) investigates the automatic tuning of Neural Network Model in which ML Model based on Bayesian optimization achieves high accuracy in comparison with grid and manual hyperparameter search. The application of model tuning has also been utilized in (Li et al., 2022), where BGRU with the Sparrow Search Algorithm (SSA) is applied to oil field production forecasting. It was demonstrated that BGRU outperforms other existing classical and statistical ML Models. Authors studied SVM, ARIMA, and variant RNN networks on COVID-19 dataset of the USA, UK and many more countries (Shahid et al., 2020). The results in terms of MAPE, MAE and RMSE showed that BLSTM outperformed the other models. The Benefit of the Deep learning framework has been pointed out by Chandra et al. (2022), where a study finds Encoder-Decoder LSTM produces significant performance in comparison to other RNN models on Indian COVID-19 cumulative confirmation cases. COVID-19 spread control methodology (Ardabili et al., 2020) has been demonstrated where Multi-Layer Perceptron (MLP) and Adaptive Fuzzy systems show better prediction outbreaks in various countries.

Many studies have provoked the automatic tuning of NN models. The study demonstrates that Cauchy-Exploration strategy BAS(CESBAS) & ANFIS achieve better COVID-19 time series prediction than other optimizations such as GA and PSO (Zivkovic et al., 2021). The benefit of ensemble RNN models has been applied (Haque et al., 2021) in Beijing Temperature prediction. It was demonstrated that the CNN-LSTM parallel Network obtains the lowest RMSE. In (Tirupati et al., 2021), the authors demonstrate the prediction on the daily COVID-19 cumulative confirmation, deaths, and recovery of the India dataset in which BGRU outperforms other conventional Techniques. Recent studies (Drewil and Al-Bahadili, 2022) show that population-based heuristic hyperparameter Algorithm and RNN variant models are useful in pollution prediction. The analysis shows that the Genetic Algorithm (GA)-LSTM achieves lower RMSE and MAE than BLSTM and CLSTM. Hyperparameter Optimization with variant RNN models has been demonstrated for stock price prediction (Kumar et al., 2022). The results indicate that

Adaptive particle swarm optimization (APSO) with LSTM produces better prediction efficiency than GA, particle swarm optimization (PSO) and classical LSTM. The part of the literature related to our study is presented in Table 1.

Table 1. Summary of other relevant time series prediction studies.

Reference	Model	Hyperparameter optimization	Description
Moftakhar et al., 2020	ANN, ARIMA	-	Predicting Covid-19 cases in Iran
Al-Qaness et al., 2020	Adaptive Neuro-Fuzzy Inference System (ANFIS)	-	Forecasting COVID-19 Cases in China
Arora, et al., 2020	BLSTM, Stacked LSTM and CNN-LSTM	-	Analysis of COVID-19 for the 32 Indian states.
Wang et al., 2020	LSTM	-	Estimating COVID-19 cases in Russia, Peru and Iran
Kırbaş et al., 2020	ARIMA, NARNN, and LSTM	-	Covid-19 cases predicting in Denmark, Belgium, Germany, France, United Kingdom, Finland, Switzerland and Turkey.
Gorgolis et al., 2019	LSTM-GA	Yes	Tuning for language model.
Melin et al.,2020	Ensemble NN - Fuzzy	-	Predicting the COVID-19 cases in Mexico
Bemani et al., 2020	ANFIS with Particle Swarm Optimization (PSO), Back Propagation (BP), Genetic Algorithm (GA), Differential Evolution (DE), Ant Colony Optimization (ACO)	Yes	Estimating the carbon dioxide diffusivity ahead of time.
Rashid et al., 2018	LSTM-Grey Wolf Optimizer (GWO), Simulated Annealing (SA), Ant Lion Optimization Algorithm(ALOA)	Yes	Breast Cancer Wisconsin and Epileptic Seizure Recognition

Materials and Methods

This section explores the datasets used for proposed model building and methods, including preprocessing, methodology and Bayesian hyperparameter optimization.

Dataset

In this study, two types of time series datasets are considered for evaluating the model. One is the SARS COVID-19 Daily Cumulative Confirmation Cases, Cumulative Deaths, Cumulative Recovery cases in India (Rustam et al., 2020; Tirupati et al., 2021) and Cumulative Confirmation Cases in India & USA (Abbasimehr and Paki, 2020). The other is the NIFTY 50 Stock closing price dataset (Kumar et al., 2022). The datasets are taken from the Oxford Martin Programme on Global Development - Oxford Martin School (Global Development, 2021) and NIFTY 50 Historical Data - Yahoo Finance (Yahoo Finance, 2024) for COVID-19 and stock price, respectively. Dataset1 includes Indian COVID-19 daily cumulative confirmation illness from 30th January 2020 to 11th August 2021; Dataset2 contains Indian COVID-19 daily cumulative deaths from 30th January 2020, to 11th August 2021; Dataset3 has Indian COVID-19 daily recovery cases from 30th January 2020 to 11th August 2021; Dataset4, India coronavirus cumulative cases from 30th January 2020, to 3rd August 2020; Dataset5, USA coronavirus cumulative cases from 30th January 2020, to 3rd August 2020, USA; and Dataset6, NIFTY 50 stock closing price from 1st January, 2015 to 31st March, 2021.

Preprocessing

The raw COVID-19 dataset is filtered so that it selects required attributes such as country, date, and cumulative cases, deaths, recovery cases and then Min-Max normalization, as shown in equation 1, has been applied to the dataset. The cumulative values of the dataset are in the range of 0 to 1. Further, the dataset is divided into training and testing the proposed model at 80% and 20%, respectively.

$$y' = \frac{y - \min_x}{\max_x - \min_x}(\text{new\_max} - \text{new\_min}) + \text{new\_min} \qquad (1)$$

y'= Normalized value of y
y=Observed value of x
$\min_x$ = Minimal of x
$\max_x$ =Maximal of x
new_max = Maximal of Normalized data.
new_ min = Minimal of Normalized data.

Finding the Lag value

The estimation of time lag is significant because the number of previous time series values determines the current value of a time series. It can be determined by the popular autoregressive statistical technique, in which the current value is a function of previous values. The order (lags) of autoregressive is found with the help of the Partial Autocorrelation Function (PACF) of stationary time series (Wang et al., 2020; Yakubu et al., 2022). The lag value of all datasets is determined in a two-stage process: the first stage makes the time series stationary by applying differentiation, and then the respective PACF plot number of lags outside the 95% indicates the order of the autoregressive process. Figure 1 shows the PACF plot of dataset1, in which lag 0 is strongly correlated and the value is always 1. PACF at lag 1 is significant and is followed by a drop in coefficient. It indicates the lag value is 1 for the cumulative confirmation cases in India (dataset1). Similarly, the lag values for other datasets are 1, 2, 3, 3, and 1 for dataset2, dataset3, dataset4, dataset5, and dataset6, respectively.

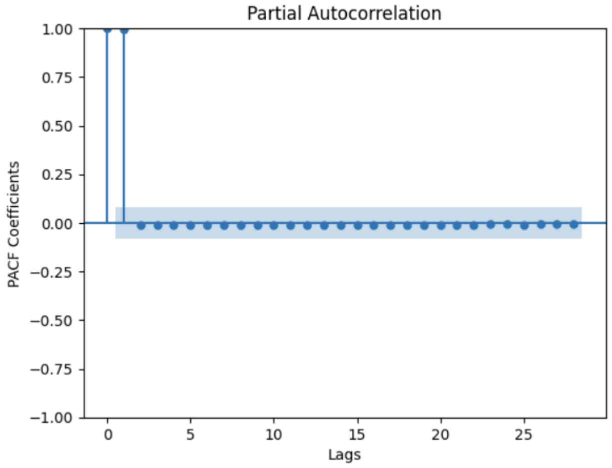


Figure 1. PACF plot of cumulative coronavirus cases (dataset1) time series in India.

Methodology

RNN variants are being used successfully in many domains of time series dataset prediction. LSTM networks can overcome the limitations of vanilla RNN such as exploding and diminishing gradient problems for long time series. However, Vanilla LSTM and bidirectional networks are limited in their performance due to the complex relationships they extract and the large forecasting window size. Moreover, Bidirectional networks have two similar networks that are in parallel; one network takes the input sequentially, and the other in an anti-chronological direction. As a result, bidirectional networks may not retain the complex relationship in some of the datasets. Therefore, recent studies show that developing hybrid RNN variant Models can produce low prediction errors and high efficiency even in disparate domains and datasets. The proposed model consists of two different LSTM networks arranged in a parallel fashion, as shown in Figure 2. Each LSTM can uniquely retain the relationship, and then regularization is applied to each LSTM Network to avoid the model overfitting by implementing the dropout approach. Finally, individual LSTM outputs are concatenated, and a dense network is used to predict the output. The structure of the LSTM Cell (Hochreiter and Schmidhuber, 1997; Graves, 2014) is shown in Figure 3, and the internal operations of the input, forget, and output gates of the LSTM are performed according to the equations 2, 3, 4, 5, 6, and 7.

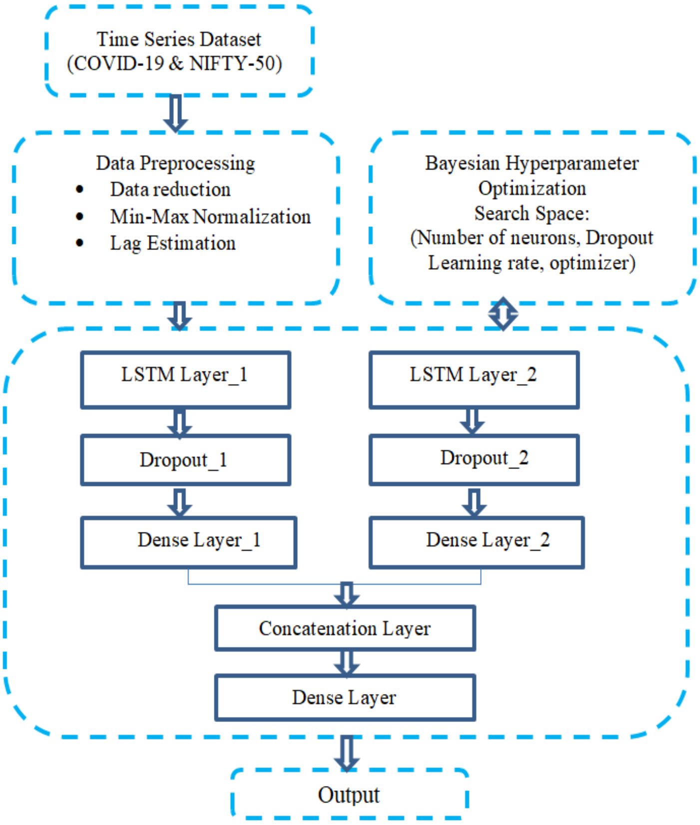


Figure 2. Architecture of the proposed model.

$$f_t = \sigma\left(W_f x_t + W_F h_{t-1} + b_f\right) \qquad (2)$$

$$i_t = \sigma(W_i x_t + W_i h_{t-1} + b_i) \qquad (3)$$

$$\tilde{C}_t = tanh(W_c x_t + W_c h_{t-1} + b_c) \qquad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \qquad (5)$$

$$O_t = \sigma(W_O x_t + W_O h_{t-1} + b_O) \qquad (6)$$

$$h_t = O_t * \tanh(C_t) \qquad (7)$$

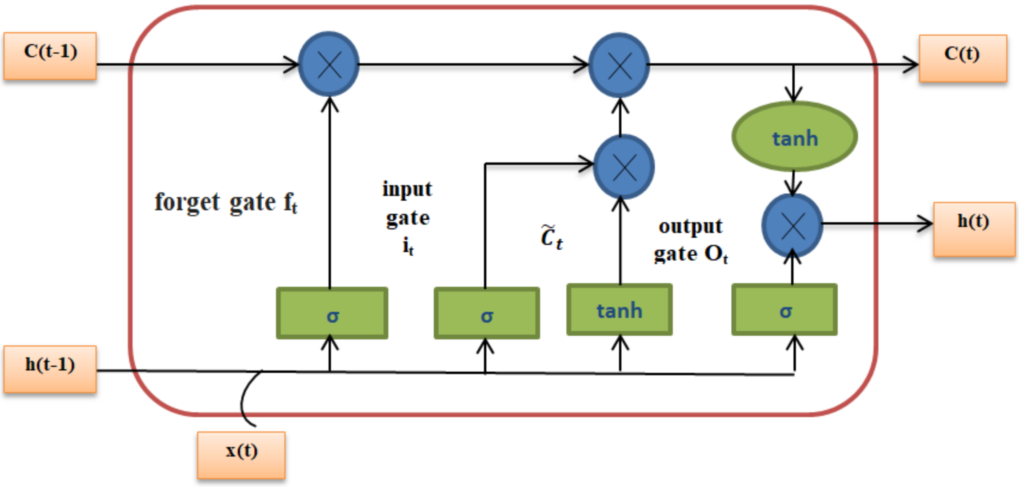
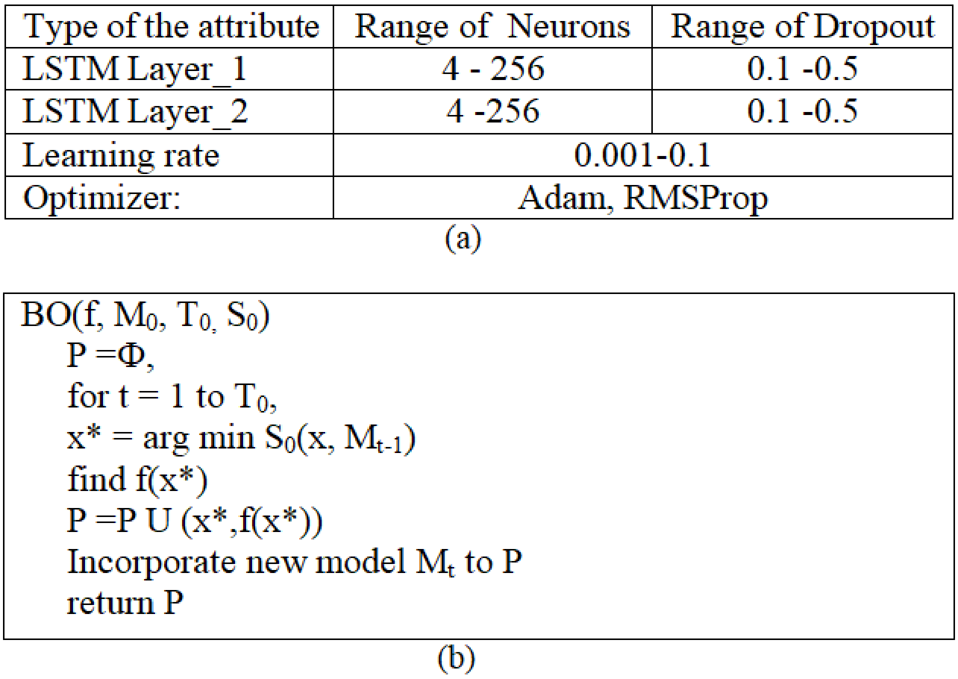


Figure 3. Structure of LSTM Cell.



Type of the attribute	Range of Neurons	Range of Dropout
LSTM Layer_1	4 - 256	0.1 -0.5
LSTM Layer_2	4 -256	0.1 -0.5
Learning rate	0.001-0.1	
Optimizer:	Adam, RMSProp	

(a)

```
BO(f, $M_0$, $T_0$, $S_0$)
    P =Φ,
    for t = 1 to $T_0$,
    x* = arg min $S_0$(x, $M_{t-1}$)
    find f(x*)
    P =P U (x*,f(x*))
    Incorporate new model $M_t$ to P
    return P
```

(b)

Figure 4. (a). Search space of proposed model (b) BO Algorithm.

Bayesian Hyperparameter optimization

It is a probabilistic method that selects current model parameters based on past evaluation.BO is an iterative procedure that is usually used to detect near-optimal hyperparameters combinations in just a few iterations. The limitations of both Grid search and Random search are that each evaluation is independent of the previous evaluation. Thus, they spent more time evaluating inadequately performing regions of search space. Bayesian Hyperparameter optimization is more effective than grid search, random search and even manual optimization by domain experts. In Genetic Algorithm, randomly initialized values for genetic operations like crossover, selection, and mutation often do not find the optimal parameters. The main limitation of particle swam optimization is that it requires correct population initialization; otherwise, it may get local optimum instead of global. The proposed model is tuned using Automatic Bayesian optimization (Hyperopt, 2019; Yang and Shami, 2020), in contrast to other hyperparameter optimizations (Bergstra et al., 2011; Muhammed, 2023). The present model is tuned with respect to search space, as shown in Figure 4(a). The other specifications of the model are epochs, batch size, activation, and recurrent activation functions of 200, 32, tanh, and sigmoid, respectively. Hyperparameter optimization is represented by as follows:

$$x^* = \underset{x \epsilon X}{\arg\min} f(x) \qquad (8)$$

where f(x) denotes a goal of minimizing an objective score, such as RMSE or MSE calculated on the validation set. The collection of hyperparameters called x* can have any value in the domain X and produces the score with the lowest value. The Bayesian Optimization (BO) algorithm, as shown in Figure 4(b), differs in how the model f via observation history P and in the criteria they optimize to acquire x* given a model (surrogate) of f. The next set of hyperparameters from the surrogate function is chosen using the Expected Improvement (EI) as illustrated in eq. 9.

$$EI_{y*}\,(x) = \int_{-\infty}^{\infty} \max(y^{*} - y, 0)\, p(y|x)\, dy \qquad (9)$$

y* is the objective function's threshold value, x is the suggested set of parameters, y is the real value of the objective function, and p(y | x) is the surrogate model probability describing the probability of y given x.

Advantages

The superiority of the current model hyperparameter optimization is given below when compared to other techniques.

a) Model hyperparameters are selected based on past evaluation.

b) It gives optimal solutions within a few evaluations than Genetic and Grid Search.

Evaluation criteria

The proposed model performance is quantified in terms of Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), R-squared (R^2) Score, Standard Deviation of Residuals (S_{res}) and Root Mean Square Error (RMSE).

a. Mean Absolute Percentage Error (MAPE): MAPE more accurately captures the prediction error (Willmott & Matsuura, 2005). It computes the difference between the actual data and the model's forecast based on test data, then divides that difference by the actual data. A lower MAPE reflects a model prediction that is reasonably close to the observed value. MAPE is characterized as follows;

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100 \qquad (10)$$

y_i, $\hat{y}_i$ and n is the actual value, prediction values and size, respectively.

b. Mean Absolute Error (MAE): Another metric is used to assess the regression model, along with MAPE. Sometimes MAE is very different depending on whether if it is predicted or the actual value that is largest. The best score is zero and worst is infinity. MAE is calculated as follows

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (11)$$

c. R-Squared ($R^{2)}$ Score: It is known as the coefficient of finding and conveys how well data values fit the curve. The R^2 score illustrates how the data values are dispersed over the regression curve. Typically, it lies between 0 and 1. A high R^2 score suggests that the prediction model is reliable. Using eq.(12), the R2 score can be calculated as follows:

$$R^2 score = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right] \qquad (12)$$

$\bar{y}$ is the mean of real data.

d. Standard Deviation of Residuals (Sres): Standard deviation of residuals is the difference between standard deviation of predicted and actual data. The S_{res} shows the spread around the regression line. The small S_{res} indicates a model is good at predicting the output. The S_{res} can be calculated as given in equation (13).

$$S_{res} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} \qquad (13)$$

e. Root Mean Square Error (RMSE): It is referred to as the residual, and it calculates prediction error based on the separation between best-fit values and actual data. The formula for RMSE is given eq.(14).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (14)$$

Results and Discussion

The proposed model PLSTM-BO, along with other models such as LSTM, BLSTM, and BGRU, is used on coronavirus Cumulative Confirmation (CC) cases in India (dataset1), Cumulative Deaths (CD) in India (dataset2), cumulative Recovery Cases (RC) in India (dataset3) (Rustam et al., 2020; Tirupati et al., 2021), cumulative confirmation cases in India (dataset4), cumulative confirmation cases in USA (dataset5) (Abbasimehr and Paki, 2020), and NIFTY 50 stock closing price time series (dataset6)(Kumar et al., 2022). For the coronavirus cumulative confirmation cases in India (dataset1), PLSTM with BO model uses search space and other parameters mentioned in the previous section. The proposed model is tuned with train data within 10 evaluations, and then the best set of configuration parameters is chosen for the subsequent prediction. The proposed model produces MAPE, MAE, RMSE, S_{res}, and

R^2 score values of 0.1032, 0.0009, 32501, 0.0010 and 0.9999, respectively. Table 2 shows the performance metrics of the present and existing models on the cumulative confirmation (CC) cases test dataset. It shows that the present model is superior to other existing models and also resembles BLSTM, which performs good in comparison with conventional RNN variants. The obtained results of the current model indicate a significant development in prediction as compared to the DL model with manual tuning (Tirupati et al., 2021). Figure 5 exhibits MAPE bar charts of proposed and other models on dataset1. The proposed and existing models prediction is shown in figure 6(a).

Table 2. Performance metrics of PLSTM-BO and existing models on test data.

Type of dataset	Model/Metrics	MAPE(%)	MAE	RMSE	S_{res}	R^2 Score
CC(Dataset1)	BGRU-BO	0.3412	0.0030	102610	0.0013	0.9994
	BLSTM-BO	0.1583	0.0013	52703	0.0011	0.9998
	LSTM-BO	5.7710	0.0513	1702035	0.0135	0.8358
	PLSTM-BO	0.1032	0.0009	32501	0.0010	0.9999
CD(Dataset2)	BGRU-BO	0.2442	0.0018	1100	0.0021	0.9997
	BLSTM-BO	0.1864	0.0013	905	0.0020	0.9998
	LSTM-BO	0.2472	0.0020	1005	0.0023	0.9998
	PLSTM-BO	0.1620	0.0012	867	0.0017	0.9998
RC(Dataset3)	BGRU-BO	0.3382	0.0028	95907	0.0030	0.9995
	BLSTM-BO	0.3125	0.0024	90818	0.0024	0.9996
	LSTM-BO	0.4612	0.0040	134178	0.0021	0.9991
	PLSTM-BO	0.2240	0.0018	65613	0.0021	0.9997

The current model is evaluated on dataset2, which produces a MAPE of. 0.1620, a MAE of 0.0012, an RMSE of 867, a S_{res} of 0.0017, and an R^2 score of 0.9998. The performance metrics of proposed and existing models are shown in Table 2, which represents that the present model performs reasonably well in terms of prediction value in comparison with other existing models. Particularly, BLSTM again performs better when compared with RNN variants. MAPE bar chart of the proposed model on daatset2 is shown in Figure 5. Figure 6(b) illustrates the prediction of proposed and existing models on test dataset2. On the coronavirus cumulative Recovery Cases(RC) in India dataset, the proposed model using a BO produces a MAPE of 0.2240, a MAE of 0.0018, a RMSE of 65613, a S_{res} of 0.0021 and an R^2 score of 0.9997. The results indicate that proposed model performs better than the existing models. However, BLSTM-BO produces reasonably good compared to LSTM and BGRU. Table 2 denotes the performance metrics of the proposed model and other RNNs. MAPE bar chart representation of models is shown in Figure 5. Figure 6(c) illustrates the prediction on dataset3 of proposed and existing models.

The current model is evaluated on dataset4 and produces a MAPE of. 0.9831, a MAE of 0.0059, an RMSE of 13473, a S_{res} of 0.0071, and an R^2 score of 0.9986. The performance metrics of proposed and existing models are shown in Table 3, which indicates that the current model performs reasonably well in terms of prediction value in comparison with the proposed and other existing models. Figure 7 shows the MAPE representation in the form of bar charts. Figure 8(a) illustrates the prediction of the proposed and other models on dataset4.

Table 3 denotes the performance metrics of PLSTM-BO and other models on dataset5. The proposed model obtains MAPE of 0.2061, MAE of 0.0015, a RMSE of 8745, S_{res} of 0.0018 and R^2 score of 0.9997. The obtained results indicate that PLSTM-BO is better than others in terms of MAPE, RMSE, MAE, S_{res} and R^2 score. The present model obtains low prediction error in relation to model (Abbasimehr and Paki, 2020) and existing models. MPAE bar chart of the current model and others is given in Figure 7. Figure 8(b) illustrates the prediction on dataset5.

The scope of existence of the proposed model is also verified in other domains, like stock price time series prediction, in which the model obtains a MAPE of 1.1901, a MAE of 0.0159, a RMSE of 194, a Sres of 0.0232, and an R2 score of 0.9896. It was observed that the proposed model achieves commendable results with respect to the BGRU and LSTM model. However, the current model performs better than the model (Kumar et al., 2022), and BLSTM as shown in Table 3. The proposed model and other model's predictions on NIFTY 50 closing price are shown in Figure 8(c). MAPE bar chart representation of models is given in Figure 7.

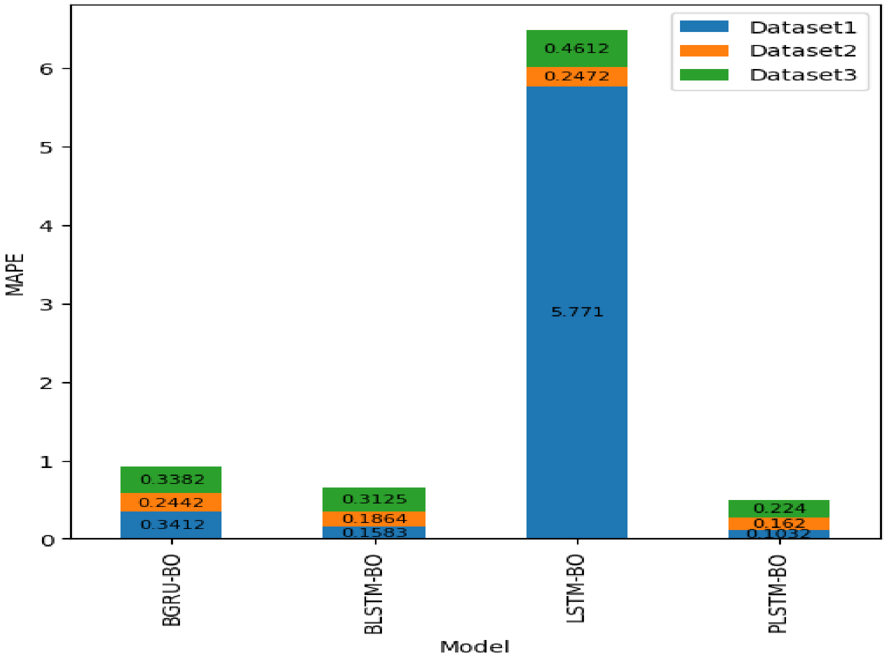


Figure 5. MAPE bar charts of present and existing models.

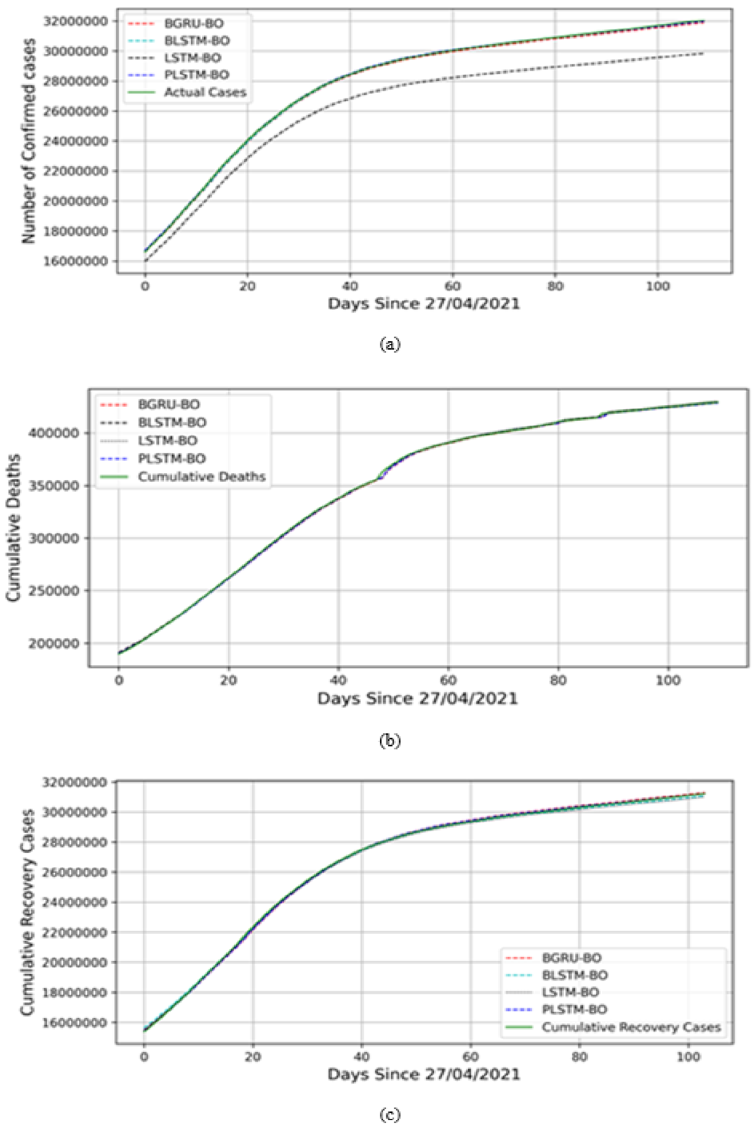


Figure 6. (a). Prediction on coronavirus cumulative confirmation cases test data in India (b) Prediction on cumulative deaths test data in India (c) Recovery cases test data prediction using PLSTM-BO and existing models.

Table 3. Performance metrics of PLSTM-BO and existing models on test data.

Type of dataset	Model/Metrics	MAPE(%)	MAE	RMSE	S_{res}	R^2 Score
CC(Dataset4)	BGRU-BO	4.2016	0.0250	56036	0.0230	0.9497
	BLSTM-BO	5.2568	0.0387	101080	0.0390	0.9224
	LSTM-BO	3.2420	0.0167	34112	0.0173	0.9911
	PLSTM-BO	0.9831	0.0059	13473	0.0071	0.9986
CC(Dataset5)	BGRU-BO	0.2865	0.0021	11603	0.0024	0.9996
	BLSTM-BO	1.9037	0.0161	96993	0.0165	0.9750
	LSTM-BO	0.3713	0.0028	15668	0.0028	0.9993
	PLSTM-BO	0.2061	0.0015	8745	0.0018	0.9997
Nifty50(Dataset6)	BGRU-BO	1.1962	0.0158	195	0.0231	0.9894
	BLSTM-BO	1.2263	0.0162	197	0.0237	0.9893
	LSTM-BO	1.2057	0.0159	196	0.0232	0.9896
	PLSTM-BO	1.1901	0.0157	194	0.0230	0.9897

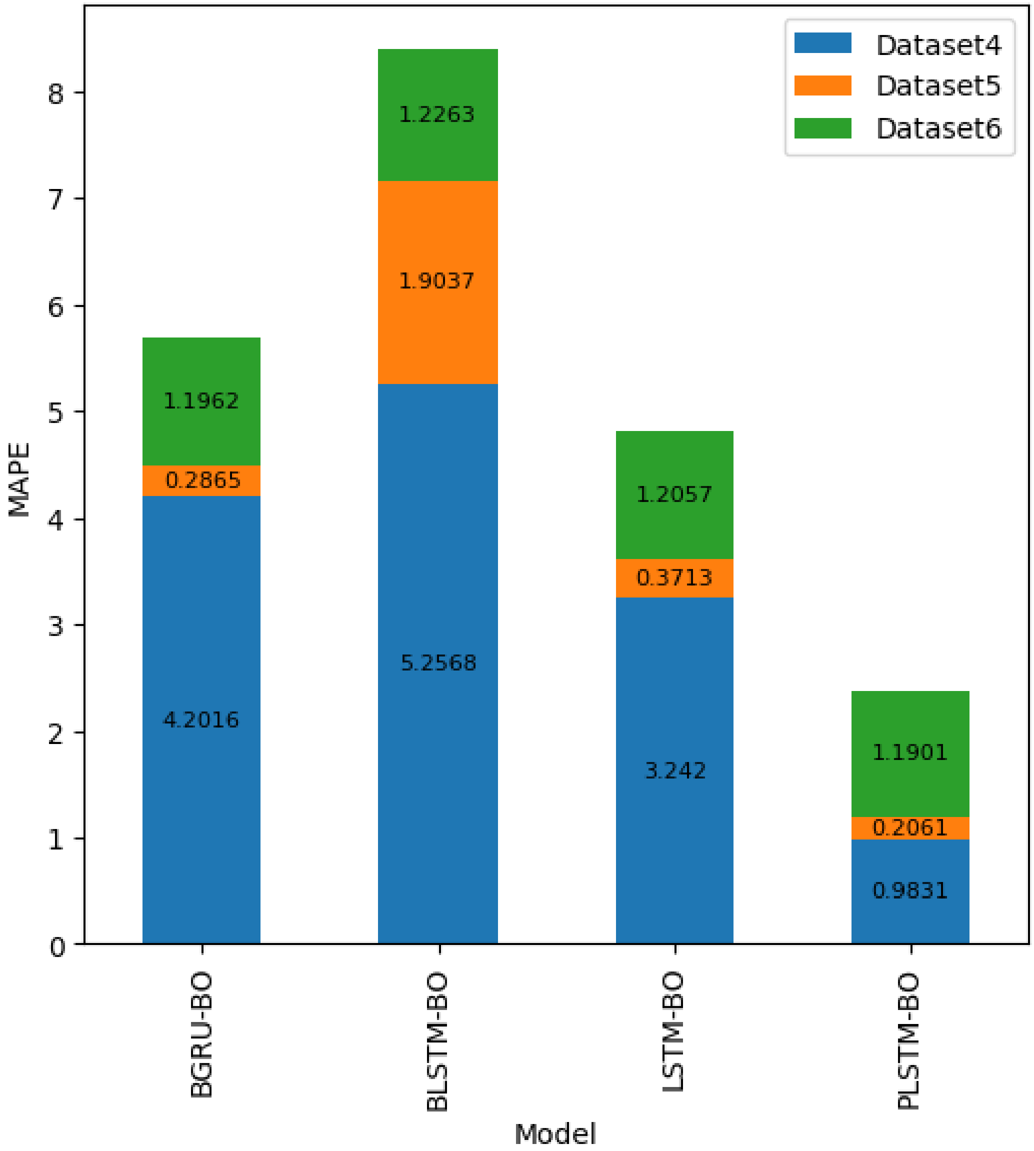
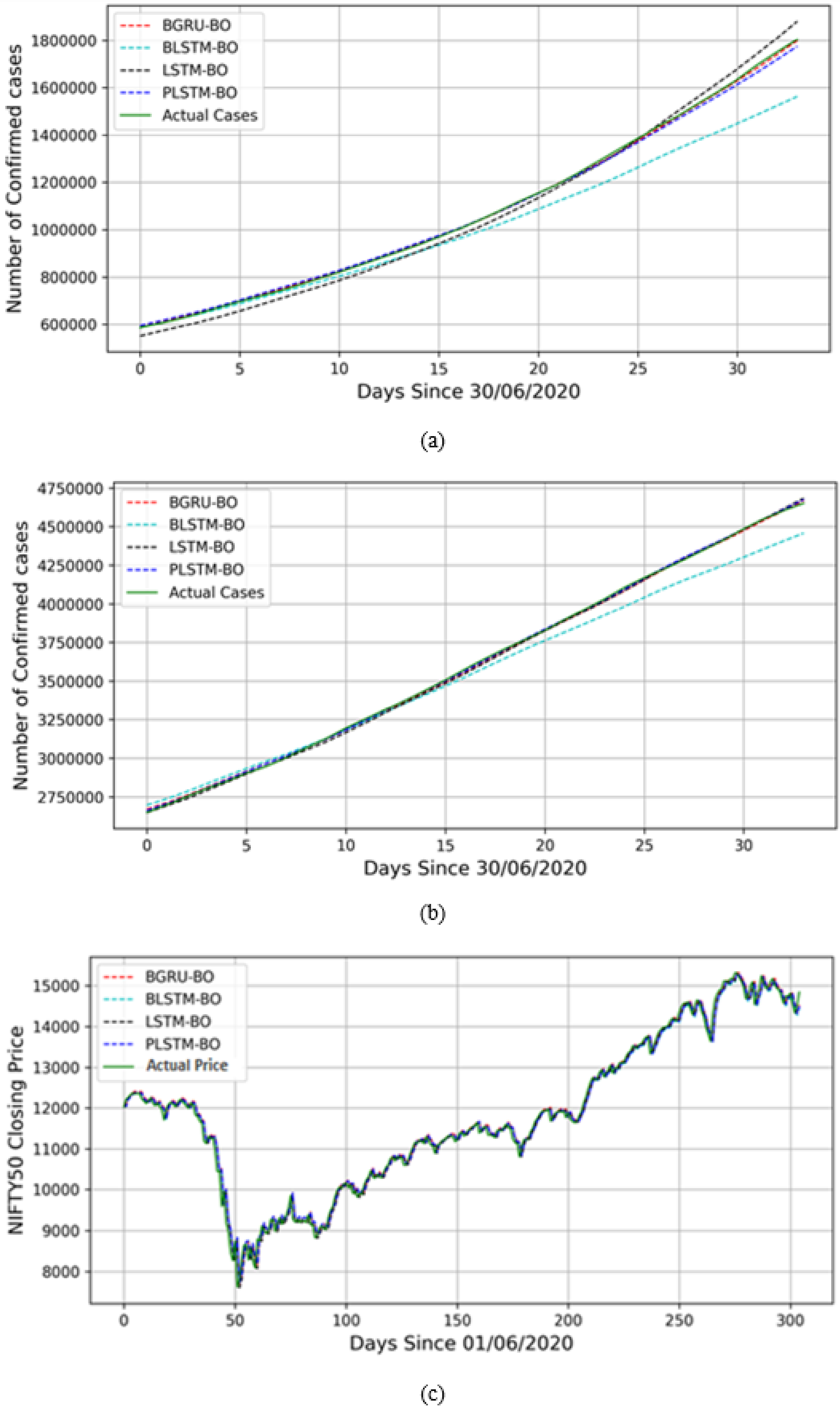


Figure 7. MAPE bar charts of proposed and existing models.



(a)

(b)

(c)

Figure 8. (a) Prediction of coronavirus cumulative cases test data in India (dataset4) (b) Prediction of coronavirus cumulative cases test data in USA (dataset5) (c) NIFTY 50 closing price test data prediction using proposed and existing models (dataset6).

Conclusion

In this paper, a parallel heterogeneous LSTM network Integrated with automatic tuning like Bayesian hyperparameter optimization obtains low prediction error with respect to COVID-19 daily cumulative confirmed cases, cumulative deaths, recovery cases patient's data in India and the cumulative confirmation cases in the USA. Moreover, the proposed model achieves Commendable results on the NIFTY 50 stock closing price dataset. The study concludes with the following findings.

The Proposed model is unique and obtains a low prediction error on the SARS-Cov-2 time series datasets and Commendable performance on NIFTY 50 stock closing price.

Two different LSTM networks are configured in parallel fashion to retain the complex relationship in time series, which is different from a bidirectional RNN network.

The present Model gives remarkable prediction efficiency, which can be used in different domains for forecasting with variable window size.

The Study promotes the scope for developing a new model with global search optimization approaches.

Future Scope

The model achieves notable prediction efficiency with respect to RNN variants. However, a few hyperparameters of the model have been considered for automatic tuning due to limited resources. In the future, work can be extended to consider long-term forecasting with different window sizes, and concurrently, it can also overlap other global search optimization approaches.

Acknowledgement

Not Applicable.

Conflict of Interest

The authors declare that there are no conflicts of interest.

References

Abbasimehr, H., & Paki, R. (2020). Prediction of COVID-19 Confirmed Cases Combining Deep Learning Methods and Bayesian Optimization. *Chaos, Solitons & Fractals*, 110511. https://doi.org/10.1016/j.chaos.2020.110511

Al-qaness, M. A. A., Ewees, A. A., Fan, H., & Abd El Aziz, M. A. E. (2020). Optimization Method for Forecasting Confirmed Cases of COVID-19 in China. *Journal of Clinical Medicine*, *9*(3), 674. https://doi.org/10.3390/jcm9030674

Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi- Koczy, A. R., Reuter, U., Rabczuk, T., & Atkinson, P.M. (2020). COVID-19 Outbreak Prediction with Machine Learning *Algorithms*, *13*(10), 249. https://doi.org/10.3390/a13100249

Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons & Fractals*, *139*, 110017. https://doi.org/10.1016/j.chaos.2020.110017

Bergstra, J., Remi, B., Yoshua, B., & Balazs, K. (2011). Algorithms for Hyper-Parameter Optimization, *25th Annual Conference on Neural Information Processing Systems*.

Bemani, A., Baghban, A., Mosavi, A., & S., S. (2020). Estimating CO_2-Brine diffusivity using hybrid models of ANFIS and evolutionary algorithms. *Engineering Applications of Computational Fluid Mechanics*, *14*(1), 818–834. https://doi.org/ 10.1080/19942060.2020.1774422

Chandra, R., Jain, A., & Singh Chauhan, D. (2022). Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLOS ONE*, *17*(1), e0262708. https://doi.org/10.1371/journal.pone.0262708

Di Gennaro, F., Pizzol, D., Marotta, C., Antunes, M., Racalbuto, V., Veronese, N., & Smith, L. (2020). Coronavirus Diseases (COVID-19) Current Status and Future Perspectives: A Narrative Review. *International Journal of Environmental Research and Public Health*, 17(8), 2690. https://doi.org/10.3390/ijerph17082690

Global Development. (2021). Oxford Martin School. http://oxfordmartin.ox.ac.uk/global-development

Gorgolis, N., Hatzilygeroudis, I., Istenes, Z., & Gyenne, L. (2019). Hyperparameter Optimization of LSTM Network Models through Genetic Algorithm. *10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. https://doi.org/10.1109/iis a.2019.8900675

Drewil, G. I., & Al-Bahadili, R. J. (2022). Air pollution prediction using LSTM deep learning and meta heuristics algorithms. *Measurement: Sensors*, 100546. https://doi.org/10.1016/j.measen.2022.100546

Graves, A. (2014). Generating sequences with recurrent neural networks. https://doi.org/10.48550/arXiv.1308.0850

Dubey, A., Dubey, S.M., Kumari, J., Yadav, P., & Sharma, G. (2024). Forecasting Wind Speed Using

Clustering of Trend-Based Time Series Data. *International Journal of Experimental Research and Review*, *42*, 40-49. https://doi.org/10.52756/ijerr.2024.v42.004

Haque, E., Tabassum, S., & Hossain, E. (2021). A Comparative Analysis of Deep Neural Networks for Hourly Temperature Forecasting. *IEEE Access*, *9*,160646–160660. https://doi.org/10.1109/access.2021.3131533

Huang, R., Wei, C., Wang, B., Yang, J., Xu, X., Wu, S., & Huang, S. (2022). Well performance prediction based on Long Short-Term Memory (LSTM) neural network. *Journal of Petroleum Science and Engineering*, *208*, 109686. https://doi.org/10.1016/j.petrol.2021.109686

Hochreiter, S., & Schmidhuber, J. (1997). Long Short Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyperopt. (2019).hyperopt/hyperopt. Retrieved from https://github.com/hyperopt/hyperopt

Kırbaş, I., Sozen, A., Tuncer, A. D., & Kazancıoglu, F. Ş. (2020). Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos, Solitons& Fractals*, *138*, 110015. https://doi.org/10.1016/j.chaos.2020 .110015

Kumar, G., Singh, U. P., & Jain, S. (2022). An adaptive particle swarm optimization-based hybrid long short-term memory model for stock price time series forecasting. *Soft Computing*. https://doi.org/10.1007/s00500-022-07451-8

Li, X., Ma, X., Xiao, F., Xiao, C., Wang, F., & Zhang, S. (2022). Time-series production forecasting method based on the integration of Bidirectional Gated Recurrent Unit (Bi-GRU) network and Sparrow Search Algorithm (SSA). *Journal of Petroleum Science and Engineering*, *208*, 109309. https://doi.org/10.1016/j.petrol.2021.109309

Malki, Z., Atlam, E.-S., Ewis, A., Dagnew, G., Alzighaibi, A. R., ELmarhomy, G., Elhosseini, M. A., Hassanien, A. E., & Gad, I. (2020). ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. *Neural Computing and Applications*, *33*(7), 2929–2948. https://doi.org/10.1007/s00521-020-05434-0

Melin, P., Monica, J. C., Sanchez, D., & Castillo, O. (2020). Multiple Ensemble Neural Network Models with Fuzzy Response Aggregation for Predicting COVID-19 Time Series: The Case of Mexico. *Healthcare*, *8*(2), 181. https:// doi.org/10.3390/healthcare8020181

Moftakhar, L., Seif, M., & Safe, M. S. (2020). Exponentially Increasing Trend of Infected Patients with COVID-19 in Iran: A Comparison of Neural Network and ARIMA Forecasting Models. *Iranian Journal of Public Health*. https://doi.org/10.18502/ijph.v49is1.3675

Muhammed, M.O. (2023). Hyperparameter Optimization of a Parallelized LSTM for Time Series Prediction. 1–26. https://doi.org/10.1142/s2196888823500033

Rashid, T. A., Fattah, P., & Awla, D. K. (2018). Using Accuracy Measure for Improving the Training of LSTM with Metaheuristic Algorithms. *Procedia Computer Science*, *140*, 324–333. https://doi.org/10.1016/j.procs.2018.10.307

Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B., Aslam, W., & Choi, G. S. (2020). COVID-19 Future Forecasting Using Supervised Machine Learning Models. *IEEE Access*, *8*, 101489–101499. https://doi.org/10.1109/ACCESS.2020.2997311

Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, *140*, 110212. https://doi.org/10.1016/j.chaos.2020.110212

Tirupati, G., Krishna Prasad, M.H.M., & Srinivasa Rao, P. (2021). COVID-19 Prediction Modeling Using Bidirectional Gated Recurrent Unit Network Model. *Journal of Webology*, *18*(5), 15-41.

Wang, Y.W., Shen, Z. Z., & Jiang, Y. (2018). Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China. *PLOS ONE*, *13*(9), e0201987. https://doi.org/10.1371/journal.pone.0201987

Wang, P., Zheng, X., Ai, G., Liu, D., & Zhu, B. (2020). Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru & Iran. *Chaos, Solitons & Fractals*, *140*, 110214. https://doi.org/10.1016/j.chaos.2020.110214

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82. https://doi.org/10.3354/cr030079

Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., & Deng, S.H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *Journal of Electronic Science and Technology*, *17*(1), 26–40.

https://doi.org/10.11989/JEST.1674-862X.80904120

Yahoo Finance. (2024). *Yahoo Finance - Business Finance, Stock Market, Quotes, News*. Yahoo Finance. https://finance.yahoo.com/

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

Yakubu, U. A., & Saputra, M. P. A. (2022). Time Series Model Analysis Using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for E-wallet Transactions during a Pandemic. *International Journal of Global Operations Research*, *3*(3), 80–85. https://doi.org/10.47194/ijgor.v3i3.168

Zhou, Q., Zhou, C., & Wang, X. (2022). Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and feature selection. *PloS One*, *17*(2), e0262501. https://doi.org/10.1371/journal.pone.026250

Zivkovic, M., Bacanin, N., Venkatachalam, K., Nayyar, A., Djordjevic, A., Strumberger, I., & Al-Turjman, F. (2021). COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustainable Cities and Society*, *66*, 102669. https://doi.org/10.1016/j.scs.2020.102669

How to cite this Article:

Tirupati G, Krishna Prasad MHM and Srinivasa Rao P. (2024). An Improved Parallel Heterogeneous Long Short-Term Memory Model with Bayesian Optimization for Time Series Prediction. *International Journal of Experimental Research and Review*, *45*, 106-118.

DOI : https://doi.org/10.52756/ijerr.2024.v45spl.009