Enhanced bat algorithm for COVID-19 short-term forecasting using optimized LSTM

Hafiz Tayyab Rauf · Jiechao Gao · Ahmad Almadhor · Muhammad Arif · Md Tabrez Nafis

Abstract

The highly infectious COVID-19 critically affected the world that has stuck millions of citizens in their homes to avoid possible spreading of the disease. Researchers in different fields are continually working to develop vaccines and prevention strategies. However, an accurate forecast of the outbreak can help control the pandemic until a vaccine is available. Several machine learning and deep learning-based approaches are available to forecast the confirmed cases, but they lack the optimized temporal component and nonlinearity. To enhance the current forecasting frameworks’ capability, we proposed optimized long short-term memory networks (LSTM) to forecast COVID-19 cases and reduce mean absolute error. For the optimization of LSTM, we applied bat algorithm. Furthermore, to tackle the premature convergence and local minima problem of BA, we proposed an enhanced variant of BA. The proposed version utilized a Gaussian adaptive inertia weight to control the individual velocity in the entire swarm. In addition, we substitute random walk with the Gaussian walk to observe the local search mechanism. The proposed LSTM examines the personal best solution with the swarm’s local best and preserves the optimal solution by combining the Gaussian walk. To evaluate the optimized LSTM, we compared it with the non-optimal version of LSTM, recurrent neural network, gated recurrent units, and other recent state-of-the-art algorithms. The experimental results prove the superiority of the optimized LSTM over other recent algorithms by obtaining 99.52% accuracy.

Keywords COVID-19 · Gaussian distribution · Gaussian inertia weight · LSTM

1 Introduction

The entire world is experiencing a continuous pandemic called the coronavirus (COVID-19) disease due to severe acute respiratory syndrome coronavirus two (SARS-CoV-2) (Abrams et al. 2020). It has been arisen from Wuhan, the capital of Hubei Province in China, through December 2019 (WH Organization et al. 2020). The virus has been discovered on 7th January and found that it is distributed by human-to-human transmission through direct contact or droplet (Wang et al. 2020; Cucinotta and Vanelli 2020). Covid-19 was estimated to be an average incubation period of 6.4 times and a first reproduction number of 2.24–3.58. It has been spread over the entire world, and so the World Health Organization (WHO) had announced COVID-19, a worldwide outbreak on 11th March 2020 (Huang et al. 2020). COVID-19 contains a few taxonomy symbols as it belongs to the coronavirus family. All such viruses hold several essential proteins fastened in the viral membrane. As it is well worth discovering, the viral plot displays a large diameter, nearly double of a standard organic layer (Bárcena...
et al. 2009). The genome of SARS-CoV-2 includes six notable open-reading structures (ORFs), usually investigated in several CoVs. A number of the genes received less than 80% nucleotide chain identification to SARS-CoV (Zhou et al. 2020). With ultraviolet warmth and rays, COVID-19 is fragile. There is a common misconception that at 27°C, this virus might have disappeared. Additionally, Covid-19 may be inactivated by chloroform, peroxymonosulfuric acid, chlorine-containing disinfectant, ether (75 percent), except for chlorhexidine (Cascella et al. 2020).

In 1995, a large-scale study proved that primary clinical symptoms are dyspnea (21.9 percent of cases), expectoration (28.2 percent of cases), fatigue or myalgia (35.8% of cases), cough (68.6% of cases), and ever (88.5 percent of cases). In contrast, the minor ones contain vomiting and nausea (3.9% of cases), fatigue (4.8% of cases), headache, or nausea (12.1% of cases) (Lq et al. 2020). The frequency of novel coronavirus, like many pathogens, is thought to transpire by respiratory droplets. Thus, the immense bulk of scattering cases is restricted to the adjacent spaces (Cascella et al. 2020).

The SARS-CoV-2 is a pathogenic human coronavirus below the beta coronavirus genus. In the last decade, the two pathogenic species MERS-CoV and SARS-CoV were outbreaks in 2012 and 2002 in the Middle East and China, respectively (Lu et al. 2020; Cui et al. 2019). The laboratory of China put at the NCBI GenBank by discovering the whole genomic sequence (Wuhan-HU1) of the massive RNA virus (SARS-CoV-2) on 10th January (Yang 2020). Hence, SARS-CoV-2 is one positive-stranded RNA virus (Lu et al. 2020).

Following the WHO, no anti-inflammatory medicines and vaccines are not yet prepared for this pandemic (Basu and Chakraborty 2020), and medical industries are looking hard to acquire the vaccine. The vaccine may take at least 18–24 months until it is available, following the quick tracking of the normal vaccine intervals. 5–10 decades, and may take additional time to make it appropriate for the large organizations of the world (Grenfell and Drew 2020). Additionally, we do not understand just how long a vaccine could remain successful since the virus mutates. Every attempt was adopted to slow down the coronavirus spread and prepare reasonable medical systems to protect front-line medical staff with sufficient supplies of protective equipment such as personal protective equipment (PPE) masks and other essentials. Consequently, if we knew ahead of the number of new coronavirus cases for the next ten days, we could plan our necessary actions. As compared to Asian countries, the USA has been greatly affected by COVID-19. USA COVID-19 cases summary from Feb 2020 to Sep 2020 is illustrated in Fig. 1.

The success of healthcare technologies is a key to artificial intelligence (Panch et al. 2019). Data is structured in smart devices and increases the efficiency of healthcare machine learning (Knight et al. 2016). Several COVID-19 forecasting approaches have been proposed based on machine learning, deep learning, and statistical learning in the past few weeks. However, the primary issue is they lack the temporal components and nonlinearity in terms of machine learning where deep learning approaches are limited to comparative analysis, and uni-model forecasting (Benvenuto et al. 2020; Wieczerzek et al. 2020a). Furthermore, some studies considered epidemiological models that need to make hypothesis-based parameter initialization. That model tends to low the net precision due to its under-fitting data nature (Wieczerzek et al. 2020a; Gao et al. 2019).

Several optimization algorithms have been used in previous studies to solve time series problems for the weight optimization of neural networks, such as the arithmetic optimization algorithm (Abualigah et al. 2021), group search optimizer (Abualigah 2020), dragonfly algorithm (Alshinwan et al. 2021), genetic algorithm (Momani et al. 2016), reproducing kernel algorithm (Arqub et al. 2017; Arqub 2017) and fuzzy conformable fractional approaches (Arqub and Al-Smaedi 2020).

To predict the distribution of COVID-19 in various regions, the authors used Google trend and ECDC data term frequency (Prasanth et al. 2021). To pick the successful COVID-related search words, they used Spearman correlation. The optimization of hyperparameters through the LSTM network proposed a new technique based on a meta-heuristic GWO algorithm.

Three approaches are suggested (Abbasimehr and Paki 2021) that combine Bayesian optimization and deep learning. The optimized values for hyperparameters are effectively chosen by Bayesian optimization in their process. The system architecture is considered to be a process of multiple-output forecasting. Their proposed methods performed better than the reference model on data from the COVID-19 time series.

In order to forecast the COVID-19 outbreak in Saudi Arabia, a study of various deep learning models is proposed (Elsheikh et al. 2021). Officially recorded data was used to evaluate the model. The optimal values of the parameters of
the model that optimize the accuracy of forecasting have been determined. They used seven statistical evaluation parameters to forecast the accuracy of the model.

Likewise, the previous studies on COVID-19 did not consider the hyperparameter optimization of neural networks that can help boost the performance of models.

To overcome the issue as mentioned above, we proposed a deep learning model that predicts real-time transmission using optimized LSTM. For the optimization of LSTM, we employed BA. To further deals with the premature convergence (Perwaiz et al. 2020; Rauf et al. 2020b), and local minima problem (Rauf et al. 2020a) of BA, we proposed an enhanced variant of BA. The proposed version consists of two significant enhancements. Firstly, we carried out Gaussian adaptive inertia weight to control the individual velocity in the entire swarm. Secondly, we substitute the random walk with the Gaussian walk to explore the local search mechanism.

2 Methodology

2.1 Proposed BA

The real-world challenges are becoming more complicated every day. Swarm intelligence (SI) is the subset of meta-heuristic algorithms employed to tackle complex optimization problems of continuous nature. We used the self-learning nature of this meta-heuristic to optimize the neural network training parameters. Such features clearly state that local interaction is essential between the swarm-based system components to preserve their potential.

In this research, we have carried out an enhanced version of BA to optimize LSTM training weights. The optimized LSTM dynamically adopt optimal training parameter and decide the execution cycle timeline based on the global convergence manner of enhanced BA. We bring two modifications to classical BA. Firstly, we proposed Gaussian adaptive inertia weight to improve the velocity updating mechanism. Finally, we update each individual’s local searching strategy to retain local solutions based on the weighted mean of the personal best and the current global solution of the entire swarm.

Properties of standard BA are as follows:

- Every micro-bat estimates distance within surroundings and prey by utilizing its property of echolocation.
- Frequency of fixed range is utilized to find micro-bat’s velocity from location beside different loudness and distinct wavelength while searching for prey.
- Emission pulse rate increases to adjust its pulse frequency while estimating distance among prey and micro-bat.
- Loudness will decrease from a considerable positive value to a smaller value.

BA follows three fundamental rules to converge toward an optimal solution.

- Each bat is represented by $\tau_i^t$ for $i = \{1, 2, 3, \ldots, N\}$ with the whole population $\mathcal{N}_p$ in an entire search space $S$ and use sonar echolocation to sense the prey and measure the estimated difference of the distance to the prey.
- During the convergence process, each bat’s moves with velocity $v_i^t$ and the frequency $f_i^t$. The current position of individual can be represented by $x_{ip}^t$, where $p$ represents the partial coordinate of the current search space. The frequency $f_{\text{min}}^t$, consolidates with bat wavelength $\omega$ and variation of loudness $A_o$.
- The variation of loudness $A_o$ depends on the current location $x_{ip}^t$ and the estimated distance $D_{ip}^t$.

Population of fixed size $S_p$, in our case $S_p = 40$, is initialized with the random initial values following the uniform distribution $\tau_i^1 \in [\tau_l, \tau_u]$, where $l$ and $u$ are lower and upper limits of uniformly distributed sequence. After population initialization, the mutation operators are used to encourage the bats’ movement in the multidimensional search space. The ultimate objective of this phase is to obtain the new local solution, while the frequency $f_{\text{min}}^t$ factor controls the step-size of the solution. For each individual $\tau_i^t$, the current frequency $f_i^t$, current velocity $v_i^t$ and the current bats potion $x_{ip}^t$ can be updated using the following equations.

$$f_i^t = f_{\text{min}}^t + (f_{\text{max}}^t - f_{\text{min}}^t) \cdot R$$
$$v_i^{t+1} = v_i^t + (x_{ip}^t - x_{ig}^t) \cdot f_i^t$$
$$x_{ip}^{t+1} = x_{ip}^t + v_i^{t+1}$$

Referred to equation 1, $f_{\text{max}}^t - f_{\text{min}}^t$ are the difference of lower and upper corresponding frequency where $R$ indicates the random number over the interval of $[0, 1]$. Velocity of each individual $\tau_i^t$ can be updated using equation 2, where $x_{ip}^t - x_{ig}^t$ is the mean difference of local solution $x_{ip}^t$ of entire swarm and global solution $x_{ig}^t$ of all swarms. Likewise, the new vector solution $x_{ip}^{t+1}$ can be determine using equation 3.

In the proposed BA, we introduced Gaussian adaptive inertia weight to update the velocity in such a manner to avoid more long jumps leading to exploration and to avoid more short jumps leading to exploitation. The proposed Gaussian adaptive inertia weight can help the velocity updating mechanism achieve each individual’s optimal convergence steps. The Gaussian function can be defined as:

$$f(x) = xe^{-\frac{(a-y)^2}{2c^2}}$$
Table 1 Recent related works with their dataset details and results

| Ref.           | Dataset                                      | Model                          | Results                                      |
|----------------|----------------------------------------------|-------------------------------|---------------------------------------------|
| Wieczorek et al. (2020b) | Government repositories          | NAdam training model          | Accuracy above 99%                          |
| Chowdhury et al. (2020)  | Bangladesh COVID-19                   | Neuro-fuzzy inference system  | Correlation coefficient 0.75, MAPE 4.51, and RMSE 6.55 |
| Dutta et al. (2020)      | WHO official                        | CNN and RNN                    | CNN-LSTM approach outperforms               |
| Chimmula and Zhang (2020)| Dataset Canadian Health Authority     | LSTM                          | Gained highly accurate results              |
| Pathan et al. (2020)     | The patient’s dataset of different countries | RNN and LSTM                 | Obtained optimum results                    |
| Alakus and Turkoglu (2020)| Laboratory data              | Clinical predictive models     | Accuracy of 86.66% and F1-score of 89%      |
| Tuli et al. (2020)       | Data by Hannah Ritchie               | ML-based improved model        | Yields high accuracy                        |
| Kavadi et al. (2020)     | Indian dataset                      | Linear regression model        | Outperformed state-of-the-art methods       |
| Pinter et al. (2020)     | Data from Hungary                   | Multi-layered perceptron-imperialist competitive algorithm (MLP-ICA) | Obtained promising results                  |
| Prasanth et al. (2021)   | Google trend and ECDC data            | A hybrid GWA algorithm         | Reduce MAPE by 74% results                  |
| Abbasimehr and Paki (2021)| Live time series data               | Bayesian optimization based algorithm | Mean SMAPE is 0.25 results                    |
| Elsheikh et al. (2021)   | Official data from Saudi Arabia      | LSTM and other variants        | Obtained highly accurate results            |

where \((x, y, z)\) are real constant that can be varied over the nature of the problem. A bell shape curve in the Gaussian distribution indicates the height of bell curves and can help the population control the exploration process with the following probability density function.

\[
 g(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.
\] (5)

In equation 5, \(\mu = y\), and can be interpreted as the expected value with variance \(\sigma^2 = z^2\).

In order to generate optimal location vectors \(\overline{g}_i^{t+1}\) through Gaussian distribution over \(t\) iterations and \(D\) dimensions, the mathematical definition following the adaptive process can be:

\[
\overline{g}_i^{t+1} = \overline{g}_{\text{min}} + (\overline{g}_{\text{max}} - \overline{g}_{\text{min}}) \cdot \frac{\overline{v}_i^t}{\sigma}.
\] (6)

where \(\overline{g}_{\text{max}} - \overline{g}_{\text{min}}\) are upper and lower intervals \([0, 1]\) of Gaussian distribution. The proposed BA utilized the following equation to update the velocity of each bat \(\overline{v}_i^{t+1}\).

\[
\overline{v}_i^{t+1} = \overline{v}_i^t + \left(\overline{x}_{ip} - \overline{x}_{ig}\right) \cdot f_i^t.
\] (7)

In equation 7, \(\overline{g}_i^{t+1}\) shows the proposed Gaussian adaptive inertia weight factor, controlling the exploration and exploitation during the entire convergence process. Gaussian bell curves in the adaptive inertia weight dynamically select each bat’s speed to help the local best vector holder bat to escape local minima. Apart from velocity \(\overline{v}_i^{t+1}\), updated local solutions \(\overline{x}_{ip}^{\text{new}}\) play an essential role in the exploitation of bats. Consider the speed is regulated, but the newly generated local solutions \(\overline{x}_{ip}^{\text{new}}\) are not robust enough to limit the boundary of the entire swarm’s global worst \(\overline{x}_{ig}\). In that case, premature convergence can be held. Standard BA uses the following equation to select the best solution among all existing vectors in the swarm:

\[
\overline{x}_{ip}^{\text{new}} = \overline{x}_{ig}^t + \varepsilon A_i^t.
\] (8)

\(\varepsilon\) is a random walk generator throughout \([0, 1]\) and \(A_i^t\) represents the average loudness factor. The random walk can produce the best solution in the current iteration \(t\) and build the worst one in the next iteration \(t + 1\). The entire local best holder individual will likely follow the best solution \(\overline{x}_{ig}\), which is the worst in the next iteration \(t + 1\) and leads to the local minima and premature convergence problem. To avoid this random selection that leads to the worst local best solution and effect exploitation, we replace this random walk
with a Gaussian walk and propose a local search mechanism. Our proposed variant of BA will use the following equation to attain the local best solution $x_{iG}^{new}$.

$$x_{iG}^{new} = x_{ig}^{t-1} + \frac{1}{iG} (x_{ig}^{t} - \overline{P}_{ig}) + \epsilon A_{i}^{t}. \quad (9)$$

In the proposed equation 9, $x_{ig}^{t+1}$ is previously computed Gaussian distribution where $x_{ig}^{t} - \overline{P}_{ig}$ is the mean difference of local best of swarm $x_{ig}^{t}$ and the personal best $\overline{P}_{ig}$ of each bat. The proposed solution will iteratively evaluate the current best and the local best solution $\overline{P}_{ig}$ for each swarm $x_{ig}^{t}$ in the population and check the following condition to use the iterative difference.

$$x_{iG}^{new} = \begin{cases} x_{ig}^{t} & if \ (x_{ig}^{t} > x_{ig}^{t-1}) \\ x_{ig}^{t-1} - \overline{P}_{ig} & Otherwise \end{cases}. \quad (10)$$

Referred to equation 10, the new local best will be selected $x_{ig}^{t}$ if the bats’ personal best is less than the swarm local best otherwise, the weighted mean of local best $x_{ig}^{t}$ and global best $\overline{P}_{ig}$ will be chosen as new local best.

New $N$ local bests $x_{iG}^{new}$ will likely control by the convergence rate, which can be defined by two critical factors loudness $A_{i}$ and pulse emission rate $p_{i}^{t}$ which can be update thought the following two equations.

$$A_{i}^{t+1} = \alpha A_{i}^{t} \quad (11)$$

$$p_{i}^{t} = p_{i}^{t-1}[1 - \exp(-\gamma^{t})]. \quad (12)$$

### 2.2 Optimized Long Short-Term Memory (LSTM)

Recurrent neural networks (RNN) has turned out to be the most reliable algorithm for prediction as essential features are extracted automatically from samples of training (Jiang and Schotten 2020). RNN performed well at data processing, and ensured encouraging outcomes for time series prediction while keeping immense information in the internal state (Conn et al. 2019). Nevertheless, it might take much training time due to gradient detonate and evanescence problems (Tomar and Gupta 2020). Hence, in 1997 a long short-term memory (LSTM) structure was designed by Schmidhuber and Hochreiter (Hochreiter and Schmidhuber 1997) to overcome that flaw by administering long-term dependency through multiplicative gates that will handle memory cells and flow of information in the recurrent hidden layer. LSTM’s architecture comprises four gates, i.e., input gate, output gate, control gate, and forget gate (Tomar and Gupta 2020).

Input can be defined as:

$$i_{t} = \sigma(W_{i} *[h_{t-1}, x_{t}] + b_{i}). \quad (13)$$

The information extracted from the above equation can be transferred to the cell. Forget gate decides data that will be ignored from the previous layer’s input by utilizing the following equation:

$$f_{t} = \sigma(W_{f} *[h_{t-1}, x_{t}] + b_{f}). \quad (14)$$

The input from the entire memory cell is controlled by control gate through following equations:

$$C_{t} = \sigma(W_{c} *[h_{t-1}, x_{t}] + b_{c}) \quad (15)$$

$$C_{t} = f_{t} * C_{t-1} + i_{t} * C_{t} \quad (16)$$

Output and hidden layer $h_{t}$ is updated as following:

$$O_{t} = \sigma(W_{o} *[h_{t-1}, x_{t}] + b_{o}) \quad (17)$$

$$h_{t} = O_{t} * h(C_{t}) \quad (18)$$

The interval [-1 to 1] is normalized by using tanh, where $W$ is the weight matrices and $\sigma$ shows activation function taken as sigmoid.

We feed the learning rate, momentum rate, and dropout rate in each of the LSTM dropout layers to the BA for automatic optimization of the hyperparameters. Each parameter is examined before the classification layer of LSTM to determine BA’s best optimal global solution. If the fitness function produces the same values, the proposed algorithm will check in the next generation to see if it avoids premature convergence.

Hyperparameters of each hidden layer $h_{t-1}$ for $t = \{1, 2, 3 \ldots N\}$ are optimized by providing global solution $x_{iG}^{new}$ obtained using equation 9. The output layer of optimized LSTM can be interpreted as:

$$O_{t} = \sigma(W_{o} *[h_{t-1}, x_{t}] + b_{o}) \quad (19)$$

where each hidden layer choose global best of the entire population $x_{ig}^{t}$ or mean of personal best and local best of swarm $x_{ig}^{t} - \overline{P}_{ig}$. The pseudocode of proposed Algorithm is presented in Algorithm 1.

We also checked single parameter optimization impact on the proposed technique, and we observed that only learning rate optimization produces a negligible impact on the performance of the proposed LSTM. However, the collective optimization of the learning rate, momentum rate, and dropout rate tends to increase the overall performance of the proposed LSTM.
WHO accounted for the outbreak of COVID-19 in states and regions around the world. Several areas of South and North America, in particular, witness the adverse effects of a massive COVID-19 explosion. The operation of huge air traffic between each state of the USA has entirely encouraged COVID-19 to propagate from its source to the next infected states; individual-to-individual spread has thus been reported among travelers worldwide. The primary goal of this

---

**Fig. 2** Proposed architecture of optimized LSTM

**3 Experiments**

WHO accounted for the outbreak of COVID-19 in states and regions around the world. Several areas of South and North America, in particular, witness the adverse effects of a massive COVID-19 explosion. The operation of huge air traffic between each state of the USA has entirely encouraged COVID-19 to propagate from its source to the next infected states; individual-to-individual spread has thus been reported among travelers worldwide. The primary goal of this
Algorithm 1 Pseudocode for proposed BA

Require: Initial fitness values of LSTM parameters: \( \bar{x}_i \) for \( i = 1, 2, 3 \ldots N_p \)

Ensure: Optimal solution \( \bar{x}_{igw}^{new} \) for \( i = 1, 2, 3 \ldots N_p \)

1: Population initialization \( \bar{x}_i \) Random(0,1)
2: Determine initial fitness (\( \bar{x}_i \))
3: Determine initial global fitness (\( \bar{x}_{igw} \))
4: while \( t \leq \) termination do
5: \( i = 1 \) to \( N_p \) do
6: Update \( f_i \) through equation 1.
7: Update \( \bar{x}_i^{t+1} \) through equation 7.
8: Update \( \bar{x}_{igw}^{t+1} \) through equation 3.
9: if \( r(0, 1) > \bar{x}_i \) then
10: Update local solution through equations 9-10.
11: end if
12: \( f(\text{new}) \) \( \bar{x}_{igw}^{t+1} \)
13: Generation = Generation + 1
14: if \( f(\text{new}) < f(\text{old}) \) and \( R(0, 1) < \bar{x}_i \) then
15: \( \bar{x}_{igw}^{t+1}, \bar{x}_i^{t+1} \) and \( f(\text{old}) \) \( f(\text{new}) \)
16: end if
17: \( \bar{x}_{igw}^{t+1} = \min(\bar{x}_{igw}^{t+1}) \)
18: end for
19: end while

The experiments are conducted using specific python packages, namely Keras, TensorFlow, NumPy, and iplot using python language to compare the performance of the proposed optimized LSTM, tested other standard forecasting algorithms, i.e., Simple LSTM, GRU, and RNN.

3.1 Results

This study provides an optimized deep-learning model for COVID-19’s time series analysis of the USA. The proposed framework dynamically selects optimal training parameters and determines the execution cycle based on enhanced BA’s global convergence manner.

The forecasting of COVID-19 was achieved in two preliminary stages: data training and evaluation. To compare the proposed variant with existing algorithms, we used five evaluation metrics; namely root mean absolute error (RMSE), mean absolute percentage error (MAPE), standard deviation (Stdev), prediction interval, and accuracy. The following equations can define RMSE, MAPE, and Stdev:

\[
RMSE = \left[ \frac{1}{N} \sum_{i=1}^{N} (a_i - \hat{a}_i)^2 \right]^{1/2}
\]  
(20)

where \( a_i - \hat{a}_i \) represents squared difference for forecasted and actual values.

\[
MAPE = \frac{1}{n} \sum_{i=1}^{N} \left| \frac{e_i}{d_i} \right|
\]  
(21)

where \( |e| \) indicates absolute error and \( d \) shows demand for each period.

\[
Stdev = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.
\]  
(22)

In the above equation, \( \bar{x} \) is mean of \( i \)th sample and \( N \) indicates total number of instance.

The raw data is pre-processed and standardized in the initial stages and subsequently used to develop the optimized predictive model based on LSTM. The model’s boundary parameters are selected so that the MAPE can be minimized. From a particular stage on, the optimized LSTM with the optimal learning parameters is used in the testing phase to predict the extent of COVID-19 cases in the USA.

Table 2 presents the empirical results for confirmed and predicted cases obtained through GRU, RNN, LSTM, and optimized LSTM. RMSE shows the root mean square errors in each network during the training. MAPE is total loss subtracted from precision, where Stdev shows the significant difference between confirmed and predicted COVID-19 cases. Prediction interval represents the difference in response to confirmed cases between each day of the forecasted cases.

We presented a statistical test called Kruskal–Wallis test for the experimental results, comparing the results with other published methods. The average rank, median value, and Z-score obtained through Kruskal–Wallis test for each employed algorithm is presented in Table 5.

Likewise, training and validation loss minimization curves using GRU, RNN, LSTM, and optimized LSTM are illustrated in Figs. 3, 4, 5, and 6. The convergence curves of real and forecasted COVID-19 cases through optimized LSTM in the USA are presented in Fig. 7.

A comparison of the proposed optimized LSTM with other standard deep learning forecasting models is tabulated in Table 4. We take the forecasting dates from 1/9/20 to 10/9/20, and to validate the predicted values, we retain previous ten-day cases 22/8/20 to 31/8/20. Referenced to Table (4), actual confirmed cases do not appear yet in the USA from 31/8/20 to 1/9/20, predicted shows the forecasted cases through existing
Table 2 Comparison of proposed optimized LSTM with other standard deep learning forecasting models

| Model      | RMSE   | MAPE   | Stdev  | Prediction interval | Accuracy |
|------------|--------|--------|--------|---------------------|----------|
| GRU        | 1786.613 | 30.01539 | 3261.895 | 6393.313572 | 70       |
| RNN        | 531.3041  | 8.817398 | 970.0242 | 1901.247   | 91       |
| LSTM       | 751.2309  | 12.12951 | 1371.554 | 2688.245   | 88       |
| Optimized LSTM | 32.99262 | 0.483875 | 60.23602 | 118.0626   | 99.52    |

3.2 Analysis

Table 2 shows that GRU obtained the worst accuracy with 1786.613 RMSE and 3261.895 Stdev, which shows a significant difference between actual and predicted COVID-cases. After GRU, standard LSTM performed better with 2688.245 prediction intervals and 12.12 MAPE. The performance of RNN is relatively good compared to GRU and LSTM with 91% accuracy and 1371.55 Stdev. Lastly, it can be seen that the proposed version of optimized LSTM outperformed all other deep learning models with 32.99 RMSE better than...
GRU, 0.4838 MAPE better than LSTM, and only 60.23 significant difference among confirmed and predicted cases.

Furthermore, the validation loss in the case of GRU and RNN is not stable throughout the learning process and meets greater than 0.5 and 0.7 (refer Figs. 3 and 4). From Fig. 5, the validation loss of LSTM is stable compared to GRU and RNN throughout the learning process with a greater 0.40. As opposed to GRU, LSTM, and RNN, the proposed model minimized the validation loss up to 0.04 and shows the better capability of loss minimization (refer Fig. 6).

The performance of the proposed optimized LSTM can be confirmed through Fig. 7, where the USA’s actual cases on 31/8/20 were 6030587, and the predictions were 3734918, 5328279, 7653031, and 6097641 using GRU, RNN, LSTM, and optimized LSTM, respectively.

From Table 5, it can be observed that the proposed LSTM obtained the best mean rank of 17.0 through Kruskal–Wallis test as compared to others. Advanced algorithms such as NAdam with 41 mean rank, two LSTM variants with 16 and 13 mean ranks, respectively. Similarly, the proposed LSTM outperformed other published results by obtaining the best positive Z-score of 163.

We can conclude that using the proposed optimized framework can help the USA and other governments predict the actual cases with 99% accuracy and take precautionary measures in advance.
Table 5 Kruskal–Wallis test: proposed LSTM vs recent state-of-the-art algorithms

| Model                      | Median | Ave rank | Z    |
|----------------------------|--------|----------|------|
| Adagrad Wieczorek et al. (2020b) | 40.100 | 2.5      | −1.33|
| Adam Wieczorek et al. (2020b)   | 87.530 | 9.0      | 0.00 |
| AdamaX Wieczorek et al. (2020b) | 87.470 | 8.0      | −0.20|
| ANN Alakus and Turkoglu (2020) | 86.900 | 6.0      | −0.61|
| CNN Alakus and Turkoglu (2020) | 87.350 | 7.0      | −0.41|
| CNNLSTM Alakus and Turkoglu (2020) | 92.300 | 13.0     | 0.82 |
| CNNRNN Alakus and Turkoglu (2020) | 86.240 | 5.0      | −0.82|
| Ftrl Wieczorek et al. (2020b)  | 40.100 | 2.5      | −1.33|
| LSTM-1 Chimmula and Zhang (2020) | 93.400 | 15.0     | 1.22 |
| LSTM-2 Chimmula and Zhang (2020) | 92.670 | 14.0     | 1.02 |
| LSTM Alakus and Turkoglu (2020) | 90.340 | 12.0     | 0.61 |
| LSTM Wieczorek et al. (2020b)  | 93.560 | 16.0     | 1.43 |
| NAdam Wieczorek et al. (2020b) | 87.730 | 11.0     | 0.41 |
| RMSprop Wieczorek et al. (2020b) | 87.650 | 10.0     | 0.20 |
| RNN Alakus and Turkoglu (2020) | 84.000 | 4.0      | −1.02|
| SGD Wieczorek et al. (2020b)   | 9.800  | 1.0      | −1.63|
| Optimized LSTM               | 99.520 | 17.0     | 1.63 |

4 Conclusion

This research offers the optimized LSTM to forecasts COVID-19 cases in the USA. Many machine learning and deep learning approaches are available to forecast concerned cases, but they lack both the optimized temporal aspect and nonlinearity. To overcome this issue, we applied the BA for the optimization of LSTM. Besides, we implemented an enhanced BA variant to tackle BA's premature convergence and local minima problem. The proposed version of BA used Gaussian adaptive inertia weight to control the individual velocity in the swarm. In addition, we replace the random walk with the Gaussian walk to observe the local search. The robust local search mechanism assists LSTM hyperparameter optimization during the training process. The proposed optimized LSTM is compared with GRU, RNN, and LSTM. Empirical results reveal that optimized LSTM minimized MAPE by 0.48, which is far better than the existing algorithms.

In future work, we intend to adopt other evolutionary models such as the Genetic Algorithm and Differential evolution algorithm in the regression-based deep learning model for multivariate forecasting of a pandemic.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.
Connor J, Martin R, Atlas L (1994) Recurrent neural networks and robust time series prediction. IEEE Trans Neural Netw 5(2):240–254. https://doi.org/10.1109/72.279188

Cucinotta D, Vanelli M (2020) Who declares covid-19 a pandemic. Acta Bio-Medica Atenei Parmensis 91(1):157–160

Cui J, Li F, Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17(3):181–192

Dutta S, Bandypadhyay SK, Kim TH (2020) Cnn- lstm model for verifying predictions of covid-19 cases. Asian J Res Comput Sci, pp 25–32

Elsheikh AH, Saba AI, Abd Elaziz M, Lu S, Shanmugan S, Muthuramalingam T, Kumar R, Mosleh AO, Essa F, Shehabeldeen TA (2021) Deep learning-based forecasting model for covid-19 outbreak in saudi arabia. Process Safety Environ Protect 149:223–233

Gao J, Wang H, Shen H (2019) Task failure prediction in cloud data centers using deep learning. In: 2019 IEEE International Conference on Big Data (Big Data), IEEE. https://doi.org/10.1109/bigdata47900.2019.9006011

Grenfell R, Drew T (2020) Here’s why its taking so long to develop a vaccine for the new coronavirus. Science Alert Archived from the original on 28

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506

Jiang W, Schotten HD (2020) Deep learning for fading channel prediction. IEEE Open J Commun Soc 1:320–332

Kavadi MDP, Patan R, Ramachandran M, Gandomi AH (2020) Partial derivative nonlinear global pandemic machine learning prediction of covid 19. Chaos Solitons Fractals, p 110056

Knight GM, Dharan NJ, Fox GJ, Stennis N, Zwerling A, Khurana R, Kavadi MDP, Patan R, Ramachandran M, Gandomi AH (2020) Partial derivative nonlinear global pandemic machine learning prediction of covid-19. Chaos Solitons Fractals, p 110056

Knight GM, Dhuran NJ, Fox GJ, Stennis N, Zwerling A, Kavadi MDP, Dowdy DW (2016) Bridging the gap between evidence and policy for infectious diseases: how models can aid public health decision-making. Int J Infect Dis 42:17–23. https://doi.org/10.1016/j.ijid.2015.10.024

Li L, Huang T, Wang Y, Wang Z, Liang Y, Huang J, Zhang H, Sun W, Wang Y (2020) Covid-19 patients’ clinical characteristics, discharge rate, and fatality rate meta-analysis. J Med Virol 92(6):577–583

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N et al (2020) Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395(10224):565–574

Momani S, Abo-Hamad ZZ, Alzaidi OM (2016) Solution of inverse kinematics problem using genetic algorithms. Appl Math Inf Sci 10(1):225

WH Organization et al (2020) Covid-19 in suspected human cases: interim guidance, 1 January 2020. WHO

Panch T, Mattie H, Celi LA (2019) The inconvenient truth about AI in healthcare. NPJ Digital Med. https://doi.org/10.1038/s41746-019-0125-3

Pathan RK, Miswas M, Khandaker MU (2020) Time series prediction of covid-19 by mutation rate analysis using recurrent neural network-based lstm model. Chaos Solitons Fractals, p 110018

Perwaiz U, Younas I, Anwar AA (2020) Many-objective bat algorithm. PLoS ONE 15(6):e0234625

Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R (2020) Covid-19 pandemic prediction for hungary: a hybrid machine learning approach. Mathematics 8(6):890

Prasanth S, Singh U, Kumar A, Tikkivala VA, Chong PH (2021) Forecasting spread of covid-19 using google trends: a hybrid gwo-deep learning approach. Chaos Solitons Fractals 142:110356

Rauf HT, Malik S, Shoab U, Irfan MN, Lali MI (2020a) Particle swarm optimization with sugeno-function fuzzy search. Appl Soft Comput 90:106159. https://doi.org/10.1016/j.asoc.2020.106159

Rauf HT, Shoab U, Lali MI, Alhaisoni M, Irfan MN, Khan MA (2020b) Particle swarm optimization with probability function for global optimization. IEEE Access 8:110535–110549. https://doi.org/10.1109/access.2020.3002725

Tomar A, Gupta N (2020) Prediction for the spread of COVID-19 in india and effectiveness of preventive measures. Sci Total Environ 728:138762. https://doi.org/10.1016/j.scitotenv.2020.138762

Tuli S, Tuli S, Tuli R, Gill SS (2020) Predicting the growth and trend of covid-19 pandemic using machine learning and cloud computing. Internet of Things, p 100222

Wang C, Horby PW, Hayden FG, Gao GF (2020) A novel coronavirus outbreak of global health concern. Lancet 395(10223):470–473

WHO (2020) Git-hub live data. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_daily_time_series/time_series_covid19_confirmed_global.csv

Wieczorek M, Silka J, Woźniak M (2020a) Neural network powered COVID-19 spread forecasting model. Chaos Solitons Fractals 131:110203. https://doi.org/10.1016/j.chaos.2020.110203

Wieczorek M, Silka J, Woźniak M (2020b) Neural network powered covid-19 spread forecasting model. Chaos Solitons Fractals, p 110203

Yang J (2020) Inhibition of sars-cov-2 replication by acidizing and rna lyase-modified carbon nanotubes combined with photodynamic thermal effect. J Exploratory Res Pharmacol, pp 1–6

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579(7798):270–273

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.