

Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks

Khondoker Nazmoon Nabi^{a,*}, Md Toki Tahmid^b, Abdur Rafi^b, Muhammad Ehsanul Kader^b, Md. Asif Haider^b

^a Department of Mathematics, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

^b Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

ARTICLE INFO

Keywords:
 Time series analysis
 Deep learning
 Convolutional neural network (CNN)
 Long short term memory (LSTM)

ABSTRACT

Though many countries have already launched COVID-19 mass vaccination programs to control the disease outbreak quickly, numerous countries around worldwide are grappling with unprecedented surges of new COVID-19 cases due to a more contagious and deadly variant of coronavirus. As the number of new cases is skyrocketing, pandemic fatigue and public apathy towards different intervention strategies pose new challenges to government officials to combat the pandemic. Henceforth, it is indispensable for the government officials to understand the future dynamics of COVID-19 flawlessly to develop strategic preparedness and resilient response planning. In light of the above circumstances, probable future outbreak scenarios in Brazil, Russia, and the United kingdom have been sketched in this study with the help of four deep learning models: long short term memory (LSTM), gated recurrent unit (GRU), convolutional neural network (CNN) and multivariate convolutional neural network (MCNN). In our analysis, the CNN algorithm has outperformed other deep learning models in terms of validation accuracy and forecasting consistency. It is unearthed in our study that CNN can provide robust long-term forecasting results in time-series analysis due to its capability of essential features learning, distortion invariance, and temporal dependence learning. However, the prediction accuracy of the LSTM algorithm has been found to be poor as it tries to discover seasonality and periodic intervals from any time-series dataset, which were absent in our studied countries. Our study has highlighted the promising validation of using convolutional neural networks instead of recurrent neural networks when forecasting with very few features and less amount of historical data.

Introduction

As the northern hemisphere moved into winter, many countries have already started grappling with a sudden surge of COVID-19 cases. Though many countries have launched mass vaccination programs with a view to bringing COVID-19 transmissions to a halt, unprecedented spikes of COVID-19 infections are posing new challenges to the respective government officials. Recently, more contagious and potentially more deadly variants have already been detected in different parts of the world [1], and multiple questions have already been raised about the efficacy of vaccines against those emerging deadly variants [2]. As the crisis deepens, prudent experts have already warned about the gruesome nature of imminent upsurges of COVID-19 as the number of new COVID-19 cases is skyrocketing at unprecedented levels. In the face of the another wave of infection, pandemic fatigue and public apathy towards

various intervention strategies have prompted people to act impetuously, imposing new challenges to the government officials. Different non-pharmaceutical intervention strategies such as wearing efficacious face coverings, closure of educational institutions, strict travel restrictions, and strict containment measures are the most powerful intervention strategies to flatten the epidemic curve [3–5]. In addition, mass level testing and tracing program are also indispensable for breaking the continuous transmission chain. Government officials need to ensure facilitated access to affordable rapid tests. With the help of effective case detection, aggressive contact tracing, and local follow-ups and support, the probable risks of imminent surges of COVID-19 in numerous countries can be controlled. Different mathematical and statistical paradigms have performed really well in predicting the epidemic scenario in different countries since the outbreak of COVID-19. Multiple studies [6–9] have provided considerable insights to gain deeper

* Corresponding author.

E-mail address: khnabi@math.buet.ac.bd (K.N. Nabi).

understandings of the transmission dynamics of COVID-19. In addition, numerous studies [10] have enlightened us about the most effective control strategies to battle against this pandemic. In the absence of a widely available COVID-19 vaccine, some promising studies [11,12] have enlightened the importance of different non-pharmaceutical intervention strategies to promote effective public health policies in different worst-hit countries when battling against this pandemic. However, it is often really challenging to incorporate all essential real-life interactions in a single mathematical model [13]. Hence, results are generally based on numerous assumptions and often fail to predict the true outbreak scenario. Nevertheless, a synergy of explicit mathematical models and deep learning networks could be a great way to achieve robust forecasting results in time-series analysis [14,15]. However, this is still an emerging research area that requires more study and experiments. Deep learning algorithms play a vital role in the analysis and prediction of different epidemic outbreak scenarios [14,16,15]. These methods can often provide robust forecasting results, which can be used by government officials to deploy appropriate intervention strategies with a view to quelling the spread of any highly transmissible virus. Wang et al. used the LSTM model in their study [17] to forecast the new COVID-19 cases for 120 days horizon. However, in our study, we have found that the LSTM model has failed to capture the

right trend with promising accuracy. For instance, it has estimated continuous downward trends in Iran, Russia, and Peru. However, it is evident that these countries have already encountered the hit of the second wave of COVID-19. In another study [18], Arora et al. developed a stacked forecasting model named as “Convolutional LSTM” using the concepts of recurrent neural network and convolutional neural network to predict the daily new cases in India. However, that model worked well for a forecasting horizon of only three days. This suggests that the LSTM architecture does not provide considerable insights about long-term time series forecasting with commendable accuracy. Long-term forecasting is undoubtedly a crucial part in terms of policy making and strategic planning. From another point of view, the LSTM algorithm works well with a large dataset, when it is required to find seasonality in time series analysis. However, with relatively fewer data points, it fails to find any pattern and makes the model too sensitive for forecasting. In such a case, a CNN architecture performs well in learning local data patterns and extracting essential features from datasets. In light of the above concerns, a rigorous comparative study of four deep learning algorithms: long short term memory (LSTM), gated recurrent unit (GRU), convolutional neural network (CNN), and multivariate convolutional neural network (MCNN) has been presented to forecast the number of daily new COVID-19 cases in Brazil, Russia, and the UK. Mean absolute percentage error (MAPE) and normalised root mean square error (nRMSE) have been used as performance metrics to evaluate the

Table 1
Model parameters

Method	Parameter	Values
LSTM/GRU	Number of layers	2
LSTM/GRU	Number of units	{50, 20}
LSTM/GRU	Epochs	1000
LSTM/GRU	Batch size	300
LSTM/GRU	Learning rate	0.001
LSTM/GRU	Optimizer	Adam
LSTM/GRU	Loss function	MSE
CNN/Multivariate CNN	Number of Convolutional layers	4
CNN/Multivariate CNN	Number of Units in CNN layers	{32, 64, 128, 256}
CNN/Multivariate CNN	Kernel size	3×3
CNN/Multivariate CNN	Number of Dense layers	2
CNN/Multivariate CNN	Number of Units in Dense Layers	{10, 1}
CNN/Multivariate CNN	Epochs	1000
CNN/Multivariate CNN	Batch size	300
CNN/Multivariate CNN	Learning rate	0.001
CNN/Multivariate CNN	Optimizer	Adam
CNN/Multivariate CNN	Loss function	MSE

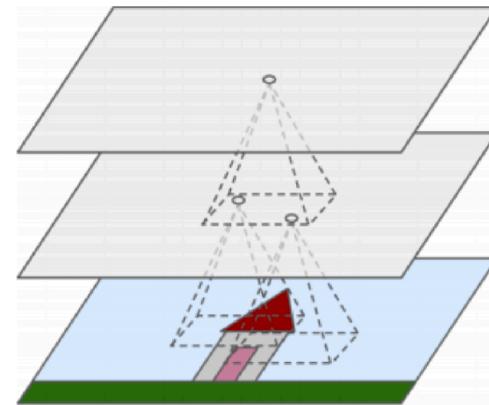


Fig. 2. The concept of Convolutional Neural Network: Shrinking down data dimensions to extract most relevant features.

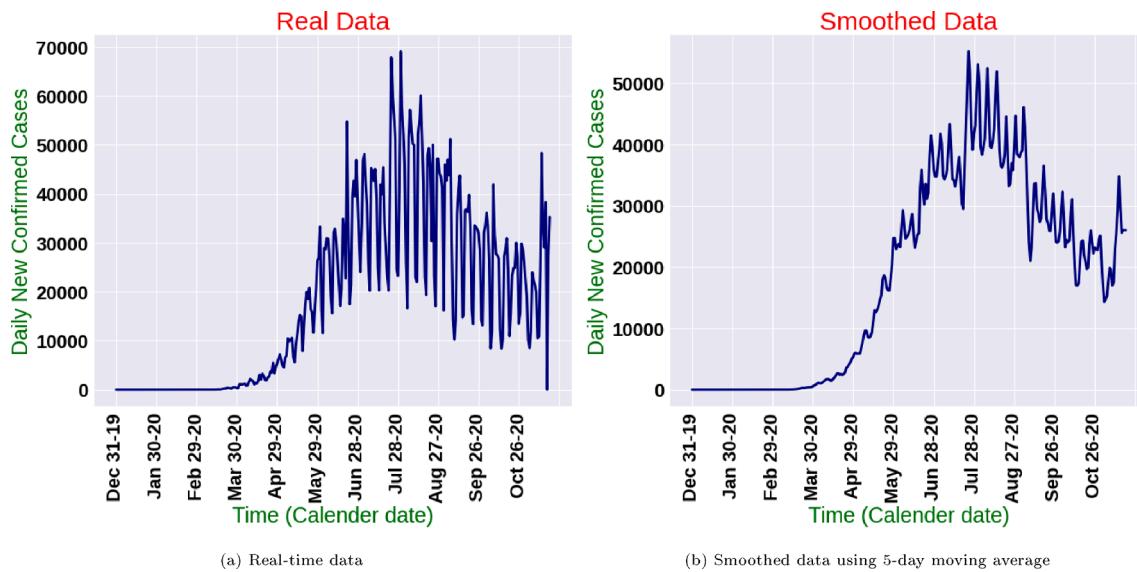


Fig. 1. Comparative scenario analysis between real-time data and smoothed real-time data.

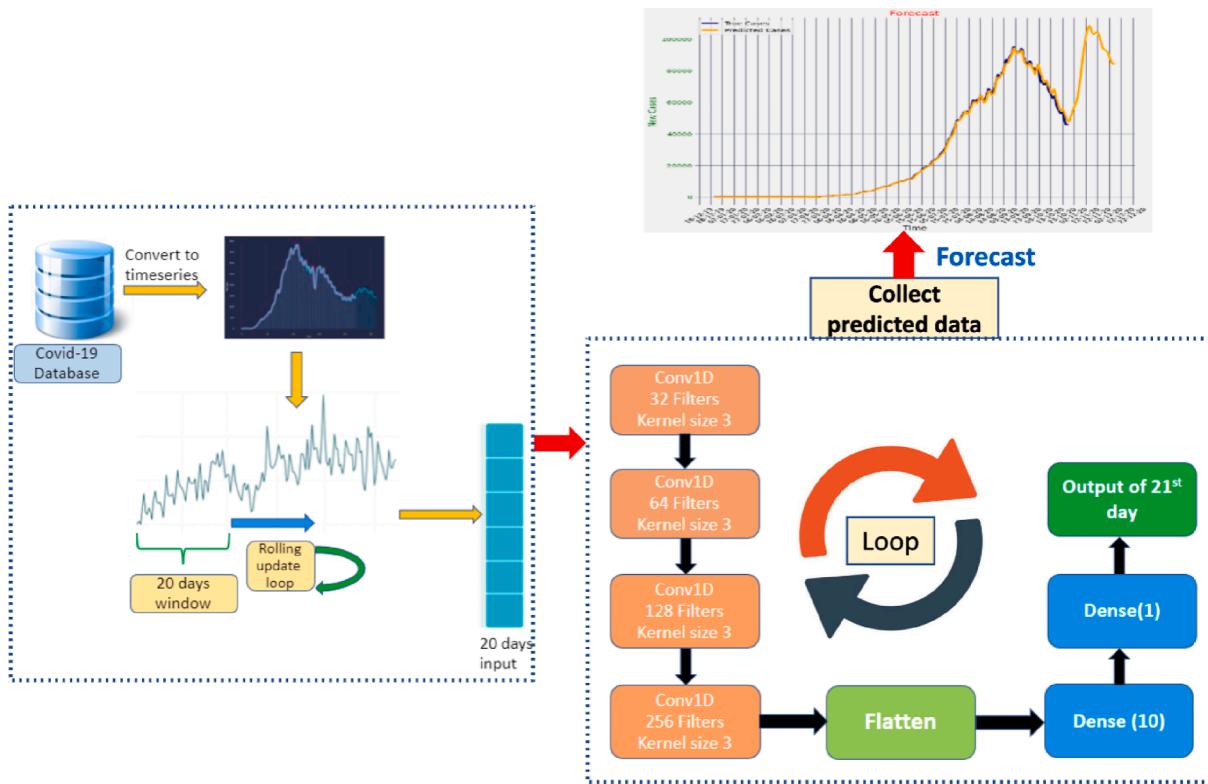


Fig. 3. Convolutional Neural Network architecture used in this study.

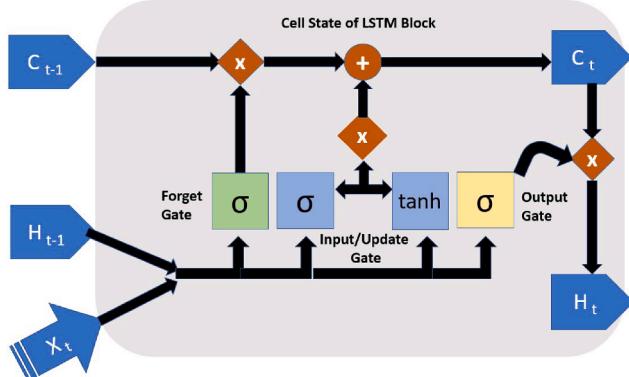


Fig. 4. Internal architecture of LSTM network.

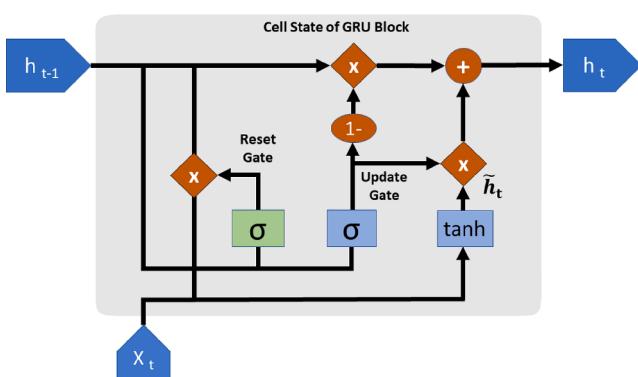


Fig. 5. Internal architecture of GRU network.

robustness of model forecasting results. Results based on these metrics suggest that CNN performs better than all other deep learning models. Considering all the scenarios, CNN has turned out to be the best forecasting model due to its power of feature learning. Two deep learning methods: CNN and multivariate CNN, have been proposed for forecasting as both methods have shown promising consistency and prediction accuracy in terms of long-term forecasting. Importantly, a strong probability of forthcoming second wave of COVID-19 has been unearthed in the above-mentioned countries. Robust planning and strategic management can be ensured in these countries based on our forecasting results to avoid disasters in healthcare systems. The entire chapter is organized as follows. Materials and methods are presented in Section "Materials and methods". In Section "Results and discussion", forecasting results have been discussed using real-time COVID-19 data

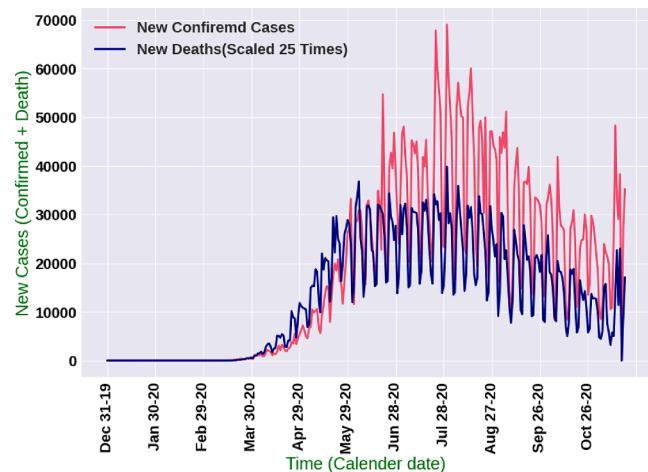
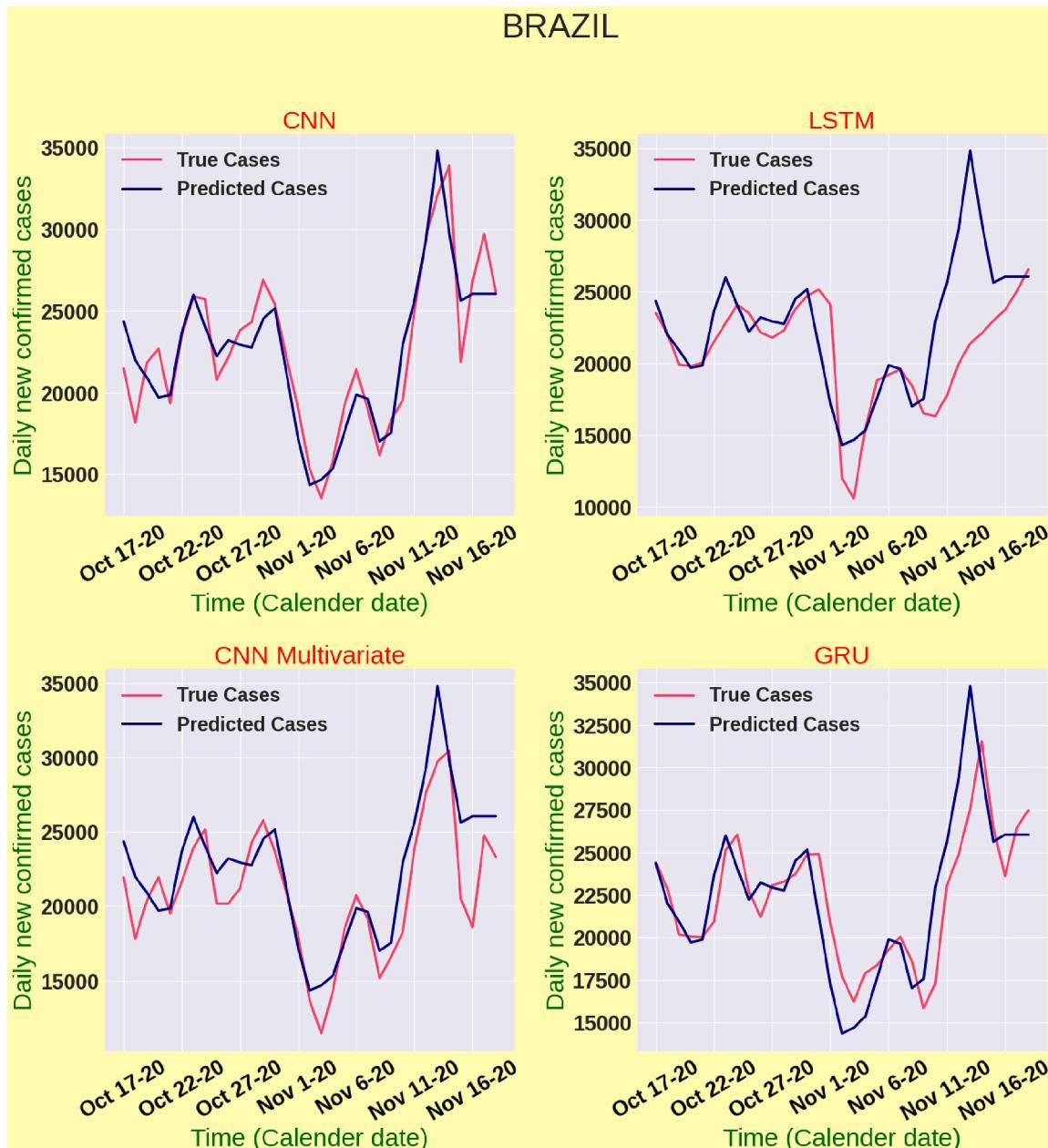


Fig. 6. Correlation between new cases and new deaths. Here the new deaths data has been scaled 25 times to visualize the trends more clearly.

Table 2

Values of evaluation metric from various deep learning models on different countries.

Country	Deep learning model	Evaluation metric values			
		$nRMSE_{train}$	$nRMSE_{val}$	$MAPE_{train}$	$MAPE_{val}$
Brazil	LSTM	0.16	0.197	14.92%	13.33%
	GRU	0.103	0.124	8.39%	10.267%
	CNN	0.054	0.086	8.20%	6.94%
	Multivariate CNN	0.068	0.109	16.73%	8.21%
Russia	LSTM	0.203	0.05	14.03%	4.64%
	GRU	0.05	0.01	5.76%	1.199%
	CNN	0.06	0.014	6.61%	0.85%
	Multivariate CNN	0.05	0.01	4.79%	0.86%
United Kingdom	LSTM	0.26	0.065	16.36%	5.25%
	GRU	0.11	0.036	7.0%	2.997%
	CNN	0.12	0.048	12.0%	3.75%
	Multivariate CNN	0.15	0.094	14.48%	7.797%

**Fig. 8.** Model validation for Brazil.

of Brazil, Russia, and the UK. The paper ends with some insightful experimental results found in this study.

Materials and methods

In this paper, deep learning models (CNN, LSTM, GRU, multivariate CNN) are used to forecast new cases using the COVID-19 dataset for three countries, including Brazil, Russia, and the United Kingdom. For Brazil, Russia, and the United Kingdom data of 291 days are used for training, and data of 33 days are used for validation. The forecast is made for 40 days. The performance of the models is evaluated using MSE, nRMSE, and MAPE. Models are implemented and trained using the Keras module of TensorFlow on the Google Collaboratory platform. Recurrent and convolutional neural networks (CNN) are verified by numerous studies to be used in the public health sector, medical data analysis, and biological investigation. In this section, we discuss some of those advancements. Forecasting plays an essential role in the public

health sector and decision-making. Traditional data analysis, and forecasting models have been widely used in predicting COVID-19 cases. Along with traditional machine learning models, deep Learning techniques have shown colossal prospects because of their automatic feature extraction ability. Recurrent neural networks (RNN) have been widely used for epidemic forecasting [19], weather condition predictions [20], predicting the stock market [21], and other time series analysis tasks. As recurrent neural networks primarily deal with time-series data and can extract features from previous data, it provides a long-term dependency. Over the years, modifications have been made in the architecture of RNN networks. The two most common recurrent neural networks are long short term memory (LSTM) and gated recurrent unit (GRU). Both of these networks are used in forecasting and analyzing time-series data. LSTM is used to compare the time series trends of COVID-19 between India and the USA in [16]. Not only in time series forecasting, but LSTM networks have also been used to classify ECG signals [22], emotion recognition from brain waves [23], localization of subcellular protein

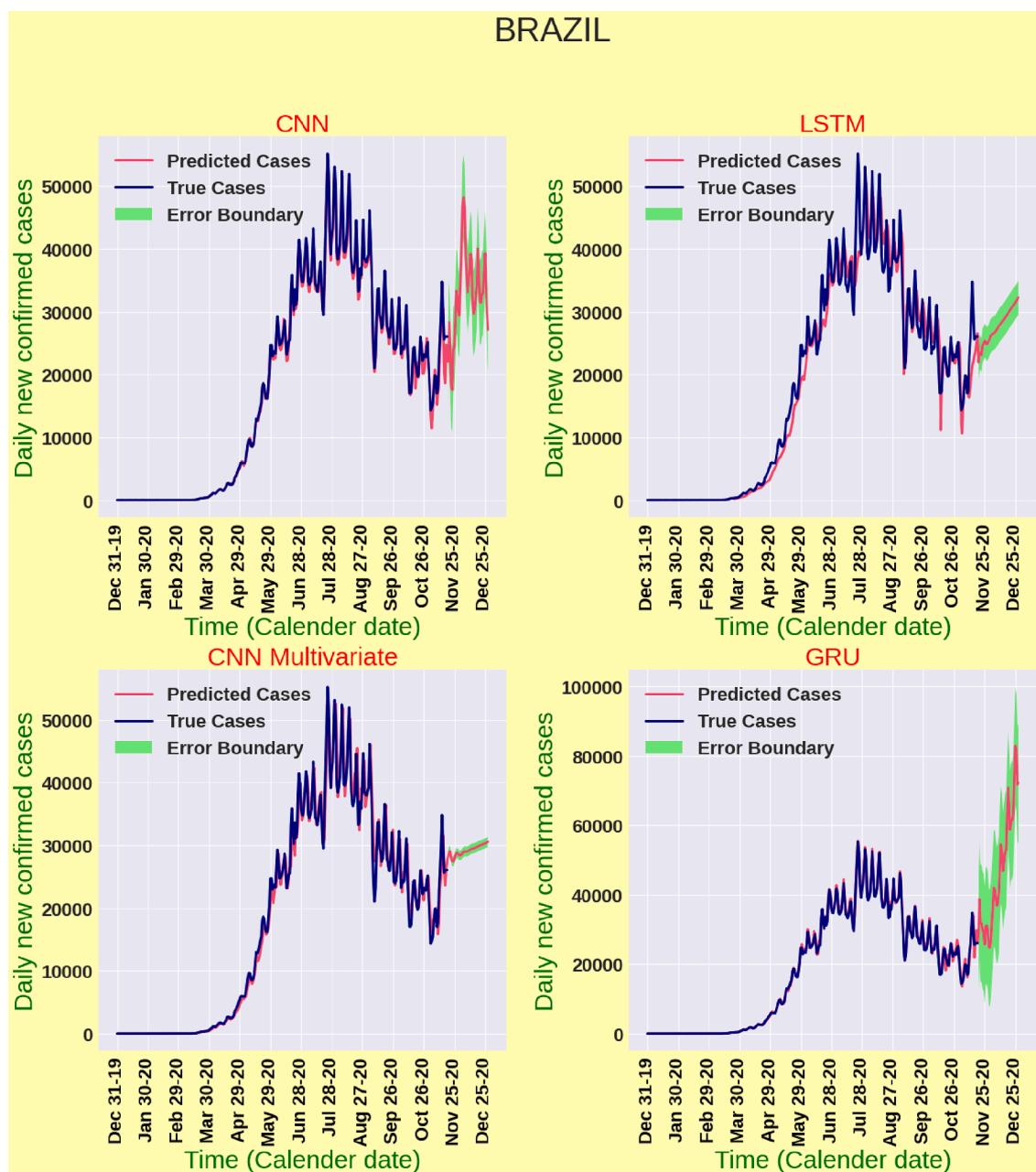


Fig. 9. Forecasting results for daily new cases in Brazil.

from amino acid sequences [24]. Convolutional neural network (CNN), which was proposed primarily to handle and analyze image data, has also been used in time series analysis. With its ability to focus on local data patterns, CNN has shown promising results in time series trend analysis and forecasting. CNN is used in [25] to forecast COVID-19 cases in China, and it is shown that the CNN model has outperformed other Deep learning models. Temporal CNN, a specialized form of CNN, can work with sequence data, auto-regressive prediction, and very long-term memory has been used in [26] for multivariate time series analysis. Convolutional neural networks have revolutionized the field of genetic sequence analysis. It is widely used in protein–protein interaction analysis [27], protein's secondary structure prediction [28], and localization [29].

Data sources

We have collected the data up to November 18, 2020 for Brazil, Russia, and the UK from a trusted online repository developed and maintained by Johns Hopkins University [30]. We have extracted new cases and new deaths data for the countries mentioned above using that repository.

Methods

In this study, specialized four deep learning algorithms: convolutional neural network (CNN), long short term memory (LSTM), gated recurrent unit (GRU), and multivariate convolutional neural network (MCNN) have been applied to understand the future transmission dynamics of COVID-19. Among the mentioned four techniques, LSTM and GRU belong to the recurrent neural networks family, whereas CNN and multivariate CNN belong to the convolutional neural networks family. As the COVID-19 dataset fluctuates so frequently, the moving average method has been used to smoothen our dataset. Dataset for each country was considered as a time series. Our prediction is performed on the moving averaged data. We used a rolling mechanism [17] by incorporating windowed data in the time series. By the trial-and-error method, we have chosen the best value of window and assigned each country with corresponding best window size. To be precise, a window size of 3 days has been used for Russia. Due to the high volatility in Brazil's dataset, a window size of 10 days has been used. Moreover, a window size of 5 days has been taken for the UK. For COVID-19 time series prediction, recurrent neural networks and their sophisticated variations named LSTM and GRU are widely used. However, the use of the convolutional neural networks (CNN) in time series analysis is not widespread [25]. In this study, CNN performed very well on both training and validation data as we hyper-tuned the parameter values. It turns out that CNN worked better than LSTM and GRU on both training and validation data. For forecasting with MCNN, daily confirmed cases and daily deaths have been used to predict the number of newly confirmed cases for the next day. To the best of our knowledge, multivariate convolutional neural network time series analysis is not used before in any study for COVID-19 forecasting. We have listed the parameter values for each model in Table 1. We have used root mean squared error (RMSE) as our loss function during the training time. Normalized root mean squared error (nRMSE) and mean average percentage error (MAPE) are used as insightful metrics to evaluate model prediction accuracy and to compare performance between models. We have also incorporated an error boundary for our forecasting zone, where we demonstrated the possible standard deviations from our predicted data.

Moving average

The Moving average is one of the most used techniques for understanding trends in time series analysis. It is a way to smooth out fluctuations in time series data and to help distinguish between noise and trends in a dataset. Often in spite of having clear trends and patterns, there might be huge fluctuations found in the dataset, which may cause

any forecasting model to perform poorly. In that case, the value of any day is replaced by the average of some of its neighboring values. Mathematically, a simple moving average is the mean of the previous n data:

$$P(n) = \frac{1}{k} \sum_{i=n-k}^{n-1} P(i)$$

, where k is the previous number of days for determining the i^{th} value. To ensure that the variation in mean corresponds to the variation of data, we have taken the mean of the values of any particular date, the previous two days and the next two days to calculate the moving averaged value of that day Fig. 1. Mathematically,

$$P(n) = \frac{P(n-2) + P(n-1) + P(n) + P(n+1) + P(n+2)}{5}$$

Here, $P(n)$ is the value for n^{th} day.

Deep learning methods

Convolutional neural network (CNN)

Convolutional neural network (CNN) is a more generalized version of multilayer perceptron [31]. However, the CNN architecture causes overfitting in the model. Each neuron in the one convolutional layer is connected only to neurons located within a small rectangle in the previous layer. This architecture allows the network to concentrate on low-level features in the first hidden layer Fig. 2, then assemble them into higher-level features in the next hidden layer, and so on. To use the CNN model in time series forecasting Fig. 3, four convolutional layers are stacked to extract complex feature and patterns from the time series. Finally, one flatten layer and two dense layers are used to produce the necessary output.

Long short term memory (LSTM)

As traditional artificial neural networks fail to provide the desired accuracy while working with sequential data, recurrent neural networks (RNN) or neural networks with loops come into play by maintaining a vector of activation for each time-step and carrying them forward Fig. 4. Although RNNs are much better at handling sequential data by connecting previous information to the present inputs, often there might be problems while training deeper models. Due to RNN's repetitive nature, parameter matrices related to calculating the gradients, tend to shrink

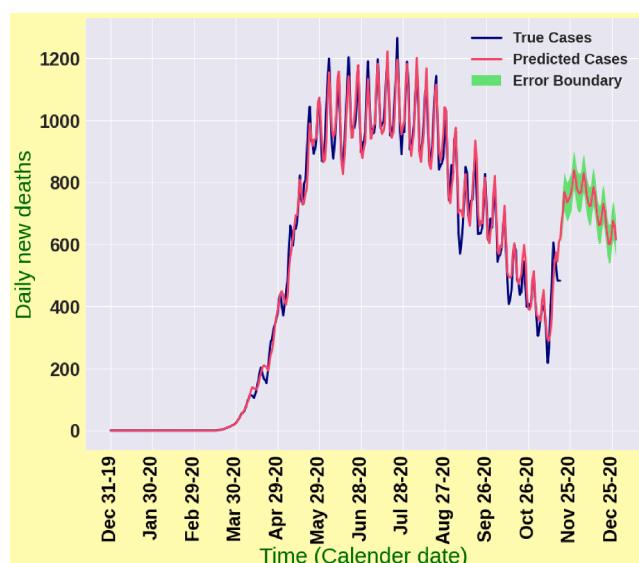


Fig. 7. Forecasting results for daily death cases in Brazil.

down exponentially during backpropagation, thus making the model training process much slower and less effective. This phenomenon is known as the Vanishing Gradient problem. As a result, typical RNNs are found to be a poor option while practically dealing with this long-term dependency. The LSTM or Long Short-Term Memory is a well-known variant of traditional RNNs, that addresses this issue by effectively tackling long-term dependencies. Structurally, each LSTM block in the network module consists of a set of carefully designed vectorized operations between new inputs and previous outputs, along with the usage of mathematical functions like sigmoid and tanh where necessary, instead of a single layer neural network. In short, the LSTM block contains a memory cell and three operational gates with their own weights and bias vectors through which information passes serially. First of all, a combination of current block input and previous activation values are passed through a nonlinear sigmoid function to filter out which values to store and which values to forget or erase out, as sigmoid function outputs just values in between 0 and 1. This layer is known as the Forget

Gate. Later on, another set of parameters in the Input/Update Gate are used to filter out the same combination of activation from the previous layer and current block input, eventually to be used as the updated input. Then, the primary set of input is passed through a tanh (hyperbolic tangent) function to produce new candidate values of the current block. Finally, the linear Vector-Matrix addition of the values from the previous block, scaled by Forget Gate, and the current candidate values scaled by the Input/Update Gate is served as the final values of the current block. Third, the final values of the current block are passed through another tanh activation layer and then scaled by an Output Gate, consisting of another parameterized version of the initial combination of block input and previous activations to produce the newly updated activations of the current block. Thus, using sigmoid activations to decide what to forget and what to store in each block with respect to the previous blocks, the LSTM blocks can increase accuracy significantly. Mathematically,

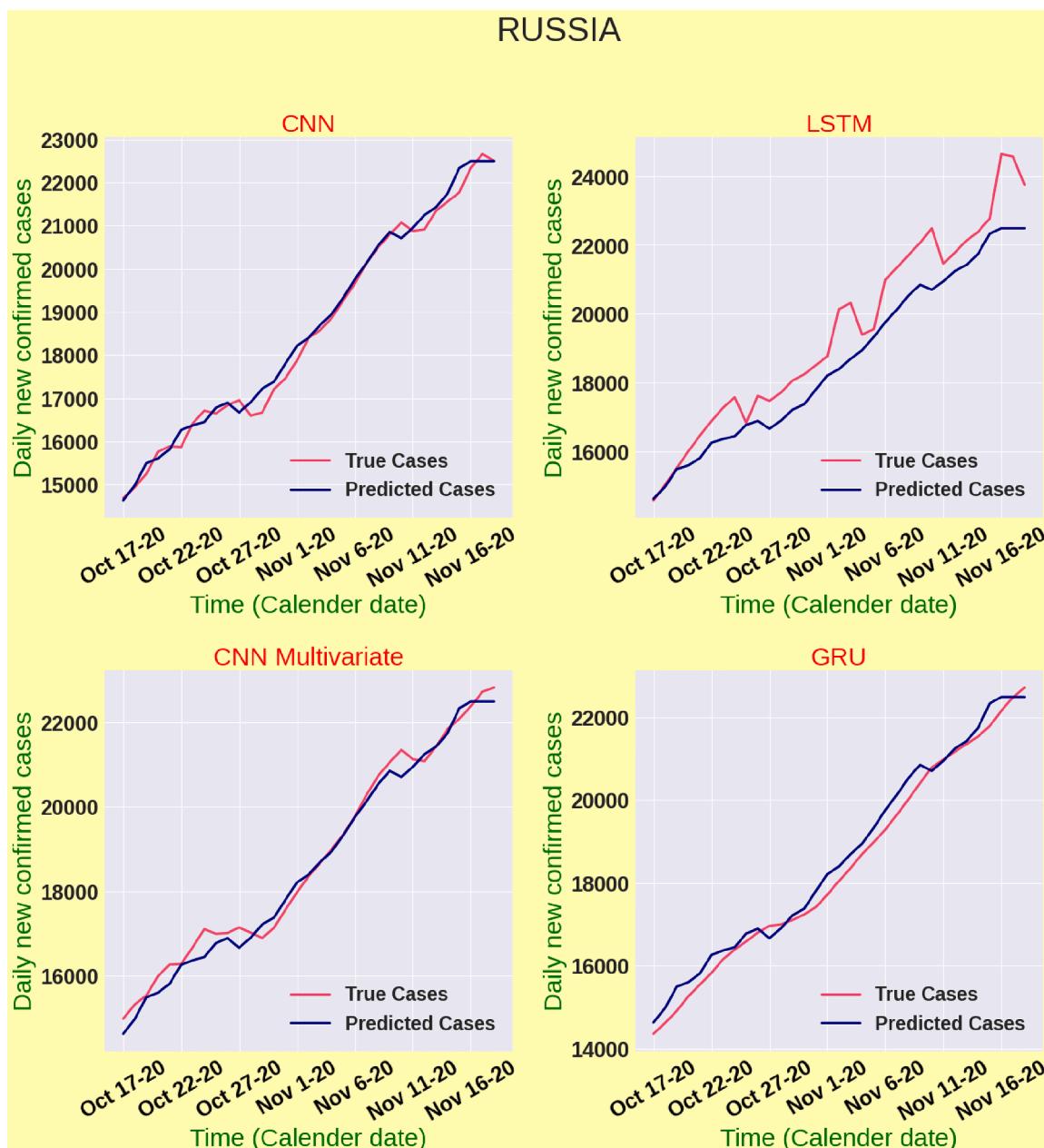


Fig. 10. Model validation for Russia.

$$F_t = \sigma(W_{fx}X_t + W_{fh}H_{t-1} + B_f) \quad (\text{ForgetGate}) \quad (1)$$

$$I_t = \sigma(W_{ix}X_t + W_{ih}H_{t-1} + B_i) \quad (\text{InputGate}) \quad (2)$$

$$O_t = \sigma(W_{ox}X_t + W_{oh}H_{t-1} + B_o) \quad (\text{Outputgate}) \quad (3)$$

$$\tilde{C}_t = \tanh(W_cX_t + W_cH_{t-1} + B_c) \quad (\text{IntermediateCellStatewithCandidateValue}) \quad (4)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (\text{CellStateorFinalCurrentBlockValue}) \quad (5)$$

$$H_t = O_t \odot \tanh(C_t) \quad (\text{NewStateorNewBlockInput}) \quad (6)$$

Where, W and B are relevant weights and biases in respective gates associated with each LSTM block, and X is the input vector for each block.

Gated recurrent unit (GRU)

GRU or Gated Recurrent Unit can be portrayed as a variant of LSTM, and it has many similarities with LSTM [32]. Like the LSTM, GRU is also primarily used to solve the “Vanishing Gradient” problem found in typical RNNs and thus improving the learning of long-term dependencies in the network. GRU blocks also use tanh and sigmoid functions to compute necessary values. But unlike the LSTM block, a GRU block does not have separate memory cells. This type of block also does not have a separate Forget gate and the Input/Update gate is responsible for controlling the flow of information. Because of these two structural differences with LSTM, it has fewer parameters, and the design of it is less complex, which eventually makes it more computationally efficient and easier to train. tanh and sigmoid functions are used inside to compute necessary values. Other than Update gate, GRU blocks have a Reset gate. Inside a GRU block, four values are computed, these are - Update gate, Reset gate, candidate activation, and output activation Fig. 5. Each gate and candidate activation have their separate weights and biases. Along with their weights and biases, the current block input and previous activation values are used as input to compute these three values. In the first step, sigmoid function is used to calculate the values of the gates.

$$U_t = \sigma(W_{ux}X_t + W_{uh}H_{t-1} + B_u) \quad (\text{UpdateGate}) \quad (7)$$

$$R_t = \sigma(W_{rx}X_t + W_{rh}H_{t-1} + B_r) \quad (\text{ResetGate}) \quad (8)$$

$$\tilde{H}_t = \tanh(W_{hx}X_t + W_{hh}(R_t \odot H_{t-1}) + B_h) \quad (\text{CandidateActivation}) \quad (9)$$

$$H_t = U_t \odot \tilde{H}_t + (1 - U_t) \odot \tilde{H}_{t-1} \quad (\text{OutputActivation}) \quad (10)$$

Then, candidate activation is calculated by applying tanh function to a combination of input of Reset gate value and previously mentioned inputs. Finally, the linear vector addition of the values of candidate activation, scaled by Update gate and previous activation values scaled by the result of subtraction between vector of ones and Update gate produces the values of output activation.

Multivariate convolutional neural network

In multivariate time series analysis, multiple features are provided as an input for any particular timestep [33]. Along with the targeted feature, this method also predicts the values for other features used as an input. In this study, we have used daily confirmed COVID-19 cases and daily confirmed COVID-19 deaths as our multivariate features as we have observed a strong correlation between their trends Fig. 6. LSTM networks have previously been used for COVID-19 prediction using multivariate time series analysis [34]. However, we are the first to implement CNN models for multivariate COVID-19 forecasting, which provided us with a promising result Table 2. Normalized root mean

squared error (nRMSE) and mean absolute percentage error (MAPE) have been implemented to understand the relative performance of different deep learning models. Following are the formulas to determine the values of these metrics:

$$\text{nRMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (11)$$

$$\text{MAPE} = \frac{\sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i}}{N} \times 100 \% \quad (12)$$

Here, y_i are the actual values, \hat{y}_i are the values predicted by the model and N is the number of observations. In the nRMSE formula, \bar{y} is the mean of all the predicted values. As the difference between actual values and predicted values are squared in nRMSE and absolute values are taken in MAPE, the problem of positive and negative errors canceling each other out is avoided.

Results and discussion

In this analysis, we have used four deep learning architectures to forecast the new cases of Brazil, Russia, and the United Kingdom for the next 40 days. We have summarized the values of normalized root mean squared error train ($nRMSE_{train}$), normalized root mean squared error validation ($nRMSE_{val}$), mean average percentage error train ($MAPE_{train}$), and mean average percentage error validation ($MAPE_{val}$) in Table 2. Along with new COVID-19 cases, we have also implemented our best validated CNN model to forecast new death cases for the above countries with a forecasting horizon of 40 days. In our study, we have found that CNN performed the best in terms of consistency among all four deep learning models. In this section, study findings are summarized below.

Due to the highly inconsistent daily test numbers, it has been found that Brazil's daily data is highly volatile and noisy. To address this issue, the moving average technique has worked quite well to capture the real trends illustrated in Fig. 1. Although Brazil had already passed its peak of the first wave with approximately 69,000 daily cases, our analysis is showing that it headed towards another wave. It is evident from Fig. 8 that our CNN model has the lowest metrics value with $nRMSE_{val} = 0.086$ and $MAPE_{val} = 6.94\%$. In addition, the model projected that the country had witnessed another wave by December with a tentative peak of 80,000 new cases on a daily basis illustrated in Fig. 9. Multivariate CNN

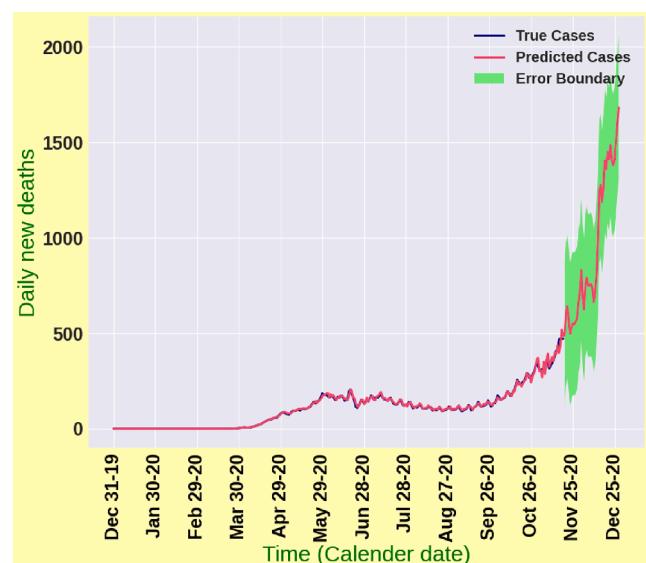


Fig. 12. Forecasting results for daily death cases in Russia.

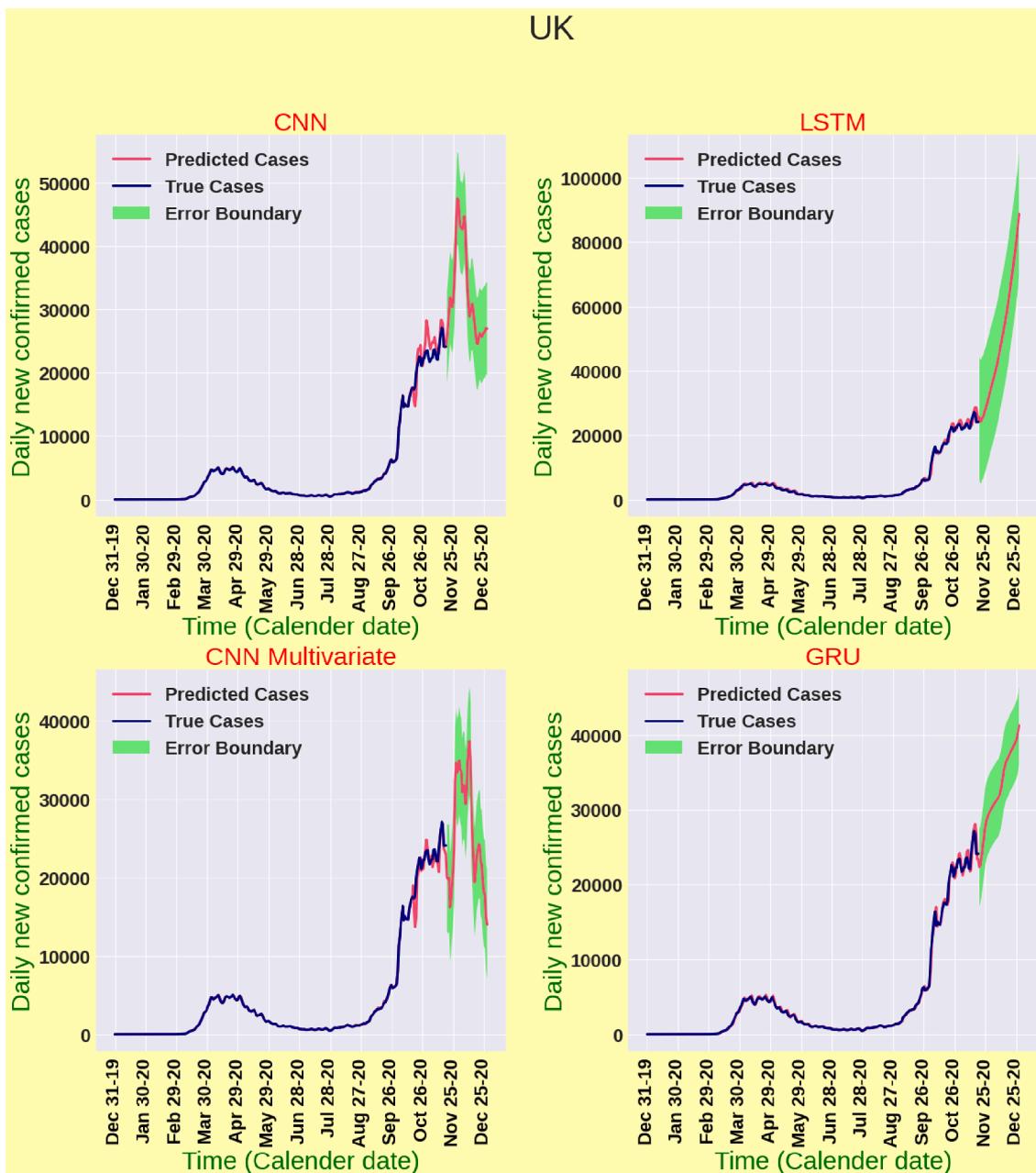


Fig. 15. Forecasting results for daily new cases in the United Kingdom.

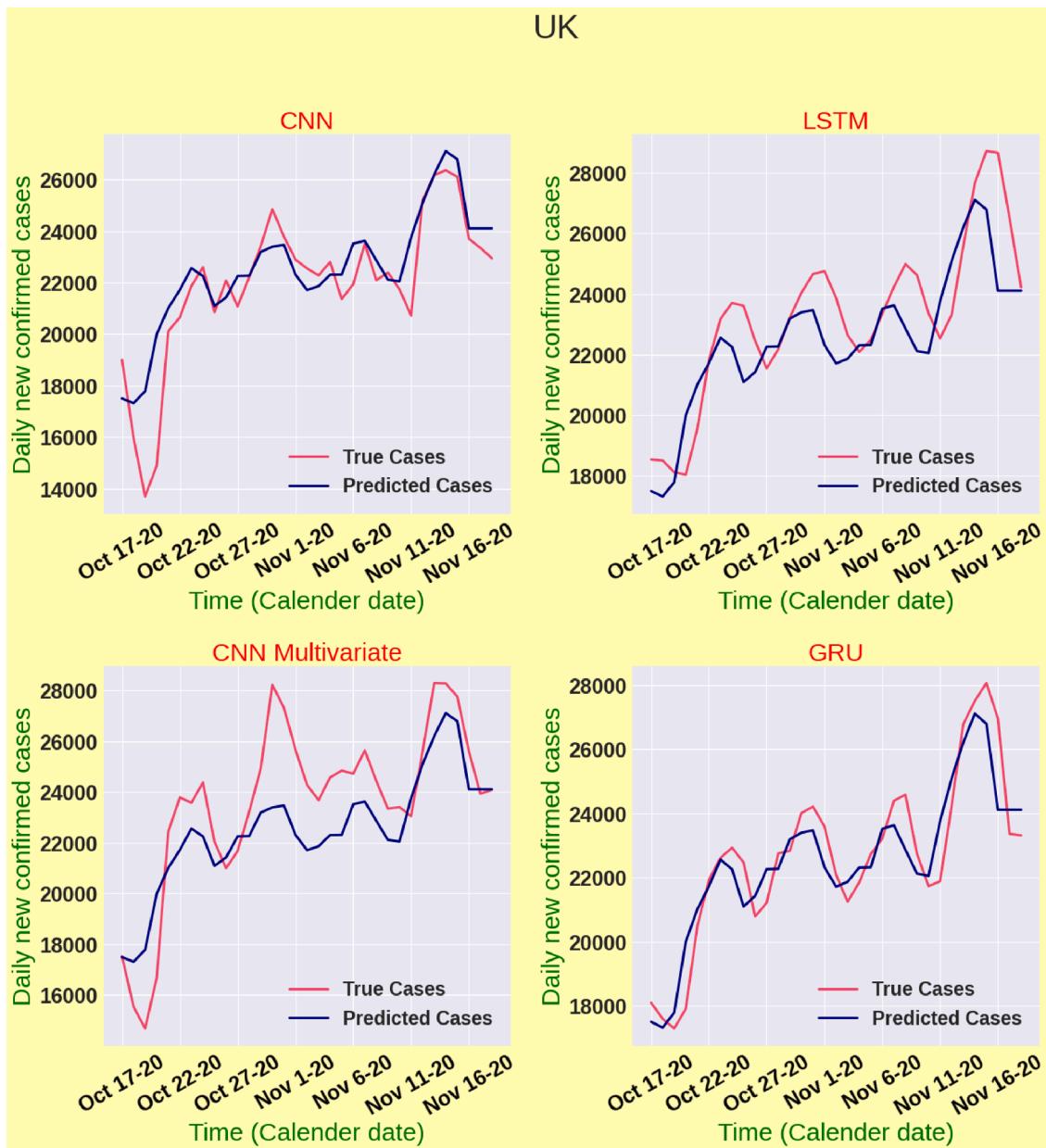


Fig. 14. Model validation for the United Kingdom.

model was the second-best performer in validation data with $nRMSE_{val} = 0.109$ and $MAPE_{val} = 8.2\%$ and it also suggests that another wave of COVID-19 daily cases is imminent. It also predicts that the number of cases might decrease soon after reaching the peak of the second wave. Two other models: LSTM ($nRMSE_{val} = 0.2$ and $MAPE_{val} = 13.33\%$) and GRU ($nRMSE_{val} = 0.12$ and $MAPE_{val} = 10.267\%$) have also predicted that the number of new cases might increase without showing lucid indication of impending second wave. Moreover, Fig. 7 shows that the tally of daily deaths in Brazil is going to increase gradually until early December, although the intensity might not be as high as the first wave with a peak of around 800 new deaths per day.

Russia has already witnessed the first wave of COVID-19. Among the four models, CNN ($nRMSE_{val} = 0.01$, $MAPE_{val} = 0.85\%$) performs the

best in terms of validation data for Russia depicted in Fig. 10. It predicts that the number of daily cases might mount to nearly 30,000 by late December. Multivariate CNN ($nRMSE_{val} = 0.01$, $MAPE_{val} = 0.86\%$) shows that the cases might mount to 25,000 and then might follow a downward trend. In contrast, LSTM ($nRMSE_{val} = 0.05$ and $MAPE_{val} = 4.64\%$) and GRU ($nRMSE_{val} = 0.014$ and $MAPE_{val} = 1.2\%$) are indicating a downturn of cases. Fig. 12 enlightens that the number of daily deaths could skyrocket alarmingly and reach 1500 on a daily basis by late December if this trend continues. It is often really challenging to capture the outbreak scenario as it is highly dynamic and depends on deploying various intervention strategies.

The United Kingdom (UK) has already witnessed two waves of COVID-19 illustrated in Fig. 15. In the first wave, the transmission chain

was quite well-controlled with well-implemented and maintained lockdown policies. Moreover, the situation was comparatively better than most of the European countries until September. However, due to restriction fatigue and public apathy towards lockdown, the country is grappling with new surges of cases after September and already witnessed a recent peak of about 33,000 daily cases. It has been found in our analysis that the number of daily cases may hold an upward trend in upcoming days. Fig. 14 enlightens the fact that, the CNN model achieved an $nRMSE_{val}$ score of 0.04809 and $MAPE_{val}$ score of 3.75%, while the Multivariate CNN model shows $nRMSE_{val}=0.09$ and $MAPE_{val}=7.8\%$. On the other hand, $nRMSE_{val}=0.06481$ and $MAPE_{val}=5.25\%$ have been found for the LSTM model. Finally, the GRU model hits $nRMSE_{val}=0.04$ and $MAPE_{val}=2.997\%$ metric. Comparing with the other four models, it is found that the GRU model best validates the data with the lowest percentage error. The CNN model also performs quite well in this case. However, the CNN and the multivariate CNN models have also projected a downtrend of cases soon after the peak is reached. According to our model projection illustrated in Fig. 13, the number of daily deaths is going to escalate by overshooting the intensity of prior waves. see Fig. 11.

We used data up to 18 November 2020, to train and test the model devised for our analysis. The motivation behind selecting this time period lies in the fact that many countries of the world started to get hit by the second wave of COVID-19. We wished to analyze and compare the capability of recurrent and convolutional neural networks to interpret the complex trend in time series that appeared at that time period by measuring our results with a set of validation data of 1 month and finally, forecasting new cases of the next 40 days. We have shown that convolutional neural networks are more promising in keeping the forecasting consistent, when the trend might show an intricate upward tendency. According to real-time data [30] in Brazil, three days moving average around 24 December, 2020 was 53,790 complementing our prediction by the CNN model within our error boundary. According to the same reference, the new death count in Brazil during 25 December, 2020, with three days moving average was around 750 per day. Our prediction for the death rate in that course of time agrees with the real data with a tentative number between 600 to 700 deaths per day. This shows the aptitude of our CNN model to project the trend for a long period with explicit consistency. In Russia, new confirmed cases jumped

up to 28,000 per day by December 25, 2020. The proposed CNN model also forecasts cases around 30,000 by that time, which confirms the robustness of the network to be applied for a horizon of a long period. The forecasting for death cases did not work as well as the other predictions by our model. This issue can be well visualized by looking at the error boundary of prediction for Russia. The non-uniformity of death cases and the sudden rise of numbers within a very short period, reduced the sensitivity and accuracy of prediction to a large extent, which resulted in a wide-ranging error boundary and also failed to properly project the death cases for a long period. When the data changes rapidly within a short period of time, other factors must be taken into account along with daily cases in order to properly forecast the real scenario. The UK has experienced a continual uprise in daily cases and 3 days moving average around 24 December was 36,999. Our CNN model predicts this number between 35,000 to 40,000 within the error boundary. Death rate predicted by our model depicts that the number should lie between 600 to 750 at the end point of our forecasting horizon. According to real-time data [30], the death toll was 631 with a moving average of 3 days. All these results suggest that CNN models are pretty helpful in case of long-term forecasting. These results might help further investigations of the imminent rise of COVID-19 cases due to the new variant, as the studied model can handle complex patterns that appeared at the onset of the previous second wave. It is worth mentioning that researchers have used the LSTM and its variants overwhelmingly, with a view to forecasting COVID-19 cases in an early outbreak scenario [18,19,35]. Nevertheless, our analysis enlightens that CNN and multivariate CNN algorithms have performed really well compared to the LSTM algorithm. A probable reason for this might be explained from [36]. Our dataset lacks a large amount of data, whereas sufficient amount of data is preferred for the LSTM algorithm to capture the essential features in time series data. For instance, the performance of the LSTM is really laudable in terms of dealing with stock price data while historical data is available [37]. On the other hand, CNN fundamentally aims at focusing on local features with more attention [38]. This algorithm has successfully unearthed the complex nonlinear dependencies regarding epidemic transmission. The number of daily death cases in Brazil, Russia, and the UK has been projected with a forecasting horizon of 40 days with our best-validated CNN model.

Conclusions

In this paper, four deep learning architectures: LSTM, GRU, CNN, and multivariate CNN have been implemented with a view to forecasting new COVID-19 cases and new deaths in Brazil, Russia, and the UK until late December 2020. We have used MSE as our loss function and used MAPE and nRMSE as the performance indicators for the studied deep learning models. Among the four deep learning models, CNN has performed extremely well in terms of validation accuracy and forecasting consistency. Moreover, the concept of multivariate CNN architecture has been implemented for the first time in this study, which showed robust forecasting performance for the studied countries. CNN algorithm has been proposed for long-term forecasting in the absence of seasonality and periodic patterns in time-series datasets. Moreover, failure of the LSTM architecture has also been disclosed in analysing such datasets. We have shown that the cases predicted by the CNN model after the 40 days horizon promisingly match the real-time scenario within our error boundary. Due to pandemic fatigue and public apathy towards different useful non-pharmaceutical intervention strategies, dreadful upsurges in daily COVID-19 cases could be witnessed in aforesaid countries. Public health officials in those countries should act immediately and deploy stringent intervention strategies to battle strategically against the COVID-19 pandemic.

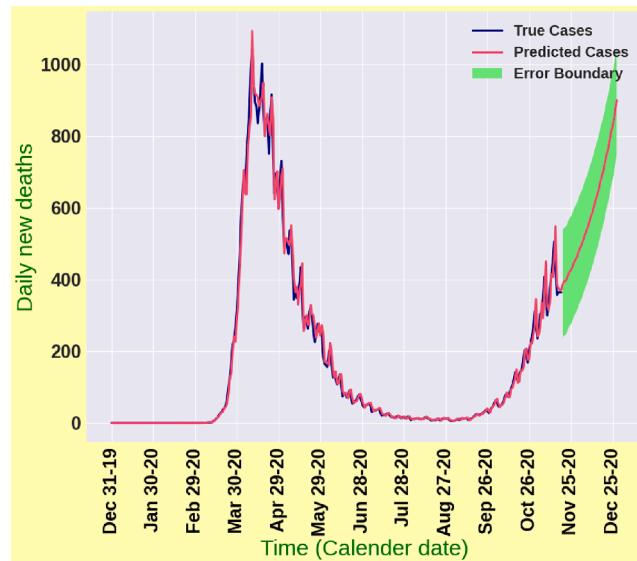


Fig. 13. Forecasting results for daily death cases in the United Kingdom.

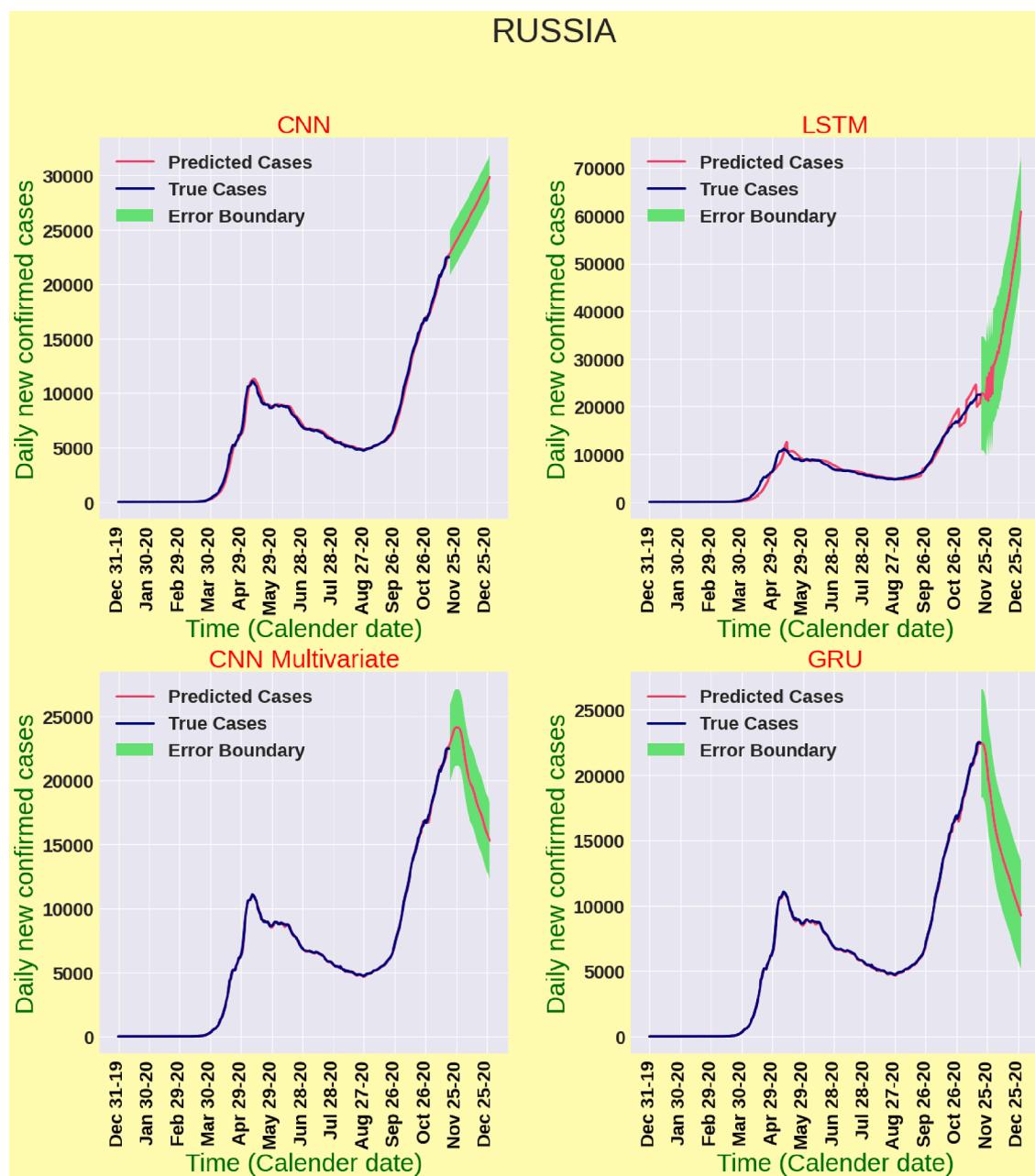


Fig. 11. Forecasting results for daily new cases in Russia.

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.

References

- [1] Wise J. Covid-19: new coronavirus variant is identified in uk. *BMJ* 2020;371. <https://doi.org/10.1136/bmj.m4857>.
- [2] Madhi SA, Baillie V, Cutland, et al. Efficacy of the chadox1 ncov-19 covid-19 vaccine against the b.1.351 variant. *New Engl J Med* 2021. <https://doi.org/10.1056/NEJMoa2102214>.
- [3] Eikenberry SE, Mancuso M, Iboi E, Pham T, Eikenberry K, Kuang Y, Kostelich E, Gumel AB. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the covid-19 pandemic. *Infect Disease Model* 2020;5: 293–308. <https://doi.org/10.1016/j.idm.2020.04.001>.
- [4] Nabi KN, Kumar P, Erturk VS. Projections and fractional dynamics of covid-19 with optimal control strategies. *Chaos Solitons Fractals* 2021;145:110689. <https://doi.org/10.1016/j.chaos.2021.110689>.
- [5] Nabi KN. *Epidemic prediction and analysis of COVID-19: a mathematical modelling study*. Springer; 2021.
- [6] Kucharski AJ, Klepac P, Conlan, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of sars-cov-2 in different settings: a mathematical modelling study. *Lancet Infect Dis* 20(10); 2020: 1151–1160. [https://doi.org/10.1016/S1473-3099\(20\)30457-6](https://doi.org/10.1016/S1473-3099(20)30457-6).
- [7] Nabi KN. Forecasting covid-19 pandemic: a data-driven analysis. *Chaos Solitons Fractals* 2020;139:110046. <https://doi.org/10.1016/j.chaos.2020.110046>.
- [8] Nabi KN, Abboubakar H, Kumar P. Forecasting of covid-19 pandemic: from integer derivatives to fractional derivatives. *Chaos Solitons Fractals* 2020;141:110283. <https://doi.org/10.1016/j.chaos.2020.110283>.
- [9] Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, Wang MH, Cai Y, Wang W, Yang L, et al. A conceptual model for the outbreak of coronavirus disease 2019 (covid-19) in wuhan, china with individual reaction and governmental action. *Int J Infect Dis* 2020. <https://doi.org/10.1016/j.ijid.2020.02.058>.
- [10] Asamoah JKK, Jin Z, Sun, et al. Sensitivity assessment and optimal economic evaluation of a new covid-19 compartmental epidemic model with control interventions. *Chaos Solitons Fractals* 147; 2021: 110885. <https://doi.org/10.1016/j.chaos.2021.110885>.
- [11] May T. Lockdown-type measures look effective against covid-19. *BMJ* 2020;370. <https://doi.org/10.1136/bmj.m2809>.

- [12] Ferguson N, Laydon D, Nedjati Gilani G, Imai N, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. Imperial College COVID-19 Response Team; 2020. <https://doi.org/10.1016/j.idm.2020.04.001>.
- [13] Panovska-Griffiths J. Can mathematical modelling solve the current covid-19 crisis? BMC Public Health 2020;20:551. <https://doi.org/10.1186/s12889-020-08671-z>.
- [14] ArunKumar KE, Kalaga DV, Kumar CMS, Kawaji M, Brenza TM. Forecasting of covid-19 using deep layer recurrent neural networks (rnnns) with gated recurrent units (grus) and long short-term memory (lstm) cells. Chaos Solitons Fractals 2021; 110861. <https://doi.org/10.1016/j.chaos.2021.110861>.
- [15] Fokas AS, Dikaios N, Kastis GA. Mathematical models and deep learning for predicting the number of individuals reported to be infected with sars-cov-2. J R Soc Interface 2020;17(169):20200494. <https://doi.org/10.1098/rsif.2020.0494>.
- [16] Shastri S, Singh K, Kumar S, Koush P, Mansotra V. Time series forecasting of covid-19 using deep learning models: India-usa comparative case study. Chaos Solitons Fractals 2020;140:110227. <https://doi.org/10.1016/j.chaos.2020.110227>.
- [17] 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>.
- [18] 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>.
- [19] Chimmula VKR, Zhang L. Time series forecasting of covid-19 transmission in canada using lstm networks. Chaos Solitons Fractals 2020;135:109864. <https://doi.org/10.1016/j.chaos.2020.109864>.
- [20] Karevan Z, Suykens JA. Transductive lstm for time-series prediction: an application to weather forecasting. Neural Networks 2020;125:1–9. <https://doi.org/10.1016/j.neunet.2019.12.030>.
- [21] Baek Y, Kim H. Modaagnet: a new forecasting framework for stock market index value with an overfitting prevention lstm module and a prediction lstm module. Expert Syst Appl 2018;113:457–80. <https://doi.org/10.1016/j.eswa.2018.07.019>.
- [22] ÖY. A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification. Comput Biol Med 2018; 96: 189–202. <https://doi.org/10.1016/j.combiomed.2018.03.016>.
- [23] Alhagry S, Fahmy AA, El-Khoribi RA. Emotion recognition based on eeg using lstm recurrent neural network. Int J Adv Comput Sci Appl 2017; 8(10). <https://doi.org/10.14569/IJACSA.2017.081046>.
- [24] Sønderby SK, Sønderby CK, Nielsen H, Winther O. Convolutional lstm networks for subcellular localization of proteins. Algorithms Comput Biol 2015. https://doi.org/10.1007/978-3-319-21233-3_6.
- [25] Huang C-J, Kuo P-H. Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting. IEEE Access 2019;7:74822–34. <https://doi.org/10.1109/ACCESS.2019.2921238>.
- [26] Shih SY, Sun FK, Lee HY. Temporal pattern attention for multivariate time series forecasting. Mach Learn 2019;108:1421–41. <https://doi.org/10.1007/s10994-019-05815-0>.
- [27] Wang L, Wang HF, Liu, et al. Predicting protein-protein interactions from matrix-based protein sequence using convolution neural network and feature-selective rotation forest. Scientific Rep 2019; 9(9848). <https://doi.org/10.1038/s41598-019-46369-4>.
- [28] Uddin MR, Mahbub S, Rahman MS, Bayzid MS. Saint: self-attention augmented inception-inside-inception network improves protein secondary structure prediction. Bioinformatics 2020;36(17):4599–608. <https://doi.org/10.1093/bioinformatics/btaa531>.
- [29] Shao Y, Chou KC. ploc_deep-mvirus: A cnn model for predicting subcellular localization of virus proteins by deep learning. Natural Sci 2020; 12(6). <https://doi.org/10.4236/ns.2020.126033>.
- [30] Repository CG. Center for systems science and engineering at johns hopkins university. <https://github.com/CSSEGISandData/COVID-19> (Date accessed December 19, 2020).
- [31] LeCun Y, Haffner P, Bottou L, Bengio Y. Shape, contour and grouping in computer vision 1999. <https://doi.org/10.1007/3-540-46805-6>.
- [32] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).
- [33] Cao L, Mees A, Judd K. Dynamics from multivariate time series. Physica D: Nonlinear Phenomena 1998;121(1–2):75–88. [https://doi.org/10.1016/S0167-2789\(98\)00151-1](https://doi.org/10.1016/S0167-2789(98)00151-1).
- [34] Alhirmizy S, Qader B. Multivariate time series forecasting with lstm for madrid, spain pollution, in: International Conference on Computing and Information Science and Technology and Their Applications (ICCISTA) 2019;2019:1–5. <https://doi.org/10.1109/ICCISTA.2019.8830667>.
- [35] 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>.
- [36] Wang L, Chen J, Marathe M. Defsi: deep learning based epidemic forecasting with synthetic information. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019. pp. 9607–9612. <https://doi.org/10.1609/aaai.v33i01.33019607>.
- [37] Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock price prediction using lstm, rnn and cnn-sliding window model. In: 2017 international conference on advances in computing, communications and informatics (icaci); 2017. p. 1643–7.
- [38] Zhang B, Li J, Lü Q. Prediction of 8-state protein secondary structures by a novel deep learning architecture. BMC Bioinf 2018;19(1):293. <https://doi.org/10.1186/s12859-018-2280-5>.