Computers in Biology and Medicine 144 (2022) 105342

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Forecasting COVID-19 new cases using deep learning methods

Lu Xu [b], Rishikesh Magar [a], Amir Barati Farimani [a,*]

[a] Department of Mechanical Engineering, Carnegie Mellon University, PA, 15213, Pittsburgh, United States
[b] Department of Biomedical Engineering, Carnegie Mellon University, PA, 15213, Pittsburgh, United States

ARTICLE INFO

Keywords:
Sars-Cov-2
Deep Learning
Forecasting
Infected Population

ABSTRACT

After nearly two years since the first identification of SARS-CoV-2 virus, the surge in cases because of virus mutations is a cause of grave public health concern across the globe. As a result of this health crisis, predicting the transmission pattern of the virus is one of the most vital tasks for preparing and controlling the pandemic. In addition to mathematical models, machine learning tools, especially deep learning models have been developed for forecasting the trend of the number of patients affected by SARS-CoV-2 with great success. In this paper, three deep learning models, including CNN, LSTM, and the CNN-LSTM have been developed to predict the number of COVID-19 cases for Brazil, India and Russia. We also compare the performance of our models with the previously developed deep learning models and notice significant improvements in prediction performance. Although our models have been used only for forecasting cases in these three countries, the models can be easily applied to datasets of other countries. Among the models developed in this work, the LSTM model has the highest performance when forecasting and shows an improvement in the forecasting accuracy compared with some existing models. The research will enable accurate forecasting of the COVID-19 cases and support the global fight against the pandemic.

1. Introduction

The rampant spread of the COVID-19 pandemic has resulted in huge economic, human life loss and disruption of normal public life across the globe [1]. According to the World Health Organization (WHO), over 200 million people have been infected by the SARS-CoV-2 virus worldwide [2]. The virus is known to transmit between people through respiratory routes during human mobility [3], increasing its transmissibility and making the general public susceptible. This correlation between human mobility and transmissibility of the virus has led to measures such as mandatory face coverings, social distancing, closing public transportation, schools, restaurants, and avoiding gathering have been imposed by governments across the world [4]. The enforcement of such policies has helped in arresting the spread of the virus, yet its highly contagious nature coupled with the evolution of dangerous mutations has continued to ravage public human health.

With the increasing number of patients, medical supplies are usually short of demand burdening the health care systems and professionals in many countries [5]. Thus, understanding the spread and reliably forecasting the trends is one of the most crucial elements to prevent the spread of the pandemic, particularly in countries with a large population like India. Reliability in forecasting trends of the COVID-19 spread can help predict the pandemic outbreak and increase the preparedness of governments in tackling the pandemic. Moreover, accurate forecasting can provide feedback on whether the undertaken policy is effective in alleviating the stress on the healthcare system of that country. It also allows governments to evaluate mitigation strategies and regulate policies based on the forecasts of the areas in concern. For example, by applying mathematical models, such as SIR and SEIR models, researchers have successfully predicted the reproduction parameter of the COVID-19 in Indonesia for the early prevention of the pandemic, reinforcing the need for reliable forecasting models [6].

Recently, machine learning models have been extensively used for forecasting and can be especially useful in terms of pandemic planning. In this study, we develop a deep learning approach to forecast the pandemic trend for three countries including Brazil, India and Russia. These are among the top-10 most heavily affected countries worldwide and have been widely studied by healthcare experts. In this paper, we implement three different deep learning models, including the Convolutional neural network (CNN), Long short-term memory (LSTM) and Convolutional neural network-Long short-term memory (CNN-LSTM), to predict the number of cases and forecast the spread of COVID-19. The

* Corresponding author.
E-mail address: barati@cmu.edu (A. Barati Farimani).

https://doi.org/10.1016/j.compbiomed.2022.105342
Received 9 December 2021; Received in revised form 20 February 2022; Accepted 20 February 2022
Available online 23 February 2022

prediction performances of the three models are evaluated using mean absolute error (MAE), R^2 score and explained variance (EV) score. The LSTM model and the CNN-LSTM models perform comparably and have the lowest MAE for the countries that we consider in our study. Moreover, the LSTM model we developed outperforms some of the previously developed models [7] and hence we use it for forecasting the COVID-19 cases a week into the future. Using our model, we also reliably forecast the number of cases for the next 14 days, outside the training and test datasets. Our ML models incorporate the additional features like the different governmental policies in an effective manner developing a more informed deep learning-based forecasting model than the previous works. Our models contribute to the variety of tools available for COVID-19 forecasting, we believe that our models can help us improve our pandemic preparedness and tackle it more effectively.

2. Related work

Machine learning models have been successfully used to understand the various aspects of the pandemic from developing machine learning models that can design antibodies [8], using medical image datasets, notably chest X-rays [9,10], modeling and understanding mutations [11,12,33], to detecting whether a patient is infected by SARS-CoV-2 to forecasting the trends of the pandemic. In addition, some short-term forecasting methods, including SutteIndicator, which is widely used to predict the stock price based on the previous days' data [13]; SutteARIMA, which averages the forecasting results of the α-Sutte Indicator and ARIMA [14]; and Holt-Winters, which can capture three important aspects of the time-series data: the average, trend and seasonality [15] have been applied to predict the development of the pandemic. In this work, we focus on forecasting the pandemic trends for different countries namely Brazil, India and Russia, because these are the countries that have been widely studied. In this case, we can compare our results with those in previous studies. Previously conducted forecasting studies using machine learning pertinent to these countries have been noted in this section.

Brazil being one of the most heavily affected countries due to the pandemic, has been widely studied by researchers. Ribeiro et al., used autoregressive integrated moving average (ARIMA), cubist regression, random forest, ridge forest, SVR and stacking-ensemble learning, respectively to analyze the cumulative confirmed cases in Brazil [16]. With the comparison of forecasting performance, they concluded that SVR performs the best with an error of less than 6.9%. Another study using training on limited data of 30 days and 40 days, respectively was conducted to predict the rate of spread in Brazil using the Gated Recurrent Unit (GRU) [17]. They observed that the highest accuracy of 85% has been achieved on the time-step of 30 days using the validation data from 4/7/2020 to 6/13/2020. However, the accuracy drops markedly (a maximum of 68%) as the predicting period increases, indicating that the model behaves relatively poorly in a long-time range forecasting.

In another study, Da Silva et al. analyzed the number of infections of the top 27 affected Brazilian cities using the single ARIMA and the hybrid model, which is the integration of the Ensemble Empirical Mode Decomposition (EEMD) method and the ARIMA, respectively. Their results show that the ensemble model performed 26.73% better than the single model [18].

Researchers have expressed considerable interests in analyzing the dynamic spread of India as well. Swaraj et al. developed a model integrating ARIMA and nonlinear autoregressive neural network (NAR) for predicting the COVID-19 outbreak in India. The result shows a significant reduction in evaluation metrics (RMSE: 16.23%, MAE: 37.89% and MAPE: 39.53%) with the hybrid model compared to the single ARIMA model [19]. Besides, Wadhwa et al. studied the effects of lockdown policy on disease transmission by predicting the number of active cases all over India. Based on the Linear Regression (LR) model, they generated a graphical representation of the COVID-19 cases of three months ahead [20]. In another study, Khan et al. built three machine learning models (Decision Tree (DT), SVM and Gaussian Process Regression (GPR) to analyze the time point where the number of cases stops rising in India, and thus were able to analyze policy regulations. According to their results, the GPR model outperforms the other models with an accuracy of 95% [21].

Apart from Brazil and India, another country that has been widely studied is Russia. Wang et al. developed an LSTM model to forecast trends of the pandemic in 150 days ahead using the daily new confirmed cases in Russia, Peru and Iran ([12]. In another study, the Bayesian model has been applied to investigate the effects of lockdowns on the COVID-19 transmission using the data from March 1 to June 29, 2020 in the top five countries (India, Brazil, Russia, USA and UK). It was demonstrated that the outbreak pace will significantly increase in Brazil, India and Russia once loosening the lockdowns [22]. Dairi et al. has compared the prediction performance of machine learning methods (LR and SVR) and deep learning methods (the hybrid LSTM-CNN, the hybrid GAN-GRU, CAN, CNN, LSTM) [23] Data used in this study was from Russia, Brazil, India, US, France, Mexico and Saudi Arabia. It was reported that deep learning tools outperform the conventional machine learning tools in terms of forecasting performance, especially LSTM-CNN exhibiting the most accurate prediction with a MAPE of 3.718%. More deep learning methods, specifically RNN, LSTM, BiLSTM, GRU and VAE, were analyzed to predict the COVID-19 cases in different countries (Italy, France, Spain, China, Australia and the USA) [7].

3. Methods

3.1. Data preprocessing

The Center of System Science and Engineering (CSSE) at John Hopkins University has aggregated the COVID-19 cases data from 22 January 2020 till date for around 210 countries across the world [24]. In our study, we analyze the data of three highly impacted countries: Brazil, India and Russia. The trend of the cumulative number of cases for the countries that we study is shown in Fig. S1. To account for the delay between the COVID test and report results and updating of cases on the portal, we apply a smoothening 7-day average (Fig. S2) and assign it to the day where 0 cases were reported. This way, we ensure that the data is stable and the days where there were no cases reported will be eliminated. To get the features, such as face coverings, restrictions on gatherings, closing public transportation and staying at home, and the

Table 1
Specific policies used as training features.

	Face Covering	No Gathering	Closure of Public Transportation	Stay at Home
0	Not required	Not required	Not required	Not required
1	Recommended	Gatherings should not be greater than 1000 people	Recommended	Recommended
2	In some public places where other people are nearby, the policy is enforced.	Gatherings can between 100 and 1000 people	Required	The policy is enforced except for daily exercise, food purchasing and indispensable trips
3	In all public places where other people are nearby, the policy is enforced	Gatherings can between 10 and 100 people	–	The policy is enforced with minimal exceptions
4	The policy is enforced at all times no matter whether people are nearby	Gatherings should be less than 10 people	–	–

overall stringent index, which is from 0 to 1000, were considered. We use data from Our World in Data Server [32], as shown in Table 1 ("Our World in Data," 2021).

Finally, to ensure stability in numerical prediction we normalized the cases data to [0, 1] using MinMaxScaler, defined as

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x_{min} and x_{max} refer to the minimum and maximum of input data. Apart from the features enlisted in Table 1 we also used previous day's data as a feature to the model.

3.2. Models

After the data preprocessing, three deep learning models including CNN, LSTM, CNN-LSTM are implemented. The performances of the three models are compared and the best performing model is selected to forecast cases of future 7 days. The flowchart of this work is illustrated in Fig. 1.

3.3. CNN

Convolutional neural network (CNN) models are capable of automated feature extraction from the data given a prediction task. Previous studies have validated the performance of the convolutional neural network in analyzing time-series data due to its strong capability of extracting the features from data like the stock price predictions, air quality forecasts and energy load forecasting [16,25,26]. Following previous research, we decide to use CNN to predict the spreading of the COVID-19. In this work, we develop a CNN model with 3 convolutional layers and two fully connected layers. The detailed description of the architecture is given in Table 2.

We use ReLU as the activation function for the non-linear transformation and two fully connected layers are implemented at the end of the model [27]. To train the CNN model, we use the data from January 1 to July 13, 2021. We train the model for 500 epochs and observe convergence as the loss does not decrease substantially when we train for more than 300 epochs. The details for the hyperparameter optimization are provided in SI (Table S1 & Fig. S4) and the final architectural parameters are presented in Table 2. The plot of loss in CNN training vs the number of epochs is shown in Fig. 3. We notice that the training loss for Brazil is higher than the other two countries. This may be due to the fact that the number of cases in Brazil fluctuates more than that of Russia and India, causing the higher loss.

3.4. LSTM

In addition to CNN, multiple studies have used the long short-term memory [28] (LSTM) framework for forecasting the transmission of COVID-19, because of its memory capacity [29–31]; P [12]. Since the prediction of COVID-19 cases is a time series problem and involves capturing time dependencies in the data, we develop an LSTM model that can predict the COVID-19 cases. The detailed description of the architecture is given in Table 2. Similar to the CNN model we train the model from January 1 to July 13, 2021. It is observed that training loss does not substantially improve after 300 epochs, so we train only for 300 epochs and generate the predictions. The complete details about the hyperparameter optimization and model architecture are available in Table S2 and Fig. S4.

Table 2
Parameters of architectures for each model.

	CNN	LSTM	CNN-LSTM
Kernel Size	2	–	2
Hidden Layers	3	4	3 CNN & 3 LSTM
Hidden Units	–	130	175
Convolutional Filters	32, 45, 64	–	32, 45, 64
Stride	1	–	1
Padding	0	–	0
Learning Rate	0.005	0.005	0.005
Optimizer	Adam	Adam	Adam
Epoch	300	300	300

3.5. CNN-LSTM

We also investigate the CNN-LSTM model that takes advantages of both the CNN and LSTM models, where the CNN part is extracting important features from the data and the LSTM is designed to learn sequence patterns in time-series data. Specifically, CNN first extracts features from the training set through convolutional and pooling layers and generates an embedding. This embedding from CNN is then fed as an input to the LSTM. LSTM with its ability to capture the time dependencies in the input data takes the features extracted by the CNN as input and predicts the number of cases. The architecture of the CNN-LSTM model is shown in Fig. 2 and detailed parameters for the model are available in Table 2. We train the CNN -LSTM model for 300 epochs. The training loss curve for the model is shown in Fig. 3.

3.6. Evaluation metrics

To compare the performance of three models quantitively, evaluation metrics: Mean Absolute Error (MAE), Coefficient of determination (R^2) are calculated as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3}$$

where y_i is the actual case and $\hat{y}_i$ is the predicted cases. In addition, we also use the explained variance as a metric to evaluate the performance of the models. The model with the least MAE and highest R2 score and EV score is considered the best architecture and prepared to forecast the COVID-19 transmission. The results of the three models are presented with details in the following sections.

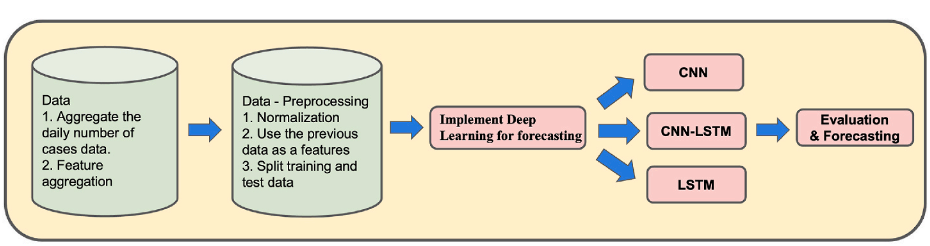

Fig. 1. Flowchart capturing the workflow that we used in this study. Green cylinders indicate the data preprocessing step; red boxes represent the training, evaluation and forecasting phase.

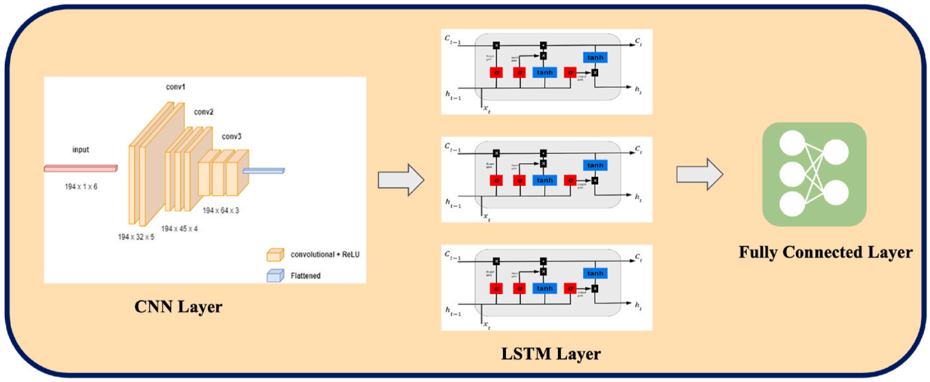

Fig. 2. Illustration of the architecture of CNN-LSTM model that we developed in this work. We have 3 convolutional layers and 3 LSTM layers in the architecture.

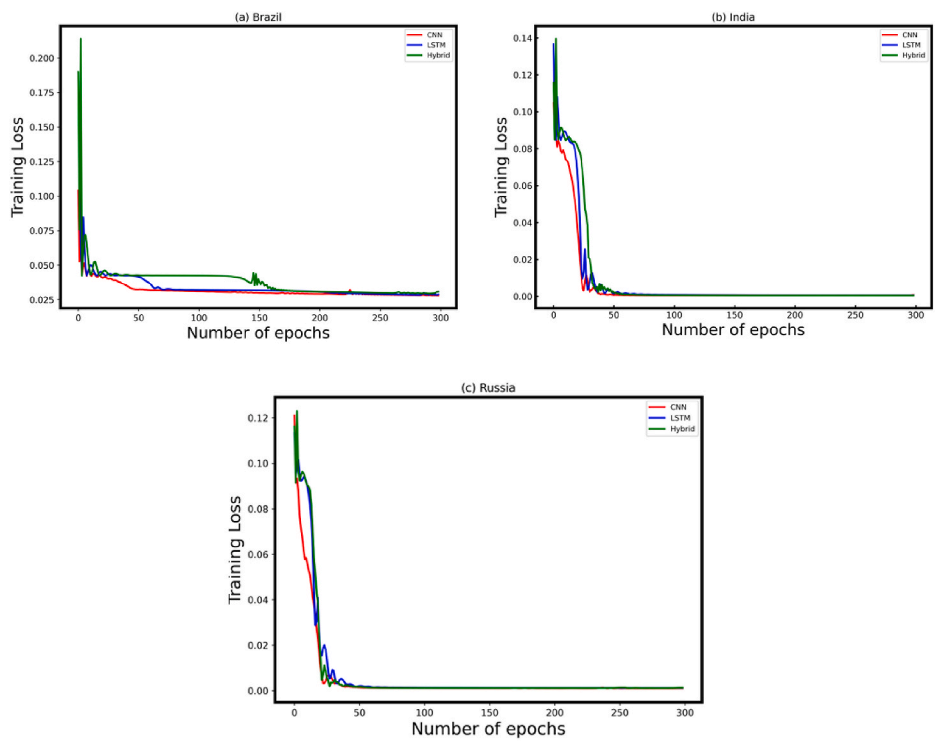

Fig. 3. Training loss vs epochs for each country (a) Brazil, (b) India and (c) Russia. We observe that the training loss decreases with epochs for all models, and we observe convergence around 300 epochs.

4. Results and discussion

4.1. Comparison of models

We analyze the prediction performance of the three deep learning models on data from three countries - India, Brazil and Russia. The model performances are trained on data from 1st January to 13th July 2021 and evaluated using the test data from 14th July to 31st July 2021. The prediction performance of the models on the test data for all the models is shown in Fig. 4.

The prediction performances are evaluated quantitatively using metrics (MAE, R^2, EV) and also validate models in detail. As summarized in Table 3, among all the trained models the CNN-LSTM model and LSTM have relatively better performance than CNN for different countries. For India, the CNN-LSTM model performs the best with an MAE of 5245 cases considering the average number of cases is 39426 for the test set. The error is within 13.30%, making it highly accurate and reliable. Similarly, for Russia, the CNN-LSTM also achieves the least error about 4.20% with an MAE of 986 cases and the average number of cases 23502 in the testing period of 18 days. For Brazil, the best model is the LSTM model where the error is 36.90% with an MAE of 15275 and the average number of cases in the testing period being 42547. It is observed that all models perform relatively poorer for Brazil. This is because there are large variations in the data, causing models trained on this largely varying training data unable to generalize on the test. The CNN model, though it has an acceptable prediction performance generally fares poorly when compared to LSTM and CNN-LSTM, as the CNN model is not explicitly designed to capture the time dependencies in the model. In addition, it is observed that R^2 of LSTM and CNN-LSTM is relatively high for cumulative cases prediction, especially for India and Russia (near 1), indicating that the predicted cases closely follow the trend of the true cases (Fig. 5(a)). Similarly, we also evaluate the EV score (Fig. 5(b)), and it was observed that EV for LSTM and CNN-LSTM models is usually higher. Although the MAE for CNN-LSTM in the prediction set is relatively lower, it must be noted that the MAE for LSTM is not very different for Brazil and Russia when compared to CNN-LSTM and it also has a

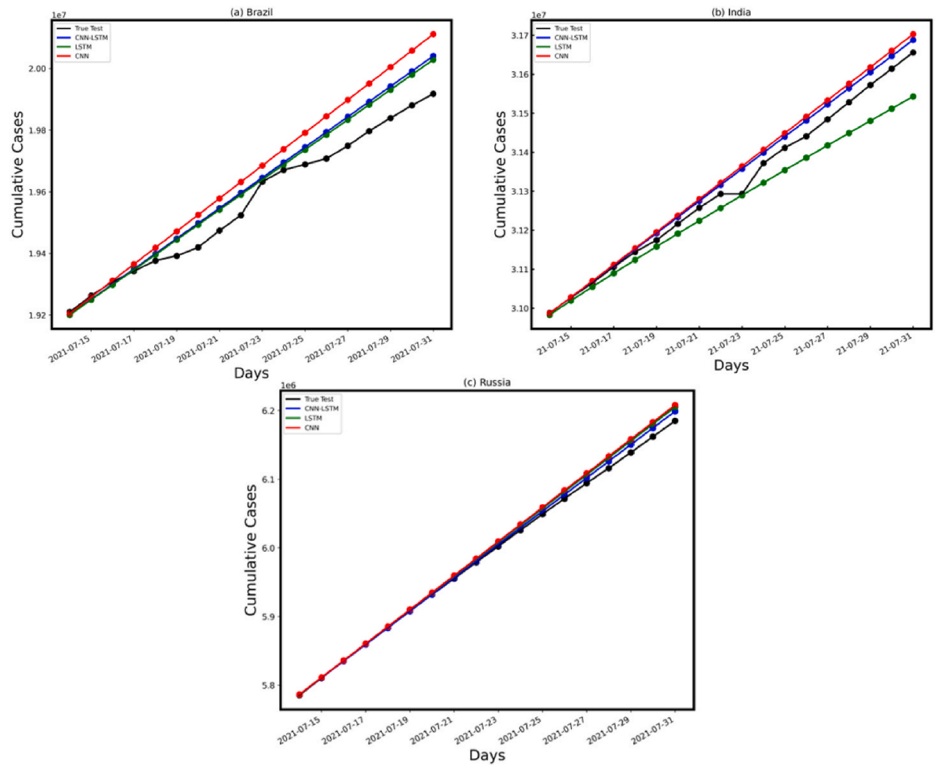

Fig. 4. Predicted COVID-19 cumulative cases from 14 July to 31 July 2021 of (a) Brazil, (b) India and (c) Russia using different models.

Table 3
Evaluation metrics for comparing the forecasting performance of three models using the daily new confirmed cases.

	Models	MAE
India	LSTM	8949
	CNN	5407
	CNN-LSTM	5245
Russia	LSTM	1198
	CNN	1342
	CNN-LSTM	986
Brazil	LSTM	15275
	CNN	17668
	CNN-LSTM	15563

higher R^2 score for cumulative cases prediction than CNN-LSTM for India and Russia indicating its strong performance.

For governments to prevent and control the pandemic, MAEs are calculated for cases in Brazil where each of the following governmental measures is not considered in the model. According to the results shown in Table S6, the order of importance is no gatherings, face coverings, closure of public transportation and stay at home. Therefore, governments are expected to emphasize no gatherings to mitigate the spread of the COVID-19.

In order to illustrate the reliability of our models, we have compared our results with those of previous studies using the same data from January 22, 2020 to July 17, 2020 [7]. The MAEs of the forecasting cases calculated using the same data of four countries are shown in Table 4. It is observed that our models show better performances with a decreased MAE compared to those in the study, especially our LSTM model (in red). Considering all metrics, we choose the LSTM model for the later forecasting analysis. The improvements in our model can be attributed to the additional features (governmental policies) used in our model, since the previous study only considered the number of cases. In addition, we use the daily new cases to train the model and calculate the cumulative cases based on the output of the model for testing, but some of the previous research directly used the cumulative cases for training and testing. This usage of daily cases instead of cumulative cases also helps our model to improve the prediction performance. Moreover, our

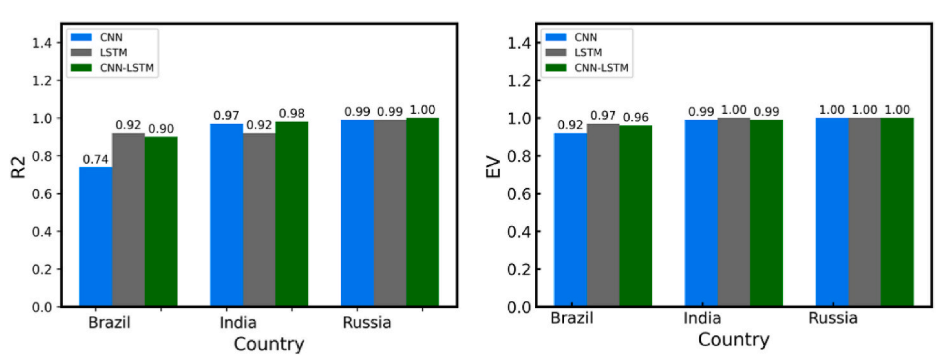

Fig. 5. a.) R^2 score calculated for all models. We use the total cases predicted by the model vs the actual total cases for calculating the R^2 score b.) We also calculate the explained variance score for all the models for the three different countries.

Table 4
MAE for comparing the forecasting performance of our models and those in the previous study. The results reported by previous studies has been denoted by previous studies (Study) and the results from this paper are denoted by (Ours).

		Models	MAE
Italy	Study	RNN	1.06E + 06
		GRU	1.13E + 06
		LSTM	1.05E + 06
		BiLSTM	1.03E + 06
	Ours	CNN	2.68E + 03
		LSTM	2.14E + 03
		CNN-LSTM	2.91E + 03
France	Study	RNN	1.28E + 06
		GRU	1.20E + 05
		LSTM	1.08E + 06
		BiLSTM	1.16E + 06
	Ours	CNN	2.52E + 03
		LSTM	2.43E + 03
		CNN-LSTM	2.52E + 03
USA	Study	RNN	5.14E + 06
		GRU	4.24E + 06
		LSTM	1.12E + 06
		BiLSTM	4.19E + 06
	Ours	CNN	4.76E + 04
		LSTM	4.38E + 04
		CNN-LSTM	2.33E + 04
Pain	Study	RNN	1.68E + 05
		GRU	1.79E + 06
		LSTM	1.24E + 06
		BiLSTM	1.19E + 06
	Ours	CNN	5.63E + 04
		LSTM	3.24E + 03
		CNN-LSTM	5.27E + 03

model architecture varies from that of the study. For example, in terms of the LSTM model, our learning rate and hidden units are 0.005 and 130, which are larger than 0.0005 and 16 in the previous studies. Such differences in model architectures can potentially improve our predicting performances when compared to the previous models.

4.2. Forecasting

With the LSTM determined to have superior performance when compared to previously published models and its comparable performance to CNN-LSTM, we use the trained model for forecasting the COVID-19 cases outside the training and the test data. Specially, we forecast the COVID-19 cases from 1 August to 7 August 2021. We use the previous day's prediction as the input data for the LSTM model during the forecasting process. In addition, we assume that the policies of the forecasting day are the same as those of the previous day. The assumption is reasonable since it is highly unlikely that the governmental policies will fluctuate in just two consecutive days. The results of the three countries are illustrated in Fig. 6, where the error percentage for each day is calculated and shown in the plots. For example, the model has an error of 2.49% in predicting the daily new cases of 8/1/2021 in India. From the graphs, the model has a good forecasting performance with cases in India and Russia, where the daily error percentages are less than 10%. However, relatively high errors occur in Brazil (8/1 and 8/2) as the actual data is very volatile in this country, causing difficulties in capturing those variations for our model.

4.3. Limitations of the model

The lag time between testing and recording of cases may lead to large swings in the data which are difficult to model and possibly can lead to some errors in the forecast. Moreover, different social-economical, geographical and political reasons can also influence governmental policies such as imposing lockdowns, mask mandates and vaccination status. These factors have not been included in our model as they are difficult to model, and datasets related to such factors are not unavailable for most of the countries.

In summary, our LSTM model can successfully forecast the trend of the cumulative cases and predict the daily new cases for countries with the relatively stable transmission. However, for countries with the rapidly changing number of cases, the model may have difficulties in capturing the most recent changing trends. In this case, we may need to train on a larger quantity of data to achieve more accurate results. To ensure that our analysis is exhaustive we also performed a similar

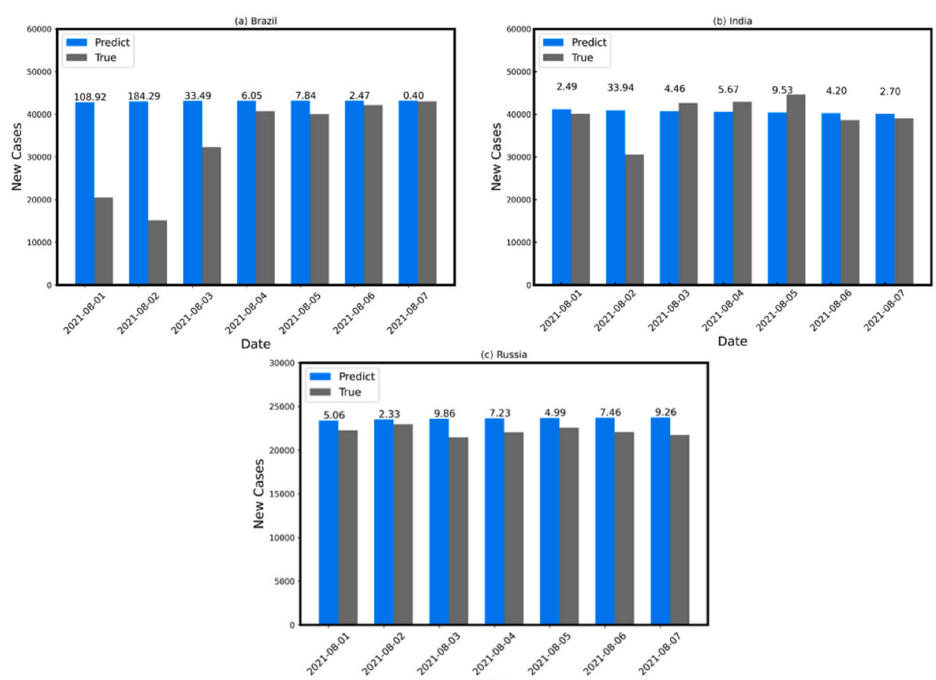

Fig. 6. The actual and predicted daily new cases of 1 August to 7 August 2021 for a) Brazil, b) India c) Russia.

forecasting analysis for CNN-LSTM, the results for which are noted in Figure S5. It was observed that the forecasting performance of LSTM model is slightly superior when compared to CNN-LSTM.

5. Conclusion

In this study, we have implemented three deep learning models and compared their predicting performances for forecasting the COVID-19 cases for three countries - Brazil, India and Russia. All three models successfully capture the transmission trend in each country. We observe that the LSTM model has the best performance based on the results of evaluation metrics MAE, R^2 and EV. We would also like to note that our model also shows an improvement with a reduced error compared with previous studies that use deep learning for predicting the SARS-CoV-2 cases. Using the best performing LSTM model, we then forecast the COVID-19 cases of 7 days outside the training and the test dataset for the three countries in the study. Developing such models can be crucial in pandemic planning and helping tackle the COVID-19 more effectively. In addition to the studied countries, the proposed models and training strategies can also be applied for the other countries and also can help in assessing the effectiveness of the policies that are being imposed to curb the spread of the virus. In the future, we aim to integrate additional information about the types of SARS-CoV-2 variants, vaccination status of citizens, healthcare infrastructure, etc. as features to further improve the model capacity and performance.

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.

Acknowledgements

This work is supported by the start-up fund provided by CMU Mechanical Engineering.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105342.

References

[1] K.H. Jacobsen, Will COVID-19 generate global preparedness? Lancet Lond. Engl. 395 (2020) 1013–1014, https://doi.org/10.1016/S0140-6736(20)30559-6.
[2] World Health Statistics, A Visual Summary, 10.17.21, 2021 [WWW Document], 2021. URL, https://www.who.int/data/stories/world-health-statistics-2021-a-visual-summary.
[3] Q. Wang, L. Liu, On the critical role of human feces and public toilets in the transmission of COVID-19: evidence from China, Sustain. Cities Soc. 103350 (2021), https://doi.org/10.1016/j.scs.2021.103350.
[4] L. Durán-Polanco, M. Siller, Crowd management COVID-19, Annu. Rev. Control (2021), https://doi.org/10.1016/j.arcontrol.2021.04.006.
[5] A. Miethke-Morais, A. Cassenote, H. Piva, E. Tokunaga, V. Cobello, F.A. Rodrigues Gonçalves, R. dos Santos Lobo, E. Trindade, L.A. Carneiro D'Albuquerque, L. Haddad, COVID-19-related hospital cost-outcome analysis: the impact of clinical and demographic factors, Braz. J. Infect. Dis. 25 (2021) 101609, https://doi.org/10.1016/j.bjid.2021.101609.
[6] S. Annas, Muh Isbar Pratama, Muh Rifandi, W. Sanusi, S. Side, Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia, Chaos, Solit. Fractals 139 (2020) 110072, https://doi.org/10.1016/j.chaos.2020.110072.
[7] A. Zeroual, F. Harrou, A. Dairi, Y. Sun, Deep learning methods for forecasting COVID-19 time-Series data: a Comparative study, Chaos, Solit. Fractals 140 (2020) 110121, https://doi.org/10.1016/j.chaos.2020.110121.
[8] R. Magar, P. Yadav, A. Barati Farimani, Potential neutralizing antibodies discovered for novel corona virus using machine learning, Sci. Rep. 11 (2021) 5261, https://doi.org/10.1038/s41598-021-84637-4.
[9] N. Zhu, D. Zhang, W. Wang, A Novel Coronavirus from Patients with Pneumonia in China, 2019, 9.22.21, NEJM, 2020 [WWW Document]. URL, https://www.nejm.org/doi/full/10.1056/nejmoa2001017.
[10] M. Toğaçar, B. Ergen, Z. Cömert, COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches, Comput. Biol. Med. 121 (2020) 103805, https://doi.org/10.1016/j.compbiomed.2020.103805.
[11] B. Mullick, R. Magar, A. Jhunjhunwala, A. Barati Farimani, Understanding mutation hotspots for the SARS-CoV-2 spike protein using Shannon Entropy and K-means clustering, Comput. Biol. Med. 138 (2021) 104915, https://doi.org/10.1016/j.compbiomed.2021.104915.
[12] P. Wang, X. Zheng, G. Ai, D. Liu, B. Zhu, 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, Solit. Fractals 140 (2020) 110214, https://doi.org/10.1016/j.chaos.2020.110214.
[13] A.M.C.H. Attanayake, S. Perera, Forecasting COVID-19 cases using alpha-sutte indicator: a comparison with autoregressive integrated moving average (ARIMA) method, BioMed Res. Int. (2020) 1–11, https://doi.org/10.1155/2020/8850199, 2020.
[14] A.S. Ahmar, E. Boj, Will COVID-19 confirmed cases in the USA reach 3 million? A forecasting approach by using SutteARIMA Method, Curr. Res. Behav. Sci. 1 (2020) 100002, https://doi.org/10.1016/j.crbeha.2020.100002.
[15] V.K. Sharma, U. Nigam, Modeling and forecasting of COVID-19 growth curve in India, Trans. Indian Natl. Acad. Eng. 1–14 (2020), https://doi.org/10.1007/s41403-020-00165-z.
[16] M.H.D.M. Ribeiro, R.G. da Silva, V.C. Mariani, L. Coelho, S. dos, Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil, Chaos, Solit. Fractals 135 (2020) 109853, https://doi.org/10.1016/j.chaos.2020.109853.
[17] M. Hawas, Generated time-series prediction data of COVID-19′s daily infections in Brazil by using recurrent neural networks, Data Brief 32 (2020) 106175, https://doi.org/10.1016/j.dib.2020.106175.
[18] T.T. da Silva, R. Francisquini, M.C.V. Nascimento, Meteorological and human mobility data on predicting COVID-19 cases by a novel hybrid decomposition method with anomaly detection analysis: a case study in the capitals of Brazil, Expert Syst. Appl. 182 (2021) 115190, https://doi.org/10.1016/j.eswa.2021.115190.
[19] A. Swaraj, K. Verma, A. Kaur, G. Singh, A. Kumar, L. Melo de Sales, Implementation of stacking based ARIMA model for prediction of Covid-19 cases in India, J. Biomed. Inf. 121 (2021) 103887, https://doi.org/10.1016/j.jbi.2021.103887.
[20] P. Wadhwa, A. Aishwarya, Tripathi, P. Singh, M. Diwakar, N. Kumar, Predicting the time period of extension of lockdown due to increase in rate of COVID-19 cases in India using machine learning, in: Mater. Today Proc., International Conference on Newer Trends and Innovation in Mechanical Engineering: Materials Science, vol. 37, 2021, pp. 2617–2622, https://doi.org/10.1016/j.matpr.2020.08.509.
[21] F.M. Khan, A. Kumar, H. Puppala, G. Kumar, R. Gupta, Projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods, J. Saf. Sci. Resil. 2 (2021) 50–62, https://doi.org/10.1016/j.jnlssr.2021.05.001.
[22] N. Feroze, Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models, Chaos, Solit. Fractals 140 (2020) 110196, https://doi.org/10.1016/j.chaos.2020.110196.
[23] A. Dairi, F. Harrou, A. Zeroual, M.M. Hittawe, Y. Sun, Comparative study of machine learning methods for COVID-19 transmission forecasting, J. Biomed. Inf. 118 (2021) 103791, https://doi.org/10.1016/j.jbi.2021.103791.
[24] CSSEGISandData, COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at, Johns Hopkins University, 2021.
[25] L. Di Persio, O. Honchar, Artificial Neural Networks architectures for stock price prediction: comparisons and applications, Int. J. Circuits Syst. Signal Process. (2016).
[26] R. Yan, J. Liao, J. Yang, W. Sun, M. Nong, F. Li, Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering, Expert Syst. Appl. 169 (2021) 114513, https://doi.org/10.1016/j.eswa.2020.114513.
[27] V. Nair, G.E. Hinton, Rectified Linear Units Improve Restricted Boltzmann Machines, vol. 8, 2010.
[28] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[29] Y. Gautam, Transfer Learning for COVID-19 cases and deaths forecast using LSTM network, ISA Trans. (2021), https://doi.org/10.1016/j.isatra.2020.12.057.
[30] J. Luo, Z. Zhang, Y. Fu, F. Rao, Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms, Results Phys. 27 (2021) 104462, https://doi.org/10.1016/j.rinp.2021.104462.
[31] Z.E. Rasjid, R. Setiawan, A. Effendi, A comparison: prediction of death and infected COVID-19 cases in Indonesia using time series smoothing and LSTM neural network, in: Procedia Comput. Sci., 5th International Conference on Computer Science and Computational Intelligence 2020, vol. 179, 2021, pp. 982–988, https://doi.org/10.1016/j.procs.2021.01.102.
[32] Our World in Data, Our World Data, WWW Document], 2021. URL, https://ourworldindata.org, 9.29.21.
[33] Y. Wang, P. Yadav, R. Magar, A.B. Farimani, Bio-informed Protein Sequence Generation for Multi-Class Virus Mutation Prediction, 2020, https://doi.org/10.1101/2020.06.11.146167.