Prediction of COVID-19 Spreading using LSTM Recurrent Neural Networks in Southeast Asia

Md. Moshiur Rahman and Sumaiya Kabir

Abstract—The world has come to a standstill due to the coronavirus. The number of confirmed cases and the number of deaths are increasing day by day. Now accurate and real-time predictions are very important. Currently, a lot of work is being done on the spread prediction of coronavirus. If we can predict the number of corona patients in a country in advance, then it will be beneficial for the government of that country to take action in advance. In this paper, we experiment with Bangladesh and India data using the deep learning method. We have used LSTM (LONG SHORT-TERM MEMORY) neural network which is a type of recurrent neural network. The LSTM method can work very well even with very little data. Since we do not currently have much data on covid19, LSTM neural networks may be a suitable model. We have got very good prediction results using this method even with very little data. The prediction of various parameters (number of confirmed cases per day, number of deaths per day, number of total confirmed cases, and number of total deaths). We also calculate MAPE, MAE, and RMSE values and compare them among different models. Our model outperforms others’ models. We think this research will be a beneficial tool for administrators and health officials.

Index Terms—Covid-19, RNN, LSTM NN, Prediction.

I. INTRODUCTION

SARS-COV-2 spreading has put our world in a threatening place. The novel coronavirus was first reported by China’s Wuhan Municipal Health Commission. 7.12 million confirmation cases and over 406,000 deaths across 188 countries whereas recovering cases are 3.29 million [1]. According to WHO [2], services like 53% of the country have been partially disrupted. The first outbreak news of Bangladesh came out on 5 January 2020 and the spreading was confirmed in March [3]. Bangladesh confirmed 38292 cases on May 27 and this number reached 368690 on October 4 [4]. The death toll counted 1.45% of the total infected. The first case of Covid-19 was reported in India on January 30, 2020. In India, a government curfew was imposed voluntarily for a maximum of 14 hours on 22 March. On June 9, it was confirmed that 274780 the death toll had risen to 7719. On June 9, the worldometer calculated death at 5.4%. Figure 1 and Figure 2 has been shown how the coronavirus is spreading day by day. Lockdown has not been effective in Bangladesh and India. Pandemic and lockdowns have hampered most economic progress, operations and business cycles. AI methods such as neural networks and other technologies are increasingly being used to predict

Fig. 1. No of total confirmed cases in Bangladesh.

Fig. 2. No of total confirmed cases in India.

DOI: https://doi.org/10.3329/gubjse.v8i1.62325

This paper was received on 3 June 2022, revised on 6 July 2022, and accepted on 27 August 2022

Md. Moshiur Rahman is with the Green university of Bangladesh, Lecturer, Bangladesh. E-mail: moshiur@cse.green.edu.bd

Sumaiya Kabir, is with Green university of Bangladesh, Assistant professor, Bangladesh. E-mail: sumaiya@cse.green.edu.bd.
cancer, cardiovascular disease, obesity, and other diseases in T2D patients [5]. Neural networks work as a biological neural network to recognize where it is in relation to the underlying dataset. Interconnected nodes are contained in layers in a neural network [6]. Predictive platforms for spreading the virus can create neural networks. Assumption analysis techniques of neural networks are conducted in classification. The recorded dataset time of the Covid-19 can be used to study the predictions of the time series applying the neural network. This paper proposes LSTM (Long Short-Term Memory) based on the Artificial Repetitive Neural Network (RNN) architecture for predicting long-term risk Covid-19. Continuous time series prediction problems can be modeled by LSTM and require a large amount of data [7]. Due to lack of an adequate dataset we used LSTM model and observed that it operates effectively and predicts the output value in close proximity to real-time observation.

II. LSTM BASED PREDICTION MODEL

LSTMs are a type of recurrent neural network, but instead of simply feeding its outcome into the next part of the network, an LSTM does a bunch of math operations so it can have a better memory. A typical LSTM NN cell is basically made up of four gates: input gate, input modulation gate, forget gate and output gate. Input gate receives a new input points from outside gate and processes the newly arrived data. The input gate controls the extent to which a new value flows into the cell, the forget gate controls the extent to which a value remains in the cell and the output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit. The activation function of the LSTM gates is often the logistic sigmoid function.

Let us input time series is \( P = (p_1, p_2, p_3, \ldots, p_n) \), hidden layer of memory cells as \( H = (h_1, h_2, h_3, \ldots, h_n) \) and output time series is \( Q = (q_1, q_2, q_3, \ldots, q_n) \). Now LSTM NN do computation that is given below:

\[
h_t = H(w_{hp}p_t + w_{hh}h_{t-1} + V_h)
\]

\[
p_t = w_{hq}h_{t-1} + V_q
\]

Where weights are denoted as \( w \) and bias vectors denoted as \( V \). The hidden state of memory cells is computed in the following formulas that is given below.

\[
I_t = \sigma(w_{gp}p_t + w_{gh}(h_{t-1} + w_{ic}c_{t-1} + V_i))
\]

\[
f_t = \sigma(w_{gp}p_t + w_{gh}(h_{t-1} + w_{ic}c_{t-1} + V_h))
\]

\[
c_t = f_t * c_{t-1} + i_t * g(w_{gp}p_t + w_{gh}(h_{t-1} + w_{cc}c_{t-1} + V_c))
\]

\[
\alpha_t = \sigma(w_{gp}p_t + w_{gh}(h_{t-1} + w_{ac}c_{t-1} + V_o))
\]

\[
h_t = \alpha_t * h(c_t)
\]

Where \( \sigma \) stands for the standard sigmoid function defined in equation 8 stands for the scalar product of two vectors. \( g \) and \( h \) are the extends of stand sigmoid function with the range changing to \([2,2]\) and \([-1,1]\).

\[
\sigma(p) = \frac{1}{1 + e^p}
\]

For the optimization of this LSTM NN, we use Adam optimizer [8] with adaptive learning rates.

III. EXPERIMENTS

Our experiments were conducted on a computer with a 3.40 GHz Intel(R) Core(TM) i5-8250 CPU with 8 GB of memory running Windows 10 Home. The program was written in the jupyter notebook using python language. This study uses tensorflow [9], pyplot [10] and Keras: The Python Deep Learning library [11] to complete the experiments.

In our experiment, we used data from March 15 2020 to October 4 2020. To minimize the prediction error we collect data from March 15. We collected data from different places. However, the most notable and reliable sources are Statista [12], Wikipedia [13], IEDCR [3], Ourworldindata [14] and Worldometers [4]. We have collected the data of four features. These are the number of cases confirmed per day, number of deaths per day, total number of confirmed cases and total number of deaths. We have predicted these four indexes in our experiment. We used a total of 204 days of data in our experiments. In these 204 days, we used 80% of the data in model training and the remaining 20% of data we used in model evaluation. We have used the following formula for error calculation.

\[
Error = \frac{OfficialValue - PredictedValue}{OfficialValue} \times 100\%
\]

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (OfficialValue - PredictedValue)^2}{n}}
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|OfficialValue - PredictedValue|}{OfficialValue}
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{OfficialValue - PredictedValue}{OfficialValue}
\]

In our experiment we used three layers. Although more layers could be used to get better results. Using more layers increase the complexity of the training. Other hyperparameter value is given in table Table I.

A. Daily Cases Prediction

Here we have used the daily data i.e. how many people are being infected by corona virus and how many people are being death every day. Below is the prediction by experimenting on Bangladesh and India data.

Figure 3 (a) shows the number of patients infected with Covid-19 per day in Bangladesh. Here we predict the data for last 40 days which is compared with the official record which
is shown in the figure. Figure 3 (b) shows the number of deaths of patients infected with Covid-19 per day in Bangladesh. Moreover we also predict the data for the last 40 days which is compared with the official record shown in the figure. The figures show that our predictions are very close to the official recorded data. In more observations, the daily confirmed case prediction is better than the daily dead prediction.

Figure 4 (a) shows the number of patients infected with Covid-19 per day in India. Here we predict the data for last 40 days which is compared with the official record which is shown in the figure. Figure 4 (b) shows the number of deaths of patients infected with Covid19 per day in India. Moreover we also predict the data for the last 40 days which is compared with the official record shown in the figure. The figures show that our predictions are very close to the official recorded data. In more observations, the daily confirmed case prediction is better than the daily dead prediction. For model evaluation, we have compared our predicted data with the official data of last 10 days.

Table II shows the prediction data and the official Record data in Bangladesh for september 25 to October 4 and the percentage errors are also shown. Here we see that Daily Confirmed Case Prediction Error is very low. But the Daily Confirmed Death Prediction error is much higher.

Table III shows the prediction data and the official Record data in India for September 25 to October 4 and the percentage errors are also shown. Here we see that Daily Confirmed Case Prediction Error is very low. But the Daily Confirmed Death Prediction error is much higher.

**B. Total Case prediction**

Here we have used the total data i.e. how many people are being infected and how many people are dying of covid-19 overall across the countries. Below is the prediction by experimenting on Bangladesh and India data.

Figure 5 (a) shows the number of patients infected with Covid-19 in total in Bangladesh. Here we predict the data for the last 40 days which is compared with the official record which is shown in the figure. Figure 5 (b) shows the number of deaths of patients infected with Covid-19 in total in Bangladesh. Moreover, we also predict the data for the last
Fig. 4. Predicted value on the basis of daily data compared with official recorded data of India.

Fig. 5. Comparison of Official Recorded data and our predicted value on Bangladeshi total data.

Fig. 6. Comparison of Official Recorded data and our predicted value on Indian Total data.
40 days which is compared with the official record shown in the figure. The figures show that our predictions are very close to the officially recorded data. In more observations, the Confirmed case prediction and dead prediction both are very good.

Figure 6(a) shows the number of patients infected with Covid-19 in India. Here we predict the data for the last 40 days which is compared with the official record which is shown in the figure. Figure 6 (b) shows the number of deaths of patients infected with Covid-19 in India. Here also we predict the data for the last 40 days which is compared with the official record shown in the figure. The figures show that our predictions are very close to the officially recorded data. In more observations, the Confirmed case prediction and dead prediction both are very good.

Table IV shows the prediction data and the official Record data in Bangladesh for September 25 to October 4 and the percentage errors are also shown. Here we see that the total Confirmed Case Prediction Error is around -0.65% and the total death prediction error is around -0.22%.

Table V shows the prediction data and the official Record data in India for September 25 to October 4 and the percentage errors are also shown. Here we see that the total Confirmed Case Prediction Error is around -0.80% and the total death prediction error is around -0.45%.

Table VI shows that LSTM achieve the lowest error rate among other models such as Autoregressive Moving Average(ARIMA), Autoregressive Model(AR), and Moving Average(MA). LSTM is very suitable for very short data. Recently LSTM and ensemble methods are used in short-term and long-term prediction respectively. LSTM can run on very low data and provide satisfactory accuracy. It is very efficient that’s the main reason we believe. For evaluation, we use RMSE(Root Mean Square Error), MAE(Mean Absolute Error), and MAPE (Mean Absolute Percentage Error). For calculating those values we use equations 10, 11, 12.

IV. CONCLUSION AND FUTURE WORK

According to a research report, [15], the effects of the Covid-19 epidemic are threatening to bring Bangladesh’s poverty rate back to 15 years ago. In India, an estimated 140 million people lost their jobs during the lockdown [16], and many more lost their salaries. A prediction gives an estimation of the future. Predictions can be effective in helping to plan for future uncertain and potential developments. Our proposed LSTM method predicts covid-19 infection on a daily basis and total infection all over in two countries. The necessary steps for the coming days can be taken through these predicted values. Despite having a small dataset, the LSTM method is making accurate estimations. Proposed LSTM is also a handful for country-specific datasets and predictions.

ACKNOWLEDGMENT

This work was supported in part by the Center for Research, Innovation, and Transformation (CRIT) of Green University of Bangladesh (GUB).

REFERENCES


Md. Moshiur Rahman was born in Kashipur, Barishal, in 1998. He received his B.Sc. from Patuakhali science and technology University and now studying M.Sc. at the Institute of Information technology, University of Dhaka. At present, he is working as an Assistant Professor in the Department of Computer Science and Engineering at Green University of Bangladesh (GUB) from February 2021 to present. He is a memb-er of the Machine learning and artificial intelligence research cell. His research interest includes data science, unsupervised learning, NLP, and software engineering.

Ms. Sumaiya Kabir was born in Barisal, Bangladesh, in 1989. She received the B.Sc. in Computer Science and Engineering degree from Patuakhali Science and Technology University (PSTU) in 2012 and M.Sc. in CSE from East West University (EWU) in 2017. At present she is working as an Assistant Professor of department of Computer Science and Engineering in Green University of Bangladesh (GUB) from 2013 to present. She is a member of Systems and Security research cell. Her research interest includes semantic web, web 3.0 architecture, ontology designing and semantic knowledge engineering.