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 confrmed 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 benefcial 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 confrmed cases per day, number of deaths per day, number of total confrmed 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 benefcial tool for administrators and health offcials.

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

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.
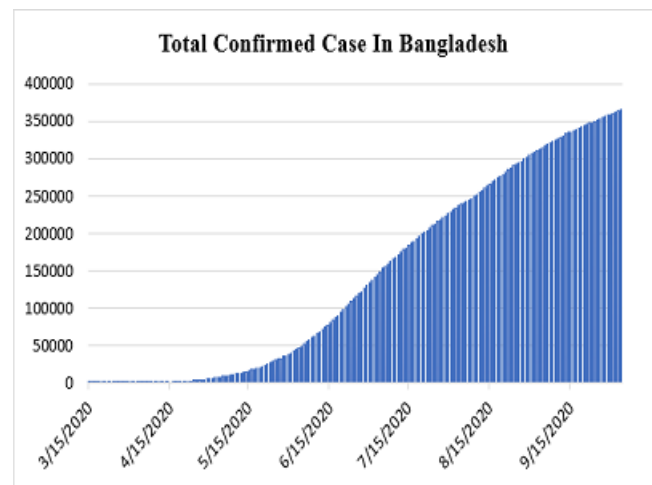

Fig. 1. No of total confirmed cases in Bangladesh.

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
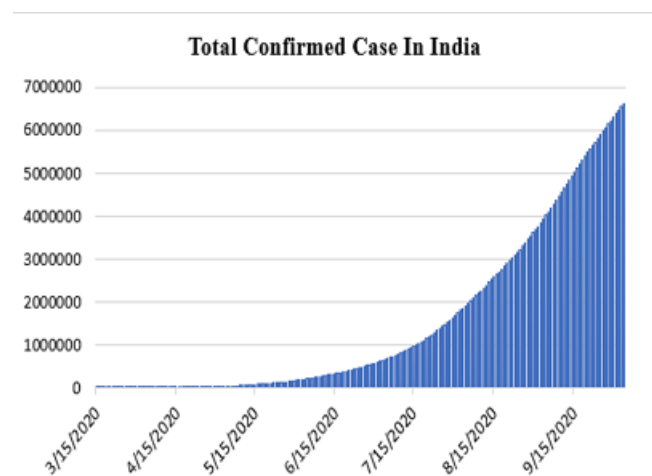

Fig. 2. No of total confirmed cases in India.

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

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:

$$h_t = H(w_{hp}p_t + w_{hh}h_{t\text{-}1} + V_h) \tag{1}$$

$$p_t = w_{hq}q_{t\text{-}1} + V_q \tag{2}$$

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.

$$I_t = \sigma(w_{ip}p_t + w_{hh}h_{t\text{-}1} + w_{ic}c_{t\text{-}1} + V_i) \tag{3}$$

$$f_t = \sigma(w_{hp}p_t + w_{hh}h_{t\text{-}1} + w_{fc}c_{t\text{-}1} + V_h) \tag{4}$$

$$c_t = f_t * c_{t\text{-}1} + i_t * g(w_{cp}p_t + w_{hh}h_{t\text{-}1} + w_{cc}c_{t\text{-}1} + V_c) \tag{5}$$

$$o_t = \sigma(w_{op}p_t + w_{hh}h_{t\text{-}1} + w_{oc}c_{t\text{-}1} + V_o) \tag{6}$$

$$h_t = o_t * h(c_t) \tag{7}$$

Where σ 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].

$$\sigma(p) = \frac{1}{1 + e^p} \tag{8}$$

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.

$$Error = \frac{OfficialValue - PredictedValue}{OfficialValue} \times 100\% \tag{9}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(OfficialValue - PredictedValue)^2}{n}} \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{OfficialValue - PredictedValue}{OfficialValue}\right| \tag{11}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|OfficialValue - PredictedValue\right| \tag{12}$$

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
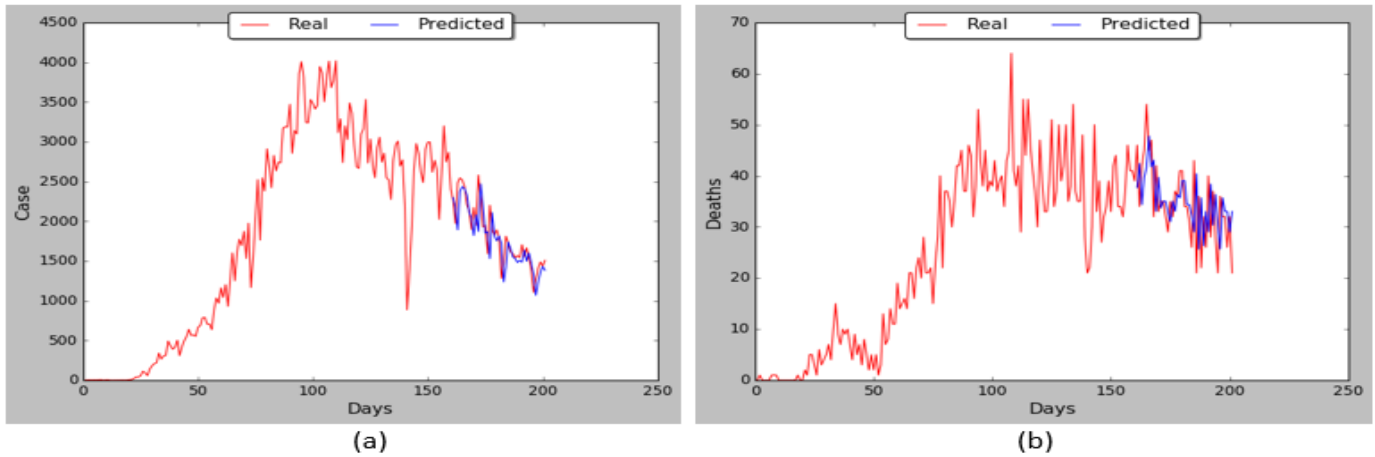
Fig. 3. Predicted value on the basis of daily data compared with official recorded data of Bangladesh.

TABLE I
THE PARAMETERS FOR LSTM RNN TRAINING

Parameter name	Value
Learn rate	0.001
Number of epochs	5000
Batch size	10
Units in LSTM hidden layer	3
Max iteration	50
Momentum	0.8

TABLE II
COMPARISON BETWEEN OFFICIAL RECORD (OR) AND PREDICTED VALUE
(PV) FOR BANGLADESH.

Daily Confirmed Case				Daily Deaths		
Date	OR	PV	Error	OR	PV	Error
25/9/2020	1383	1301	1.95	21	20	4.76
26/9/2020	1106	1069	3.34	36	29	19.44
27/9/2020	1275	1399	-1.96	32	34	-6.25
28/9/2020	1407	1489	-5.82	32	27	15.62
29/9/2020	1488	1504	-1.07	26	22	15.38
30/9/2020	1436	1476	-2.78	32	33	-3.12
1/10/2020	1508	1498	0.66	21	31	-47.61
2/10/2020	1396	1450	-3.86	33	33	0.0
3/10/2020	1182	1239	-4.82	20	25	-25.0
4/10/2020	1125	1189	-5.68	23	29	-26.08

TABLE III
COMPARISON BETWEEN OFFICIAL RECORD (OR) AND PREDICTED VALUE
(PV) FOR INDIA

Daily Confirmed Case				Daily Deaths		
Date	OR	PV	Error	OR	PV	Error
9/25/2020	85468	85057	0.48	1093	1067	2.34
9/26/2020	89010	84611	4.94	1124	1099	2.14
9/27/2020	82767	88117	-6.46	1040	1049	-0.92
9/28/2020	69671	71937	-3.25	777	821	-5.66
9/29/2020	80500	75975	5.62	1178	1095	6.98
9/30/2020	86748	79694	8.13	1179	1097	6.89
10/1/2020	81693	85878	-5.12	1096	1098	-0.24
10/2/2020	79974	80874	-1.12	1071	1024	4.33
10/3/2020	75479	79173	-4.89	937	1002	-6.96
10/4/2020	74767	74724	0.06	902	882	2.15

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
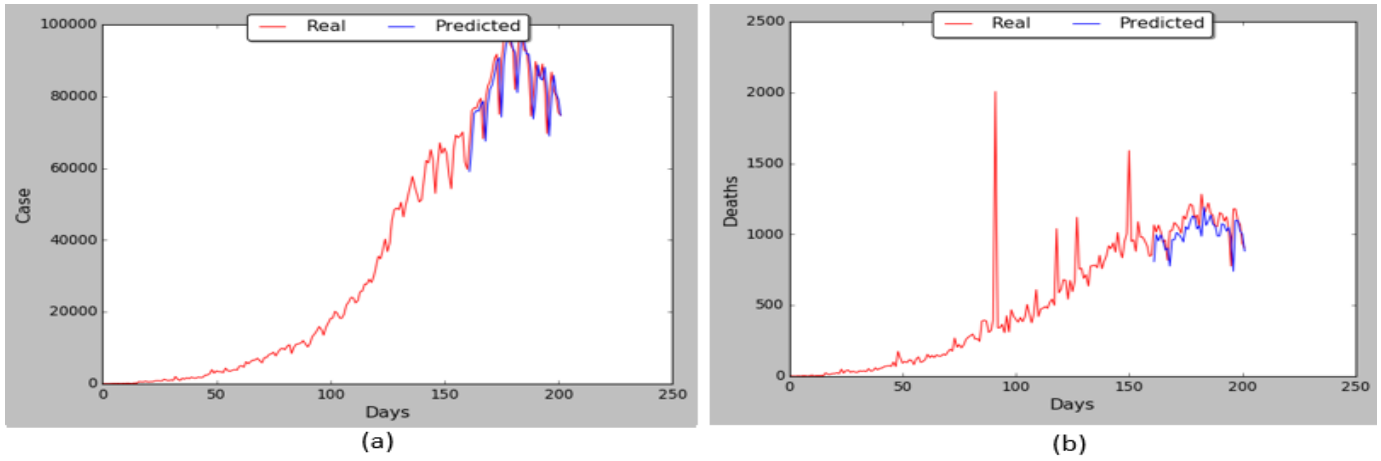
Fig. 4. Predicted value on the basis of daily data compared with official recorded data of India.
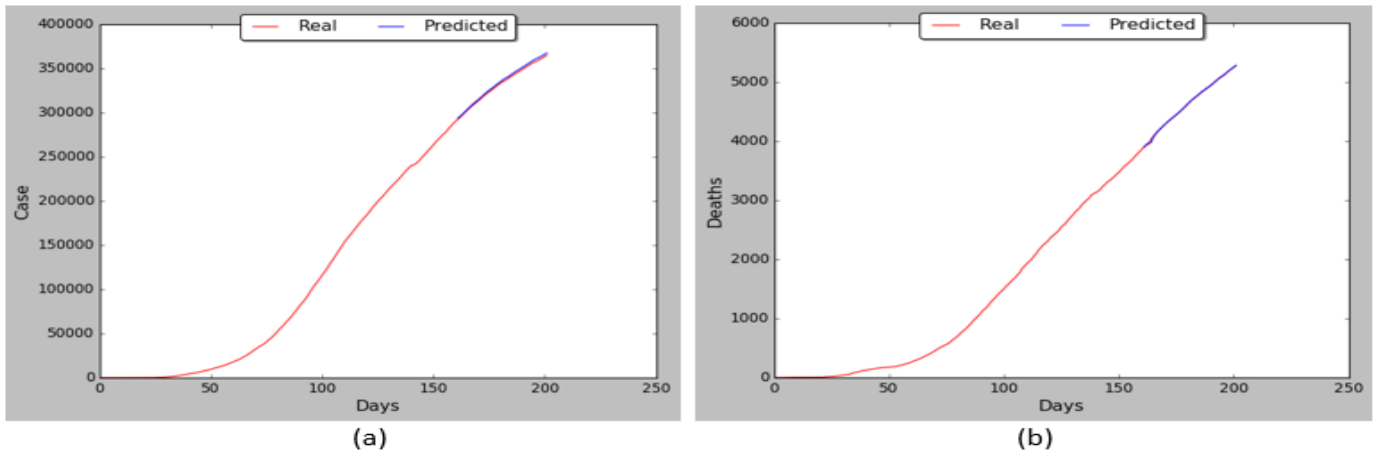

Fig. 5. Comparison of Official Recorded data and our predicted value on Bangladeshi total data.
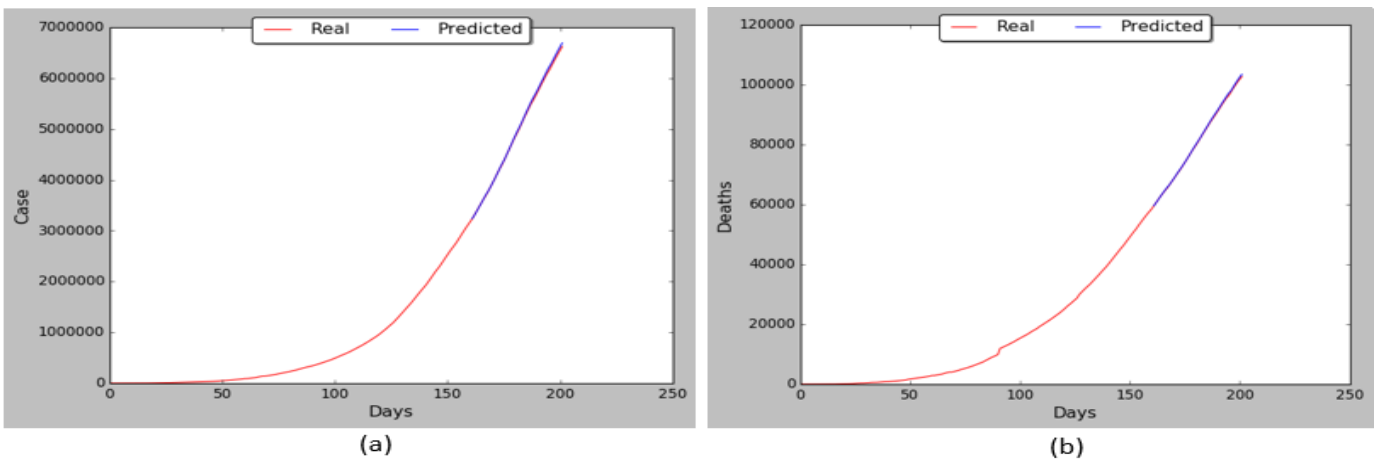

Fig. 6. Comparison of Official Recorded data and our predicted value on Indian Total data.

TABLE IV
COMPARISON BETWEEN OFFICIAL RECORD (OR) AND PREDICTED VALUE
(PV) FOR BANGLADESH.

	Total Confirmed Case			Total Deaths		
Date	OR	PV	Error	OR	PV	Error
9/25/2020	356767	359431	-0.74	5093	5105	-0.23
9/26/2020	357873	360829	-0.82	5129	5126	-0.05
9/27/2020	359148	361948	-0.77	5161	5162	-0.02
9/28/2020	360555	363237	-0.74	5193	5194	-0.02
9/29/2020	362043	364660	-0.72	5219	5226	-0.47
9/30/2020	363479	366164	-1.13	5251	5252	-0.02
10/1/2020	364987	367616	-0.72	5272	5284	-0.22
10/2/2020	366383	368512	-0.58	5305	5375	-1.32
10/3/2020	367565	368432	-0.24	5325	5388	-1.18
10/4/2020	368690	368543	0.04	5348	5369	-0.39

TABLE V
COMPARISON BETWEEN OFFICIAL RECORD (OR) AND PREDICTED VALUE
(PV) FOR INDIA

	Total Confirmed Case			Total Deaths		
Date	OR	PV	Error	OR	PV	Error
9/25	5901571	5945817	-0.74	93410	93817	-0.43
9/26	5990581	6033189	-0.71	94534	94927	-0.41
9/27	6073348	6124183	-0.83	95574	96067	-0.51
9/28	6143019	6208794	-1.07	96351	97123	-0.80
9/29	6223519	6280018	-0.91	97529	97911	-0.39
9/30	6310267	6362311	-0.82	98708	99107	-0.40
10/1	6391960	6450993	-0.92	99804	100304	-0.50
10/2	6471934	6534506	-0.96	100875	101416	-0.53
10/3	6547413	6616262	-1.05	101812	102503	-0.67
10/4	6622180	6693423	-1.07	102714	103454	-0.72

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 VI
ALL PROPOSED MODELS' PERFORMANCE TO FORECAST THE TOTAL
NUMBER OF CONFIRMED CASES

Model	RMSE	MAE	MAPE
ARIMA	115.5	102.5	4.88
AR	145.34	88.45	6.49
MA	88.95	79.34	7.45
LSTM	65.12	56.1	4.37

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%.

TableVI 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.

REFERENCES

[1] "Coronavirus disease 2019," wikipedia, 2020, Accessed on : June 11, 2020. [Online]. Available: {https://en.wikipedia.org/wiki/Coronavirusdisease2019/09-01-2020}

[2] "Covid-19 Data," World Health Organization: Geneva, Switzerland, 2020, Accessed on : June 4, 2020. [Online]. Available: {https://www.who.int/data/gho/09-01-2020-who-Covid-19Data}

[3] "covid-19-free-online-scientific-resources," the Graduate Institute of International and Development Studies (IHEID) and the University of Geneva (UNIGE), 2020, Accessed on : June 5, 2020. [Online]. Available: {https://cerahgeneve.ch/resources/covid-19-free-online-scientific-resources/09-01-2020}

[4] "Covid -19 Data," Worldometers,Dover, Delaware, U.S.A, 2020, Accessed on : June 4, 2020. [Online]. Available: {https://www.worldometers.info/coronavirus/country/bangladesh/}

[5] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Applied energy*, vol. 212, pp. 372–385, 2018.

[6] A. Tomar and N. Gupta, "Prediction for the spread of covid-19 in india and effectiveness of preventive measures," *Science of the Total Environment*, p. 138762, 2020.

[7] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016, pp. 324–328.

[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[9] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[10] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in science & engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[11] D. Zeng, "Workshop 3: Deep learning with python," 2019.

[12] "covid-19-coronavirus," statista, 2020, Accessed on : 31 May, 2020. [Online]. Available: {https://en.wikipedia.org/w/index.php?title= COVID-19pandemicinBangladesholdid=961944330}

[13] "COVID-19 pandemic in Bangladesh," Wikipedia contributors, 2020, Accessed on : 30 May, 2020. [Online]. Available: {https://www.statista. com/page/covid-19-coronavirus}

[14] "COVID-19 pandemic in India," Ourworldindata, 2020, Accessed on : 2 June, 2020. [Online]. Available: {https://ourworldindata.org/ coronavirus/country/india?country=~IND}

[15] R. E. Hall, C. I. Jones, and P. J. Klenow, "Trading off consumption and covid-19 deaths," National Bureau of Economic Research, Tech. Rep., 2020.

[16] B. Gururaja and N. Ranjitha, "Socio-economic impact of covid-19 on the informal sector in india," *Contemporary Social Science*, vol. 17, no. 2, pp. 173–190, 2022.

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 mem- ber 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 Pat- uakhali 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 As- sistant 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.