Research Article

Koyel Datta Gupta, Rinky Dwivedi, and Deepak Kumar Sharma*

Predicting and monitoring COVID-19 epidemic trends in India using sequence-to-sequence model and an adaptive SEIR model

https://doi.org/10.1515/comp-2020-0221
received June 15, 2020; accepted January 18, 2021

Abstract: In the year 2019, during the month of December, the first case of SARS-CoV-2 was reported in China. As per reports, the virus started spreading from a wet market in the Wuhan City. The person infected with the virus is diagnosed with cough and fever, and in some rare occasions, the person suffers from breathing inabilities. The highly contagious nature of this corona virus disease (COVID-19) caused the rapid outbreak of the disease around the world. India contracted the disease from China and reported its first case on January 30, 2020, in Kerala. Despite several counter measures taken by Government, India like other countries could not restrict the outbreak of the epidemic. However, it is believed that the strict policies adopted by the Indian Government have slowed the rate of the epidemic to a certain extent. This article proposes an adaptive SEIR disease model and a sequence-to-sequence (Seq2Seq) learning model to predict the future trend of COVID-19 outbreak in India and analyze the performance of these models. Optimization of hyper parameters using RMSProp is done to obtain an efficient model with lower convergence time. This article focuses on evaluating the performance of deep learning networks and epidemiological models in predicting a pandemic outbreak.

Keywords: COVID-19, sequence to sequence, SEIR, transmission rate

1 Introduction

As on May 30, 2020, India has the highest number of SARS-CoV-2 cases in Asia, with total cases counting up to 173,763 cases, 4,971 deaths, and 82,370 recoveries [1]. The fatality rate is near 3.09%. The first confirmed case of India began in Kerala at the end of January 2020 and has since reported a total of 33,050 cases till April 30, 2020. In the month of March, the basic reproduction rate (R0) of the SARS-CoV-2 in India was around 1.7 [2], which was on the lower side with respect to the major affected countries. Now, the reproductive rate ranges between 1.22 and 1.23. However, according to some beliefs, the lower transmission rate of the virus can be due to the lack of mass testing facility in India. Till February 3, 2020, three confirmed COVID-19 cases were reported in India. On February 20, 2020, the patients in Kerala recovered and were discharged. However, by March 9, 2020 the number of confirmed SARS-CoV-2-infected patients in India became 44 with 4 from New Delhi, the capital of India. Six new cases were reported in Kerala, 14 in Haryana, 9 in Uttar Pradesh, 2 cases each in Ladakh and Rajasthan, and 1 each in Tamil Nadu, Punjab, Telangana, and Jammu including 16 cases of foreign patients. The World Health Organization declared the outbreak of COVID-19 as a pandemic. By March 14, 2020 India reported 84 corona positive cases with two fatalities. India being the second highest populous country needed to act fast to control the COVID-19 outbreak. Initially government started by shutting down educational institutes and placed a stop on tourism. A voluntary curfew was observed on March 22, 2020, on the request of Shri Narendra Modi, the Prime Minister of India. With the continuous rise in the number of COVID-19 patients, the Prime Minister of India ordered a complete India-wide lockdown, under section 6(2)(i) of the disaster management Act, 2005, effective from March 25, 2020. By April 12, 2020, the number of confirmed cases rose to 8,447 with 273 deaths. On April 14, the lockdown was extended till May 3. The four states with a maximum number of COVID-19 cases are Maharashtra,

---

*** Corresponding author: Deepak Kumar Sharma,** Department of Information Technology, Netaji Subhas University of Technology, New Delhi, India, e-mail: dk.sharma1982@yahoo.com
Koyel Datta Gupta: Department of Computer Science and Engineering, Maharaja Surajmal Institute of Technology, New Delhi, India, e-mail: koyel.dg@msit.in
Rinky Dwivedi: Department of Computer Science and Engineering, Maharaja Surajmal Institute of Technology, New Delhi, India, e-mail: rinkydwivedi@msit.in

Tamil Nadu, Delhi, and Gujarat. The lockdown has further been extended till May 31, 2020. However, since April 20, 2020, certain relaxation has been provided to the places that have restricted the spread of the virus. The lockdown may have reduced the rate of spread of SARS-CoV-2, but has hit the economy of India badly. Even with the tremendous effort from government and other sources to ease the impact of lockdown on daily wagers, small-scale entrepreneurs, and others who have lost their source of income, it has become an extremely difficult situation. India is facing huge financial challenges due to the loss of business in international and domestic markets, loss in revenues, crisis in tourism and other industries, and the economic liability for controlling COVID-19 in terms of management and treatment. The trajectory of the pandemic already has a critical impact on the Indian economy. Now with the restrictions of the lockdown being removed gradually in India, the outbreak is observing a steep rise in the number of COVID-19 cases. The forecast of the number of cumulative COVID-19 patients in India is the primary objective of this article. India being a densely populated country, it is important to understand the pattern of the outbreak under extreme measures taken by Government to curb the virus menace. To tackle the financial difficulties, states of the nation have eased the restrictions on certain business activities. It is important to foresee the results of these actions on the pandemic spread in the country. This article focuses on the prediction of the basic transmission rate of COVID-19 in India, projecting the possible number of cases by mid-June and the challenges involved. For predicting any temporal data, machine learning algorithms is usuallya primary choice. In this article, we have modeled an efficient Seq2Seq model designed optimally by deploying root mean square propagation (RMSProp) technique. This model relies on the training from the previous records of COVID-19 cases. However, to understand the reproduction rate of the virus, we use an adaptive SEIR model to forecast the reproductive number of SARS-CoV-2 in India. With the help of the results obtained, an analysis of the performance of both the methods is estimated in terms of root-mean-squared error (RMSE). The next task is analyzing the performance of a deep learning model and a disease model and how the change in Government policies is affecting the pandemic curve. The rest of this article is organized as follows. The related works are presented in Section 2. Section 3 proposes the methodology. Section 4 shows the simulation. The research work is concluded in Section 5.

2 Related works

The last few decades have witnessed the outbreak of many zoonotic diseases, which were highly infectious and fatal in many cases. Several studies were presented for predicting the outbreak of these diseases. The researchers have used machine learning models, traditional disease models, and even statistical models to estimate the outbreak of the epidemics. An adaptive neuro-fuzzy inference system combining the properties of ANN and fuzzy logic systems is proposed in ref. [3] for temporal data prediction. A mathematical model is presented in ref. [4] to monitor the infection trends of the SARS epidemic during 2003. A comparative study of four temporal data forecasting models is proposed in ref. [5]. The study uses 13 years records of Hepatitis A infection in Turkey and estimates the performance of prediction of the disease using different ANN models (MLP, time-delay neural networks, and RBF) and statistical model (ARIMA). The article concludes that MLP performs best in the given condition. The article [6] presents a Bayesian inference-based dynamic model to anticipate the deadly Ebola virus spread in African nations. In ref. [7], the authors proposed a model to forecast West Nile virus (WNV) transmission risk using the records of the WNV spread in long Island during 2001–2014. For monitoring the outbreak of influenza A (H1N1-2009), a forecasting model is proposed in ref. [8]. In ref. [9], a probability-based model is deployed to predict the worldwide spread of the MERS. The authors in ref. [10] emphasized the use of IoT, blockchaining, big-data analytics, AI, and deep learning to build a prediction model for COVID-19 and estimate its effect on the healthcare sector. In ref. [11], an ARIMA model is deployed in to forecast the spread of COVID- 2019. To estimate the disease reproductive rate, many researchers have deployed several epidemiological models. The SIR and SEIR is a disease model [12] with independent variable time and dependent variables which is the population expressed as a function of time. The population can be categorized as susceptible, infected and removed and in some cases as exposed. A logistic and exponential model is presented in ref. [13] for disease prediction. In ref. [14], the serial interval of MERS and SARS to assess the serial interval of COVID-19. The authors in ref. [15] assess the rate of virus reproduction using a Reservoir-People transmission network model for analyzing the transmission pattern of the disease. The work [16] proposes a delay adaptive SIR model that and evaluates the performance with various refrainment strategies. In ref. [17], the nature of exposure to the virus is utilized in computing its impact on patients. A modified SEIR model is used in ref. [18] with a moving population for susceptible and exposed category. Both the articles [19,20] propose modified mathematical and statistical models to forecast COVID-19 trends. The authors in ref. [21] modified the SIR model by including variable denoting the death of patients. The aims of ref. [22] are to predict COVID-19 and to identify individuals who may exhibit severe symptoms. The authors in ref. [23] build two models to improve the performance of CNN and ConvLSTM for COVID-19 detection, using data augmentation. A flower pollination-based optimized GRNN model is proposed in ref. [24] for the prediction of COVID-19 in Europe. The authors in ref. [25] presented deep learning techniques based on CNN to determine SARS-CoV-2 using X-ray and CT images and claims to achieve an accuracy of 100.

3 Methodology

The highly contagious characteristic of the novel corona virus makes it even more deadly. To study the future of this pandemic in India, the dataset has been collected from several sources. After the first confirmed case in January 2020, there were only three confirmed cases of COVID-19 in India during the entire month of February 2020. Even with the government-imposed social distancing, the number of cases started rising abruptly from the end of March and is still continuing. To estimate the pattern of this time series, we have used two approaches. The first approach tries to estimate the reproduction rate of the virus. The varied scenario in India and diverse measure, taken to restrict the epidemic outbreak have inspired us to choose an epidemiological model (adaptive SEIR) to predict the transition of different stages of SARS-CoV-2 transmission. The second approach uses an efficient Seq2Seq model optimally configured using the RMSProp for the faster convergence rate.

3.1 Adaptive SEIR

The SEIR model used in his work is modified to accommodate the changing conditions in India. Unlike many countries, India has started taking measures like isolation of infected and exposed population and lockdown for maintaining social distancing among entire population to reduce the rate of COVID-19 transmission. Another important factor is that the testing for the disease is quite low in India, which may mean that many asymptotic patients are undetected for the virus. We have incorporated the concept of herd immunity. The infected person in this case does not exhibit any symptoms of the virus and eventually develops antibodies for the virus. But these undetected patients can still spread the infection to other susceptible individuals. The overall population (75% of the total population of the country) is divided (shown in Figure 1) into susceptible ($S \in$ population who have neither been exposed nor contracted the disease), exposed ($E \in$ population who have been in contact with an infected person, may have been identified for isolation, not yet infected, considering incubation period of the virus), infected ($I \in$ population who are infected, may be asymptomatic or otherwise, may have been detected or otherwise), and removed ($R \in$ population who have recovered or succumbed to the disease). The birth and death rates have been assumed to be a constant.

To explain the working of the adaptive SEIR model, the essential symbolic notations used in the equations are described in Table 1.

Definition 1. The rate of change in susceptible population is given by the product of the rate of transmission from stage S to E stage and subsequently to I stage due to contact between a susceptible subject (S) and an infected (I).

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{-\beta SI}{N}. \tag{1}$$

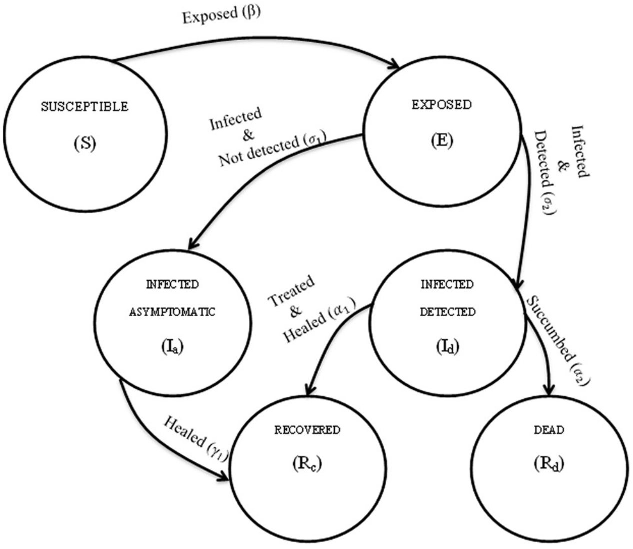

Figure 1: State transition diagram depicting transmission stages of COVID-19.

Table 1: Symbolic notation for adaptive SEIR

S. No	Symbol	Details
1	β	Rate of transmission from S to E stage
2	σ_1	Rate of incubation and the developing asymptomatic infection
3	σ_2	Rate of incubation and the developing of infection with symptoms
4	γ_1	Rate of recovery for asymptomatic cases
5	α_1	Rate of recovery for detected symptomatic cases
6	α_2	Rate of death for detected symptomatic cases
7	R_0	Basic rate of reproduction
8	N	Total population

Definition 2. The rate of change of exposed population is given by the fraction of susceptible population transforming to exposed (equation (1)) minus the fraction of exposed person who becomes infective (asymptomatic and symptomatic).

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\beta \mathrm{SI}}{N} - \sigma_1 E - \sigma_2 E. \tag{2}$$

Definition 3. The rate of change of infected population (asymptomatic) is given by the fraction of the exposed population who becomes infected without exhibiting any symptoms (equation (2)) minus the fraction of asymptomatic infected persons who have become resistant/healed (considering the concept of herd immunity).

$$\frac{\mathrm{d}I_a}{\mathrm{d}t} = \sigma_1 E - \gamma_1 I_a. \tag{3}$$

Definition 4. The rate of change of infected population (symptomatic) is given by the fraction of the exposed population who becomes infected, exhibiting symptoms of viral infection (equation (2)) minus the fraction of these infected persons who either become resistant/healed or succumbs to the disease.

$$\frac{\mathrm{d}I_d}{\mathrm{d}t} = \sigma_2 E - \gamma_2 I_d \; (I = I_a + I_d). \tag{4}$$

Definition 5. The rate of change of resistant/recovered population is given by the fraction of asymptomatic infected population and detected infected population of viral infection who gets cured and become resistant to the virus (equations (3) and (4)).

$$\frac{\mathrm{d}R_c}{\mathrm{d}t} = \gamma_1 I_a + \alpha_1 I_d. \tag{5}$$

Definition 6. The rate of change of the deceased population is given by the fraction of the detected infected population of viral infection who succumbs to the disease (equation (4)).

$$\frac{\mathrm{d}R_d}{\mathrm{d}t} = \alpha_2 I_d \quad (\gamma_2 = \alpha_1 + \alpha_2). \tag{6}$$

Definition 7. The rate of reproduction/transmission of the virus is given by the ratio of the rate of transmission from S to E and subsequently to I and the rate of recovery and fatality of the infected population.

$$R_0 = \frac{\beta}{\gamma_1 + \gamma_2}. \tag{7}$$

3.1.1 Algorithm 1

Algorithm 1: Adaptive SEIR

```
1: Initialize the variable $N$ to the total population of India. The initial values of susceptible population is computed as $S = (N - I)/N$.
2: Set $maxdays = 132$
3: Initialize the values of $\beta$, $\gamma_1$, $\gamma_2$, $\sigma_1$, and $\sigma_2$
4: The change in the number of susceptible, exposed, infected, and recovered/dead population is calculated using equations (1)–(6)
5: Step 3 is repeated for $maxdays$ and fitted to existing data and the values of $\beta$, $\gamma_1$, $\gamma_2$, $\sigma_1$, and $\sigma_2$ are updated.
6: The reproduction rate $R_0$ is computed using equation (7).
7: The values of are used to predict the values of infected people in the month of June.
8: The method is validated using "holdout."
```

3.2 Efficient sequence-to-sequence algorithm

Deep learning is a machine learning technique that builds models to imitate human reasoning and learning. Models are trained to perform the task of classification or prediction by using a large dataset and multi-layered neural network architectures. The recurrent neural network (RNN) is a category of deep neural networks that has feedback loops within the network enabling them to store past record to train previous series [26]. The basic RNN model experiences the problems of vanishing and exploding gradient. In the case of long sequences, the back-propagated gradient values get reduced, which causes failure in updates of the weights and large gradients may explode for long temporal sequences generating unstable weights. The long-short-term-memory (LSTM) cell was introduced in ref. [27], which are capable of saving long-term dependencies in temporal sequences. These RNNs with LSTM core are capable of automatic feature extraction and are efficient in time series prediction since they integrate the knowledge of related information from the previous inputs. These networks work well for non linear patterns. This type of RNNs exhibit better prediction accuracy due to their extended range dependence [28]. A single LSTM unit consists of a memory unit/cell and three gates, the input, the output, and the forget. Activation functions are applied at the gates. A high activation value at the input signifies that the input information needs to be saved in the memory, a high value at the output triggers the release of the stored data to the next neuron, and a high value at the forget gate erases the data from the memory unit. Now, sequence-to-sequence (Seq2Seq) [29] architecture (Figure 2) was initially used for neural machine translation. The model is suitable for ML tasks for time series prediction. A Seq2Seq model has three modules including an encoder, an intermediate vector, and a decoder. LSTM cells constitute the encoder. The input sequence is presented to the encoder which extracts features of the input elements, and the last hidden state is called the context vector. This vector encodes the entire information from the input data. Multiple LSTM units constitute the decoder, and each unit calculate its individual hidden state and generates an output data. This article proposes applying the Seq2Seq model to predict the COVID-19 curve in India. To explain the encoder–decoder model used in this study, the essential symbolic notations used in the equation are presented in Table 2.

3.2.1 Encoder

It consists of one or more LSTM units each accepting input sequence, assimilating information, and forwarding it. The equation for the encoder [30] is given as follows:

$$h_t = f(w^{hh} * h_{t-1} + w^{hx} * x_t). \tag{8}$$

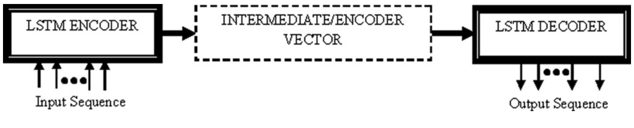

Figure 2: Basic sequence to sequence model.

Table 2: Symbolic notation for sequence-to-sequence

S. No	Symbol	Details
1	x_t	Input vector to the LSTM encoder
2	h_t	Hidden state vector of the encoder
3	s_t	Hidden state vector of the decoder
4	y_t	Output vector of the decoder
5	w^{hx}	Weights integrating input vector information
6	w^{hh}	Weights integrating information from the previous time-step
7	w^s	Weights at the last hidden state
8	v, w	Weight vectors
9	c_t	Semantic vector at time t
10	e_{ti}	Degree of influence of the hidden state h_i at time t
11	ζ_{ti}	Normalized value of the Softmax function

3.2.2 Intermediate vector

The last hidden state of the encoder generates the intermediate vector using equation (8). This vector aims to assimilate the source sequence information and acts as the decoder component's first hidden state

3.2.3 Decoder

It consists of one or more LSTM units. Each LSTM cell is provided the trailing hidden state and generates output and current hidden state. The hidden state at current step, *t* i.e., *ht* is computed by the following equation [30]:

$$s_t = f(w^{hh} * s_{t-1}). \tag{9}$$

The output at time step is given by the equation [30]:

$$y_t = f(w^s * s_t). \tag{10}$$

Softmax is used to generate the output sequence. An attention mechanism can be used to save the output sequence of the encoder's source. Bahdanau attention [31] arranges the input and output sequences, with an alignment value. It ensures that the most important information in the input receives attention.

$$c_t = \sum_{i=1}^{T}(\zeta_{ti} * h_i). \tag{11}$$

$$\zeta_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{T}\exp(e_{tk})}. \tag{12}$$

$$e_{ti} = v^T * \tanh(w[s_{t-1}, h_i]). \tag{13}$$

$$s_t = \tanh(w[s_{t-1}, y_{t-1}, c_t]). \tag{14}$$

RMSprop is used to optimize the learning rate of the LSTM. Moving average of gradient-square to normalize. This enables to increase the step size vanishing gradient and to decrease the step for large gradient.

3.2.4 Algorithm 2

Algorithm 2: Sequence-to-sequence

```
1: Load input data.
2: Built the Seq2Seq model using equations (8)–(10).
3: Bahdanau attention (equations (11)–(14)) is applied
   on the encoder outputs.
4: Split the input data into training and testing set
   as 3:1.
5: During training RMSProp is used to optimize the
   LSTM units.
6: After the completion of the training phase, testing
   data is loaded and the corresponding output is
   analyzed. The validation is done using “holdout”
   method.
7: The final model is used for prediction of the pan-
   demic curve in India in the month of June.
```

4 Simulation

With the first COVID-19 patient detected on January 30, India as on 9 June has 2, 70,876 confirmed cases. Figure 3 shows the number of total, currently active, recovered, and death cases in India from January 30, 2020, to June 9,2020.

Figure 4a and b depicts the recovery rate and death rate till June 9, 2020.

The dataset of COVID-19 patients (records till June 9,2020) of India has been used to analyze the percentage of death among infected person on the basis of the age group. However, many patient records have no information related to age or gender. We have tried to find the impact of the age of a COVID-19 patient on their likelihood of death. Figure 5 shows the percentage of death of the infected population (based on only those patient records where information of their age was present). As evident, the population over 60 years is most susceptible to succumb to COVID-19. In India, the patients whose age is below 40 showed less probability of dying die due to SARS-CoV-2 disease. However, due to the lack of complete age information about every corona patient in India, it is difficult to draw a concrete supposition about the relation of the infected individual’s age and the

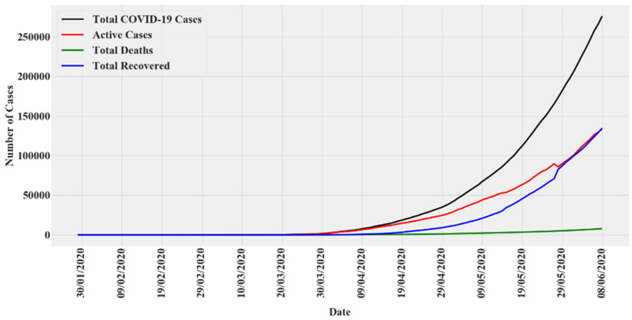

Figure 3: COVID-19 cases in India.

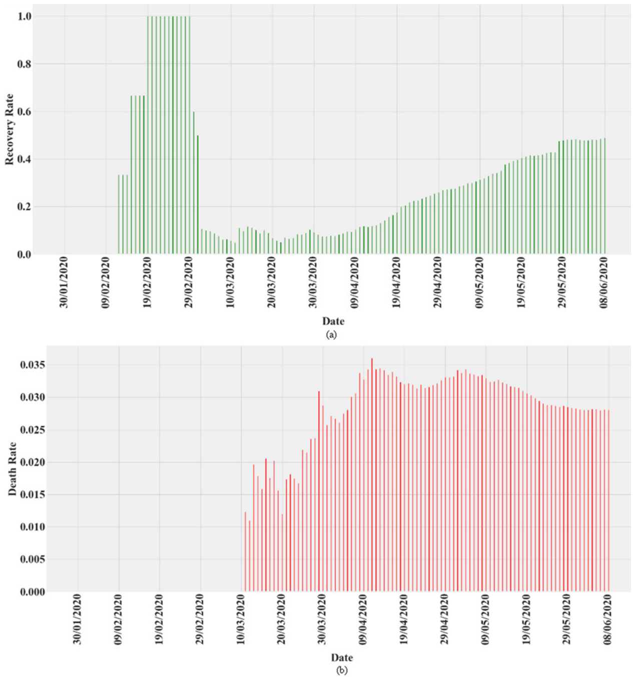

Figure 4: (a) Recovery rate and (b) death rate of COVID-19 in India.

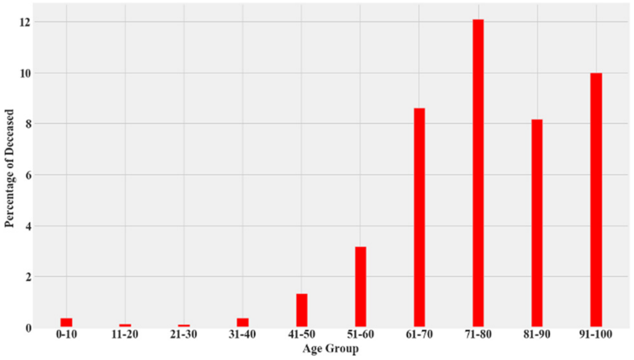

Figure 5: Age-group wise percentage of death of COVID-19 patients in India.

probability of death due to the disease. For building the adaptive SEIR model, the effects of lockdown, isolation of exposed, or infected population and unlocking are considered while forecasting the pandemic pattern in India.

4.1 Adaptive SEIR setup

The proposed model is implemented in Python. The model predicts the daily reported COVID-19 cases in India. The parameters are calculated to fit the COVID-19 data of India. It is assumed that initially the count of the susceptible population $S(N-I)/N \approx N$, where the total population is N and count of the infected population $I = 1$ (as on January 30, 2020). The initial value of R_0 is set to 3.2 at time $t = 0$. The value of β was computed equal to 0.63 and value of 0.31. Equation (7) gives the rate of spread of $R_0 = 2.03$.

4.2 Sequence-to-sequence model set up

The sequence-to-sequence model built in this work has two LSTM encoder units each in the encoder and two LSTM decoder units constitute the decoder. Bahdanau attention is used on the encoder's output. The dropout was set at 0.3. The initial learning rate of the experiment was set to 0.01, and RMSProp was used for optimizing the hyper parameters.

4.3 Experimental results

To estimate the value of the parameters of the modified SEIR model and sequence-to-sequence model, datasets of COVID-19 cases in India are collected from various sources [32–35]. The dataset starts from January 30, 2020 and recorded cases till June 9, 2020. The dataset is split for training and validation for the Seq2Seq model. For training, we used cases from January 30, 2020 till May 9, 2020. The validation of the model is done using data from 10th January to June 8. The adaptive SEIR curve is fitted to the available dataset. The data is also used to train and test LSTM and support vector regression (SVR) model. The RMSE and root mean squared log error values are presented in Table 3.

The prediction is done from June, 9 to June 28, 2020. The number of cases predicted by the SVR, LSTM, and adaptive SEIR and Seq2Seq models is shown in Figure 6.

Table 3: Comparison of adaptive SEIR and Seq2Seq models

Model	RMSE	RMSLE
Seq2Seq	0.69436	0.023
Adaptive SEIR	1316.2575	1.81
LSTM	954.28	0.029
SVR	6843.24	1.69

The projected values of the cumulative number of COVID-19 cases by SEIR and sequence-to-sequence model are 528557 and 470783, respectively.

The number of new cases after June 8, 2020, is depicted in Figure 7. The Seq2Seq model successfully identifies the ups and downs in the number of daily counts of patients. The models predict a value of more than 13,000 COVID-19 by the last week of June.

The cumulative number of predicted death cases due to COVID-19 is predicted in Figure 8. The first death in India was reported on March 12, 2020.

4.4 Discussion

From the results shown, it is seen that the adaptive SEIR model cannot properly depict the actual pandemic curve unlike deep learning algorithms. The simplicity of the disease model causes it to discard information over the span of time. This causes severe lack of ability of epidemiological model to reproduce the pattern of a real pandemic outbreak. The adaptive SEIR model used in this work is able to show the overall pattern of the graph but fails to identify the exact points of daily increase (or decrease) in the number of COVID-19 patients. Though the modified SEIR model showed the change in the curve during periods of restrictions (lockdown) and the effect of

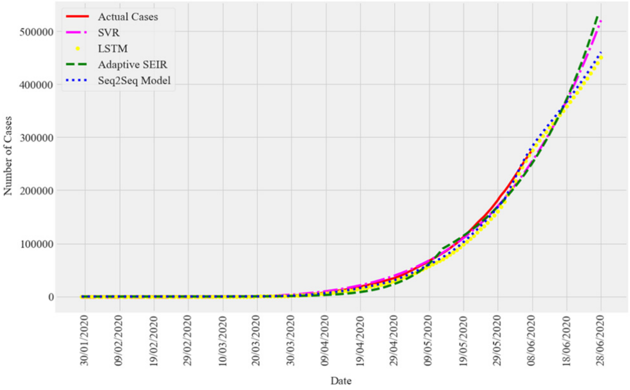

Figure 6: Prediction of total number of COVID-19 cases in India.

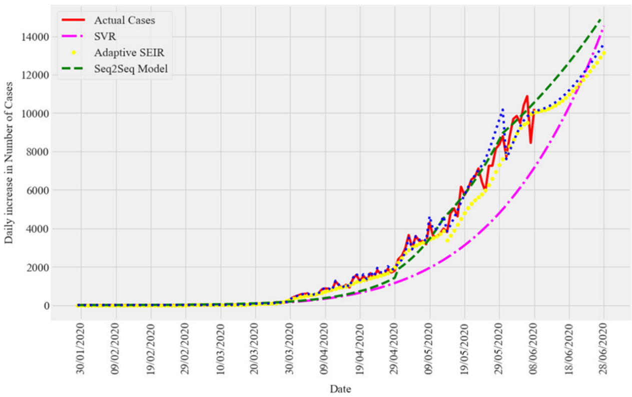

Figure 7: Prediction of day-wise new cases of COVID-19 cases in India.

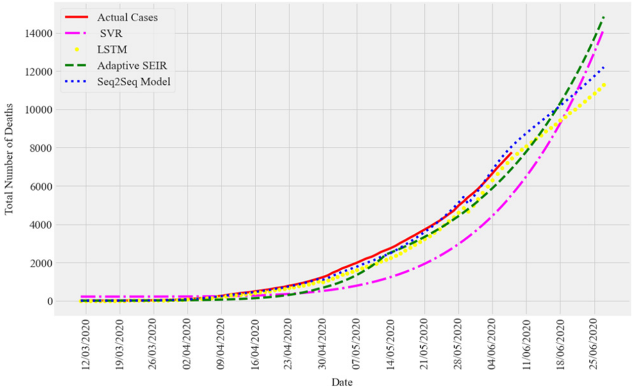

Figure 8: Prediction of cumulative number of death cases caused by COVID-19 in India.

the unlocking phase, the RMSE value of the Seq2Seq model proves a more precise prediction. Seq2Seq converts one data sequence into one more form of sequence. This conversion is achieved by the use of GRU or LSTM encoder–decoder. The context for every input is the output from the earlier stage. The encoder stage converts an input and its context into a hidden vector. The decoder converts the hidden vector into corresponding output data, using the prior output. The Seq2Seq model thus has the capacity to predict a time series data, by studying the influence of past record on present input and output and predict the output sequence. It outperforms other machine learning techniques like LSTM and SVR. The COVID-19 dataset is used to train the LSTM encoder–decoder model to predict the future trend. However, size of the dataset was small, and the accuracy of the deep learning model may be enhanced with more training data. Due to the enhanced learning capacity of the model, the Seq2Seq model has a higher accuracy than the SEIR model and has an RMSE value of 0.69436 as opposed to 1316.2575 of SEIR.

5 Conclusion

The restrictions imposed by Indian government have curbed the exponential growth of SARS-CoV-2 transmission. However, these measures could not flatten the curve as expected because of practical challenges. With the present economic situation, extending the lockdown beyond May 17, 2020, can lead to severe financial implication for the nation and its people. Moreover, people may violate the norms of lockdown and isolation, heightening the chance of community transmission. The study presented here predicted the pattern of the COVID-19 outbreak course and provided the comparison of a disease model and an LSTM-based encoder–decoder model. Modification has been made to the existing SEIR model in terms of the Indian situation. One very important fact in India is the lack of testing kits, and this has made it difficult to detect asymptomatic cases. These asymptomatic population is, however, capable of infecting other susceptible population. The lockdown and social isolation definitely contributed to the decay in the growth of the disease spread pattern. The data of COVID-19 cases obtained are small in dimension, which might be insufficient to fully estimate the underlying disease spread rate. The proposed models were deployed to predict the data for 20 days using the epidemiological model and deep neural networks, respectively. The models were analyzed based on the results, and it is concluded that epidemiological models cannot precisely follow the trends of a disease spread curve and can predict a curve depending on the disease information. The prime reason behind this behavior of the disease model can be attributed to the fact that it cannot save information over a time-length. On the contrary, the LSTM encoder–decoder model has the capability to emphasize past relevant information. The Bahdanau mechanism enables one to focus on the most vital source information. This further enhances the capacity of the network to predict the pandemic curve precisely. The sequence-to-sequence model with attention can be used to forecast the COVID-19 spread curve and take precautionary measures. The change in the future transmission of COVID-19 extensively depends on the government policies and the extent of social restrictions followed by people. The dynamic of COVID-19 is varying frequently due to changing social situations. The outcome of the research shown here relies on the current data and rules imposed by the government. The actual count may differ if there is a change in the current the socio-economic condition.

Funding information: No Funds received.

Author contributions: All authors have accepted for the entire content of this manuscript and approved its submission.

Conflict of interest: There is no conflict of interest.

Ethical approval: The conducted research is not related to either human or animal use.

Data availability statement: All data generated or analyzed during this study are mentioned in this published article. The reference for the same is also provide. https://www.mohfw.gov.in/, https://twitter.com/CMODelhi, https://twitter.com/MoHFW_GUJARAT, https://arogya.maharashtra.gov.in/1175/Novel–Corona-Virus.

References

[1] https://www.who.int/docs/default-source/wrindia/situation-report/india-situation- report-6606711da860b4-d38b266c91265952977.pdf.

[2] One COVID-19 positive infects 1.7 in India, lower than in hot zones, The Indian Express, 19 March 2020.

[3] J. S. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man Cybern.*, vol. 23, pp. 665–685, 1993.

[4] E. Massad, M. N. Burattini, L. F. Lopez, and F. A. Coutinho, "Forecasting versus projection models in epidemiology: The case of the SARS epidemics," *Med. Hypotheses*, vol. 65, pp. 17–22, 2005.

[5] M. Ture and I. Kurt, "Comparison of four different time series methods to forecast hepatitis A virus infection," *Expert Syst. Appl.*, vol. 31, pp. 41–46, 2006.

[6] J. Shaman, W. Yang, and S. Kandula, "Inference and forecast of the current West African Ebola outbreak in Guinea, Sierra Leone and Liberia," *PLoS Curr.*, vol. 6, pp. 1–16, 2014.

[7] N. B. DeFelice, E. Little, S. R. Campbell, and J. Shaman, "Ensemble forecast of human West Nile virus cases and mosquito infection rates," *Nat. Commun.*, vol. 8, pp. 1–6, 2017.

[8] J. B. S. Ong, I. Mark, C. Chen, A. R. Cook, H. C. Lee, V. J. Lee, et al., "Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore," *PLoS One*, vol. 5, pp. 1–11, 2010.

[9] K. Nah, S. Otsuki, G. Chowell, and H. Nishiura, "Predicting the international spread of Middle East respiratory syndrome (MERS)," *BMC Infect. Dis.*, vol. 16, p. 356, 2016.

[10] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, "Digital technology and COVID-19," *Nature Med.*, vol. 26, pp. 459–461, 2020.

[11] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the ARIMA model on the COVID-2019 epidemic dataset," *Data Brief*, vol. 29, p. 105340, 2020.

[12] A. Rachah and D. F. M. Torres, "Analysis, simulation and optimal control of a SEIR model for Ebola virus with demographic effects," *Commun. Fac. Sac. Univ. Ank. Series A1*, vol. 67, pp. 179–197, 2018.

[13] M. Junling, "Estimating epidemic exponential growth rate and basic reproduction number," *Infect. Disease Model.*, vol. 5, pp. 129–141, 2020.

[14] S. Zhao, S. S. Musa, Q. Lin, J. Ran, G. Yang, W. Wang, et al., "Estimating the unreported number of novel Coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak," *J. Clinical. Med.*, vol. 9, no. 2, p. 388, 2020.

[15] T. M. Chen, J. Rui, Q. P. Wang, Z. Y. Zao, J. A. Cui, and L. Yin, "A mathematical model for simulating the phase-based transmissibility of a novel coronavirus," *Infect Dis Poverty*, vol. 9, p. 24, 2020.

[16] F. Casella, *Can the COVID-19 epidemic be managed on the basis of daily data?* Preprint at https://arxiv.org/abs/2003.06967. p. 06967, 2020.

[17] J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, et al. "Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan," *China. Nat. Med.*, vol. 26, pp. 506–510, 2020.

[18] Z. Yang, Z. Zeng, K. Wang, S. S. Wong, W. Liang, M. Zanin, et al., "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *J Thorac Dis.*, vol. 12, no. 3, pp. 165–174, 2020.

[19] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russel, et al. "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *Lancet Global Health*, vol. 8, pp. e488–e496, 2020.

[20] A. J. Kucharski, T. W. Russel, C. Diamond, Y. Liu, J. Edmunds, S. Funk, et al."Early dynamics of transmission and control of COVID-19: a mathematical modelling study," *Lancet Infect Dis.*, vol. 20, no. 5, pp. 553–558, 2020.

[21] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the COVID-19 outbreak," *PLoS One*, vol. 15, p. e0230405, 2020.

[22] X. Jiang, M. Coffee, A. Bari, J. Wang, X. Jiang, J. Huang, et al., "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *CMC-Comput Mater Continua*, vol. 63, no. 1, pp. 537–55, 2020.

[23] A. Sedik, A. M. Iliyasu, B. El-Rahiem, M. E. Abdel Samea, A. Abdel-Raheem, M. Hammad, et al., "Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections," *Viruses*, vol. 12, no. 7, p. 769, 2020.

[24] K. DattaGupta, R. Dwivedi, and D. K. Sharma, "Prediction of Covid-19 trends in Europe using generalized regression neural network optimized by flower pollination algorithm," *J. Interdisc Math.*, vol. 24, no. 1, pp. 35–51, 2021, doi: 10.1080/09720502.2020.1833447.

[25] A. Sedik, M. A Hammad, F. ElSamie, B. B. Gupta, and A. Abd El-Latif, "Efficient deep learning approach for augmented detection of Coronavirus disease," *Neural. Comput. Appl.*, S.I.: Healthcare Analytics, pp. 1–18, 2021, doi: 10.1007/s00521-020-05410-8.

[26] M. Assaad, R. Bone, and H. Cardot, "A new boosting algorithm for improved time-series forecasting with recurrent neural networks," *Inf. Fusion*, vol. 9, no. 1, p. 41–55, 2008.

[27] S. Hochreiter and J. Schmidhuber, "Long short term memory," *Neural. Comput.*, vol. 9, pp. 1735–1780, 1997.

[28] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, *Recurrent neural networks for short term load forecasting: An overview and comparative analysis*, Cham, Switzerland: Springer, 2017.

[29] J. Huang, Y. Sun, W. Zhang, H. Wang, and T. Liu, "Entity highlight generation as statistical and neural machine

translation," *IEEE/ACM Trans. Audio Speech Lang. Process*, vol. 26, pp. 1860–1872, 2018.

[30] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," In *Advances in neural information processing systems, NIPS*, Cambridge, MA, United States: MIT Press, 2014, pp. 3104–3112.

[31] D. Bahdanau, K. Cho, and Y. Bengio "Neural machine translation by jointly learning to align and translate," In *Proceedings of 3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, 2015.

[32] https://www.mohfw.gov.in.

[33] https://twitter.com/CMODelhi.

[34] https://twitter.com/MoHFW_GUJARAT.

[35] https://arogya.maharashtra.gov.in/1175/Novel-Corona-Virus.