Research Article

Open Access

Swati Tyagi, Shaifu Gupta, Syed Abbas*, Krishna Pada Das, and Baazaoui Riadh

Analysis of infectious disease transmission and prediction through SEIQR epidemic model

<https://doi.org/10.1515/msds-2020-0126>

Received November 3, 2020; accepted March 25, 2021

Abstract: In literature, various mathematical models have been developed to have a better insight into the transmission dynamics and control the spread of infectious diseases. Aiming to explore more about various aspects of infectious diseases, in this work, we propose conceptual mathematical model through a SEIQR (Susceptible-Exposed-Infected-Quarantined-Recovered) mathematical model and its control measurement. We establish the positivity and boundedness of the solutions. We also compute the basic reproduction number and investigate the stability of equilibria for its epidemiological relevance. To validate the model and estimate the parameters to predict the disease spread, we consider the special case for COVID-19 to study the real cases of infected cases from [2] for Russia and India. For better insight, in addition to mathematical model, a history based LSTM model is trained to learn temporal patterns in COVID-19 time series and predict future trends. In the end, the future predictions from mathematical model and the LSTM based model are compared to generate reliable results.

Keywords: Infectious disease; Mathematical model; Stability analysis; Long Short Term Memory networks (LSTM); Parameter estimation

MSC: 00A71; 92B20; 34D20

1 Introduction

In history, various pandemics and epidemics have emerged out several times, which have demolished humanity, sometimes resulting into end of civilizations and tremendous change in the course of history. Mathematical modelling plays an important role in understanding the complexities of such infectious diseases and their control. The initial models of infectious diseases primarily focused on the theoretical disease research. More recently, there have been efforts to broaden the field still further by incorporating many aspects of complexity of natural systems. Provided this rich history in development of fundamental theory, the applied aspect of the questions being asked, and the extensive data collected on many disease systems, the field of infectious disease modelling represents one of the richest areas of research at the interface of pure theory and data. Mathematical modelling can be used to study the mechanisms underlying observed epidemiological patterns, assessing the effectiveness of control strategies, and predicting epidemiological trends. The investigation of dynamics of infectious disease transmission can be reached from various frameworks, such as relatively simple curve fitting techniques to standard SIR, SEIR etc. compartmental models to complex stochastic models using simulations. To model the outbreaks of infectious diseases, there are two major components, namely;

Swati Tyagi: Department of Mathematics, Amity University, Punjab, 140306, India

Shaifu Gupta: Department of Computer Science & Engineering, Indian Institute of Technology Jammu, Jammu, Jammu and Kashmir, 181221, India

***Corresponding Author: Syed Abbas:** School of Basic Sciences, Indian Institute of Technology Mandi, Mandi, 175005, H.P., India, E-mail: sabbas.iitk@gmail.com; abbas@iitmandi.ac.in

Krishna Pada Das: Mahadevananda Mahavidyalaya Monirampore, P.O.-Barrackpore, Kolkata 700120, India

Baazaoui Riadh: College of Sciences and Humanities, Prince Sattam bin Abdulaziz University, Saudi Arabia

(i) the basic reproduction number (R_0 , the average number of secondary infections produced by primary infection while coming in contact with the entire susceptible population) and (ii) the generation time (average time from the onset of symptom in primary infectious individual to symptom onset in secondary case). These two components jointly determine the likelihood and speed of the outbreaks of epidemic. Furthermore, all the modelling techniques largely depend on the availability of data for the corresponding infectious disease under investigation, in order to provide the robust estimate the parameters and their distribution. In case of emergence of a new infectious disease, rapid growth of literature on mathematical modelling also plays as a key factor. Mathematical modelling also helps in assessment of potential long-term impacts of such diseases and provides means to evaluate and predict the effects of possible interventions, even with the availability of limited data. Some typical examples are profound early models for HIV in 1980s, pandemic influenza, Creutzfeldt–Jakob Disease, outbreak of severe acute respiratory syndrome (SARS) in 2003, Nipah virus etc.

Very recently, the novel coronavirus, also known as COVID-19, 2019-nCoV or SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) has provided the world with an unparallel challenge and is crucially threatening the mankind [1, 2]. The first case of COVID-19 was identified in the Wuhan city, Hubei Province of China on December 31, 2019, and since then at this stage, it has spread widely in 215 countries and territories around the whole world [1, 2] with the infections increasing exponentially. The coronavirus is a zoonotic disease, where primary host was animals, from there it transmitted to humans [3]. The infected people face flu-like symptoms, which leads to critical illness, respiratory problems, primarily to those which have other diseases like diabetes, asthma, blood pressure, heart disease etc. [2].

On 30 January 2020, the epidemic of coronavirus disease 2019 (COVID-19) was declared as a Public Health Emergency of International Concern, the highest level in the World Health Organization's (WHO) emergency response for infectious diseases. The number of cases was accelerating in China and subsequently all over the world; however, the publications of the COVID-19's potential transmission and effectiveness of government interventions were limited. It is urgent to provide more scientific information for a better understanding of the novel coronavirus and further containment of the outbreak. Till date, more than 9.9 million people are infected with COVID-19 according to WHO and 0.49 million people have died due to the infection. As a preventive measure, various countries have implemented a strict lock-down and their administrations are encouraging and advising people to stay at home and maintain social distancing and proper hygiene. This pandemic has impacted immensely in a negative manner on world politics, education, economy, sociality, and various other global aspects [4, 5].

Knowledge of early spread dynamics of infection and figuring out the possible control measures is critical to assess the extensive outspread. Besides medical and biological research, it becomes an emergency to formulate a mathematical model which effectively describes the development and transmission of disease. This will help to make significant decisions and preventive measures based on the effective assumptions provided by the model.

This paper is organized as follows. In Section 2, we propose a SEIQR mathematical model to predict the dynamics of infectious disease. In Section 3, the positivity and boundedness of the solutions are described. We also compute the basic reproduction number and investigate the stability of the equilibria. In Section 4, we perform the data analysis as a special case for COVID-19 in comparison with the model solution and also predict the disease dynamics. We train the proposed mathematical model using the data available on [2] to estimate the parameters, as a case for Russia and India. In Section 5, we present the deep learning model to be used for learning past trends and to predict future trends. The experimental results of this model are presented in Section 6. Section 7 presents a comparative evaluation of results of mathematical and temporal analysis based machine learning model. At the end we conclude our paper in Section 8.

2 Model formulation

In this work we are investigating SEIQR model based on the fact that infected/symptomatic individuals are put into quarantine for treatment (hospital or home remedies), from which they move into recovered class. The

SEIQR, an extensively used epidemic model, can reflect the flows of people between five states: susceptible (S), exposed (E), infectious (I), quarantine (Q) and recovered (R). The total host population is partitioned into susceptibles, exposed (in the latent period), infectious, quarantined and recovered, with the densities, respectively, denoted by $S(t)$, $E(t)$, $I(t)$, $Q(t)$ and $R(t)$. The total population size at time t is denoted by $N(t)$, with $N = S + E + I + Q + R$.

The model is described by the following system of differential equations:

$$\begin{aligned}\frac{dS}{dt} &= A - \beta_{10}SE - \beta_{20}SI - \mu S \\ \frac{dE}{dt} &= \beta_{10}SE + \beta_{20}SI - \gamma_0 E - \mu E \\ \frac{dI}{dt} &= \gamma_0 E - k_0 I - \alpha_0 I - \mu I \\ \frac{dQ}{dt} &= k_0 I - k_1 Q - \alpha_0 Q - \mu Q \\ \frac{dR}{dt} &= k_1 Q - \mu R.\end{aligned}\tag{1}$$

with non-negative initial conditions; the positive parameters μ is the rate of natural death; α_0 is non-negative constant and denote the rate of disease caused death. The parameter γ_0 denote the transfer rates between the exposed and the infectious. The constant k_0 is the rate at which the infectious individuals are quarantined; the constant k_1 is the rate at which the quarantine individuals recover; β_{10}, β_{20} are the rate of the efficient contact in the latent and infected; A are constant recruitment of susceptibles respectively. The initial conditions are $S(0) > 0$; $E(0) \geq 0$; $I(0) \geq 0$; $Q(0) \geq 0$ and $R(0) \geq 0$.

3 Analysis of the system

3.1 Non-negativity and boundedness of solution

Theorem 3.1. *The solutions of the system (1) are non-negative and bounded.*

Proof. From (1), we have $\frac{dS}{dt} \geq -\mu S$. This implies

$$S(t) \geq S(0)\exp(-\mu t) > 0.$$

Similarly we have

$$E(t) \geq E(0)\exp(-(\gamma_0 + \mu)t) \geq 0,$$

$$I(t) \geq I(0)\exp(-(k_0 + \alpha_0 + \mu)t) \geq 0,$$

$$Q(t) \geq Q(0)\exp(-(k_1 + \alpha_0 + \mu)t) \geq 0,$$

$$R(t) \geq R(0)\exp(-\mu t) \geq 0.$$

Next, let $N = S + E + I + Q + R$. Then we have

$$\frac{dN}{dt} + \mu N \leq A - \alpha_0 I - \alpha_0 Q \leq A.$$

Integrating both sides of the above inequality and using theory of differential inequalities, we obtain

$$0 < N(S, E, Q, I, R) \leq \frac{A}{\mu} \left(1 - e^{-\mu t}\right) + N(S(0), E(0), Q(0), I(0), R(0)) e^{-\mu t}$$

and letting $t \rightarrow \infty$, we have $0 < N \leq \frac{A}{\mu}$. Thus all solutions of (1) that initiate in $\{R_+^5 \setminus 0\}$ are confined in the region $\Omega = \left\{(S, E, Q, I, R) \in R_+^5 : N = \frac{A}{\mu} + \epsilon\right\}$ for any $\epsilon > 0$ and for $t \rightarrow \infty$. This proves the theorem. $\square$

3.2 Computation of Reproduction number

Following the similar approaches as in [6], it can be verified that the system (1) always has the disease free equilibrium (DFE) $P_0 = (S^0, 0, 0, 0, 0) = (\frac{A}{\mu}, 0, 0, 0, 0)$ in the absence of infection. Furthermore, let $R_0 = \rho(FV^{-1})$ represents the spectral radius of the matrix FV^{-1} , then R_0 is given by

$$R_0 = \rho(FV^{-1}) = \frac{\gamma_0 \beta_{20} S^0}{(k_0 + \alpha_0 + \mu)(\gamma_0 + \mu - \beta_{10} S^0)}. \quad (2)$$

The parameter R_0 is referred to as the Basic reproduction number. The number R_0 indicates the transmission of infectious disease, for $R_0 > 1$, the outbreak is expected to continue and expected to end for $R_0 < 1$. It can be observed from (2) that R_0 significantly depends upon the average rate of contact between susceptible and infected, and transfer rate between susceptible and latent individuals. Thus, we can say lockdown/quarantining individuals plays a significant role in controlling the spread of infectious disease.

3.3 Stability analysis

Stability analysis of DFE P_0 :

Theorem 3.2. *If $\beta_{10} S^0 < \gamma_0 + \mu$, then the disease-free equilibrium point is locally asymptotically stable.*

Proof. In order to check the stability of the equilibrium points, we calculate the eigenvalues using Jacobian matrix. The Jacobian at DFE is given by

$$J_{DFE} = \begin{pmatrix} -\mu & -\beta_{10} S^0 & -\beta_{20} S^0 & 0 & 0 \\ 0 & \beta_{10} S^0 - (\gamma_0 + \mu) & \beta_{20} S^0 & 0 & 0 \\ 0 & \gamma_0 & -(k_0 + \alpha_0 + \mu) & 0 & 0 \\ 0 & 0 & k_0 & -(k_1 + \alpha_0 + \mu) & 0 \\ 0 & 0 & 0 & k_1 & -\mu \end{pmatrix}$$

The characteristic roots of the Jacobian matrix J_{DFE} are $-\mu$, $-\mu$, $-(k_1 + \alpha_0 + \mu)$ and $\frac{\beta_{10} S^0 - \gamma_0 - (k_0 + \alpha_0 + 2\mu)}{2} \pm \frac{\sqrt{(k_0 + \alpha_0 + \mu + \beta_{10} S^0 - \gamma_0 - \mu)^2 + 4\gamma_0 \beta_{20} S^0}}{2}$. Clearly all five roots are negative. Thus the disease-free equilibrium point (P_0) is locally asymptotically stable.

Further to prove the global asymptotic stability of DFE, we use theorem by Castillo-Chavez et.al. [7].

Theorem 3.3. [8] *The given mathematical model (1) can be represented in the form*

$$\frac{dX}{dt} = F(X, Y), \quad \frac{dY}{dt} = G(X, Y), \quad G(X, 0) = 0 \quad (3)$$

where $X = S$; $Y = (E, I, Q)^T$ denote the classes of uninfected and infected individuals respectively.

The DFE is represented by $P_0 = (X_0, 0) = (\frac{A}{\mu}, 0)$.

For global asymptotic stability of P_0 , condition (H1) and (H2) must be satisfied.

(H1) For $\frac{dX}{dt} = F(X_0, 0)$, X_0 is globally asymptotically stable.

(H2) $G(X, Y) = MY - \hat{G}(X, Y)$, $\hat{G}(X, Y) \geq 0$. Here $M = D_Y G(X_0, 0)$ is M-matrix, that is all off-diagonal elements are non-negative.

If given system (1) satisfies (3), then $P_0 = (X_0, 0)$ is a globally asymptotically stable equilibrium of given mathematical model.

Theorem 3.4. *The equilibrium $P_0 = (X_0, 0)$ of system (1) is globally asymptotically stable provided the conditions in (3.3) are satisfied.*

Proof: Using Theorem 3.3 to our model system (1), we obtain

$$F(X_0, 0) = A - \mu S, \quad G(X, Y) = MY - \hat{G}(X, Y)$$

where

$$M = \begin{bmatrix} -(\gamma_0 + \mu) & 0 & 0 \\ \gamma_0 & -(k_0 + \alpha_0 + \mu) & 0 \\ 0 & k_0 & -(k_1 + \alpha_0 + \mu) \end{bmatrix}, \hat{G}(X, Y) = \begin{bmatrix} \beta_1 SE + \beta_2 SI \\ 0 \\ 0 \end{bmatrix}.$$

We can observe that $\hat{G}(X, Y) \geq 0 \forall (X, Y)$. Furthermore, the matrix A is M-matrix. Thus the DFE P_0 is globally asymptotically stable. $\square$

Existence of Endemic equilibria and its stability:

In addition to DFE P_0 , the model (1) admits a unique endemic equilibrium point $P^* = (S^*, E^*, Q^*, I^*, R^*)$, where

$$\begin{aligned} S^* &= \frac{\gamma_0 + \mu}{\beta_{10} + d_1 \beta_{20}} \\ E^* &= \frac{A d_1 \beta_{20} (R_0 - 1)}{R_0 (\gamma_0 + \mu) (\beta_{10} + d_1 \beta_{20})} \\ I^* &= d_1 E^* \\ Q^* &= d_2 d_1 E^* \\ R^* &= \frac{d_1 d_2 k_1}{\mu} E^* \\ d_1 &= \frac{\gamma_0}{k_0 + \alpha_0 + \mu} \\ d_2 &= \frac{k_0}{k_1 + \alpha_0 + \mu} \end{aligned}$$

From above expressions, we can observe that P^* is positive only when $R_0 > 1$. Thus the endemic equilibrium point P^* exists whenever $R_0 > 1$.

Theorem 3.5. *The endemic equilibrium point (EE) $P^*(S^*, E^*, I^*, Q^*, R^*)$ is locally asymptotically stable.*

Proof. The Jacobian matrix at EE is given by

$$J_{EE} = \begin{pmatrix} -\beta_{10}E^* - \beta_{20}I^* - \mu & -\beta_{10}S^* & -\beta_{20}S^* & 0 & 0 \\ \beta_{10}E^* + \beta_{20}I^* & \beta_{10}S^* - (\gamma_0 + \mu) & \beta_{20}S^* & 0 & 0 \\ 0 & \gamma_0 & -(k_0 + \alpha_0 + \mu) & 0 & 0 \\ 0 & 0 & k_0 & -(k_1 + \alpha_0 + \mu) & 0 \\ 0 & 0 & 0 & k_1 & -\mu \end{pmatrix}$$

Solving the above for $\det(J_{EE} - \lambda I) = 0$, we obtain the following characteristic equation

$$(-\mu - \lambda)(-k_1 - \alpha_0 - \mu - \lambda)(\lambda^3 + A\lambda^2 + B\lambda + C) = 0, \quad (4)$$

where $A = -(a + d + h)$, $B = ad + ah + dh - bc - eg$, $C = aeg + bch - adh - fbe$, with $a = -\beta_{10}E^* - \beta_{20}I^* - \mu$, $b = \beta_{10}E^* + \beta_{20}I^*$, $c = -\beta_{10}S^*$, $d = \beta_{10}S^* - (\gamma_0 + \mu)$, $e = \gamma_0$, $f = -\beta_{20}S^*$, $g = \beta_{20}S^*$ and $h = -(k_0 + \alpha_0 + \mu)$.

Since roots of the characteristic equation (4) corresponding to Jacobian matrix evaluated at P^* are $-\mu < 0$, $-(k_1 + \alpha_0 + \mu) < 0$, and remaining roots satisfy the characteristic equation $\lambda^3 + A\lambda^2 + B\lambda + C = 0$. Applying Routh-Hurwitz criterion, the endemic equilibrium P^* is locally asymptotically stable, provided $A > 0$, $C > 0$ and $AB - C > 0$. This completes the proof. $\square$

4 Parameter estimation, model validation and prediction

In this section, we perform numerical simulations to support our analytical results to observe the dynamics of the proposed nonlinear system. In order to do so, we first estimate the parameters by analyzing the present

Figure 1: Fitting of model for prediction and validity for the COVID-19 outbreak in Russia

state in the country. In this work ‘*fmincon*’ function of MATLAB is used as an optimization function to perform numerical simulations. It is used to optimize the unknown parameters of the model equations. The format of the *fmincon* function is as follows:

$$[b, fmin, flag] = fmincon(@fun, b_0, [], [], [], [], lb, ub, options);$$

Here, ‘*fun*’ is the objective function. In this study, the objective function is $N = S + E + I + Q + R$; b_0 refers to the initial state of parameters. The initial state of parameters are chosen randomly. The terms *lb* and *ub* represent the lower bound and upper bounds of the parameters that are chosen randomly. In the options set, the maximum number of iterations required for attaining global optimum are specified.

Case study: Russia

We first check for the validity of model and predictions for COVID-19 in Russia. The model parameters have been estimated considering the total infectious cases from May 1 to September 30 as provided by [2] using minimization techniques. We have considered the initial population size as on 1 May as $S(0) = 140000000$; $E(0) = 140000$; $I(0) = 106498$; $Q(0) = 100000$; and $R(0) = 13220$. The parameter values are given in the following Table:

Parameter	Value
β_{10}	0.020
β_{20}	0.254
γ_0	0.456
k_1	0.645
α_0	1.292
A	100000.649
μ	0.007
k_0	0.650

Figure 2: Fitting of model for prediction and validity for the COVID-19 outbreak in India

In Figure 1, we provide the time-series analysis of total cases (infected, quarantined and recovered). It is clear from the figure that disease will possibly grow above 1300000 by end of November 2020 in Russia, if all the conditions remain same with a controllable number of infections.

Case Study: India

We further perform rigorous numerical simulation to analyze the scenario of COVID-19 epidemic in India. We first estimate the model parameters using the total infectious cases from June 27 to September 30 as provided by [2] using minimization techniques. We have considered the initial population size as $S(0) = 130000000$; $E(0) = 100000$; $I(0) = 508953$; $Q(0) = 100000$; and $R(0) = 295881$. The estimated parameter values are given in the following Table:

Parameter	Value
β_{10}	0.020
β_{20}	0.001
γ_0	0.023
k_1	0.650
α_0	0.020
A	100000.650
μ	0.001
k_0	0.650

In Figure 2, we have presented the best fitted curve for the total number of COVID-19 and the same of real cases [2] of the best fitted curve and the model predictions. It can be observed from the model predictions that the disease may continue upto even after end of November, if all the conditions remain same and the total infections may reach 11000000.

5 Predictions using Long short term memory (LSTM) model

Estimation of future COVID-19 cases can be viewed as a time series analysis problem where past trends in the increase of COVID-19 cases are used to predict future trend. The time series to be used for prediction of COVID-19 in future can be formulated as given in Eq. 5. In Eq. 5, let $\hat{x}^{t+1}$ denote the predicted number of COVID-19 cases at time instant $t + 1$. The number of cases are predicted as a function $f(\cdot)$ of the cases at previous time steps. In this Eq. p represents the past time lags under consideration for the study and the terms x^t represent actual cases at any time instant t .

$$\hat{x}^{t+1} = f(x^t, x^{t-1}, \dots, x^{t-p+1}) \quad (5)$$

Multi-step ahead prediction includes predicting two or more values in future, In this study, they are generated using an iterative method [9] where the predicted value is repeatedly passed back to the model to generate the next step forecast. For example $\hat{x}^{t+2}$ can be predicted as:

$$x^{t+2} = f(\hat{x}^t + 1, x^t, \dots, x^{t-p+1}) \quad (6)$$

Using the iterative process given in Eq. 6, the multi-step ahead prediction $\hat{x}^{t+M}$ can be obtained. An M step ahead forecast requires M one-step ahead forecasts. The term M is called prediction horizon. .

The nature of the function $f(\cdot)$ in Eq. 5 can be linear or non-linear. There are different models in literature for learning this linear or non-linear function for learning patterns in data. Autoregressive integrated moving average (ARIMA) method uses a linear function for prediction [10]. However it is seen in different predictive analytic applications such as future resource workload prediction [11] that non-linear models perform better in comparison to linear models. This is due to the fact that most of the real life patterns are non-linear in nature. Neural networks overcome the limitation of ARIMA models and can learn non-linear relationships in past time steps and predicted values [12]. A neural network uses non-linear activation function in its layers which enables to learn non-linear relations. Different works in literature have also used hybrid of ARIMA and neural networks for better modelling of patterns in a time series [13, 14]. Recently, deep neural networks are well favoured over conventional neural networks as they have shown promising results. Among deep neural networks, recurrent neural network and its variation long short term memory model (LSTM) is very popular.

LSTM has shown its efficacy in dealing with short-term and long-term dependency problems. It has been successfully applied in different domains for modelling sequential and temporal dependencies. For example, LSTM is used for estimating future resource usage prediction in cloud data centres in [15, 16], resource contention failure identification [17], traffic flow prediction [18] and many other use cases. These applications from literature indicate that LSTM is playing significant role for modelling temporal and sequential dependencies to predict COVID-19 patterns. The architecture of LSTM model is explained in next subsection.

5.1 Architecture of LSTM model

In this section, we provide basic introduction to use LSTM for time series prediction. LSTM is a variant of recurrent neural network (RNN). RNNs are used for modelling sequential dependencies in data. A basic RNN consists of neurons which are arranged in the form of an array in different layers of the RNN. Each neuron in a given layer is connected to every neuron in the next layer. In these networks, the output from the previous time step is passed as input to the next time step which allows the network to retain memory and hence enables the model to learn sequential and temporal dependencies in data [17]. The prediction is computed as:

$$\hat{x}^{t+1} = \mathbf{W}_{hy} \mathbf{h}^t + \mathbf{b}_y \quad (7)$$

The terms $\mathbf{W}_{hy}$ and $\mathbf{b}_y$ represent the weights and bias respectively between output layer and last RNN layer whose output is $\mathbf{h}^t$. The network is trained using a mean squared error loss function.

LSTM model is better than basic RNN as unlike RNN it does not have vanishing gradient problem. The time series prediction architecture of LSTM is similar to RNN except that a neuron of basic RNN model is

replaced by a cell in LSTM. Figure 3 shows the architecture of basic LSTM cell. Each cell has three gates namely, input gate, output gate and the forget gate. The forget gate identifies the information that is to be forgotten by the cell memory. The term $\tilde{c}^t$ denotes the vector of new candidate values. The role of the input gate is to decide the values that are to be updated and the output gate controls the flow of information out of the cell. Together these three gates control the flow of information in and out of cell memory enabling it learn long and short term dependencies [17]. The rectangular boxes in Figure 3 represent the layer of neurons with a desired activation function.

Estimated COVID-19 cases can be predicted by analysing its temporal dependencies with other measurable metrics as well. We explain this prediction model in the next subsection.

5.2 Univariate and multivariate modelling of COVID-19 cases

Univariate prediction models analyse temporal patterns in the past history of the number of COVID-19 cases itself to predict future trends. However it is seen in real life scenarios that each decision involves analysing multiple variables together. This is due to the fact that studying multiple related variables together can aid in the understanding of variations in the trend of desired metric to be predicted [19]. Multivariate analysis involves analysing multiple variables together to analyse the relationships between them and their combined effect on the prediction of desired metric. It is shown in the context of prediction of resource usage and performance in cloud data centres [15, 20] and prediction of mortality rates [21] that better prediction capabilities in forecasting new values are achieved by analysing metrics which are related with the desired metric. This indicates that multivariate analysis provides better statistical power to the model and aids in achieving more realistic predictions for the desired metric. Therefore, we analyse both univariate as well as multivariate prediction models. Here the desired metric is COVID-19 cases. To build multivariate prediction model, temporal dependencies in deaths as well as recovered cases are also analysed in addition to the number of COVID-19 cases.

In the next section, we analyse the results of LSTM based model for estimating COVID-19 cases for India and Russia.

6 Experimental analysis using LSTM model

In this section, we present the results of the experimental studies of COVID-19 cases. Here, we first analyse the nature of COVID-19 cases in the next subsection which is followed by its modelling and prediction by LSTM models in subsequent sections. The data of COVID-19 cases, deaths and recovered cases in India and Russia is obtained from [2].

Figure 3: Structure of LSTM cell.

6.1 Pre-processing

We analysed the time varying trends of COVID-19 cases in India and Russia. It is seen that time series of COVID-19 cases is non-stationary in nature. This is due to the reason that the series is a monotonically increasing series and hence the statistical properties of the series are not constant. We also experimentally verified this fact by using augmented Dickey-Fuller test [22]. Since the series is non-stationary, we use Box-Cox transformation to pre-process the series. Many machine learning and deep learning models are based on the assumption of normality in a time series. Therefore, here Box-Cox transformation is used so that the transformed series is closer to the normal distribution where the mean, mode and median lie at the centre. This pre-processing helps LSTM models to learn the trends in the non-stationary time series better.

6.2 Model specifications and results

To train our model we have used univariate as well as multivariate data for model training and prediction. To train the LSTM based neural network, we have used $\tanh(\cdot)$ activation function. It is seen empirically that for India, a two layered LSTM with 4 and 8 cells in individual layers performs the best. For Russia, a two layered LSTM network with 32 cells in each layer performs the best. The model is trained for the data from May 1 to September 30 for Russia and from June 27 to September 30 for India and then predictions are made using the test results. We have predicted the total COVID-19 cases 60 days ahead for both India and Russia. The obtained prediction results for India are given in Fig. 4(a)-4(b) and for Russia are given in Fig. 5(a)-5(b).

Figure 4: Model training and Prediction analysis for India

7 Comparative discussion of results

From the predictions and obtained results in Section 4 and 6, we can observe that the virus infection will increase rapidly with time resulting into a vast infected population and the situation may worsen, if no necessary steps are taken or vaccine is not invented. Thus it is high time to take necessary precautions and preventive measures. People should themselves need to be aware and follow strict precautionary measures. There should be minimum public mobility along with strict hygienic and social distancing norms. Implementing Government restrictions mathematically, we can say that the total cases may peak upto 1300000 for Russia

Figure 5: Model training and Prediction analysis for Russia

and 11000000 for India by end of November. We can also observe that LSTM is giving more precise predictions. Since, Government is changing its restriction measures and policies from time to time, so the system parameters are also changing after that period of time. Therefore, in the current position we are not interested about infected population for long time of period, but through these estimates, we can say that we need to follow strict policies to reduce the infection.

8 Conclusion

In this work, we have formulated and proposed a SEIQR epidemic model to study the dynamics of spread of infectious diseases. The proposed model is an attempt to provide a deep insight to analyze the disease spread and design possible control strategies. We have derived various analytical results, such as positivity and boundness of the solutions, local stability of equilibria, basic reproduction number. There exist two equilibria, namely the disease-free equilibrium (DFE) and the endemic equilibrium (EE). We have estimated the parameters using the real world data as a special case for COVID-19. Our main aim is to predict mathematically the number of infected individual due COVID-19 virus in Russia and India in this study. To fulfil our aim, we have performed a detailed numerical simulation of our proposed model through minimization techniques and long short term memory networks. We have tried to establish an acceptable agreement between the data analysis and numerical solutions. Analysing the results of our proposed model, we can say that India may be in a big trouble due to COVID-19 virus in near future. Indian Government should take stricter measures and policies other than lock down, quarantine etc. Since the government policies are changing continuously to protect India, thus parameters are also changing with time. Thus, we can assume that if the Indian Government and all state government may take proper steps time to time and people also follow strict social distancing measures along with proper hygiene, yoga, and etc., then the infected number of population may differ from our predictions with time and India will recover in recent future from this pandemic. At the end, we can summarize and conclude that mathematical modelling is an efficient method to estimate the situation and dynamics of this global pandemic COVID-19, if the parameters are estimated properly. Due to limited data and short time span, we have made certain assumptions while framing model. But with the release of more epidemic data, the key parameters may undergo significant changes on account of health, education, religion, per capita income, influencing the pandemic spread among the masses.

Acknowledgements—: The authors are thankful to the reviewers for their constructive comments and suggestions, which have helped in improving the manuscript considerably.

Conflict of interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

- [1] <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [2] <https://www.worldometers.info/coronavirus>
- [3] Kamrujjaman, M., Mahmud, M. S., & Islam, M. S. (2020). Coronavirus outbreak and the mathematical growth map of Covid-19. *Annual Research & Review in Biology*, 72-78.
- [4] Verdière, N., Lanza, V., Charrier, R., Provito, D., Dubos-Paillard, E., Bertelle, C., & Alaoui, A. (2014, June). Mathematical modeling of human behaviors during catastrophic events.
- [5] Ivorra, B., & Ramos, A. M. (2020). Application of the Be-CoDiS mathematical model to forecast the international spread of the 2019–20 Wuhan coronavirus outbreak. *Research Gate Preprint*, 9, 1-13.
- [6] Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7(47), 873-885.
- [7] Li, M. Y., & Muldowney, J. S. (1996). A geometric approach to global-stability problems. *SIAM Journal on Mathematical Analysis*, 27(4), 1070-1083.
- [8] Srivastav, A. K., & Ghosh, M. (2016). Modeling and analysis of the symptomatic and asymptomatic infections of swine flu with optimal control. *Modeling Earth Systems and Environment*, 2(4), 1-9.
- [9] Zhang, L., Zhou, W. D., Chang, P. C., Yang, J. W., & Li, F. Z. (2013). Iterated time series prediction with multiple support vector regression models. *Neurocomputing*, 99, 411-422.
- [10] Caglar, F., & Gokhale, A. (2014, June). iOverbook: intelligent resource-overbooking to support soft real-time applications in the cloud. In *2014 IEEE 7th International Conference on Cloud Computing* (pp. 538-545).
- [11] Gupta, S., & Dinesh, D. A. (2017). Online adaptation models for resource usage prediction in cloud network. In *2017 Twenty-third National Conference on Communications (NCC)* (pp. 1-6).
- [12] Gupta, S., Dileep, A. D., & Gonsalves, T. A. (2016). Fractional difference based hybrid model for resource prediction in cloud network. In *Proceedings of the Fifth International Conference on Network, Communication and Computing* (pp. 93-97).
- [13] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- [14] Chaâbane, N. (2014). A hybrid ARFIMA and neural network model for electricity price prediction. *International journal of electrical power & energy systems*, 55, 187-194.
- [15] Gupta, S., Dileep, A. D., & Gonsalves, T. A. (2018). A joint feature selection framework for multivariate resource usage prediction in cloud servers using stability and prediction performance. *The Journal of Supercomputing*, 74(11), 6033-6068.
- [16] Song, B., Yu, Y., Zhou, Y., Wang, Z., & Du, S. (2018). Host load prediction with long short-term memory in cloud computing. *The Journal of Supercomputing*, 74(12), 6554-6568.
- [17] Gupta, S., & Dileep, A. D. (2020). Relevance feedback based online learning model for resource bottleneck prediction in cloud servers. *Neurocomputing*. Vol. 402, 307-322.
- [18] Zhao, F., Zeng, G. Q., & Lu, K. D. (2019). EnLSTM-WPEO: short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization. *IEEE Transactions on Vehicular Technology*, 69(1), 101-113.
- [19] Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. (1992). Forecasting the behavior of multivariate time series using neural networks. *Neural networks*, 5(6), 961-970.
- [20] Hirwa J.S. & Cao J (2014) An ensemble multivariate model for resource performance prediction in the cloud. In: *Network and Parallel Computing NPC 2014, Lecture Notes in Computer Science*, Vol. 8707, pp 333–346.
- [21] Peña, D., & Sánchez, I. (2007). Measuring the advantages of multivariate vs. univariate forecasts. *Journal of Time Series Analysis*, 28(6), 886-909.
- [22] Harris, R. I. (1992). Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics letters*, 38(4), 381-386.