Article

COVID-19 Pandemic Outbreak in the Subcontinent: A Data Driven Analysis

Bikash Chandra Singh [1,2], **Zulfikar Alom** [3], **Haibo Hu** [2], **Mohammad Muntasir Rahman** [4],
Mrinal Kanti Baowaly [5], **Zeyar Aung** [6], **Mohammad Abdul Azim** [3] and **Mohammad Ali Moni** [7,*]

1   Department of Information and Communication Technology, Islamic University, Kushtia 7003, Bangladesh;
    bikash.singh@ice.iu.ac.bd
2   Department of Electronic and Information Engineering, The Hong Kong Polytechnic University,
    Kowloon, Hong Kong; haibo.hu@polyu.edu.hk
3   Department of Computer Science, Asian University for Women (AUW), Chattagram 4000, Bangladesh;
    zulfikar.alom@auw.edu.bd (Z.A.); azim@ieee.org (M.A.A.)
4   Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh;
    muntasir@cse.iu.ac.bd
5   Department of Computer Science and Engineering, Bangabandhu Sheikh Mujibur Rahman Science and
    Technology University, Gopalganj 8100, Bangladesh; baowaly@bsmrstu.edu.bd
6   Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi 127788,
    United Arab Emirates; zeyar.aung@ku.ac.ae
7   School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of
    Queensland, St. Lucia, QLD 4072, Australia
*   Correspondence: m.moni@uq.edu.au

Abstract: Human civilization is experiencing a critical situation that presents itself for a new coronavirus disease 2019 (COVID-19). This virus emerged in late December 2019 in Wuhan city, Hubei, China. The grim fact of COVID-19 is, it is highly contagious in nature, therefore, spreads rapidly all over the world and causes severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Responding to the severity of COVID-19 research community directs the attention to the analysis of COVID-19, to diminish its antagonistic impact towards society. Numerous studies claim that the subcontinent, i.e., Bangladesh, India, and Pakistan, could remain in the worst affected region by the COVID-19. In order to prevent the spread of COVID-19, it is important to predict the trend of COVID-19 beforehand the planning of effective control strategies. Fundamentally, the idea is to dependably estimate the reproduction number to judge the spread rate of COVID-19 in a particular region. Consequently, this paper uses publicly available epidemiological data of Bangladesh, India, and Pakistan to estimate the reproduction numbers. More specifically, we use various models (for example, susceptible infection recovery (SIR), exponential growth (EG), sequential Bayesian (SB), maximum likelihood (ML) and time dependent (TD)) to estimate the reproduction numbers and observe the model fitness in the corresponding data set. Experimental results show that the reproduction numbers produced by these models are greater than 1.2 (approximately) indicates that COVID-19 is gradually spreading in the subcontinent.

Keywords: COVID-19 pandemic; SARS-COV-2; coronavirus; reproduction number; modelling infectious disease

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic has spread so rapidly around the globe that the modern world could not have imagined its severity at the beginning of the outbreak that started in Wuhan, China in late December 2019. On 5 January 2020, the world health organization (WHO) first released information on the disease outbreak that a cluster of cases of pneumonia caused by an unknown reason was detected in Wuhan city [1]. After that, the Chinese authorities identified the disease caused by a novel coronavirus [2] on

7 January 2020. WHO temporarily termed it as 2019-nCoV [3] on 12 January 2020; was later officially named as "severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)" by the international committee on taxonomy of viruses (ICTV) based on genetic analysis [4,5].

Initially, WHO declared the outbreak as a public health emergency of international concern (PHEIC) [6] on 30 January 2020. Eventually by evaluating the severity and alarming levels of the transmission of its high contagiousness, WHO characterized the COVID-19 as a pandemic [7] on 11 March 2020. With this declaration COVID-19 becomes the fifth recorded pandemic after the 1918 Spanish flu pandemic [8].

As of 30 June 2020, COVID-19 is still in a pandemic with 10,434,890 confirmed cases and 508,843 deaths worldwide [9], and continues to climb globally. Unlike all other pandemics recorded in history, large amounts of data and news about COVID-19 are rapidly spreading and widely reported, and scholars in various fields have been mobilized to concentrate on analyzing these data and proposing solutions. Since the governments of different countries have responded to the COVID-19 pandemic seriously, it is important that the researchers estimate: (i) the pandemic regionally based on the basic reproduction number, (ii) the arrival of the peak time, and forecast the time course of the epidemic by analyzing the data on the total number of infected cases, (iii) the total number of confirmed cases, (iv) the total number of deaths, and (v) the total number of cases recovered, etc. Many researchers around the globe have estimated the prediction of the COVID-19 spreading and the end of the epidemic in different countries [10–18]. Some of these methods are based on statistical models [17,19–23] and some other methods use deterministic epidemic models known as susceptible infectious recovery (SIR) with different forms [10,24–28].

Recently, some countries in South Asia, especially in Bangladesh, India and Pakistan, cases of COVID-19 are increasing rapidly. In particular, the COVID-19 cases became a new hot spot after countries began to relax locking restrictions. In this study, we use SIR, EG, SB, ML and TD models to analyze data to determine the reproduction number and predict the epidemic trend of COVID-19 in Bangladesh, India and Pakistan. More particularly, this article investigates the basic reproduction number R_0 and effective reproduction number $R(t)$ using these models. Previous studies show that R_0 is applicable when an exponentially increasing epidemic starts in the case of a completely susceptible population [29].

Therefore, the contribution of this article is twofold: (i) estimating of R_0 for Bangladesh, India and Pakistan, (ii) estimating $R(t)$ which quantifies the transmission potential over time. As the epidemic progresses, this parameter will track the average number of secondary cases per case over time periods. After calculating the R_0 and $R(t)$s, we compare the spread of COVID-19 in the region of concern. Experimental results show that the R_0 and $R(t)$s in all these countries are higher than about 1.2.

The rest of the paper is organized as follows. Section 2 discusses related work, while Section 3 illustrates the overall architecture of this work. Section 4 outlines the concept of reproduction numbers. Section 5 describes several models that can be used to generate reproduction numbers, whereas Section 6 explains information about the data source. Section 7 explains the experimental results. Finally, Section 8 summarizes the paper.

2. Related Work

A number of models have been proposed and applied to the area of infectious COVID-19. The majority of the mathematical models fall into the categories: (i) standard statistical models applied to the COVID-19 dataset, (ii) deterministic epidemic models such as SIR, SEIR etc. (iii) modified variants of the well-known models attempting to incorporate specific criteria such as incorporating social distancing into SEIR, and (iv) effectiveness of the contact tracing and isolation.

In [30], authors choose auto regressive integrated moving average (ARIMA) to predict the spread of COVID-19 based on Johns Hopkins University dataset [31]. Authors intend to avoid potential biases of the evaluating model and select a simple and economical model therefore chosen ARIMA. The ARIMA model consists of an autoregressive (AR) model, moving average (MA) model, and seasonal autoregressive integrated moving average

(SARIMA) model [32]. Results demonstrate the autocorrelation function (ACF), and the partial autocorrelation function (PACF) prevalence and incidence of COVID-19 are not influenced by the seasonality. However the claim is arguable as the COVID-19 may have seasonality influences; it is just too early to claim.

Calafiore et al. [33] presents a modified susceptible-infected-recovered (SIR) model for the contagion analysis of COVID-19 in Italy. Besides the regular SIR parameters the authors incorporate the initial number of susceptible individuals. Most interestingly authors consider a proportionality factor i.e., the ratio of the detected number of COVID positives to the actual number of infected individuals. Two distinct algorithms (i) tuning of the model parameters and (ii) predicting the number of infected, recovered and deaths are presented in this article.

Kucharski et al. [34] reported how transmission in Wuhan varied over a time period (January–February 2020). The article also attempts to put foresight on how human-to-human transmissions may occur outside Wuhan by travellers. The article reported the median day-to-day reproduction number over the time in Wuhan. The susceptible exposed infected and removed (SEIR) model is used for the evaluations. Besides, the study attempts to incorporate the uncertainty of the detection and observation utilizing Poisson process and binomial observation process on infection prevalence on flights.

Peng et al. [35] attempted to modify the generalized SEIR model by incorporating quarantine into account and studied the COVID-19 pandemic in 24 provinces in Mainland and 16 counties in Hubei province in China. The authors estimate the latent time, the quarantine time, the reproduction number, the possible ending time and the final total infected cases.

Wangping et al. [36] proposed an extended susceptible infected removed (eSIR) model essentially an extension to the SIR model attempting to address the effects of different intervention measures in dissimilar periods. The eSIR is applied to model the pandemic COVID-19 analysis in Italy. The markov chain monte carlo (MCMC) algorithm is utilized to obtain the posterior estimation of the unknown parameters in the SIR model.

Chatterjee et al. [37] modeled a variant of SEIR for COVID-19 epidemic in India. In this stochastic modeling approach Monte Carlo simulation is used to model the concept with a 1000 run. The article shed into the hospitalization and intensive care unit (ICU) requirements along with deaths. Article also presents the impact of the lockdown and social distancing.

Liang el al. [38] presents infection kinetic analysis of SARS, MARS and COVID-19. Author attempts to acquire a propagation growth model by utilizing the growth rate, and inhibition constant of the aforementioned diseases. The inhibition constant depends on the prevention and control measures adopted for the population. The article assumes during the inception of the disease cycle no effective measure is in place. It figures out the growth rate, the multiplication cycle and infection inhibition constant.

Ndairou et al. [39] provides a mathematical model i.e., essentially an extension to the SIR model of COVID-19 taking eight epidemiological classes into consideration. They are (i) susceptible, (ii) exposed, (iii) symptomatic and infectious, (iv) super spreaders, (v) infectious but asymptomatic, (vi) hospitalized, (vii) recovery, and fatality class. Wuhan number of cases and deaths are compared with the model.

Roda et al. [40] argued that these model predictions have shown a wide range of variations. And the variation of the performances of the models is due to the non-identify ability in model calibrations. The qualities of the statistical models are studied utilizing the Akaike information criterion (AIC). Authors' finding the performance of SIR over SEIR is explained as the failure of the even more sophisticated model is due to the more challenging realization of a comparatively more complex model.

Contrary to the other school of analytical models where the models used to predict the number of cases and deaths. Hellewell et al. [40] attempts to quantify the effectiveness of the contact tracing and isolation to control the COVID-19. The study considers a number of scenarios containing (i) initial cases, (ii) R_0, (iii) delay in-between symptom and isolation,

(iv) probability of traced contacts, (v) proportion of transmission before symptom, and (vi) proportion of sub-clinical infections.

Contrarily this study attempts to find the transmission potential in the subcontinent. Instead of relying on just one specific model this exploration employs SIR, EG, SB, ML and TD that realize definitive regional R_0 and $R(t)$ that get a conclusive outcome of the containment measure in place.

3. Overall Architecture

Figure 1 shows the architecture we followed to estimate the reproduction number $(R_0/R(t))$ for the COVID-19 pandemic. In particular, various epidemic methods are being used to predict the reproduction number $R_0/R(t)$ over time scale. To do this, we use a real dataset of COVID-19 cases spanning approximately 4 months, and the epidemic model utilizes this dataset to predict future $R_0/R(t)$. To this end, each model uses different parameters that can be computed from the dataset (see Section 5 for details). In particular, we calculated the predicted value of $R(t)$ based on time, so if $R(t)$ increases, we can easily take preventive measures such as lockdown, social distancing, etc. Eventually, it is important to note that in order to control infected cases, we have to control the reproduction number in affected areas. However, it is true that some predicted models may have good prediction results, and others may not. With this in mind, we compare several models (e.g., SIR, EG, ML, TD, SB) to see which one give the best results in sub-continent countries. The predicted results and comparative analysis among the models are described in detail in Section 7.
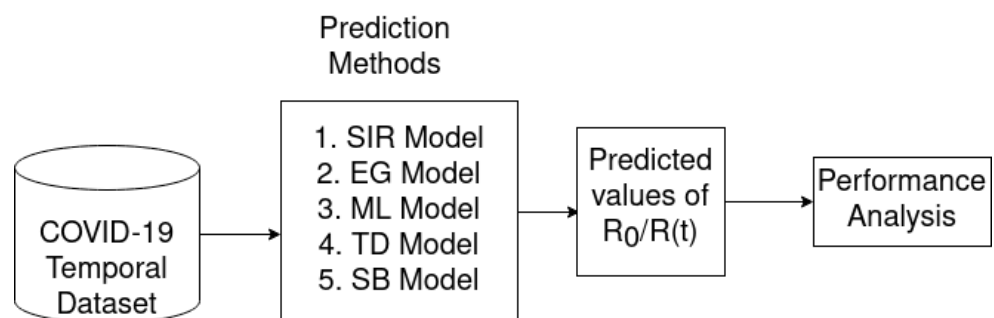

Figure 1. Overall architecture to estimate reproduction number $(R_0/R(t))$ of COVID-19 pandemic.

4. The Reproduction Number

Infectious diseases can be analyzed with a so-called reproductive number (R) that quantify the invasion or extinction of diseases in a population [41]. More precisely, the R is mainly used to determine the infectivity of a contagious disease. Alternatively, we can say that R represents the speed with which a disease spreads in a population. So, the primary task is to impose policies to control the R in order to control the contagious disease. This can be achieved by zonal shut down, social distancing and other factors that lead to control the outbreak of the disease in a particular geographic region. Fundamentally there are two types of Rs: (i) basic reproduction number R_0 and (ii) effective reproduction number $R(t)$. The R_0 refers to a measurement of the average number of cases that an infected person can spread over the person's infection period in a population [42]. Let, the R_0 of COVID-19 in a region x is y. Then, each infection may lead to y number of new secondary cases in the region x i.e., each infected person may infect y new individuals in the x zone. However, R_0 refers to a value that can indicate that the rate of the infected population has fallen or increased or remains constant.

Mathematically, $R_0 < 1$ indicates that the epidemic is in decline and it can be considered as under control. Contrarily, $R_0 > 1$ implies that the epidemic is on the rise and therefore cannot be considered as under control. And finally $R_0 = 1$ demonstrates that the infection rate remains constant.

Contrarily, the $R(t)$ is used to measure the infected cases when there is a certain immunity or certain interventions are taken place. In other words, $R(t)$ is the number of infected cases calculated in a certain population over the period of time t, taking into account that infected people are immune to infectious diseases at any given time. Therefore, we use the actual reproduction number $R(t)$ in order to measure the number of newly infected individuals, on average, infected by a single person at time t in a population. In effect, $R(t)$ represents the time variant R of the susceptible population where the change may be in decline or on the rise or remains constant. These three aforementioned conditions can be expressed as $R(t) \geq 0$. Likewise R_0, $R(t) < 1$, suggests the epidemic is on the decline and can be considered under control at time t. At $R(t) > 1$ the epidemic is on the rise and not in control at t. Finally, $R(t) = 1$ implies the infection rate remains constant. Interestingly, $R(t)$ and R_0 can be related to utilizing a simple relationship as $R(t)$ can be measured with $R_0 * S$ where S represents the number of infected people in a particular population. In case the immunity to the disease of a particular population is high then S becomes low. Consequently, $R(t)$ becomes below 1. The implication is that as herd immunity is achieved, the number of new cases in the population will decrease to zero over time [43].

In this paper, we are going to estimate R, i.e., R_0 and $R(t)$ through a variety of methods using the dataset of the selected countries such as Bangladesh, India and Pakistan. In the next section, we have explained the five most popular methods for estimating of R_0 and $R(t)$.

5. Epidemic Forecasting Models

There are numerous models that have been proposed and applied to the area of infectious COVID-19. In order to find the transmission potential in the subcontinent, we used five models, namely, SIR model, Exponential growth, Sequential Bayesian method, Maximum likelihood estimation, and Time-dependent estimation, briefly described in what follows.

5.1. SIR Model

The SIR model is a basic mathematical model for describing the dynamics of infectious diseases. It is also called the *compartmental model* because the model divides the population into different compartments. More particularly, the population of size N is divided into three compartments [33]: **S**usceptible, **I**nfectious, and **R**ecovered, which will be detailed in what follows.

Susceptible: is the number of people who are vulnerable to exposure with infectious people around at time t, denoted by $S(t)$.

Infectious: is a group of people who are infected with the disease. Moreover, they can spread the disease to susceptible people and can be recovered from it, in a specific time t, denoted by $I(t)$.

Recovered: is a number of people who get immunity in a time t, denoted by $R'(t)$. Therefore, they are not susceptible to the same disease anymore.

However, we can write the SIR model as a differential equation of each compartment [33]:

$$\frac{\partial s}{\partial t} = -\beta * s(t) * i(t) \tag{1}$$

$$\frac{\partial i}{\partial t} = \beta * s(t) * i(t) - \gamma * i(t) \tag{2}$$

$$\frac{\partial r}{\partial t} = \gamma * i(t) \tag{3}$$

where t defines the time, $s(t) = \frac{S(t)}{N}$, $i(t) = \frac{I(t)}{N}$, $r(t) = \frac{R'(t)}{N}$, and $N = (S(t) + I(t) + R'(t))$. Likewise, β is a controlling parameter that defines the number of people infected by exposure in a specific time t, and γ defines the ratio of the infected individuals who can

recover in a time t. Using these two parameters (i.e., β, γ) we can estimate the (R_0), mathematically, $R_0 = \frac{\beta}{\gamma}$, which defines the average number of people infected from single disease exposure. Hence, if the R_0 value is higher, the probability of the pandemic is also higher.

5.2. Exponential Growth (EG)

The exponential growth (EG) rate is an important measure to see the speed of the spread of an infectious disease. As introduced in Section 5.1, the exponential growth rate can be written as $r = \beta - \gamma$, where, β defines the number of people infected by exposure in a specific time t, and γ defines the ratio of the infected individuals who may recover in time t. However, r is a disease threshold value when r is at zero (i.e., $r = 0$). If r is positive (i.e., $r > 0$), the disease can invade a population, whereas it cannot invade a population, if r is negative (i.e., $r < 0$) [44]. The relationship between R_0 and growth rate r is not simple. For a specific distribution over generation time (e.g. gamma distribution), it can sometimes be simplified. Assuming that the generation time is completely constant, such as T, the R can be written as [45], $R_0 = e^{(r*T)}$.

5.3. Sequential Bayesian Method (SB)

The sequential bayesian approach can be used to estimate the initial reproduction number (R_0). Technically, the Bayesian method works in the context of probabilistic modeling. Therefore, the probability model of R_0 can be written as follows [46].

$$P[R_0] = \frac{P[R_0|\nabla T(t+\tau) \leftarrow \nabla T(t)] * P[\nabla T(t+\tau) \leftarrow \nabla T(t)]}{P[\nabla T(t+\tau) \leftarrow \nabla T(t)|R_0]}$$

where, $T(t)$ is the total number of cases up to time t, and the occurrence of new infected cases over the period τ, and $\nabla T(t+\tau) = T(t+\tau) - T(t)$. The probability distribution $P[\nabla T(t+\tau) \leftarrow \nabla T(t)]$ and $P[\nabla T(t+\tau) \leftarrow \nabla T(t)|R_0]$ are independent and dependent on R_0, respectively. Hence, the sequential bayesian estimation of R_0 can be made using the posterior distribution for R_0, at time t as the prior in the next estimation step at time $t + \tau$.

5.4. Maximum Likelihood Estimation (ML)

The maximum likelihood-based estimation relies on two considerations: (i) the number of secondary cases produced by an infected individual follows a Poisson distribution, and (ii) the expected value R_0. Suppose the data is a periodic incidence e.g., $\{N_0, N_1,... N_t\}$, where, t is a time unit, and N_t defines the number of new cases at time t. Therefore, the maximum likelihood can be estimated as follows [47]: $ML(R_0) = \prod_{t=1}^{T} \frac{\exp^{-\mu_t} \mu_t^{N_t}}{N_t!}$, where, $\mu_t = R_0 \sum_{j=1}^{min(k,t)} N_{t-j} * w_j$, where k is the constraint and w_j is the time distribution. Here, $k < T$. Note that the maximum likelihood-based method is used for the estimation of the basic reproductive number (R_0).

5.5. Time Dependent Estimation (TD)

Typically, the time-dependent R estimation is not straightforward, because we can only see the epidemic curve. There is no information about who infected whom. However, using likelihood-based estimates of R, the time dependant R can be written as [48]: $R_t = \frac{1}{N_t} \sum_{t_j=t} R_j$, where N_t specifies the number of new cases at time t, and R_j is the $R(t)$ for case j, that is the sum over all cases i, mathematically, $R_j = \sum_i p_{ij}$, where p_{ij} defines the relative likelihood. More precisely, the relative likelihood that case i has been infected by case j, normalized by the likelihood that case i has been infected by any other case k, which can

be written as follows [49]: $p_{ij} = \frac{w(t_i - t_j)}{\sum_{i \neq k} w(t_i - t_k)}$, where, $w(t_i - t_j)$ defines the time interval of the infectious diseases.

6. Data Source

We use a publicly available COVID-19 dataset (https://github.com/RamiKrispin/coronavirus) extracted from the coronavirus repository of the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE). This dataset contains daily observations on COVID-19 confirmed, recovered and death cases for most countries over the world. More precisely, the incidence data are provided on a daily basis. For this analysis, we consider data from South Asian countries such as Bangladesh, India, Pakistan, Nepal, Bhutan, Maldives and Sri Lanka. Nonetheless, the top three populated countries (Bangladesh, India, and Pakistan) finally have been selected for this particular article. Data dated upto 19 June 2020 is utilized. The entire dataset of COVID-19 cases of these selected countries is observed for each day, as shown in Figure 2. In addition, the Table 1 sequentially displays the date of the first confirmed case, the total number of confirmed incidences, the total number of deaths, the total number of recovery cases, population and COVID-19 tested per million people in each country.

Table 1. COVID-19 Dataset.

Country	Date of the First Case	End Date	Total Confirmed Cases	Total Death Cases	Total Recovered Cases	Population	Tested per Million People
Bangladesh	2020-03-08	2020-06-19	105,535	1388	42,945	161,376,708	4892
India	2020-01-30	2020-06-19	395,048	12,948	213,831	1,380,004,385	9995
Pakistan	2020-02-25	2020-06-19	171,666	3382	63,504	22,0695,321	6117

The surprising fact is that these countries have conducted very few COVID-19 tests in susceptible populations. Therefore, it is very likely that this data set may not cover the real scenario of the COVID-19 situation. However, we can use this data set to mainly calculate the reproduction numbers, so as to observe the spread of COVID-19.
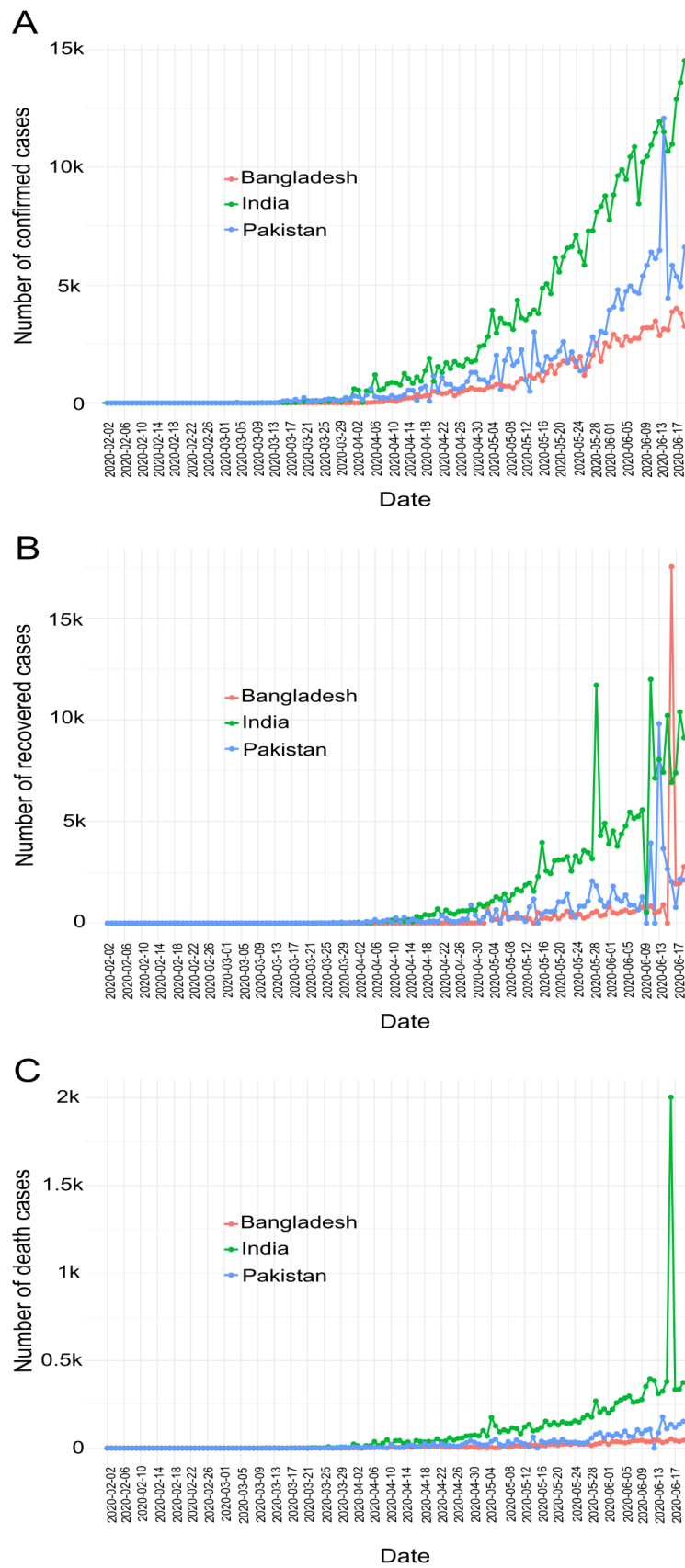
Figure 2. COVID-19 daily cases in Bangladesh, India and Pakistan: (**A**) Confirmed cases (**B**) Recovered cases (**C**) Death cases.

7. Experimental Results

We conduct several experiments to analyze the trend of COVID-19 in the concerned countries. The first experiment focuses on COVID-19 cases as confirmed, deceased and recovered to compare the COVID-19 situation in these countries. The motivation is to understand the deteriorated conditions for the COVID-19 pandemic in this region. Next, we conduct experiments through SIR, EG, SB, ML and TD models to estimate Rs for these countries and predict the COVID-19 pandemic. The result is analyzed to observe the model performances on the COVID-19 dataset. For this, we use the R0 package provided by R programming language to impletement the code for the above methods [48]. More particularly, we employ *estimate.R* function of R0 package to apply the above methods to a given epidemic curve. The result is analyzed to observe the model performances on the COVID-19 dataset.

7.1. COVID-19 Cases

This subsection analyzes the COVID-19 data in order to observe confirmed, death and recovered cases in Bangladesh, India and Pakistan. Figure 2 shows the number of confirmed (Figure 2A), recovered (Figure 2B), and death (Figure 2C) cases in Bangladesh, India and Pakistan respectively. The Figure depicts COVID-19 cases in all these countries are increasing every day. In addition Figure 2, presents that the spread of COVID-19 in India exceeds that of Bangladesh and Pakistan. The Figure also illustrates that Bangladesh has less confirmed cases and deaths apparently more control over the spread of COVID-19 than India and Pakistan as it. Intuitively, the citizens of Bangladesh comparatively better maintained medical care, lock-down and social distances, as a result COVID-19 spread slower from person to person compared to the other two countries and vice versa.

However the aforementioned argument is inconclusive and rather unlikely as Bangladesh conducted the least amount of tests per person compared to the other two. Figure 3 shows that SARS-CoV-2 tests per million people performed in Bangladesh, Pakistan and India are 4892, 6117 and 9995 respectively. Clearly, more tests reveal more infected people. With unexpectedly poor and nonuniform samples we rather not conclude as above and investigate further and focus on how COVID-19 is spreading in the subcontinent. Intuitively, the subcontinent case may reflect COVID-19 spreading in the developing countries, especially spreading in South Asia.
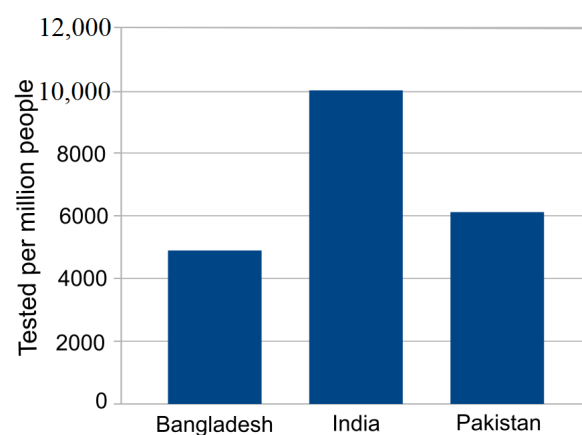

Figure 3. COVID-19 cases tested per million people in Bangladesh, India and Pakistan.

Therefore, the further analysis focuses on the transmission speed of COVID-19 using different methods.

7.2. Prediction with SIR Model

In this experiment, we use the SIR model to predict COVID-19 cases in Bangladesh, India and Pakistan.

Bangladesh. Figure 4A depicts the SIR model fitting to the number of observed confirmed cases where Figure 4B presents the same observation in the logarithmic scale in Bangladesh. We observe that the number of cases (black dotted line as shown in Figure 4A,B) follows the number of confirmed cases expected (black line as shown in Figure 4A,B) by SIR model. Note that, the observed data and predicted values overlapping with each other indicates COVID-19 clearly is in an exponential phase in Bangladesh. Figure 4B shows that the curve is flattening in between mid March (around) to 1st April. This impliedly indicates that the spreading of COVID-19 is comparatively in control in Bangladesh. Furthermore, the slope of the curve dropping down in between mid-March and April 1st conforms to the aforesaid observation (spreading of COVID-19 is in control in Bangladesh during this period). But then, the slope is going up at the steepest of all times until the end of May, i.e. in this time COVID-19 is spreading rapidly. And then, the slope remains constant with a comparatively lower value.

Furthermore, we carry out the experiment to adapt the SIR model to confirmed, death and recovered cases, as shown in Figure 4C. With this experiment, we derive the estimated values of various parameters as shown in Table 2. According to Table 2, we can see that the R_0 is around 1.23 indicating that COVID-19 is spreading in Bangladesh day by day. Based on the R, we calculate the herd immunity threshold using the equation $1 - \dfrac{1}{R_0}$ [50], i.e., 18.97%. With this trend, the epidemic will be at its peak in 2020-08-01. The SIR model estimates that the maximum infected population in Bangladesh will be 3,109,321. Of these, the serious cases will be 621,864 (assuming 20% of the infected population). It also shows that around 186,560 (assuming 6% of the infected population) people need intensive care and up to 108,826 deaths (assuming 3.5% mortality rate).

India. Figure 5A,B represents the cumulative COVID-19 infected cases in number and in logarithmic forms respectively in India. It clearly depicts that the spread of COVID-19 is in an exponential phase in this particular period.

In addition, Figure 5B shows that the slope of the curve (black dotted line) was not steep in between February 1 and mid-March, indicating that the spread of COVID-19 in India was under control during this period. But then, it rose almost exponentially, which means that this time COVID-19 is spreading rapidly. Furthermore, Figure 5C depicts the experimental results of the SIR model for confirmed, death and recovered cases. We put all the estimated values of all the different parameters calculated using the SIR model in the Table 2. The R_0 (derived from the SIR model is about 1.197) indicates that COVID-19 accelerates over time, where the computed herd immunity is 16.97%. With this R_0 the epidemic will peak in 2020-08-20. Moreover, according to the SIR model, the maximum infected population is 19,884,176, among them, the serious cases will be 3,976,835 (assuming 20% of the infected population). Furthermore, around 1,193,051 (assuming 6% of the infected population) people need intensive care and up to 695,946 deaths (assuming a mortality rate of 3.5%) (see Table 2).

Table 2. Prediction with SIR model.

| The Predicted Values for the Following Parameters	Bangladesh	India	Pakistan
Infection Rate, β	0.5524	0.5449	0.550
Recovery rate, γ	0.4475	0.4550	0.449
$R_0 = \dfrac{\beta}{\gamma}$	1.234	1.197	1.22
Herd immunity threshold $(1 - \dfrac{1}{R_0}) \times 100\%$	18.97 % of population	16.49% of population	18.18% of population
Peak of Pandemic	2020-08-01	2020-08-20	2020-08-03
Maximum Infected	3,109,321	19,884,176	3,891,427
Severe cases (assume 20% of Infected cases)	621,864	3,976,835	778,285
Patients need intensive care (assume 6% of Infected cases)	186,560	1,193,051	233,485
Deaths assumed for 3.5% fatality rate	108,826	695,946	136,200
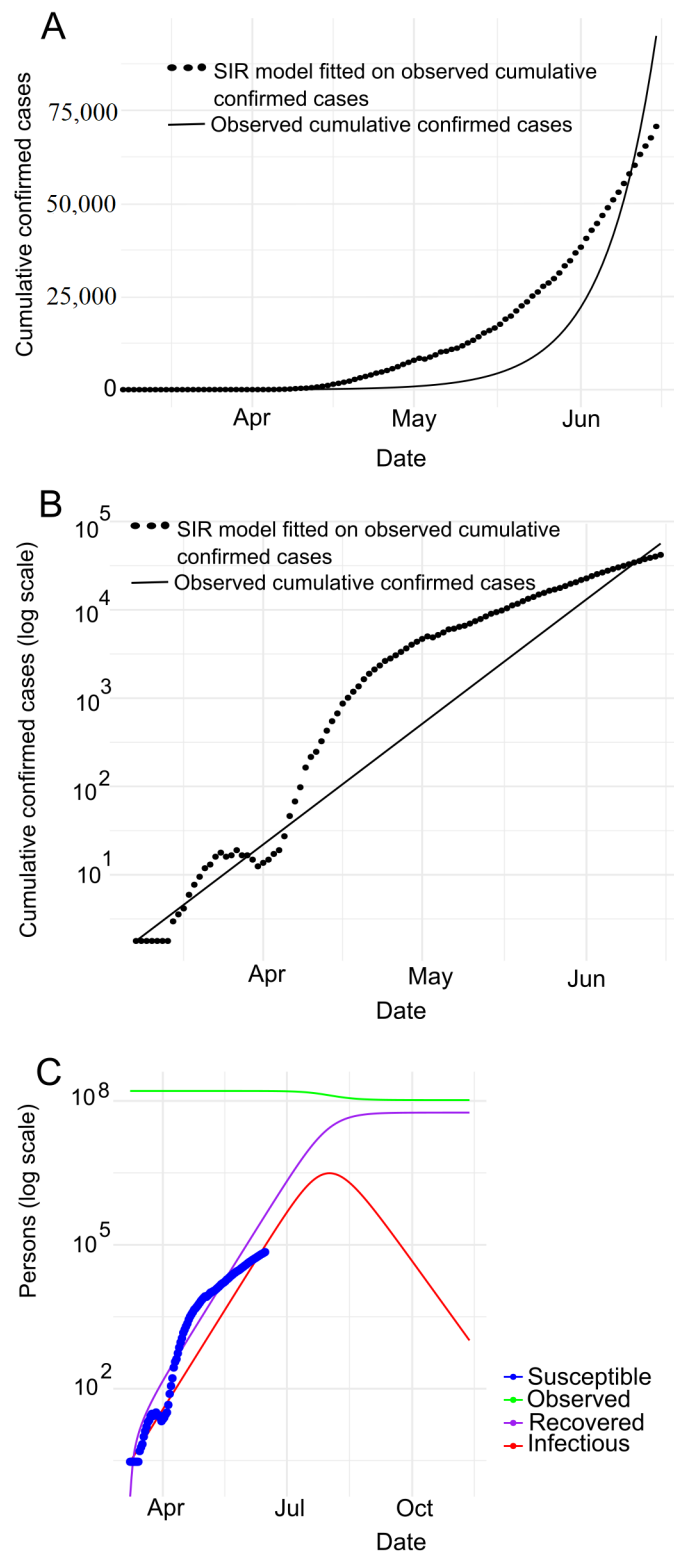
Figure 4. Prediction with SIR model for Bangladesh: (**A**) SIR model fitted on observed cumulative infected cases (**B**) SIR model fitted on observed cumulative infected cases with semi log scale (**C**) SIR model prediction with no human interaction.
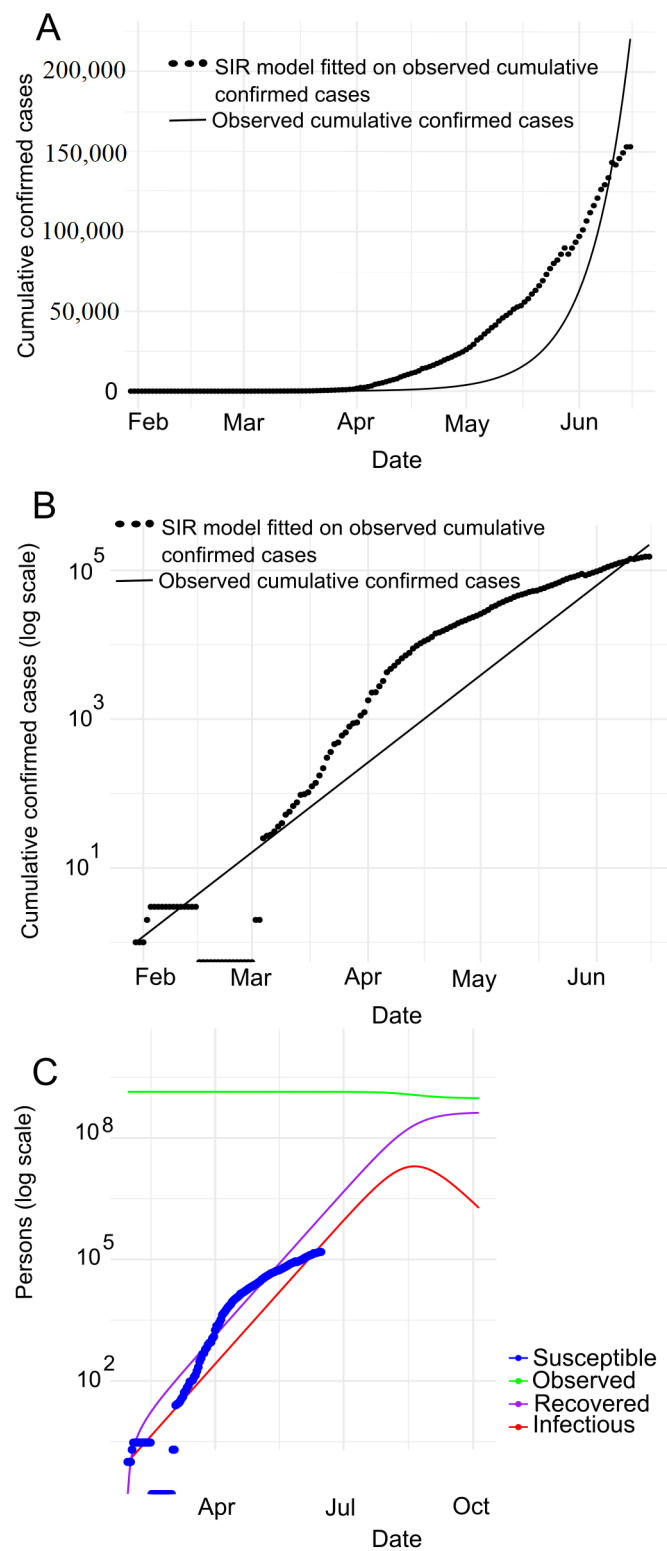
Figure 5. Prediction with SIR model for India: (**A**) SIR model fitted on observed cumulative infected cases (**B**) SIR model fitted on observed cumulative infected cases with semi log scale (**C**) SIR model prediction with no human interaction.

Pakistan. Like Bangladesh and India, we have conducted experiments on Pakistan's COVID-19 data through the SIR model. Figure 6A,B depicts the observed cumulative infected incidence in cases and in cases in logarithmic scale respectively. The figures indicate that the COVID-19 is spreading exponentially. In addition, Figure 6B shows that the slope of the curve (black dotted line) was not accentuated till mid-March, indicating that the spread of COVID-19 in Pakistan. But then it almost shot up, i.e., the COVID-19 is spreading rapidly during this time period. In order to observe deeper insights, we have conducted SIR model experiments on confirmed, death, and recovered cases, as shown in Figure 6C. The estimated values of all the various parameters of the SIR model are shown in Table 2. It is worth noting that the estimated R_0 of the SIR model is about 1.22, indicating that COVID-19 accelerates over time. We find Pakistan's herd immunity as 18.18%. The epidemic will be at its peak in 2020-08-03 with R_0 remaining at this particular rate. According to the SIR model the maximum number of infection cases is 3,891,427. Among them, the serious cases are 778,285 (assuming 20% of the infected population), the intensive care requirement is 233,485 approximately (assuming 6% of the infected population) and estimated maximum deaths is 136,200 (assuming a mortality rate of 3.5%).

Comparison. Table 2 shows all the predicted values of various parameters of Bangladesh, India and Pakistan estimated by the SIR model. It is worthy to mention that the R_0 of Bangladesh and Pakistan are very close, i.e., 1.22 and 1.23, respectively, while R_0 of India is about 1.19. It concludes that Bangladesh and Pakistan are experiencing an increasing number of infected people compared to India. In other words, India has so far somewhat stronger control over the spread of COVID-19 compared to Bangladesh and Pakistan.

Next, we will use the rest of the methods (i.e., EG, SB, ML and TD) to estimate the R in order to verify what the SIR model foresees is consistent with other models.

7.3. COVID-19 Reproduction Number (R_0 / $R(t)$) Estimation

In this experiment, we consider various methods to compute the value of the R_0 and $R(t)$. This experiment investigates how the R has changed with the various methods of the estimation process, the role of excessive dispersion in the distribution of secondary cases and the aggregation of the epidemic curve at ever larger time intervals. In this experiment, we assume that the sequence interval of COVID-19 for Bangladesh, India and Pakistan is equal to the sequence interval of COVID-19 in Wuhan, China with a mean of 5.2 days and a standard deviation of 2.8 days [51].

Figure 7 depicts an estimation of R_0 and $R(t)$ for Bangladesh (see Figure 7a) in the period from March 08 to June 19 (for 104 days), for India (see Figure 7b) in the period from January 30 to June 19 (for 142 days), and for Pakistan (see Figure 7c) in the period from February 25 to June 19 (for 116 days), respectively. Moreover, Table 3 reports the computed R_0 and $R(t)$ for Bangladesh, India and Pakistan which we have estimated with various methods.

According to the Table 3, the R_0 for Bangladesh generated with the methods of exponential growth (EG), maximum likelihood (ML) are 1.3809 [1.3803, 1.3815] and 1.2884 [1.2860 1.2908] respectively which is higher than that of India i.e., 1.3446 [1.3443, 1.34448] with EG and 1.2591 [1.2579, 1.2602] with ML and Pakistan 1.3192 [1.3189, 1.3196] with EG and 1.2645 [1.2626, 1.2662] with ML. Moreover, Table 3 also shows the estimated values of $R(t)$ for Bangladesh, India and Pakistan using the methods of sequential Bayesian and time-dependent. The results show that $R(t)$ value for Pakistan is higher than India and Bangladesh. According to the obtained results depicted in Table 3, we have observed that for different methods the R does not follow the same trend of values. Therefore, we proceed to observe the curve fitting using these Rs on the COVID-19 data. Figure 8 shows that the R generated by maximum likelihood (ML) and time-dependent methods fit well to the data compared to the other methods for all countries. Moreover, Figure 8 shows that SB method fits very poorly to the data for Bangladesh, India and Pakistan. The results show that Bangladesh and Pakistan have comparatively worse control over the spread of COVID-19, meaning $R(t)$ is getting higher than that of India. However, it is worth noting

that the R_0 and $R(t)$ in each country are greater than 1, i.e., COVID-19 is still spreading in all of these countries.
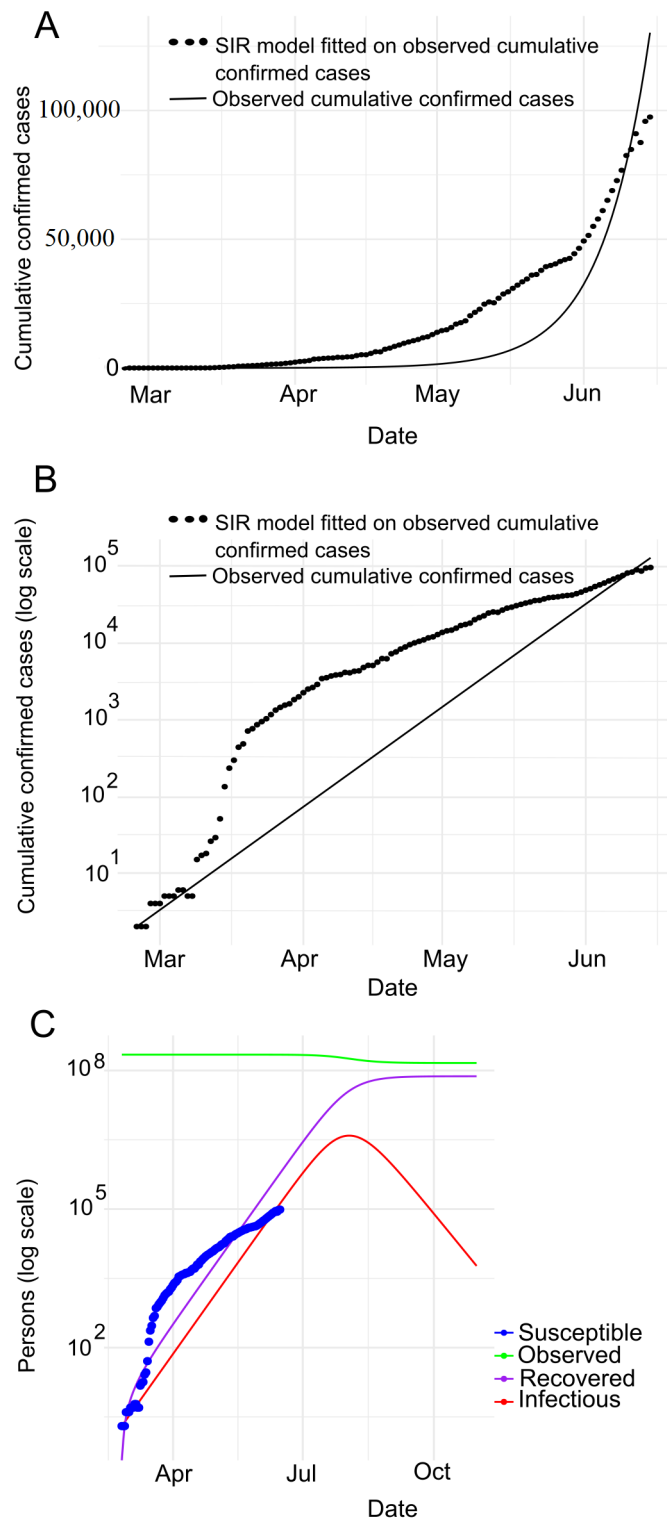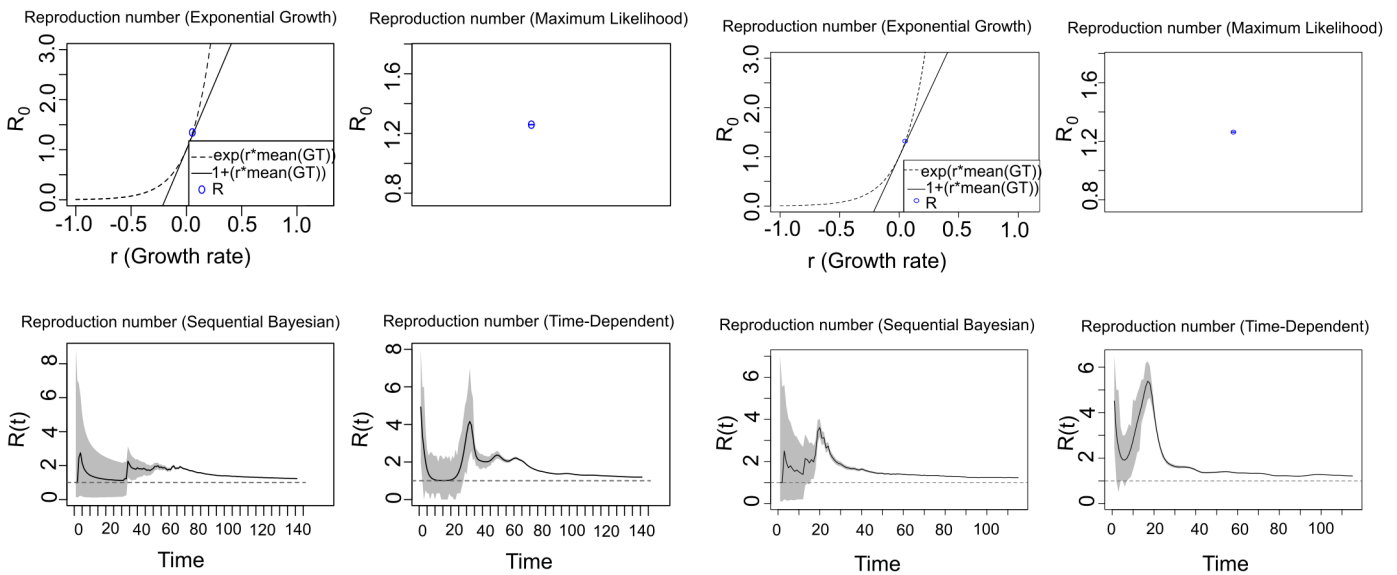

Figure 6. Prediction with SIR model for Pakistan: (**A**) SIR model fitted on observed cumulative infected cases (**B**) SIR model fitted on observed cumulative infected cases with semi log scale (**C**) SIR model prediction with no human interaction.

Table 3. $R_0/R(t)$ estimation by different methods.

Methods	SIR	EG	ML	TD	SB
Reproduction number	R_0	R_0 [CI.lower, CI.upper]	R_0 [CI.lower, CI.upper]	$R_{mean}(t)$ [$R_{low}(t)$, $R_{high}(t)$]	$R_{mean}(t)$[$R_{low}(t)$, $R_{high}(t)$]
Bangladesh	1.234	1.380 [1.380, 1.381]	1.288 [1.286, 1.290]	1.746 [1.209, 3.715]	1.555 [1.000, 2.16]
India	1.197	1.344 [1.344, 1.344]	1.259 [1.257, 1.260]	1.668 [1.007, 4.928]	1.507 [1.00, 2.75]
Pakistan	1.220	1.319 [1.318, 1.319]	1.264 [1.262, 1.266]	1.774 [1.202, 5.381]	1.560 [1.00, 3.61]
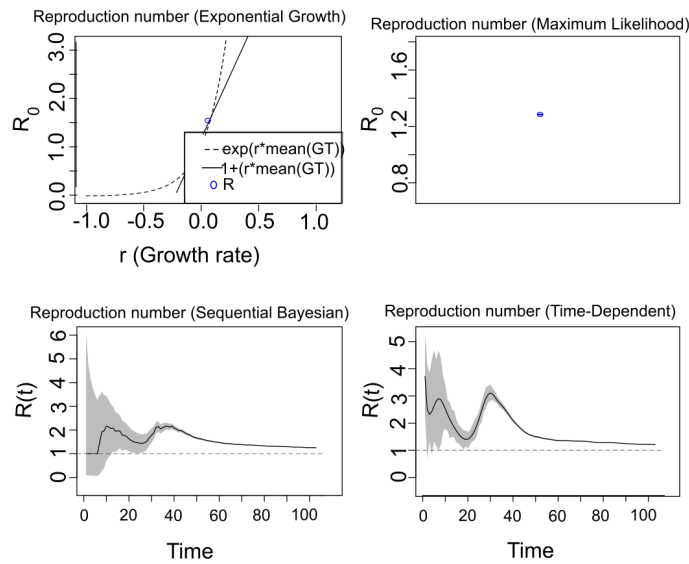
(**a**) Bangladesh

(**b**) India   (**c**) Pakistan

Figure 7. Estimate of the reproduction number $R_0 - R(t)$ in Bangladesh, India and Pakistan.

Figure 8. Models (EG, ML, SB, TD) fitting on COVID-19 confirmed cases.

8. Conclusions

In this article, we provide an assessment of COVID-19 outbreak and measure the spread rate among Bangladesh, India and Pakistan. We utilize the SIR model to predict important parameters such as infection rate, recovery rate, herd immunity threshold, peak of the pandemic, maximum infected, severe cases, patients needing intensive care, deaths of COVID-19 pandemics. We utilize EG, SB, ML and TD models to validate the effectiveness of the estimated parameters of the SIR model. For doing so, we find the basic reproduction number R_0 and effective reproduction number $R(t)$.

Experiments show that in all the considering countries, the estimated effective reproduction number $R(t)$ is much larger than the basic reproduction number R_0, which means that the containment measures implemented by Bangladesh, India and Pakistan are ineffective and inefficient. Besides that, according to different models the reproduction numbers of Bangladesh, India and Pakistan are all higher than about 1.2, indicating that the outbreak of COVID-19 is spreading rapidly.

Result directed recommendations are (i) to adopt stricter prevention and control measures, (ii) to improve the country's quarantine measures, define outcomes slowing down the spread of COVID-19. Failure in doing so the pandemic situation of the region may decline rapidly.

References

1. WHO. Pneumonia of Unknown Cause—China. 2020. Available online: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en (accessed on 30 June 2020).
2. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef]
3. WHO. Novel Coronavirus—China. 2020. Available online: https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en (accessed on 30 June 2020).
4. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **2020**, *5*, 536–544. [CrossRef]
5. WHO. Naming the Coronavirus Disease (COVID-19) and the Virus that Causes It. 2020. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it (accessed on 30 June 2020).
6. WHO. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). 2020. Available online: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) (accessed on 30 June 2020).
7. WHO. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19—11 March 2020. Available online: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (accessed on 30 June 2020).
8. Liu, Y.C.; Kuo, R.L.; Shih, S.R. COVID-19: The first documented coronavirus pandemic in history. *Biomed. J.* **2020**. [CrossRef]
9. Worldometer. COVID-19 CORONAVIRUS PANDEMIC. 2020. Available online: https://www.worldometers.info/coronavirus (accessed on 30 June 2020).
10. Sulaiman, A. On Dynamical Analysis of the Data-Driven SIR model (COVID-19 Outbreak in Indonesia). *medRxiv* **2020**.
11. Zhao, S.; Musa, S.S.; Lin, Q.; Ran, J.; Yang, G.; Wang, W.; Lou, Y.; Yang, L.; Gao, D.; He, D.; 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. Clin. Med.* **2020**, *9*, 388. [CrossRef] [PubMed]
12. Roda, W.C.; Varughese, M.B.; Han, D.; Li, M.Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect. Dis. Model.* **2020**, *5*, 271–281. [CrossRef]
13. Lin, Q.; Zhao, S.; Gao, D.; Lou, Y.; Yang, S.; Musa, S.S.; Wang, M.H.; Cai, Y.; Wang, W.; Yang, L.; et al. A conceptual model for the outbreak of Coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action. *Int. J. Infect. Dis.* **2020**, *93*, 211–216. [CrossRef] [PubMed]
14. Tang, B.; Wang, X.; Li, Q.; Bragazzi, N.L.; Tang, S.; Xiao, Y.; Wu, J. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J. Clin. Med.* **2020**, *9*, 462. [CrossRef] [PubMed]
15. Yang, Z.; Zeng, Z.; Wang, K.; Wong, S.S.; Liang, W.; Zanin, M.; Liu, P.; Cao, X.; Gao, Z.; Mai, Z.; et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* **2020**, *12*, 165. [CrossRef]
16. Fanelli, D.; Piazza, F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals* **2020**, *134*, 109761. [CrossRef]
17. Salgotra, R.; Gandomi, M.; Gandomi, A.H. Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming. *Chaos Solitons Fractals* **2020**, *138*, 109945. [CrossRef] [PubMed]
18. Djilali, S.; Ghanbari, B. Coronavirus pandemic: A predictive analysis of the peak outbreak epidemic in South Africa, Turkey, and Brazil. *Chaos Solitons Fractals* **2020**, *138*, 109971. [CrossRef]
19. Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.; Yan, P.; Chowell, G. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect. Dis. Model.* **2020**, *5*, 256–263. [CrossRef]

20. Li, L.; Yang, Z.; Dang, Z.; Meng, C.; Huang, J.; Meng, H.; Wang, D.; Chen, G.; Zhang, J.; Peng, H.; et al. Propagation analysis and prediction of the COVID-19. *Infect. Dis. Model.* **2020**, *5*, 282–292. [CrossRef]

21. Acuña-Zegarra, M.A.; Santana-Cibrian, M.; Velasco-Hernandez, J.X. Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. *Math. Biosci.* **2020**, *325*, 108370. [CrossRef]

22. Jung, S.m.; Akhmetzhanov, A.R.; Hayashi, K.; Linton, N.M.; Yang, Y.; Yuan, B.; Kobayashi, T.; Kinoshita, R.; Nishiura, H. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference using exported cases. *J. Clin. Med.* **2020**, *9*, 523. [CrossRef]

23. Powell, D.R.; Fair, J.; LeClaire, R.J.; Moore, L.M.; Thompson, D. Sensitivity analysis of an infectious disease model. In Proceedings of the International System Dynamics Conference, Boston, MA, USA, 17–21 July 2005.

24. Feng, Z.; Glasser, J.W.; Hill, A.N. On the benefits of flattening the curve: A perspective. *Math. Biosci.* **2020**, *326*, 108389. [CrossRef] [PubMed]

25. Dhanwant, J.N.; Ramanathan, V. Forecasting COVID 19 growth in India using Susceptible-Infected-Recovered (SIR) model. *arXiv* **2020**, arXiv:2004.00696.

26. Bertozzi, A.L.; Franco, E.; Mohler, G.; Short, M.B.; Sledge, D. The challenges of modeling and forecasting the spread of COVID-19. *arXiv* **2020**, arXiv:2004.04741.

27. De Castro, C.A. SIR Model for COVID-19 calibrated with existing data and projected for Colombia. *arXiv* **2020**, arXiv:2003.11230.

28. Qi, C.; Karlsson, D.; Sallmen, K.; Wyss, R. Model studies on the COVID-19 pandemic in Sweden. *arXiv* **2020**, arXiv:2004.01575.

29. Shim, E.; Tariq, A.; Choi, W.; Lee, Y.; Chowell, G. Transmission potential and severity of COVID-19 in South Korea. *Int. J. Infect. Dis.* **2020**, *93*, 339–344. [CrossRef] [PubMed]

30. Benvenuto, D.; Giovanetti, M.; Vassallo, L.; Angeletti, S.; Ciccozzi, M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data Brief* **2020**, *29*, 105340. [CrossRef] [PubMed]

31. Johns Hopkins University Center for Systems Science and Engineering. Available online: https://github.com/CSSEGISandData/COVID-19 (accessed on 21 July 2020).

32. Fattah, J.; Ezzine, L.; Aman, Z.; El Moussami, H.; Lachhab, A. Forecasting of demand using ARIMA model. *Int. J. Eng. Bus. Manag.* **2018**, *10*, 1847979018808673. [CrossRef]

33. Calafiore, G.C.; Novara, C.; Possieri, C. A modified sir model for the covid-19 contagion in italy. *arXiv* **2020**, arXiv:2003.14391.

34. Kucharski, A.J.; Russell, T.W.; Diamond, C.; Liu, Y.; Edmunds, J.; Funk, S.; Eggo, R.M.; Sun, F.; Jit, M.; Munday, J.D.; et al. Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *Lancet Infect. Dis.* **2020**, *20*, 553–558. [CrossRef]

35. Peng, L.; Yang, W.; Zhang, D.; Zhuge, C.; Hong, L. Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv* **2020**, arXiv:2002.06563.

36. Wangping, J.; Ke, H.; Yang, S.; Wenzhe, C.; Shengshu, W.; Shanshan, Y.; Jianwei, W.; Fuyin, K.; Penggang, T.; Jing, L.; et al. Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Front. Med.* **2020**, *7*, 169. [CrossRef] [PubMed]

37. Chatterjee, K.; Chatterjee, K.; Kumar, A.; Shankar, S. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Med J. Armed Forces India* **2020**, *76*, 147–155. [CrossRef]

38. Liang, K. Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS. *Infect. Genet. Evol.* **2020**, *82*, 104306. [CrossRef]

39. Ndairou, F.; Area, I.; Nieto, J.J.; Torres, D.F. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals* **2020**, *135*, 109846. [CrossRef]

40. Hellewell, J.; Abbott, S.; Gimma, A.; Bosse, N.I.; Jarvis, C.I.; Russell, T.W.; Munday, J.D.; Kucharski, A.J.; Edmunds, W.J.; Sun, F.; et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **2020**, *8*, e488–e496. [CrossRef]

41. van den Driessche, P. Reproduction numbers of infectious disease models. *Infect. Dis. Model.* **2017**, *2*, 288–303. [CrossRef]

42. Fraser, C.; Donnelly, C.A.; Cauchemez, S.; Hanage, W.P.; Van Kerkhove, M.D.; Hollingsworth, T.D.; Griffin, J.; Baggaley, R.F.; Jenkins, H.E.; Lyons, E.J.; et al. Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* **2009**, *324*, 1557–1561. [CrossRef] [PubMed]

43. Rodpothong, P.; Auewarakul, P. Viral evolution and transmission effectiveness. *World J. Virol.* **2012**, *1*, 131. [CrossRef]

44. Ma, J. Estimating epidemic exponential growth rate and basic reproduction number. *Infect. Dis. Model.* **2020**, *5*, 129–141. [CrossRef] [PubMed]

45. Roberts, M.; Heesterbeek, J. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *J. Math. Biol.* **2007**, *55*, 803. [CrossRef]

46. Bettencourt, L.M.; Ribeiro, R.M. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE* **2008**, *3*, e2185. [CrossRef]

47. Forsberg White, L.; Pagano, M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat. Med.* **2008**, *27*, 2999–3016. [CrossRef]

48. Obadia, T.; Haneef, R.; Boëlle, P.Y. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med. Informatics Decis. Mak.* **2012**, *12*, 1–9. [CrossRef] [PubMed]

49. Wallinga, J.; Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **2004**, *160*, 509–516. [CrossRef] [PubMed]
50. Fine, P.; Eames, K.; Heymann, D.L. Herd immunity: A rough guide. *Clin. Infect. Dis.* **2011**, *52*, 911–916. [CrossRef] [PubMed]
51. Ganyani, T.; Kremer, C.; Chen, D.; Torneri, A.; Faes, C.; Wallinga, J.; Hens, N. Estimating the generation interval for COVID-19 based on symptom onset data. *MedRxiv* **2020**.