Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Building a sensible SIR estimation model for COVID-19 outbreak in Kuwait

Mohammed N. Alenezi a, Fawaz S. Al-Anzi b,*, Haneen Alabdulrazzaq a

a Computer Science & Information Systems Department, The Public Authority for Applied Education and Training (PAAET), Kuwait
b Computer Engineering Department, Kuwait University, Kuwait

Received 13 June 2020; revised 24 September 2020; accepted 18 January 2021
Available online 4 February 2021

Abstract The Susceptible - Infected - Recovered (SIR) model is used in this research to analyze and predict the outbreak of coronavirus (COVID-19) in Kuwait. The time dependent SIR model is used to model the growth of COVID-19 and to predict future values of infection and recovery rates. This research presents an analysis on the impact of the preventive measures taken by Kuwait’s local authorities to control the spread. It also empirically examines the validity of various values of $R_0$ ranging from 2 to 5.2. The proposed model is built using Python language modules and simulated using official data of Kuwait in the period from February 24th to May 28th of 2020. Our results show the SIR model is almost fitted with the actual confirmed cases of both infection and recovery for the values of $R_0$ ranging from 3 to 4. The results shown indicate COVID-19 peak infection rates and their anticipated dates for Kuwait. It has been observed from the obtained prediction that if preventive measures are not strictly followed, the infection numbers will grow exponentially.

1. Introduction

These days the world faces an unparalleled challenge of the spread of an infectious virus known as Coronavirus or COVID-19. The family of coronaviruses consists of various viruses that cause illness ranging from a simple cold to more severe diseases like Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). The recently discovered COVID-19 virus belongs to this family of viruses. The first case of COVID-19 was found and reported in the city of Wuhan, China, on the 31st of December 2019, hence the name COVID-19. Since the rapid spread of this epidemic; the Chinese government introduced several measures to hinder this outbreak such as locking down the city of Wuhan and closing all routes that lead to it, in late January [1]. Up until the date of writing this paper, the pandemic has been officially acknowledged in almost 213 countries and territories around the world. The first reported case of COVID-19 outbreak outside China was in Thailand on January 13th 2020 [2]. On May 28th 2020, around 5.76 million confirmed cases,
2.39 million recovered cases, and 358 thousand deaths were registered worldwide [3].

The World Health Organization (WHO) declared COVID-19 an epidemic and Global Public Health Emergency on January 30th 2020 [4]. On that date, the total number of infected cases was 8,096. The WHO declared COVID-19 a global pandemic on March 11th 2020 because it imposed a threat to the whole world [5]. More than 40% of the World’s population have been placed under lockdowns issued by various governments in order to reduce the spread of COVID-19. Social distancing is an agreed upon measure and known to be among the best viable ways to reduce the spread of this novel coronavirus. The rapid and uncontrollable increase in the total number of cases globally has created a worldwide public health issue. This pandemic created significant challenges not only in public health but also in various areas, including politics, economics, education, and social behavior. Furthermore, this epidemic caused an intensification in poverty and unemployment globally. Due to its strong infectious nature, extended incubation period, difficulty in detection, and vagueness in transmission ways, COVID-19 is becoming a very difficult disease to control. Countries around the world have joined in tremendous efforts to decrease or prevent the outbreak of COVID-19 [6]. Many health and pharmaceutical research centers and companies are racing with time to develop a vaccine or cure to treat COVID-19. However, none of these efforts have been successful thus far. The virus has taken its toll on the world’s economy and many countries are undergoing major economic crises.

Like the rest of the world, Kuwait currently is facing the many challenges of COVID-19. The first five cases of COVID-19 infection were reported on February 24th 2020 and were associated with citizens returning from travel abroad. This sparked the threat of local transmission in the country. Upon the rapid increase in numbers of confirmed infected cases, the Kuwaiti authorities started to take preventive measures including; quarantine, banning in-bound flights from various countries, closing down retail shops, and issuing a public holiday from 12th March 2020 onward. The government issued a partial curfew on the 22nd of March 2020 from 5:00 PM to 4:00 AM daily, then amended the timings twice to be from 5:00 PM to 6:00 AM and then from 4:00 PM to 8:00 AM and finally a full lockdown that was implemented from May 10th until May 30th 2020 [7]. On May 28th 2020, the total number of confirmed infections was reported to be 24,112 from which 8,698 were recovered and 15,229 were in treatment. The overall death toll was 185, as well as 185 of the cases were considered to be under critical conditions. As of June 1st 2020, the total population size of the state of Kuwait is 4,776,407 based on data from Public Authority for Civil Information (PACI) [8]. Kuwait’s primary healthcare provider is the Ministry of Health (MOH), and there is a small number of private sector hospitals and clinics. The total bed capacity of MOH is to be around 7,118 while the private sector’s capacity is estimated at 1,082 beds [9].

In light of the uncertainties facing the world today, it is becoming essential for decision makers to have good estimates for the damages COVID-19 has inflicted so far, and what it will cause in the near future. The estimation of the epidemic spread would help the official authorities to warrant effective prevention measures and to be able to prepare for care and treatment actions. Although, a precise estimate of any pandemic is considered unachievable, nevertheless, scientists and researchers can attempt to make rough estimates by using proven scientific-based prediction techniques. Based on such estimates, official authorities can make good and informed decisions on how to proceed onwards in their efforts to control the damages caused by COVID-19.

One of the simple yet effective mathematical models to predict a pandemic is a popular model called the SIR model. The name SIR comes from Susceptible-Infected-Recovered [10]. In an SIR model, the entire population is placed in one of 3 categories: Susceptible, Infected, or Recovered. People who are not yet infected with the disease are considered susceptible. Those who are confirmed to be infected and capable of transferring the disease are placed in the infected category. The people who have recovered from the disease or are deceased are considered recovered. These three categories represent the progressive stages of a contagious disease. The SIR model is well suited to predict the number of the population who would need medical care during an epidemic spread. The SIR model assumes that a person that has recovered from a disease will attain a lifetime immunity against it and will never get infected again. SIR models are simple mathematical models and yet have been known to adequately predict a pandemic with accuracy [11].

In this paper, we use the SIR model to estimate and analyze the spread of COVID-19 pandemic in Kuwait with an assumption that the country will be almost constant in population size, which means deaths, births, and migration will cancel each other in the total population count during our analysis due to the narrow time window. Our prediction of daily confirmed cases, cumulative cases, and recovered cases are estimated with different values of basic reproduction number $R_0$ that will be described in the next sections. The data processed in this paper is based on the daily confirmed cases from 24th of February until the 28th of May 2020 and is retrieved from the official website for COVID-19 in Kuwait [7]. Our research results show a promising prediction of the peak and end dates of the COVID-19 outbreak for Kuwait.

2. Regression models

The UN Secretary-General Antonio Guterres said that currently the world is facing its most challenging crisis since World War II. People all over the world are in a state of panic due to the new coronavirus outbreak. This pandemic has shown severe effects not only on public health, but also on social, economic, and political aspects. COVID-19 has coerced countries to impose travel bans and restrictions as well as the promotion of large-scale quarantines worldwide in an attempt to restrain its spread.

Several studies, by different researchers, were done to forecast and model the COVID-19 epidemic. The studies were performed to track this epidemic’s spread and estimate its infection rate and expected ending. Various researchers utilized different models for analyzing and forecasting COVID-19, such as the logistic growth model, deterministic compartmental models (DCM), and agent-based models (ABM) [12]. Statistical models are well suited for predicting pandemics. A variety of regression models, such as linear regression with different orders (2nd and 3rd), Locally Weighted Linear Regression (loess), Generalized Linear Model (GLM), Poisson, and
logistic regression, were also used in studies. These methods are based on the type and number of dependent and independent variables, and the shape of the regression line.

Regression analysis is mainly used to analyze and predict the relationship of a dependent or target variable with an independent variable, which predicts the target variable based on previous values. It can be used to analyze the relationships between two or more dependent and independent variables. A regression is said to be linear if it contains linear parameters. Therefore, polynomial regressions are also known to be linear regressions. The dependent variable in the second-order polynomial regression with one variable can be calculated using Eq. (1). The independent variable, x can be the number of tests performed, gender, age, region, etc., whereas dependent variable(y) can be the number of confirmed cases, recovered cases, death cases, etc., which is the intercept also bias, and β₁, β₂ are the weights(slopes). In real life situations, a regression model cannot be able to predict exactly a dependent variable using an independent variable. So there exists some error which is represented by ε. It adds noise of this relationship between dependent and independent variables.

\[ y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \]  
(1)

The relationship of the dependent and independent variables which contains two explanatory variables is given the Eq. (2).

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_1 x_1^2 + \beta_2 x_2^2 + \varepsilon \]  
(2)

If \( x_i = x^i \) for \( i = 1, 2, \ldots, n \), then the polynomial regression is viewed as a linear regression of multiple independent variables such as \( x_1, x_2, \ldots, x_n \). The second-order polynomial regression is also known as a second-order model or quadratic model. \( \beta_1 \) and \( \beta_2 \) are known as linear effect parameters and quadratic effect parameters. The value of \( \beta_0 \) is y when \( x = 0 \). The dependent variable in third-order regression can be calculated by Eq. (3). Polynomial models are used to predict or estimate the values of dependent variables in which those relationships are curvilinear or nonlinear. \( \beta_1, \beta_2, \beta_3, \ldots, \beta_p \) are coefficients of variables.

\[ y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon \]  
(3)

The Generalized Linear Model (GLM) [13] is a regression model used to estimate or analyze the effect of various continuous independent variables on different dependent variables. In GLM, the data is calculated as a sum of the generated model and possible error. It can also be viewed as a flexible generalization of linear or polynomial regression models. It mainly consists of a random component, linear predictor, and a link function g(.). A random component represents the conditional distribution of dependent variable \( Y_i \) which is defined by the independent variables. A linear predictor is the linear function of dependent and independent variables which is given in Eq. (4).

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_p x_{ip} + \varepsilon_i \]  
(4)

where \( i = 1, 2, \ldots, n \) and the link function is an invertible function, which defines how the mean, \( E(Y_i) = \mu_i \) depends on the linear function.

Locally Weighted Linear Regression(loess) [14] is a non-parametric regression model that is mainly used to smooth the curve in volatile time-series using a scatter plot to get the best fitting data. It is used in local subsets to smooth their values. The loess method first identifies a smoothing parameter, then selects k nearest neighbors of \( x_0 \), which is to be smoothened. Loess algorithm assigns the weights of each point of \( x_0 \) to its nearest neighbors.

Poisson regression [15] is a regression model used to estimate the discrete dependent or response variable; it assumes response variables are positive counts which follow the Poisson distribution. Logistic regression is similar to Poisson regression. Poisson regression is mainly used to analyze the rates whose values are positive counts. In contrast, logistic regression is mainly used to calculate ratios whose values lie in the range of 0 and 1.

Logistic regression model [16] is a regression model which is used to predict or estimate a dependent variable based on the independent variable under consideration. It is a well-suited regression model to analyze the growth of epidemic diseases. The model considers that epidemics increase exponentially at the starting stage, then a steady increase occurs, and finally a decline in its growth rate. \( C \) represents the number of infected cases, \( r \) is the rate of infection, and \( K \) is the final epidemic size. The number of infected cases is calculated by Eq. (5) with initial condition \( C(0) = C_0 \) [16].

\[ C = \frac{K}{1 + \left( \frac{x - C_0}{\varepsilon} \right)^p} \]  
(5)

The changes in the number of infections at \( t \) is defined by Eq. (6) [16].

\[ \frac{dC}{dt} = Cr \left( 1 - \frac{C}{K} \right) \]  
(6)

Maximum growth rate occurring at time \( t_p \) is calculated using Eq. (7).

\[ t_p = \frac{\ln \left( \frac{x - C_0}{\varepsilon} \right)}{r} \]  
(7)

The peak number of cases and maximum growth rate at maximum peaks are defined using the following Eqs. (8) and (9) respectively.

\[ C_p = \frac{K}{2} \]  
(8)

\[ dC \]  
(9)

For fitting the actual confirmed infected population to the regression model where \( b_i = b_i(t) \) is the actual estimate for \( i \) ranging from 1 to \( n \) is given in Eq. (10).

\[ C = \frac{b_1}{1 + b_2 e^{-b_1 t}} \]  
(10)

3. Deterministic compartmental models

Deterministic compartmental models (DCM) are nonlinear models used to estimate the spread of an epidemic. In DCM, differential equations are used to model the epidemic spread. Susceptible-Infected-Recovered (SIR), Susceptible-Exposed-Recovered (SEIR), and Autoregressive Integrated Moving Average (ARIMA) model are the three mainly used DCMs, for estimating the epidemic spread.
3.1. Susceptible-Infected-Recovered (SIR) Model

The SIR model has been used in the past to estimate many diseases including HIV and Ebola [17, 18]. SIR considers the total population as a combination of three parameters: Susceptible (S), Infected (I), and Recovered (R) [19]. Susceptible holds the total population which are healthy but are at risk of being infected. Infected is the number of mildly or severely infected population. Recovered is the total number of both recovered persons from the epidemic who attained immunity and it includes persons who have deceased [20]. The total population, N can be written by Eq. (11) [10].

\[ N = S + I + R \]  (11)

With the SIR model, the population is assumed to be constant; no deaths and births during the period of epidemic prediction. The model calculates the changes in S, I, R using differential equations represented in Eqs. (12)–(14) respectively [21, 22].

\[ \frac{dS}{dt} = -\frac{\beta IS}{N} \]  (12)

\[ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \]  (13)

\[ \frac{dR}{dt} = \gamma I \]  (14)

\( \beta \) and \( \gamma \) are deterministic parameters that reflect the infection rate at which the susceptible population is infected per day and recovery rate at which they become recovered with immunity [23]. The basic reproduction number of the disease can be calculated using Eq. (15).

\[ R_0 = \frac{\beta}{\gamma} \]  (15)

The Residual Sum of Squares is a statistical method to determine the variance in a dataset which is not considered by the proposed model. It calculates the error between the dataset and estimation model. In SIR based model, RSS is used to find the optimal values of \( \beta \) and \( \gamma \), which calculates the error rate with the model using the given infection and recovery rates. RSS is computed using Eq. (16).

\[ RSS(\beta, \gamma) = \sum_{i=1}^{n} (I_{actual}(t) - I_{predicted}(t))^2 \]  (16)

The coefficient of determination\( (R^2) \) is another statistical method used as goodness-of-fit measure which is defined as the percentage of variance in the dependent variable which can be estimated using independent variables. It determines the relationship strength between the estimation model and the dependent variable. Normally, its value ranges from 0 \(-100\%\) (0 \(-1\)). \( R^2 \) can be measured using Eq. (17).

\[ R^2 = 1 - \frac{RSS}{TSS} \]  (17)

TSS is the total sum of squares which finds the sum of squared differences of predicted variable\( (y_i, i \leq n) \) with their overall mean\( (\bar{y}) \) and can be calculated using Eq. (18).

\[ TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 \]  (18)

SIR estimates the outbreak based on some initial parameters such as initial susceptible population \( (S(0)) \), initial infected population \( (I(0)) \), and the initial recovered population \( (R(0)) \).

3.2. Susceptible-Exposed-Infected-Recovered (SEIR) Model

SEIR model is an advancement of the SIR model where the total population is divided into four compartments instead of three. SEIR model assumes that the entire population is susceptible [24]. Exposed (E) refers to the population who were exposed to an infected person and have become infected, but are not yet infectious. The total population in SEIR model is represented by Eq. (19).

\[ N = S + E + I + R \]  (19)

SEIR model also considers no deaths or births during the estimation period. The changes in each time \( t \) of SEIR model is given in Eqs. (20)–(23) [25].

\[ \frac{dS}{dt} = -\frac{\beta IS}{N} \]  (20)

\[ \frac{dE}{dt} = -\frac{\beta IS}{N} - \alpha E \]  (21)

\[ \frac{dI}{dt} = \alpha E - \gamma I \]  (22)

\[ \frac{dR}{dt} = \gamma I \]  (23)

\( \alpha \) is known as incubation rate, the rate at which an individual became infectious [2]. The values of \( \beta \), \( \alpha \), and \( \gamma \) are calculated using \( \beta = \frac{\gamma}{T_i} \), \( \alpha = \frac{1}{T_i} \), and \( \gamma = \frac{1}{T_i} \). \( T_i \) and \( T_l \) are defined as serial and incubation period [25].

3.3. Autoregressive Integrated Moving Average (ARIMA) Model

ARIMA is a statistical model used to estimate or analyze time-series data and forecast future values. ARIMA is a widely used statistical model for predicting the periodic changes and analyzing time series data [26]. AR in ARIMA refers to AutoRegression which is a model used to identify the relationship of the observation with other lagged observations. Integrated, is a pre-processing step to make the time-series stationary with the help of differentiation of observations. Moving Average (MA) is applied to lagged observations using the observation and residual error dependency. The ARIMA model using lag polynomials, is given by Eqs. (24) and (25) [27].

\[ \varphi(L)(L - 1)^d y_t = \theta(L)\epsilon_t \]  (24)

\[ \left(1 - \sum_{i=1}^{p} \phi_i(L)\right)(L - 1)^d y_t = \left(1 + \sum_{i=1}^{q} \theta_i(L)\right)\epsilon_t \]  (25)

The values for \( p, d, \) and \( q \) must be greater than or equal to 0. The model with ARIMA\((p, 0, q)\) is known as ARMA\((p, q)\) model, ARIMA\((p, 0, 0)\) is an AR\((p)\) model, and ARIMA \((0, 0, q)\) is a MA\((q)\) model. In most cases, the time-series data will be differentiated once (value of \( d \) is 1). Random Walk model
is a special case of ARIMA model with $p = 0, q = 0,$ and $d = 1$, then the $y_t$ can be calculated using Eq. (26) [27].

$$y_t = y_{t-1} + e_t$$  \hspace{1cm} (26)

Several studies, from various researchers all over the world, were performed to forecast and predict the COVID-19 epidemic. Distante et al. [25] presented a paper related to COVID-19 outbreak in Italy. They estimated the peak of COVID-19 outbreak using the SEIR model. Based on their forecast, the epidemic reaches its peak value by the end of March 2020 in northern regions, and by the first week of April 2020 in southern regions. They did two different experiments to estimate the basic reproduction number $R_0$; one is based on the prediction of daily cases, and the other one is based on the studied period. Their prediction of the peak time has proven to be correct, as Italy’s numbers started going down at the end of March 2020.

Almeshal et al. [28] presented a logistic regression-based model for forecasting Kuwait’s epidemic size. They did the research based on confirmed cases in Kuwait from 24th of February 2020 to 19th of April 2020. They calculated the basic reproduction number as 2.2 based on the confirmed cases. Their study analyzed the effectiveness of COVID-19 preventive measures issued by the Kuwaiti authorities to control the epidemic. They also examined the impact of the Kuwaiti authorities’ repatriation plan.

Syed and Sibgatullah [29] offered an SIR based study on COVID-19 epidemic in Pakistan based on the reported cases from their national database. They estimated the peak value on 26th May 2020. Based on the study, they advised having policies in place to control the spread, or otherwise, 90% of the entire population of Pakistan might get infected before the 24th of July 2020.

Rahman et al. [20] performed an SIR based analysis and estimation of the coronavirus spread in Bangladesh. Their analysis focuses on showing the impact of prevention measures such as social distancing in the control of COVID-19. Based on their study, the final size of infection in Bangladesh would be 3,782,558 on the 92nd day. Based on their research, social distancing has a high impact on the epidemic spread, and strict social distancing can put the epidemic under control.

Batista [30] performed an SIR based estimate to identify the final size of the coronavirus spread in China, South Korea, and the rest of the World. He estimated the final number of infected persons and evaluated the model with $R^2$ score. In forecasting the final size, he used both a logistic and an SIR model.

Sene [31] conducted a study on SIR based epidemic model of Atangana-Baleanu Caputo fractional derivative with delay which is a non-singular fractional derivative with Mittag–Leffler kernel with delay. In this model, the positivity of the solutions was heavily based on the order of the Atangana-Baleanu-Caputo fractional derivative. They also performed a detailed analysis of the uniqueness of the suggested model based on the used fractional derivative. They used Lyapunov direct method for analyzing the global stability of this model after finding the reproduction number and equilibrium point.

Atangana [32] conducted a study on various constraints issued to account for COVID-19 and check the effectiveness of integral lockdown for saving human life. He suggested a model which considers the importance of lockdowns and the possibility of COVID-19 transmission from dead bodies. Mainly three cases were taken into account. The first case considered the transmission of the disease from a dead body to a living human during the postmortem and burial ceremonies and removed this chance in the second case. For the third case, he accounted for the effects of issued lockdowns and social distancing. A mathematical model was constructed by applying the differential and integral operators based on the data of Italy and this model agreed with the effectiveness of integral lockdown. Atangana also mentioned the detrimental effects of inadequate testing.

4. Prediction based on SIR model

Coronavirus has already spread in 213 countries in the world among which is Kuwait. The SIR model is one of the most widely used models when a prediction of a disease outbreak is required. The SIR model has been applied to estimate the COVID-19 outbreak by various researchers for different countries around the world. Brauer indicates that “In order to prevent a disease from becoming endemic it is necessary to reduce the basic reproduction number $R_0$ below one. This may sometimes be achieved by immunization”[33]. Since a vaccine is not currently available, it becomes essential to give a reasonable estimate for $R_0$ so that healthcare decision makers can take any necessary measures needed to contain the spread of COVID-19. For SIR model estimation, the time varying infection rate ($\beta(t)$) and recovery rate ($\gamma(t)$) are the two important parameters used. This research focuses on an SIR based estimation using the confirmed cases in Kuwait for the period of February 24th 2020 to May 28th 2020. The SIR model is simulated using Python programming language with the help of some predefined Python modules or tools such as sklearn, matplotlib, xld, xlswriter, and math [34]. The sklearn module is an effective machine learning platform built on NumPy, SciPy, and matplotlib; it is used for error calculation using RMSE and $R^2$. The graphs are plotted with the help of matplotlib. The data is collected from recognized sources such as the Kuwaiti government’s official COVID-19 website [7] and WHO. The collected data of infections and recoveries are plotted against time and shown in Fig. 1.

Based on the collected information the value of recovery rates($\gamma$) and infection rates($\beta$) are calculated against time (Day), which is plotted in Fig. 2. The infection and recovery rates would indicate percentage of newly infected or recovered population as a fraction of already infected population. For example, an infection rate of 0.2 indicates that 20% of the already infected population at time $t$ are infected at time $t + \Delta t$. The value of $\beta$ and $\gamma$ at any time $t$ are calculated using Eqs. (27) and (28).

$$\beta(t) = \frac{I(t + \Delta t) - I(t)}{I(t)\Delta t}$$  \hspace{1cm} (27) $$\gamma(t) = \frac{R(t + \Delta t) - R(t)}{I(t)\Delta t}$$  \hspace{1cm} (28)

The value of $S$, $I$, and $R$ are calculated at any time ($t + 1$) from the values of these populations at time $t$, given by Eqs. (29)–(31), respectively [35].

$$S(t + 1) = S(t) + \frac{dS}{dt}$$  \hspace{1cm} (29)
The value of \( \frac{dS}{dt} \), \( \frac{dI}{dt} \), and \( \frac{dR}{dt} \) are calculated using Eqs. (12)–(14), respectively.

Based on the collected data, the cumulative counts for the infection and recovery are calculated. They are analyzed and compared with the estimated count using Root Mean Square Error, \( R^2 \) score and the RSS measures. The initial values for the susceptible population, infected population, and the recovered population are 4776402, 5, and 0, respectively.

The basic reproduction rate is calculated as a fraction of infection rate with the recovery rate. The value of \( R_0 \) determines the transmission rate of emerging epidemic. The basic reproduction number detects the number of expected infections during the early stages of the outbreak [36]. The value of \( R_0 \) determines a disease will or will not become an outbreak. If \( R_0 \) value is greater than one, then the disease will grow exponentially and affect a significant portion of the entire population [36,22]. \( R_0 \) defines that an infected person is recovered in \( 1/\gamma \) days and he/she has an average of \( \beta \) contacts [37]. If the value of \( R_0 \) is below 1, there will not be an outbreak and it will suddenly decline. This study is conducted using different \( R_0 \) values.

5. Results and discussion

Our estimation is based on the confirmed cases from 24th February 2020 to 28th May 2020. February 24th 2020 is when the first five cases were reported. The values of \( \beta \) and \( \gamma \) are changed over time. In the early stages, the infected cases increased slowly, and the recovery rate is zero for initial cases. The cumulative infection and recovery numbers are forecasted using the SIR model based of different \( R_0 \) values as shown in Figs. 3 and 4.

Based on this estimation, the infection reaches its peak value between the 23rd of July 2020 and the 22nd of August.
2020. The growth rate is observed to be slow in the early stages, and increasing gradually. It then increases exponentially reaching its peak point. After this estimated period, the infection rate starts to decrease, and declines gradually.

Based on the estimated graphs shown in Fig. 4, recovery rate is slow in the initial stages and then gradually increases. Figs. 5 and 6 show the comparison with the actual and predicted cases. The simulation is done with varying $R_0$ values and is compared with the actual confirmed cases of Kuwait. The SIR model is almost fitted with the actual confirmed data for the values of $R_0$ greater than 3. In the early stages of the estimation, the actual and predicted values are almost fitted, then it varies from the predicted estimation. The values of $R_0$ less than 3, did not provide a realistic estimation considering the studied time period. The actual confirmed data is far from the estimated values.

The evaluation of the predicted estimations is performed using Root Mean Square Error, $R^2$ score and the RSS measures as listed in Tables 1 and 2 for both infection and recovery. The simulation analysis showed that the SIR model with $R_0$ values ranging from 3.38–5.2, provides almost fitted prediction. Based on the current scenario in Kuwait, these $R_0$ values are well suited.

The Kuwaiti authorities' containment measures such as imposing a partial curfew and then total lockdown have had a high impact on daily cases. Fig. 7 illustrates the number of daily cases of coronavirus spread both based on actual confirmed data and predicted data with various $R_0$ values such as 5.2, 3.84, 3.83, 3.66, 3.77, 4.96, 3.38, and 3.45. It clearly explains the effect of the Kuwaiti government’s preventive measures. Due to this imposed lockdown, the number of regular confirmed cases has gradually decreased. This demonstrates the efficacy of the lockdown as well as other preventive measures taken by the Kuwaiti government to reduce the number of cases. The preventive measures did contribute to control the increasing rate of infection in Kuwait.

The SIR model for estimating the COVID-19 spread in Kuwait used the confirmed cases from 24th of February...
Fig. 4  The forecasted cumulative Recovery highlighting the rise, peak and predicted decline of COVID-19 is plotted against time (Day) with various $R_0$ values.

Fig. 5  Comparison of estimated cumulative infection with the actual confirmed cases.
2020 to 28th of May 2020. Based on this, the total number of confirmed cases was 24,112; 15,229 cases are under treatment, 8,698 cases are recovered and 185 are deceased from COVID-19. In our study, both recovered and deceased cases are in the recovered compartment. Based on this simulation, the peak value and corresponding date and daily cases are listed in Table 3 for each $R_0$ values such as 5.2, 3.84, 3.83, 3.67, 3.77, 4.96, 3.38, and 3.45.

Table 1 Evaluation of the simulated SIR model of epidemic growth based on various $R_0$ values using Root Mean Square Error, $R^2$ score and the RSS measures: Estimation of Cumulative Infection.

| $\beta$ | $\gamma$ | $R_0$ | $R^2$ | RMSE    | RSS      |
|---------|----------|-------|-------|---------|----------|
| 0.117   | 0.0225   | 5.2   | 0.896862828 | 1467.845432 | 204684170.1 |
| 0.128   | 0.0333   | 3.84  | 0.889544204 | 1519.032168 | 219208579.2  |
| 0.129   | 0.0337   | 3.83  | 0.858856522 | 1717.129796 | 280110800.0  |
| 0.13    | 0.0355   | 3.66  | 0.897642743 | 1462.285037 | 203136365.2  |
| 0.132   | 0.035    | 3.77  | 0.70518615  | 2481.685011 | 585082247.0  |
| 0.135   | 0.04     | 3.38  | 0.876071715 | 1609.007012 | 245948383.5  |
| 0.138   | 0.04     | 3.45  | 0.558134651 | 3038.209302 | 876917997.7  |

Table 2 Evaluation of the simulated SIR model of epidemic growth based on various $R_0$ values using Root Mean Square Error, $R^2$ score and the RSS measures: Estimation of Cumulative Recovery.

| $\beta$ | $\gamma$ | $R_0$ | $R^2$ | RMSE    | RSS      |
|---------|----------|-------|-------|---------|----------|
| 0.117   | 0.0225   | 5.2   | 0.728863441 | 1102.026345 | 115373896.1 |
| 0.128   | 0.0333   | 3.84  | 0.941739843 | 510.8386664 | 24790833.59  |
| 0.129   | 0.0337   | 3.83  | 0.953076044 | 458.4534524 | 19967058.97  |
| 0.13    | 0.0355   | 3.66  | 0.953874867 | 454.5349343 | 19627143.98  |
| 0.132   | 0.035    | 3.77  | 0.944040836 | 500.6492242 | 23811716.34  |
| 0.119   | 0.024    | 4.96  | 0.79247935  | 964.1144602 | 88304085.78  |
| 0.135   | 0.04     | 3.38  | 0.950854306 | 469.1812757 | 20912451.6   |
| 0.138   | 0.04     | 3.45  | 0.795561593 | 956.9278202 | 86992531.04  |

Figs. 8–13 illustrate the comparison of predicted and actual cases of both cumulative infection and recovery in Kuwait for various values of $R_0$ such as 2.0, 2.1, 2.2, 2.4, 2.5, 2.7, 2.8, 2.9, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 4.3, 4.5, and 4.7. The figures clearly show that forecasting COVID-19 in Kuwait with values of $R_0$ ranging from 3.2 to 4.5 are an acceptable fit, whereas $R_0$ values from 2.0 to 3.1 are not fitted with the actual numbers of confirmed cases.
Fig. 7  Comparison of confirmed daily infected cases with the estimated cases with various $R_0$ values.

Table 3  Predicted Peak values of Cumulative infection, Date and corresponding daily cases of that peak day of each $R_0$ fit.

| $\beta$ | $\gamma$ | $R_0$ | Peak Value of Cumulative infection | Peak Date   | No. of cases in Peak Day |
|---------|---------|--------|-----------------------------------|-------------|--------------------------|
| 0.117   | 0.0225  | 5.2    | 2374787.102                      | 11-Aug-2020 | 630.6464222               |
| 0.128   | 0.0333  | 3.84   | 1890403.899                      | 6-Aug-2020  | 1117.877658               |
| 0.129   | 0.0337  | 3.83   | 1883554.081                      | 5-Aug-2020  | 924.802714                |
| 0.13    | 0.0355  | 3.66   | 1808058.483                      | 5-Aug-2020  | 2878.05349                |
| 0.132   | 0.035   | 3.77   | 1859013.601                      | 2-Aug-2020  | 1002.270833               |
| 0.119   | 0.024   | 4.96   | 2302212.332                      | 9-Aug-2020  | 2245.737991               |
| 0.135   | 0.04    | 3.38   | 1668007.797                      | 3-Aug-2020  | 2090.995512               |
| 0.138   | 0.04    | 3.45   | 1706922.662                      | 30-Jul-2020 | 907.7758777               |

Fig. 8  Comparison of cumulative confirmed cases of infection with the estimated values for various $R_0$ values.
Fig. 9  Comparison of cumulative confirmed cases of recovery with the estimated values for various $R_0$ values.

Fig. 10  Comparison of cumulative confirmed cases of infection with the estimated values for various $R_0$ values.

Fig. 11  Comparison of cumulative confirmed cases of recovery with the estimated values for various $R_0$ values.
6. Conclusion

Analysis and forecasting of an epidemic growth are crucial for health authorities in order to make necessary preparations and take precautions in the fight against any disease. Based on the analysis, the governments can issue proper preventive measures for controlling the disease. Mathematical models play an important role in the prediction and analysis of outbreaks such as COVID-19, Ebola, and HIV. In this research, an SIR model was used for the analysis and forecast of COVID-19 outbreak in Kuwait. The analysis is performed based on collected data of confirmed cases from the 24th of February 2020 to the 28th of May 2020. The study is performed for various values of the basic reproduction number $R_0$. Based on the study, the peak values and expected dates of COVID-19 spread were estimated. COVID-19 outbreak reaches its peak value between the second half of July and first half of August. The SIR model is almost fitted with the actual confirmed cases for the values of $R_0$ ranging from 3 to 4. Simulation of the model is performed for different values of $R_0$ ranging from 2 to 5.2 and the results showed that the values of $R_0$ from 2 to 3 were not compatible with the actual confirmed cases. The forecasted estimations of confirmed and recovered cases were compared with the actual data and the evaluation of the model is performed using various measures such as $R^2$, RMSE and RSS. This research also considered the impact of preventive measures such as lockdown which proved to be effective in controlling the disease. During the peak time, the total number of cumulative cases may increase exponentially based on this estimation, unless strict social distancing, as well as other preventive measures, are followed. The study recommends the need to extend the preventive measures to control this epidemic growth. The estimated infection and recovery numbers reflected in this study are valid only for the time period under consideration. In the future as the period under consideration may be longer, better estimates may become more possible. It should be noted that the data used does not capture external factors that may affect the resulting numbers. One factor, for example, is the policy to increase the number of COVID-19 tests conducted on random members of the population. At this time, the nature of this pandemic, as well as the responsive measures taken by the government to contain it, are continually being transformed.
Appendix A. The Tables 4 and 5 show the evaluation of simulated SIR model of epidemic growth based on various $R_0$ values plotted in Figs. 8–13 for both infection and recovery. Also the peak values and dates are listed in Table 6.

### Table 4 Evaluation of the simulated SIR model of epidemic growth based on various $R_0$ values using Root Mean Square Error, $R^2$ score and the RSS measures: Estimation of Cumulative Infection.

| $\beta$ | $\gamma$ | $R_0$ | $R^2$ | RMSE   | RSS      |
|--------|---------|-------|-------|--------|----------|
| 0.12   | 0.06    | 2     | -0.273729594 | 5158.356296 | 2527820769 |
| 0.1197 | 0.057   | 2.1   | -0.229710267  | 5068.437739  | 2440460806  |
| 0.077  | 0.035   | 2.2   | -0.416879871  | 5440.505226  | 2811914225  |
| 0.144  | 0.06    | 2.4   | 0.566613361   | 3008.918854  | 860091303.8 |
| 0.0875 | 0.035   | 2.5   | -0.35790616   | 5325.852297  | 2694646756  |
| 0.1134 | 0.042   | 2.7   | -0.01356361   | 4601.487122  | 201149954   |
| 0.1252 | 0.044   | 2.8   | 0.309215836   | 3798.777903  | 1370917788  |
| 0.1305 | 0.045   | 2.9   | 0.65010272    | 2700.481782  | 692971716.1 |
| 0.129  | 0.043   | 3     | 0.678509549   | 2591.53073   | 638024148.5 |
| 0.123  | 0.041   | 3     | 0.456145649   | 3370.652805  | 1079323532  |
| 0.1147 | 0.037   | 3.1   | 0.296778481   | 3992.988136  | 1514675655  |
| 0.128  | 0.04    | 3.2   | 0.783107017   | 2128.606907  | 430441899.5 |
| 0.1221 | 0.037   | 3.3   | 0.628913888   | 2784.244905  | 736441870.5 |
| 0.1224 | 0.036   | 3.4   | 0.70046752    | 2501.465783  | 594444938.3 |
| 0.1225 | 0.035   | 3.5   | 0.758304814   | 2247.018857  | 479669055.5 |
| 0.1333 | 0.031   | 4.3   | -0.877937395  | 6263.445049  | 372920669.6 |
| 0.126  | 0.028   | 4.5   | 0.554616352   | 3050.281008  | 883900351.4 |
| 0.1222 | 0.026   | 4.7   | 0.78993114    | 2094.853014  | 416898869.1 |

### Table 5 Evaluation of the simulated SIR model of epidemic growth based on various $R_0$ values using Root Mean Square Error, $R^2$ score and the RSS measures: Estimation of Cumulative Recovery.

| $\beta$ | $\gamma$ | $R_0$ | $R^2$ | RMSE   | RSS      |
|--------|---------|-------|-------|--------|----------|
| 0.12   | 0.06    | 2     | -0.027186437  | 2144.976681  | 437087871.3 |
| 0.1197 | 0.057   | 2.1   | 0.019451727   | 2095.719588  | 417242226.6 |
| 0.077  | 0.035   | 2.2   | -0.331360353  | 2442.00049   | 566519807.3 |
| 0.144  | 0.06    | 2.4   | 0.890644985   | 699.8698566  | 46532692.53 |
| 0.0875 | 0.035   | 2.5   | -0.263496828  | 2378.94836   | 537642553.5 |
| 0.1134 | 0.042   | 2.7   | 0.138889169   | 1963.936596  | 366414960.5 |
| 0.1232 | 0.044   | 2.8   | 0.470696779   | 1539.749974  | 22522848.33 |
| 0.1305 | 0.045   | 2.9   | 0.779554664   | 993.6842217  | 93803791.59 |
| 0.129  | 0.043   | 3     | 0.772138873   | 1010.259766  | 969593554.4 |
| 0.123  | 0.041   | 3     | 0.552275377   | 1416.13099   | 19051563.1 |
| 0.1147 | 0.037   | 3.1   | 0.298443104   | 1772.676086  | 298526148.3 |
| 0.128  | 0.04    | 3.2   | 0.81303247    | 915.1262903  | 79558332.08 |
| 0.1221 | 0.037   | 3.3   | 0.62796326    | 1290.894802  | 158308892.1 |
| 0.1224 | 0.036   | 3.4   | 0.670597801   | 1214.678051  | 140167062.8 |
| 0.1225 | 0.035   | 3.5   | 0.702955317   | 1153.476707  | 126398308.7 |
| 0.1333 | 0.031   | 4.3   | 0.711066884   | 1137.618374  | 122946678.6 |
| 0.126  | 0.028   | 4.5   | 0.948414253   | 480.6874671  | 21950741.9 |
| 0.1222 | 0.026   | 4.7   | 0.883319637   | 722.930958   | 49649771.16 |
### Table 6 Predicted Peak values of Cumulative infection, Date and corresponding daily cases of that peak day for various $R_0$ fit.

| $\beta$ | $\gamma$ | $R_0$ | Peak Value of Cumulative Infection | Peak Date | No. of cases in Peak Day |
|---------|---------|-------|-----------------------------------|-----------|-------------------------|
| 0.12    | 0.06    | 2     | 743340.762                        | 14-Oct-2020 | 400.283693             |
| 0.1197  | 0.057   | 2.1   | 826263.102                        | 6-Oct-2020  | 831.2203025            |
| 0.077   | 0.035   | 2.2   | 607444.787                        | 19-Dec-2020 | 1290.52129             |
| 0.144   | 0.06    | 2.4   | 1062892.726                       | 16-Aug-2020 | 173.694162             |
| 0.0875  | 0.035   | 2.5   | 1127632.261                       | 26-Nov-2020 | 385.8281386            |
| 0.1134  | 0.042   | 2.7   | 1268465.323                       | 18-Sep-2020 | 363.265178             |
| 0.1232  | 0.044   | 2.8   | 1335055.104                       | 30-Aug-2020 | 728.011893             |
| 0.1305  | 0.045   | 2.9   | 1398912.189                       | 17-Aug-2020 | 1815.390661            |
| 0.129   | 0.043   | 3     | 1459054.524                       | 17-Aug-2020 | 1227.613578            |
| 0.123   | 0.041   | 3     | 1457859.973                       | 25-Aug-2020 | 1694.581696            |
| 0.1147  | 0.037   | 3.1   | 1514387.873                       | 5-Sep-2020  | 653.4024175            |
| 0.128   | 0.04    | 3.2   | 1572836.563                       | 14-Aug-2020 | 2631.691249            |
| 0.1221  | 0.037   | 3.3   | 1625902.512                       | 21-Aug-2020 | 656.0364689            |
| 0.1224  | 0.036   | 3.4   | 1678119.972                       | 19-Aug-2020 | 634.9484009            |
| 0.1225  | 0.035   | 3.5   | 1728573.212                       | 17-Aug-2020 | 1771.605894            |
| 0.1333  | 0.031   | 4.3   | 2078497.393                       | 27-Jul-2020 | 907.6872101            |
| 0.126   | 0.028   | 4.5   | 2150587.236                       | 3-Aug-2020  | 1355.187671            |
| 0.1222  | 0.026   | 4.7   | 2218849.768                       | 6-Aug-2020  | 3141.318212            |

**References**

[1] W. Roda, M. Varughese, D. Han, M. Li, Why is it difficult to accurately predict the covid-19 epidemic?, Infect. Disease Model. 5 (2020).

[2] F.A. Binti Hamzah, C. Hau, H. Nazri, D. Ligot, G. Lee, M. Batista, Coronavirus disease (covid-19) situation report, Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019, Accessed on: 28-05-2020 (05 2020).

[3] WHO, Coronavirus disease (covid-19) situation report, Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019, Accessed on: 28-05-2020 (05 2020).

[4] S. Unhale, Q. Bilal, S. Sanap, S. Thakre, S. Wadatkar, R. Shaib, U. Zaidon, A. Abdullah, M. Chung, C. Ong, P. Chew, Coronavirus: What world-wide covid-19 outbreak data analysis and prediction, Bull. World Health Organ. (2020), https://doi.org/10.12471/BLT.20.255695.

[5] M. Batista, Estimation of the final size of the coronavirus epidemic by the logistic model (Update 3), medRxiv (Accessed on: 28-05-2020 (2020).

[6] P. Bhandari, analysis of prediction models in spread of ebola virus disease, Ph.D. thesis, DEAKIN University Australia (05 2019). doi:10.13140/RG.2.1.14833.436316.

[7] A. Arsath Abbasali, S. Satheesh, On predicting the novel covid-19 human infections by using infectious disease modelling method in the indian state of tamil nadu during 2020, medrxiv (04 2020). doi:10.1101/2020.04.05.20054591.

[8] M. Mandal, S. Mandal, Covid-19 pandemic scenario in india compared to china and rest of the world: a data driven and model analysis, medRxiv (2020).
[28] A. Almeshal, A. Almazrouee, M. Alenizi, S. Alhajeri, Forecasting the spread of covid-19 in Kuwait using compartmental and logistic regression models, Appl. Sci. 10 (2020) 3402.

[29] F. Syed, S. Sibgatullah, Estimation of the final size of the covid-19 epidemic in Pakistan, medRxiv (04 2020). doi:10.1101/2020.04.01.20050369.

[30] M. Batista, Estimation of the final size of the coronavirus epidemic by the sir model, medRxiv (02 2020).

[31] N. Sene, Sir epidemic model with mittag-leffler fractional derivative, Chaos, Solitons & Fractals 137 (2020) 109833.

[32] A. Atangana, Modelling the spread of COVID-19 with new fractal-fractional operators: Can the lockdown save mankind before vaccination?, Chaos, Solitons & Fractals 136 (2020) 109860.

[33] F. Brauer, Compartmental models in epidemiology, in: Mathematical epidemiology, Springer, 2008, pp. 19–79.

[34] M. Lutz, Learning Python: Powerful Object-Oriented Programming, 4th ed., O’Reilly Media, 2009.

[35] T. Burr, G. Chowell, Observation and model error effects on parameter estimates in susceptible-infected-recovered epidemic model, Far East J. Theoret. Stat. 19 (2006).

[36] P. Trapman, F. Ball, J.-s. Dhersin, V.C. Tran, J. Wallinga, T. Britton, Inferring r0 in emerging epidemics: the effect of common population structure is small, J. Royal Soc. Interface 13 (2016).

[37] Y.-C. Chen, P.-E. Lu, C.-S. Chang, T.-H. Liu, A time-dependent sir model for covid-19 with undetectable infected persons, arXiv (2020-02-28) (02 2020).