JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Coronavirus Pandemic

Analysis and modeling of COVID-19 epidemic dynamics in Saudi Arabia using SIR-PSO and machine learning approaches

Rafat Zrieq[1,2], Sahbi Boubaker[3,4], Souad Kamel[3], Mohamed Alzain[1], Fahad D Algahtani[1,2]

[1] *Department of Public Health, College of Public Health and Health Informatics, University of Ha'il, Ha'il, Saudi Arabia*
[2] *Molecular Diagnostic and Personalized Therapeutics Unit, University of Ha'il, Ha'il, Saudi Arabia*
[3] *Department of Computer and Networks Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia*
[4] *Research Unit on Study of Systems and Renewable Energy, National College of Engineering of Monastir, Tunisia*

Abstract
Introduction: COVID-19 has become a global concern because it has extensive damage to health, social and economic systems worldwide. Consequently, there is an urgent need to develop tools to understand, analyze, monitor and control further outbreaks of the disease.
Methodology: The Susceptible Infected Recovered-Particle SwarmOptimization model and the feed-forward artificial neural network model were separately developed to model COVID-19 dynamics based on daily time-series data reported by the Saudi authorities from March 2, 2020 to February 21, 2021. The collected data were divided into training and validation datasets. The effectiveness of the investigated models was evaluated by using various performance metrics. The Susceptible-Infected-Recovered-Particle-Swarm-Optimization model was found to well predict the cumulative infected and recovered cases and to optimally tune the contact rate and the characteristic duration of the illness. The feed-forward artificial neural network model was found to be efficient in modeling daily new and cumulative infections, recoveries and deaths.
Results: The forecasts provided by the investigated models had high coefficient of determination values of more than 0.97 and low mean absolute percentage errors (around 7% on average).
Conclusions: Both the Susceptible-Infected-Recovered-Particle-Swarm-Optimization and feed-forward artificial neural network models were efficient in modeling COVID-19 dynamics in Saudi Arabia. The results produced by the models can help the Saudi health authorities to analyze the virus dynamics and prepare efficient measures to control any future occurrence of the epidemic.

Key words: COVID-19 epidemic; dynamics modeling; prediction; SIR-PSO model; FF-ANN; performance metrics.

J Infect Dev Ctries 2022; 16(1):90-100. doi:10.3855/jidc.15004

(Received 05 March 2021 – Accepted 13 July 2021)

Introduction
The highly infectious disease, SARS-CoV-2, more commonly known as COVID-19, has spread worldwide causing a high number of infections and deaths and affecting the social, health and economic systems of almost every country on the planet [1-5]. In Saudi Arabia, the first positive case was reported on March 2, 2020. One year later, as of March 2, 2021, official statistics reported 377,700 total cases, 368,640 recovered cases, 2,560 active cases, 492 critical cases and 6,500 deaths in the country [6]. Local authorities implemented several measures that included a total lockdown involving the closure of universities and schools, suspension of Umrah, an airport lockdown [7,8]; and social distancing [9], but the disease continued to spread, touching almost all of Saudi territory.

Saudi health authorities started a vaccination campaign on December 17, 2020 and, at the beginning of 2021, the first wave of the disease seemed to be approaching its end with the arrival of many vaccines. However, since then, more waves have arrived.

During this ongoing pandemic, the most populated and active regions such as Riyadh, Jeddah, Makkah and Al-Madinah Al-Munawara have recorded the highest numbers of infected cases. Based on infected cases, Saudi Arabia is now ranked at number 41 among COVID-19-infected countries [5]. However, earlier, Saudi Arabia was classified for a long period at number 15. Owing to the severity of the disease and the randomness of its variable trajectory, the accurate modeling of its dynamics is a challenging issue. Yet, this challenge needs to be overcome because efficient modeling can contribute to the analysis of the dynamics and the control of any future wave of COVID-19 and of any other epidemic related to the respiratory system.

The forecasting of epidemics is mainly based on mathematical models and recorded data [10]. Modeling is a key tool that can assist health authorities to understand the transmission modes of an epidemic and to estimate the number of new cases, critical cases, recoveries and deaths, and perhaps a probable end date. Accurate models help in mitigating disease spread and in implementing suitable control actions [11]. In the case of the COVID-19 pandemic, limited information about its mode of transmission is available [12]. For instance, it has been reported that virus transmission can occur from an individual who is asymptomatic [13].

Various mathematical approaches for predicting COVID-19 dynamics have been developed, which have achieved varying levels of accuracy. These models range from the simple to the complicated. Examples of such models include the classical SIR (Susceptible, Infected and Recovered) model [14-15] and its variants (SI, SIS). The SIR model is composed of a set of continuous-time nonlinear differential equations [16]. The model considers several parameters, which requires the processing of a large amount of data. Extended versions of the SIR model such as SIER (which adds Exposed cases to the SIR model) [17,18], SIEQR (which adds Quarantined cases to the SIER model) [19,20] and SIRD (which adds Deaths to the SIR model) [21], have been investigated for different purposes and to find effective methodologies for modeling COVID-19 dynamics in different countries. However, the main difficulty faced by practitioners involved in the development of SIR models is the fact that the models are dependent on nonlinear continuous-time equations, and the solving of such equations by using exact explicit formula is impossible even when using assumptions to simplify and relax the solutions.

Practitioners have always resorted to using computer codes to get approximate solutions. Numerical algorithms are mainly based on data available in discrete-time frequency (daily in the case of COVID-19) [14,22]. As is well known, in numerical computation, smaller sampling periods may provide more accurate solutions. However, the only available choice for COVID-19 is the frequency of 1 day because the disease data are compiled and communicated daily. Moreover, as reported by [23], simpler models such as the SIR model may perform better than more complicated models such as the SIER model. Time-series econometric models have also been examined to assess their suitability for analyzing COVID-19 dynamics. Examples of such work include [24-27], which separately used the autoregressive integrated moving average (ARIMA) and reported that although the results of this approach are quite accurate, more real-time data is required in order for the approach to provide highly precise forecasts. Generalized growth models have been also assessed for estimating the growth rate and the growth scaling parameter [10,28] of this disease.

Another trend in the modeling of COVID-19 dynamics has focused on the use of machine learning (ML) and population-based algorithms. For instance, [29] developed models based on genetic evolutionary programming to calibrate parametric models for various regions in India. Meanwhile, [30] examined the utilization of artificial intelligence (AI) tools for forecasting multi-step-ahead COVID-19 numbers in China in which a modified stacked auto-encoder was used. The authors reported that the accuracy of the results was high in spatial-temporal levels. However, details of the model implementation were missing. In another China-based study, an adaptive neuro-fuzzy inference system (ANFIS) was used to forecast 10-days-ahead confirmed infected cases in Wuhan [31]. The ANFIS was also combined with the flower pollination algorithm to prevent the system from getting trapped in local solutions [32]. Elsewhere, the Long-Short-Term Memory (LSTM) network, which is a popular deep-learning algorithm, has been used for predicting COVID-19 cases in several Canadian provinces. However, although the generated forecasts are quite accurate, the authors noted that the LSTM may need a high amount of data to improve the level of accuracy.

A few works including [33] and [34] have used the particle swarm optimization (PSO) technique to determine the optimal parameters of, respectively, the SIR and the SIER models. The results presented in [34] have the drawback of having been generated by a model with given parameters. However, the results presented in [33] are based on real COVID-19 data from Hubei, China under different scenarios, and showed that the use of PSO was efficient and accurate.

Based on the previous literature concerning the importance of implementing mathematical models for predicting COVID-19 dynamics in a short-term horizon, the aim of this study is to develop models for the case study of Saudi Arabia as a representative of the Arab Gulf countries which were, for a long period of time, among the COVID-19 pandemic outbreak epicenters alongside some countries in South America and India. In Saudi Arabia, the disease has exhibited a fluctuating transmission rate, which has made its prevalence trend difficult to predict. In this study, the dynamics of the COVID-19 outbreak in Saudi Arabia are modeled by using a novel SIR-PSO approach and by using a feed-forward artificial neural network (FF-ANN) method.

The use of PSO is justified by the fact that it is a global search tool capable of operating in random and variable environments such as in the case of COVID-19. Moreover, it is useful in solving the SIR model identification problem, which is known to be nonlinear, non-convex and difficult to calibrate with discrete data because it is originally a continuous-time model. As the variables that affect the COVID-19 outbreak are difficult to quantify (disease transmission is generally caused by social behavior), the use of a FF-ANN model that expresses the next day's newly infected cases as a function of the previous days' infections would seem to offer a way to arrive at a new efficient solution. The effectiveness of the FF-ANN in using historical records as inputs is investigated owing to the high ability of neural networks to describe complex phenomena such as those related to COVID-19 dynamics. The proposed FF-ANN method is new because, to the best of the authors' knowledge, this study is the first to use several previous days' numbers of infected cases to predict the upcoming next day's number of infected cases by using a neural network. Intuitively, this idea is well founded because the infection is transmitted to new individuals from previously infected individuals. The proposed model is applied to a dataset covering the period from March 2, 2020 to February 21, 2021 as input. A comparative study of the developed models is conducted by using the following statistical performance metrics: the mean absolute percentage error (MAPE), the coefficient of determination (R^2) and the root mean squared error (RMSE).

The remainder of this paper is organized as follows: First, the methodology section presents the details of the SIR-PSO and FF-ANN models developed in this study. In the next section, the results are presented and discussed. The final section concludes the paper and includes some recommendations for future work.

Methodology

In this section, the SIR-PSO and the FF-ANN models, which are categorized, respectively, as population-based and ML approaches, are designed and investigated for modeling COVID-19 dynamics in Saudi Arabia. The SIR-PSO model employs a swarm intelligence optimization technique (PSO) to optimally determine the SIR model parameters (the contact rate and the characteristic duration of the disease) by minimizing the error between the recorded COVID-19 numbers and the same numbers generated by the model. The FF-ANN model is a ML structure that learns from knowledge. It is considered that the FF-ANN has the capacity to map data for complex phenomena such as COVID-19. The mathematical flow of the two proposed models is mainly based on dividing the data into two subsets for training (80%) and model validation (20%). The efficiency of the models is measured by using statistical performance metrics (MAPE, R^2 and RMSE).

SIR-PSO model

In this study, the SIR model is adopted because it is simpler than other models such as the SIER [23]. The SIR model is composed of a set of differential equations (in continuous-time) that relate to subsets of a population consisting of Susceptible (S), Infected (I) and Recovered (R) cases [16,33]. In the SIR model, the dynamics of COVID-19 disease are described by Eqs. 1–3 as follows:

$$\frac{dS(t)}{dt} = -KS(t)I(t) \quad (1)$$
$$\frac{dI(t)}{dt} = KS(t)I(t) - \frac{1}{\beta}I(t) \quad (2)$$
$$\frac{dR(t)}{dt} = \frac{1}{\beta}I(t) \quad (3)$$

where K is the contact rate expressing the probability of being infected and β is the characteristic duration of the disease.

As COVID-19 pandemic data are available at a discrete-time level and explicit solutions are quite difficult, the finding of SIR model solutions should be based on an efficient numerical algorithm operating on discrete-time [34]. For this purpose, the Euler first-order method is used to transform the SIR continuous-time model (Eqs. 1–3) into a discrete-time form (Eqs. 4–6) (Note that for this purpose, $dS(t) = s(t + dt) - S(t)$ and $d(t) = 1day$. The choice of 1 day as the sampling period is dictated by the availability of official data, which are reported for COVID-19 only a daily basis.):

$$S(t+1) = S(t) - KS(t)I(t) \quad (4)$$
$$I(t+1) = I(t) + KS(t)I(t) - \frac{1}{\beta}I(t) \quad (5)$$
$$R(t+1) = R(t) + \frac{1}{\beta}I(t) \quad (6)$$

where subscript t indicates the day number and (t+1) indicates the one-day-ahead variables.

Particle swarm optimization, a population-based method that was introduced by Kennedy and Eberhart

[35] in 1995 simulates the social behavior of animals. Since then, PSO has been used efficiently in various fields including water demand prediction [36], electric peak-load forecasting [37], energy market forecasting [38] and networked epidemic control [39], to cite just a few. In PSO, each member (called a particle) is assigned a position vector $X = [K\beta]$ including the parameters of the SIR model to be optimized. The principle of SIR-PSO is that each particle adjusts its position inside the search space in order to minimize the quadratic error between the actual and predicted values of the cumulative S, I and R components. Thus, the objective function is defined as follows (Eq. 7):

$$J(X) = \sum_{t_0}^{t_f} \left(S(t) - \hat{S}(t)\right)^2 + \left(I(t) - \hat{I}(t)\right)^2 + \left(R(t) - \hat{R}(t)\right)^2 \quad (7)$$

where $t_0 t_f$ are the first day and the last day of the collected COVID-19 data, respectively. $\hat{S}(t), \hat{I}(t)$ and $\hat{R}(t)$ are the estimates of S, I and R at the t^{th} day, respectively.

During the search process, each particle recalculates its current position by using Eqs. 8–10 as follows:

$$V_{k+1}^{i} = w_k V_k^i + C_1 r_1 \left(P^i - X^i\right) + C_2 r_2 \left(G^i - X^i\right) \quad (8)$$

$$X_{k+1}^{i} = X_k^i + V_{k+1}^{i} \quad (9)$$

$$W_k = W_{max} - \frac{W_{max} - W_{min}}{k_{max}} . k \quad (10)$$

where P^i and G^i are two vectors that have the same dimension as X^i and represent, respectively, the best position so far visited by the i^{th} particle and the global best position of the population. V^i is a velocity operator that represents the increment added to the current particle's position.

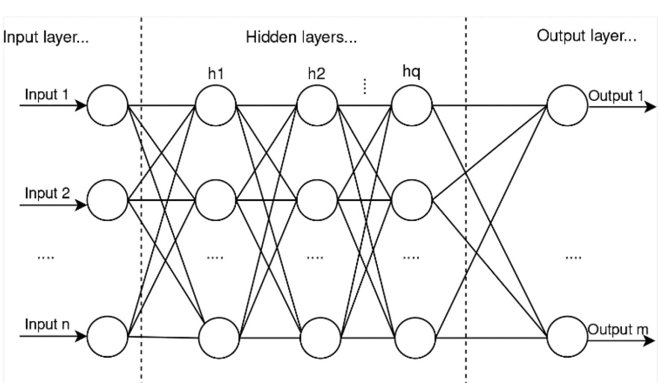
Let us consider k as the iteration index. The inertia weight function linearly decreases from an initial value w_{max} to a final value w_{min} that is reached at the end of the optimization process. C_1 and C_2 are cognitive and social factors, respectively. r_1 and r_2 are random numbers within the interval [0-1] that are used to diversify the solutions related to the swarm [35-36]. The SIR-PSO model operates as follows:

- *Step 1*: Initialize randomly the positions, velocities and personal best positions inside the search-space limits and calculate the objective function to be minimized according to Eq. 7.
- *Step 2*: Determine the personal best position and the global best position.
- *Step 3*: Update the particles' positions by using Eqs. 8–9. Confine the particles inside the search space if needed.
- *Step 4*: Calculate the objective function (Eq. 7) for all the population and determine the new personal best position for each particle and the index of the particle with the global best position.
- *Step 5*: Repeat steps 3 and 4 until a preset maximum number of iterations (k_{max}) (i.e., the stopping criterion) is reached. The position of the global best particle is taken as the solution of the SIR-PSO optimization problem.

FF-ANN model

Machine learning techniques such as artificial neural networks (ANNs) are used as modeling tools for complex phenomena in the absence of efficient parametric models. Artificial neural networks are increasingly being used in forecasting in several domains including solar energy [40], municipal water consumption [41] and the spread of COVID-19 [42]. Feed-forward artificial neural networks are currently the most widely used form of ANN (see Figure 1 for the basic FF-ANN structure). Inputs (called patterns in ANN terminology) are introduced to the input layer which, after multiplying them by weights and adding biases, transfers them to an activation function. This process is repeated until the output layer, which generates the calculated final network output. The quadratic error between the FF-ANN outputs and the predetermined outputs (called targets) is then minimized by a training algorithm that optimizes the weights and biases. In COVID-19 modeling, the patterns are the historical records of the new/cumulative

Figure 1. Basic structure of FF-ANN.



infected, recovered and dead cases of the previous days. In such a case, the FF-ANN output is calculated as follows (Eq. 11):

$$y(t+1) = f\big(y(t), y(t-1), \dots, y(t-d)\big) \quad (11)$$

where e(t) is an error function usually assumed to be normal and white.

Based on Figure 1, the behavior of the FF-ANN is governed by the following set of equations (Eq. 12):

$$\begin{aligned} a^1 &= f^1(W^{1,1}X + b^1) \quad (12)\\ a^2 &= f^2(W^{2,1}a^1 + b^2)\\ a^j &= f^j\big(W^{j,j-1}a^{j-1} + b^j\big)\\ a^N &= f^N(W^{N,N-1}a^{N-1} + b^N) \end{aligned}$$

where:

- a^j, $W^{j,j-1}$, b^j, and $j = 1{:}N$ are the outputs, weights and biases of the j^{th} layer, respectively.
- a^1 is the output of the input layer and a^N is the output of the last layer.
- f^j is the transfer function of the j^{th} layer.

The number of hidden layers and the number of neurons in each of the hidden layers are practically determined by using the trial-and-error method [40], which is the most time-consuming- step of an ANN design.

The training of an ANN involves optimally tuning the weights and biases of different layers by minimizing the error function between the ANN output and the target: d, that varies between 1 and 6, where these values represent the probable incubation periods of COVID-19 disease [31].

$$\varepsilon^2(W_i, bi) = \sum_{t=1}^{N}\Big(y(t) - f\big(y(t-1), y(t-2), \dots, y(t-d)\big) + e(t)\Big)^2 \quad (13)$$

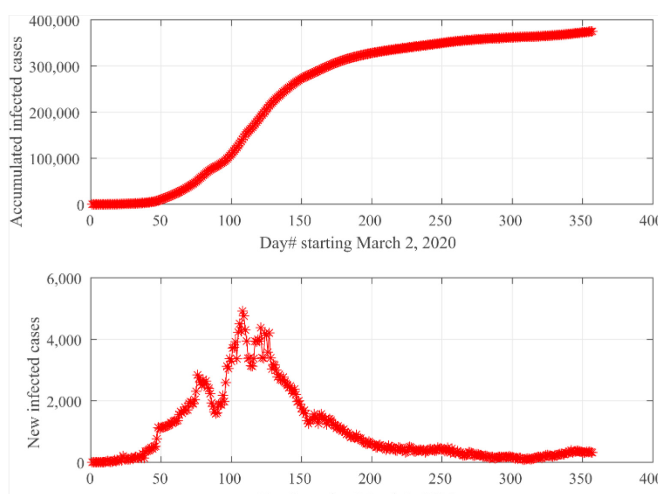
There are two main motivations for using the FF-ANN to model COVID-19 disease dynamics. First, the mechanism of infection transmission is very complex because it depends on several parameters such as the movements of infected persons, the possibility of transmission from asymptomatic infected individuals, the types of control measures in place, etc. It is known from the application of FF-ANNs in several domains that they are good universal approximators that can model complex phenomena such as COVID-19. Second, intuitively, any new infection results from previous infections. Therefore, this study uses the number of infections of previous days as inputs (patterns) of the FF-ANN and the number of infections in the next day as the output (target) for the proposed FF-ANN model. Thus, a relationship between patterns and targets through learning algorithms is established. The idea of varying the number of used inputs, which is one of the novelties of this study, allows the quantification of the contributions of the previous days reported infections to the next day's infections.

Data

The COVID-19 data for Saudi Arabia are available from the official statistics E-platform of the Saudi Ministry of Health [43], which reports on a daily basis the new confirmed cases, the cumulative cases, the new recovered cases, the cumulative recovered cases, new deaths and total deaths. This study is based on records starting on March 2, 2020 (the date of the registered first case) and ending on February 21, 2021. This dataset includes 357 records. In order to model the COVID-19 dynamics, the dataset is divided into two subsets: 285 records (80%) are used for model training (development) and the remaining 72 records (20%) are used for model validation (testing). Figure 2 depicts the plots of the infected (both new and cumulative) cases.

From Figure 2, it can be observed that the progression of newly registered cases had four notable peaks of different intensities. The most significant peak occurred around June 17, 2020 when the daily newly infected cases reached 4919. The increase and decrease phases are highly correlated with the control measures that were put in place by the Saudi authorities, which included a total lockdown at the beginning of the epidemic and the later application of a monetary penalty

Figure 2. Cumulative (top) and new (bottom) infected cases of COVID-19 in Saudi Arabia for the period from March 2, 2020 to February 21, 2021.



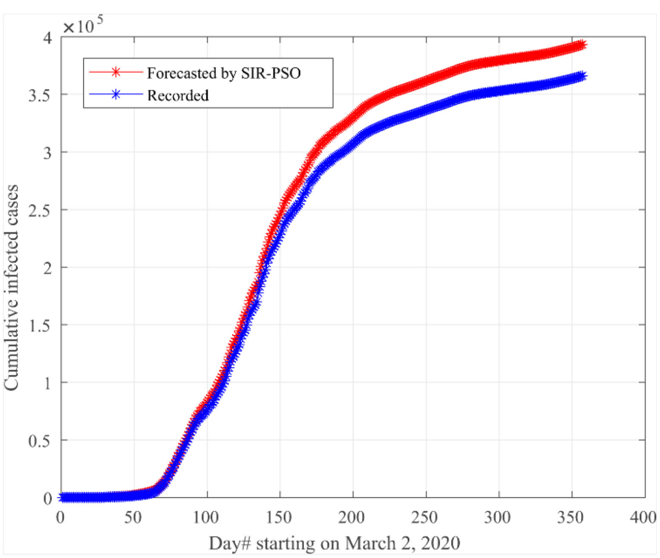
for those who did not wear a face mask or those who violated the social distancing rules. After a long period of sustained infection rates, on January 3, 2021, the daily newly infected cases finally fell to a low 82. Although Saudi Arabia flattened the curve of the first wave, the overall behavior fluctuated over the period covered by the study. The cumulative progression of the disease represents a "logistic" behavior that was also seen in many other countries such as China [20] and Italy (first wave) [27]. From January 15, 2021, Saudi Arabia saw an increase in the daily number of newly infected cases with a slightly fluctuating curve that continues until the end of the data used in this study. During that time, the country was facing the risk of a second wave although a vaccination campaign had started..

As regards the death rate due to COVID-19, this is defined by the Saudi Ministry of Health as the ratio between the cumulative number of deaths by the number of closed cases (including cases where the outcome was either recovered/discharged or dead). In Saudi Arabia, this ratio was found to be around 0.02 (2%) for the period under study. This value is relatively low when compared to the death rates of other countries. Globally, the death rate reached 3%, versus 97% of recovered cases [5].

Performance metrics

The accuracy of the two developed forecasting models was measured by comparing the predicted values to the actual (reported) values. Three performance indicators, MAPE, R^2 and RMSE, were used to assess and compare the efficiency of the models.

Figure 3. Cumulative infected cases for the period from March 2, 2020 to February 21, 2021.



These performance metrics are expressed, respectively, in Eqs. 14, 15 and 16 as follows [25,36]:

- Mean absolute percentage error ($MAPE$)

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\frac{|y(t)-\hat{y}(t)|}{\bar{y}} \quad (14)$$

- Coefficient of determination (R^2)

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{t=1}^{N}\left(y(t)-\hat{y}(t)\right)^2}{\frac{1}{N}\sum_{t=1}^{N}(y(t)-\bar{y})^2} \quad (15)$$

- Root mean square error ($RMSE$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y(t)-\hat{y}(t)\right)^2} \quad (16)$$

where $\hat{y}(t)$ and $y(t)$ represent, respectively, the predicted and the actual (reported) number of cases (infected, recovered and deaths) at day t . $\bar{y}$ is the average of the daily values under evaluation. Lower $MAPE$ and $RMSE$ and higher R^2 (value located between 0 and 1) are indicators of better data fitting [31], [42]. N is the size of the used dataset.

Results

In this section, the results of the SIR-PSO and the six versions of the FF-ANN are detailed. All models were based on time-series related to COVID-19 records in Saudi Arabia.

SIR-PSO

The SIR model is a simple model that gives the susceptible, infected and recovered cases through the use of a set of nonlinear continuous-time differential equations. As there is no efficient analytic method to solve such a problem [16] and because COVID-19 data are available on a discrete-time basis (daily), this study investigated the use of the PSO as a powerful technique to obtain a feasible and sub-optimal solution. In the model, the two parameters $K\beta$ are tuned optimally such that the quadratic error between the forecasted and reported number of cases is minimized (Eq. 7). The PSO design parameters are summarized as follows: w_{max} = 0.9, w_{min} = 0.4, number of particles = 50, C_1 = C_2 = 0.75, and k_{max} = 100 which is used as a stopping criterion. According to the Worldometer website that supplies COVID-19 data [6], the Saudi population is 34,799,212, and this number is therefore used as the susceptible initial value population, S(0)**.** The initial value of infected cases is considered as I(0) = 1 which corresponds to the first reported case on March 2, 2020. The initial number of recovered cases is

Table 1. Optimal parameters of the SIR-PSO model and performance metrics of the cumulative infected and recovered cases.

K (days^{-1})	β (day)	Cases estimation	MAPE (%)	R^2	RMSE (case)
2.8360×10^{-9}	13.6131	Infected	8.6411	0.9722	21715
		Recovered	6.9837	0.9840	19324

obviously R(0) = 0 as of the same day. The contact rate K upper limit is chosen deliberately as K_{max} = 1/35000000. The upper and lower limits of the characteristic duration of the disease, β, are chosen, respectively, as 1 day and 14 days which is considered by several studies as the incubation period of COVID-19 during which an infected individual can transfer the virus to another susceptible individual [27]. The model's two parameters are tuned optimally following the abovementioned Steps 1–5. The PSO algorithm is stochastic. Therefore, it has to be run several times. In this paper, only the results of the best run are reported. The performance metrics of the SIR-PSO model as well as the optimal values for the SIR model parameters are summarized in Table 1.

The plots of the actual and forecasted (by SIR-PSO) cumulative infected and recovered cases are shown in Figure 3 and Figure 4, respectively. It can be seen that the curves of the forecasted and recorded cases are almost similar. However, the SIR-PSO algorithm shows more success in predicting the recovered cases rather than predicting the infected cases. The predicted daily new cases (infected and recovered) were obtained by subtracting the cumulative cases of the previous day from the cumulative cases of the concerned day. Figure 5 illustrates the daily new cases for the period of study. Remarkably, good agreement can be observed for the new daily infected cases and the new daily recovered cases with greater accuracy achieved for the recovered cases. From both the plots and the performance indicators, it can be considered that the PSO is a powerful tool for solving the SIR model despite the nonlinearity and the irregularities such as nonconvexity and randomness that it possesses. However, despite the SIR-PSO model successfully fitting the Saudi Arabia COVID-19 data, the SIR-PSO model has the drawback of requiring many runs because the PSO technique is stochastic in nature.

FF-ANN

The other approach that was used to model the next day's cases (infected, recovered and dead) for both daily new cases and cumulative cases was the FF-ANN. The main motivations for using this approach are that ANNs are good approximators of complex phenomena that are random and variable and because COVID-19 dynamics are not yet well understood. The use of the FF-ANN is also intuitively justified because the infected cases of the previous days are known practically to infect new individuals on the next day. Prior to the main experiment, several FF-ANNs of various structures with various inputs were investigated. The performance metrics were calculated to assess the accuracy of each investigated FF-ANN structure. After several trials, the FF-ANN structure that was adopted for all the experiments was composed of a number of inputs varying from 1 to 6, and consisted of one input layer, two hidden layers and one output

Figure 4. Cumulative recovered cases for the period from March 2, 2020 to February 21, 2021.

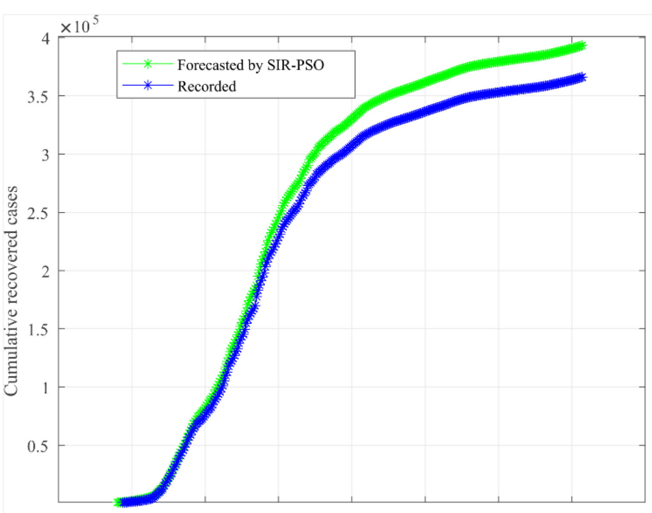
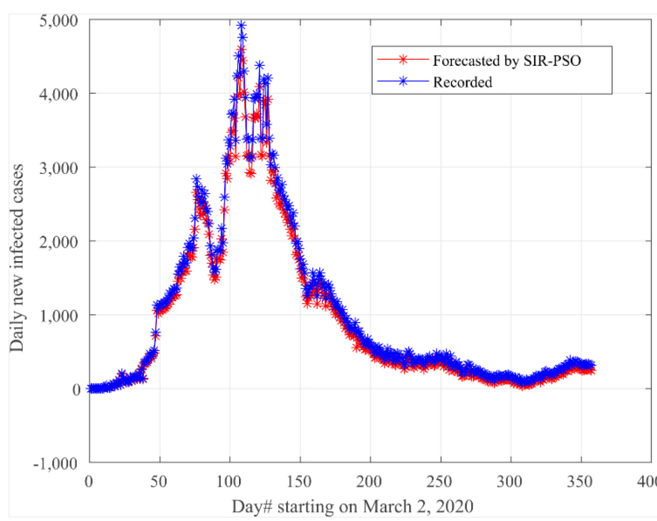


Figure 5. Daily newly infected cases for the period from March 2, 2020 to February 21, 2021.



layer. Linear transfer functions were adopted for all layers and five neurons in each layer were adopted as well. For training the model, the resilient-propagation training algorithm was used because of its fastness [40]. Each structure was used to model the new daily infected, recovered and death cases as well as their accumulated values. In total, 36 experiments were conducted. The results of the selected good runs are provided in Table 2. In the table, FF-ANN (d) means that the next day's cases are expressed as a function of the d previous days' cases.

Discussion

SIR-PSO

It is clear that the R^2 values for the infected and recovered cases were high at 0.9722 and 0.9840, respectively. As the R^2 values were close to 1, this indicates that there is a very good agreement between the actual numbers reported and the numbers predicted by the SIR-PSO model. Moreover, both the MAPE and RMSE values were quite low, which confirms that the model had a good fit to the data. The SIR-PSO model suggests that the COVID-19 initial transmission rate R_0 was around 1.3435. This value was calculated from the SIR model parameters as follows: $R_0 = K \times \beta \times S(0)$by using the investigated model parameters and the initial population of suspected cases S(0) = 34,799,212. Although the transmission rate value is greater than 1, which indicates that the virus is spreading, it can be considered as low when compared to the global overall value: several studies included in [5] reported transmission rates ranging from 1.5 to 4. (It is also worth noting that the transmission rate was 1.3 for the common flu and 2.0 for SARS).

FF-ANN

Table 2 also shows the results for the FF-ANNs. From the table, the following points arise in respect of the performance of the FF-ANN model:

- All the investigated FF-ANNs forecasted accumulated case numbers that were in good agreement with the actual reported numbers of accumulated cases. In particular, the R^2 values for the cumulative infected cases were more than 0.999 for all experiments. Moreover, the MAPE and RMSE values for all experiments were low, which proves that the FF-ANNs are efficient in forecasting cumulative infected cases.
- The accuracy of the FF-ANN forecasts increased with the increase in the number of inputs for the cumulative infected cases. This finding can be intuitively justified because the next day's cumulative infected cases are caused by the cumulative infections recorded for the previous days.
- The best R^2 (1.0000) for the forecast of accumulated infected cases was obtained by FF-ANN (5). This indicates that, in Saudi Arabia, the infections of the five previous days are the most likely to lead to an accurate prediction of the number of new infections on the next day.
- The results of the six FF-ANNs were not sufficiently meaningful in respect of the forecasted number of recovered cases because the values of the MAPE were high (overall around 28%). This may imply that the numbers of recovered cases on the previous days

Table 2. Performance metrics of FF-ANNs.

FF-ANN	Cases	MAPE (%)	R^2	RMSE (case)	FF-ANN	Cases	MAPE (%)	R^2	RMSE (case)
FF-ANN (1)	New I	8.0352	0.9803	162	FF-ANN (2)	New I	8.0283	0.9805	161.52
	Accumulated I	0.3882	0.9999	1,118		Accumulated I	0.4547	0.9999	1316
	New R	35.8561	0.7094	651		New R	31.4098	0.7466	608.15
	Accumulated R	0.4568	0.9999	1,215		Accumulated R	0.1721	1.0000	654.14
	New D	13.4056	0.9323	3.66		New D	11.6873	0.9520	3.0843
	Accumulated D	0.4081	1.0000	15.65		Accumulated D	0.1047	1.0000	4.8594
FF-ANN (3)	New I	8.0043	0.9805	161.4	FF-ANN (4)	New I	8.0263	0.9814	157.63
	Accumulated I	0.4612	0.9999	1341		Accumulated I	0.0365	1.0000	160.75
	New R	29.7187	0.7539	599.5		New R	29.2695	0.7562	596.99
	Accumulated R	1.1130	0.9996	2846		Accumulated R	0.1373	1.0000	599.75
	New D	11.7337	0.9515	3.095		New D	11.4861	0.9524	3.0649
	Accumulated D	0.8517	0.9998	32.56		Accumulated D	0.0658	1.0000	3.4543
FF-ANN (5)	New I	8.0305	0.9813	158.0	FF-ANN (6)	New I	6.0292	0.9794	154.948
	Accumulated I	0.0694	1.0000	300.5		Accumulated I	0.4182	0.9999	1123
	New R	27.567	0.7721	577.4		New R	22.7652	0.7582	570.19
	Accumulated R	0.2704	0.9999	1061		Accumulated R	1.1669	0.9995	2893
	New D	11.3622	0.9526	3.055		New D	9.2696	0.9554	2.9069
	Accumulated D	1.2289	0.9996	47.22		Accumulated D	0.8613	0.9997	36.612

I: Infected; R: Recovered; D: Dead.

probably has no effect on the recoveries of the next day.

- All the tested FF-ANNs were efficient in predicting the cumulative number as well as the daily new number of deaths because the function of deaths of the previous days had R^2 values that were almost close to 1 and the values of the MAPE and RMSE were small. Thus, the FF-ANN model may explain well the degree of lethality of COVID-19 in Saudi Arabia.
- As the negative consequences caused by COVID-19 appear, respectively, in the number of infected cases and number of deaths, the scatter plots of the best FF-ANN (i.e., FF-ANN (5) which models the next day's numbers as a function of the previous 5 days' numbers) are presented in Figure 6 and Figure 7.

Both the SIR-PSO and FF-ANN models are considered to be superior to other methods in the COVID-19 modeling literature. In fact, SIR-PSO showed good performance in providing sub-optimal solutions for the SIR model parameters with limited computational resources. Also, the FF-ANN model yielded good agreement in terms of its predicted numbers of infections (daily and accumulated) against the actual records. Importantly, both of the developed models are generic and can be used (after a few manipulations) for other case studies.

Conclusions

This study focused on the modeling of COVID-19 dynamics in Saudi Arabia for the period from March 2, 2020 to February 21, 2021. The SIR model calibrated by a PSO technique, and the FF-ANN model were implemented to predict the number of next-day infections, recoveries and deaths due to COVID-19. The results showed that both of the models were efficient in modeling COVID-19 dynamics in Saudi Arabia for the considered period. However, COVID-19 has a fluctuating behavior which makes the production of accurate forecasts for the next day (more than 1) a difficult task. The severity of COVID-19 is measured by the number of cumulative infections as well as the number of cumulative deaths. In the case of the number of infections, it is better to use the SIR-PSO model because it does not require a lot of calibration effort, computational time and memory. However, SIR-PSO model was not able to accurately predict the cumulative number of deaths. So, it would seem preferable to use the FF-ANN model to predict the number of deaths. On the other hand, the forecasting of the cumulative recovered cases can be performed by either model. Overall, the performance results of the developed models indicate that they can be used by local authorities to first understand the dynamics of COVID-19 and second implement efficient measures to eradicate the disease.

Acknowledgements

This research has been funded by the Scientific Research Deanship at the University of Ha'il, Saudi Arabia through project number RG-20 214.

Figure 6. Scatter plot of cumulative infected cases (reported and forecasted) provided by the FF-ANN (5) structure.

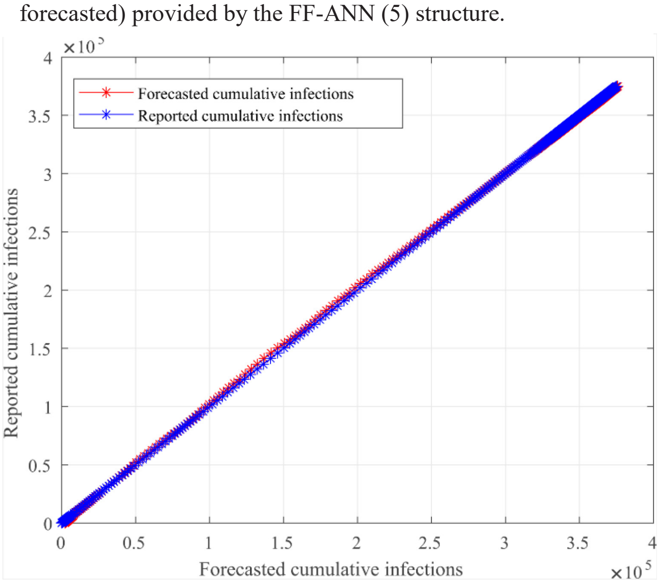
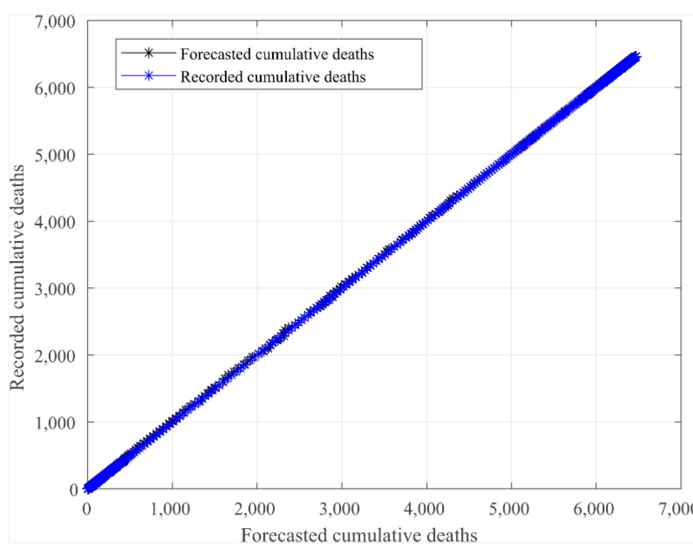


Figure 7. Scatter plot of cumulative deaths (reported and forecasted) provided by the FF-ANN (5) structure.



Supporting information

Codes and supporting information are available upon request. Please contact the corresponding author via e-mail: sboubaker@uj.edu.sa

References

1. World Health Organization (WHO) (2020) COVID-19 – China. Available: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/. Accessed: 10 March 2021.
2. Zhu Na, Zhang D, Wang W, Li X, Yang B, Song D, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W N, Zhang D, Wang W (2019) A novel coronavirus from patients with pneumonia in China, N Engl J Med 382: 727-733.
3. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Piontti APY, Mu K, Rossi L, Sun K, Viboud C, Xiong X, Yu H, Halloran ME, Jr IML, Vespignani A (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 368: 395-400.
4. World Health Organization (WHO) (2020) COVID-19 media briefing. Available: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. Accessed: 10 March 2021.
5. Worldometer (2021) Coronavirus cases. Available: https://www.worldometers.info/coronavirus/. Accessed: 10 March 2021.
6. Worldometer (2021) COVID-19 reported cases and deaths by country or territory. Available: https://www.worldometers.info/coronavirus/?#countries. Accessed: 10 March 2021.
7. Alshammari TM, Altebainawi AF, Alenzi KA (2020) Importance of early precautionary actions in avoiding the spread of COVID-19: Saudi Arabia as an example. Saudi Pharm J 7: 898-902.
8. Harapan H, Itoh N, Yufika A, Winardi W, Keam S, Te H, Megawati D, Hayati Z, Wagner AL, Mudatsir M (2020) Coronavirus disease 2019 (COVID-19): a literature review. J Infect Public Health 13: 667–673.
9. Yezli S, Khan A (2020) COVID-19 social distancing in the Kingdom of Saudi Arabia: bold measures in the face of political, economic, social and religious challenges. Travel Med. Infect Dis 37: 101692.
10. Munayco CV, Tariq A, Rothenberg R, Soto-Cabezas GG, Reyes MF, Valle A, Rojas-Mezarina L, Cabezas C, Loayza M, Chowell G, Garro DC, Vasquez KM, Castro ES, Ordinola IS, Mimbela JM, Cornejo KM, Quijano FC, Rosillo LLT, Ibarguen LO, Dominguez MV, Seminario RVG, Silva MC, Dreyfus MS, Pineda ML, Durand M, Janampa N, Chuquihuaccha J, Lizarbe SM, Cusi DE, Pilco IM, Jaramillo A, Vargas K, Cabanillas O, Arrasco J, Vargas M, Ramos W (2020) Early transmission dynamics of COVID-19 in a southern hemisphere setting: Lima-Peru: February 29th-March 30th, 2020. Infect Dis Model 5: 338-345.
11. Li L, Yang Z, Dang Z, Meng C, Huang J, Meng H, Wang D, Chen G, Zhang J, Peng H, Shao Y (2020) Propagation analysis and prediction of the COVID-19. Infect Dis Model 5: 282-292.
12. Smith BA (2020) A novel IDEA: the impact of serial interval on a modified-Incidence Decay and Exponential Adjustment (m-IDEA) model for projections of daily COVID-19 cases. Infect Dis Model 5: 346-356.
13. Liu Z, Magal P, Seydi O, Webb G (2020) A COVID-19 epidemic model with latency period. Infect Dis Model 5: 323-337.
14. Allen LJS (1994) Some Discrete-Time SI, SIR and SIS Epidemic Models. Math Biosci 124: 83-105.
15. Barmparis GD, Tsironis GP (2020) Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. Chaos Solit Fractals 135: 109842.
16. Postnikov EB (2020) Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions? Chaos Solit Fractals 135: 109841.
17. Fang Y, Nie Y, Penny M (2020) Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. J Med Virol 92: 645-659.
18. Zhao Z, Li X, Liu X, Zhu G, Ma C, Wang L (2020) Prediction of the COVID-19 spread in African countries and implications for prevention and controls: a case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. Sci Total Environ 729: 138959.
19. Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar TK (2020) A model-based study on the dynamics of COVID-19: Prediction and control. Chaos Solit Fractals 136: 109889.
20. Cui Q, Hu Z, Li Y, Han J, Teng Z, Qian J (2020) Dynamic variations of the COVID-19 disease at different quarantine strategies in Wuhan and mainland China. J Infect Public Health 13: 849–855.
21. Fanelli D, Piazza F (2020) Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solit Fractals 134: 109761.
22. Villela DAM (2020) Discrete time forecasting of epidemics. Infect Dis Model 5: 189-196.
23. Roda WC, Varughese MB, Han D, Li MY (2020) Why is it difficult to accurately predict the COVID-19 epidemic? Infect Dis Model 5: 271-281.
24. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M (2020) Application of the ARIMA model on the COVID-2019 epidemic dataset. Data Brief 29: 105340.
25. Ceylan Z (2020) Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ 729: 138817.
26. Ribeiro MHDM, Silva RGD, Cocco Mariani VC, Coelho LDS (2020) Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos Solit Fractals 135: 109853.
27. Chintalapudi N, Batineni G, Amenta F (2020) COVID-19 virus outbreak forecasting of registered and recovered cases after sixty days lockdown in Italy: a data driven model approach, J Microbiol Immunol Infect 3: 396-403.
28. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G (2020) Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model 5: 256-263.
29. Salgotra R, Gandomi M, Gandomi AH (2020) Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. Chaos Solit Fractals 138: 109945.
30. Hu Z, Ge Q, Li S, Jin L, Xiong M (2020) Artificial intelligence forecasting of Covid-19 in China. ArXive preprint 2002.07112.v2.
31. Alqaness MAA, Ewees AA, Fan H, Abd El Aziz A (2020) Optimization method for forecasting confirmed cases of COVID-19 in China. J Clin Med 9: 674.

32. Chimmula VKR, Zhang L (2020) Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solit Fractals 135: 109864.
33. Ndaïrou F, Area I, Nieto J, Torres DM (2020) Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. Chaos Solit Fractals 135: 109846.
34. Putra S, Mu'tamar K, Zulkarnain (2019) Estimation of parameters in the SIR epidemic model using particle swarm optimization. American Journal of Mathematical and Computer Modelling 4: 83-93.
35. Kennedy J, Eberhart R (1995) Particle swarm optimization. In Piscataway, NJ: IEEE Service Center. *Proceedings of the 1995 IEEE International Conference on Neural Networks* (pp. 1942-1948). Perth, WA, Australia.
36. Boubaker S (2017) Identification of monthly municipal water demand system based on autoregressive integrated moving average model tuned by particle swarm optimization. J Hydroinformatics 19: 261-281.
37. Boubaker S (2017) Identification of nonlinear Hammerstein system using mixed integer-real coded particle swarm optimization: application to the electric daily peak-load forecasting. Nonlinear Dyn 90: 797–814.
38. Ftiti Z, Tissaoui K, Boubaker S (2020) On the relationship between oil and gas markets: a new forecasting framework based on a machine learning approach. Ann Oper Res. https://doi.org/10.1007/s10479-020-03652-2.
39. Zhao TF, Chen WN, Liew AWC, Gu T, Wu XK, Zhang J (2019) A binary particle swarm optimizer with priority planning and hierarchical learning for networked epidemic control. IEEE Trans Syst Man Cybern Syst 51: 5090-5104.
40. Boubaker S, Kamel S, Kchaou M (2020) Prediction of daily global solar radiation using resilient-propagation artificial neural network and historical data: a case study of Hail, Saudi Arabia. Eng Technol Appl Sci Res 10: 5228-5232.
41. Firat M, Yurdusev MA, Turan ME (2009) Evaluation of artificial neural network techniques for municipal water consumption modeling. Water Resour Manage 23: 617-632.
42. Saba AI, Elsheikh AH (2020) Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. Process Saf Environ Prot 141: 1–8.
43. Saudi Ministry of Health (2020) COVID-19 Statistics E-Platform. Available: https://www.moh.gov.sa/en/Ministry/OpenData/Pages/platform.aspx. Accessed: 17 December 2021.
44. He S, Peng Y, Sun K (2020) SEIR modelling of the COVID-19 and its dynamics. Nonlinear Dyn 101: 1667-1680.

Corresponding authors

Dr. Sahbi Boubaker, PhD
Department of Computer and Networks Engineering, College of Computer Science and Engineering, University of Jeddah, PO Box 34, Jeddah 21959, Saudi Arabia
Phone: 00966552781848
Fax: 00966126951044
E-mail: sboubaker@uj.edu.sa

Dr. Rafat Zrieq, PhD
Department of Public Health, College of Public Health and Health Informatics, University of Ha'il, Ha'il, PO Box 2440, Ha’il City, Saudi Arabia
Phone: 00966591512445
Fax: 00966165310168
E-mail: r.zrieq@uoh.edu.sa

Conflict of interests: No conflict of interests is declared.