Modelling-based Simulator for Forecasting the Spread of COVID-19: A Case Study of Saudi Arabia

Mohammed Aseeri[1], Ahmad B. Hassanat[2,3,4], Sami Mnasri[4,*], Ahmad S. Tarawneh[6], Khaled Alhazmi[1], Ghada Altarawneh[5], Malek Alrashidi[4], Hani Alharbi[4] , Khalid Almohammadi[4], Dmitry Chetverikov[6], Murad AbdulRaheem Younis[7,8]

[1]Communication and Information Technology Research Institute, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia;
[2]Industrial Innovation and Robotics Center, University of Tabuk, Tabuk 71491, Saudi Arabia;
[3]Computer Science Department, Mutah University, Karak 61711, Jordan;
[4]Computer Science Department, Community College, University of Tabuk, Tabuk 71491, Saudi Arabia;
[5]Accounting Department, Faculty of Business Administration, University of Tabuk, Tabuk, Saudi Arabia;
[6]Department of Algorithms and Their Applications, Eötvös Loránd University, Budapest, Hungary;
[7]Public Health & Epidemiology, Unit of Biostatistics & Research Methodology, School of Medical Sciences, Universiti Sains Malaysia,16150, Kota Bharu, Kubang Kerian, Kelantan, Malaysia;
[8]Assistant Medical Science Department, Community College, University of Tabuk, Tabuk 71491, Saudi Arabia
* Corresponding author: Sami Mnasri

Abstract: In March 2020, Saudi Arabia reported that the Coronavirus disease (COVID-19) spread to its territory, originating from China. In this study, a new simulation model estimates and forecasts the number of infected subjects with COVID-19 in the upcoming weeks, based on different parameters, in two major cities in Saudi Arabia, namely Riyadh (the capital) and Jeddah, the second largest city. Unlike most of the recent simulators, our simulator attempts to focus on real data related to Saudi Arabia. Moreover, this paper investigates the parameters that can help to understand and predict the behavior of the biological curve, particularly, in Saudi Arabia. The proposed forecasting model considers several parameters, such as the infection rate, the virus lifetime, the number of infected people, the number of uninfected people, the recovery rate, the death rate, the recovery period, the period of simulation (in days), and the social distancing. The study investigates different scenarios using random test data and real data, where the results show the importance of two parameters on the pandemic spread; the infection rate and the walking distance. Therefore, this work can be used to raise the awareness of public and officials to the seriousness of the current pandemic.

Keywords: *COVID-19; Forecasting; Infection rate; Virus; Prediction; Simulation.*

1. Introduction

In December 2019, atypical pneumonia caused by the zoonotic 2019 novel coronavirus (2019-nCoV) was first reported and confirmed in Wuhan, China. 2019-nCoV is a fast-spreading virus that infects people in almost all countries. On January 25, 2020, human-to-human transmission was confirmed [1–3].

To mitigate the spread of the virus, the Saudi Government has applied a partial quarantine to prevent the spread of the virus, starting with night curfew from March 23, 2020, followed by closing three major cities, Riyadh, Mecca, and Medina. Entry and exit from these cities have been banned in accordance with the limits defined by the competent authorities.

Other extremely serious prevention procedures have been applied to mitigate the pandemic spreading: group tours, travel packages are suspended across Saudi Arabia; mass gatherings are all cancelled; schools and universities are closed and transferred to online learning. Moreover, the government has closed border crossings and cancelled all international flights.

When writing this paper, we still know little about the infectiousness of 2019-nCoV. The most worrisome aspect is the infectiousness of incubation (or exposed) individuals. That means that incubation individuals are being infected and may have the ability of infectiousness even if they do not have any symptoms [4].

The existing forecasting models, such as the Susceptible-Exposed-Identified-Recovered (SEIR) model, assume that the exposed individual is infected without infectiousness. The Susceptible-Infected-Recovered (SIR) also model does not consider the exposition process. Therefore, the traditional transmission models of pandemic spreading might not simulate COVID-19 accurately. This is a possible key reason why the infected populations estimated by the SEIR model are much larger than the official numbers [5]. The estimated number was 75815

Manuscript received October 5, 2020
Manuscript revised October 20, 2020
https://doi.org/10.22937/IJCSNS.2020.20.10.16

individuals on January 25, 2020 [6], while the official number was only 688; the estimated number was 190000 individuals in Wuhan on February 4, 2020 [7], while the official number was only 1967. Furthermore, a good forecasting model should consider other factors such as the prevention measures taken and the ones that we will focus on in this study.

The main goals of this research paper can be summarized as follows:

- To develop a new simulation model for forecasting the pandemic spreading process before and after the prevention measures.
- To provide a visual interface to educate and inform the public and officials about the speed of the pandemic spreading, and to raise their awareness in the seriousness of the current pandemic.
- This simulator will be customized for the Saudi's case; however, it can be extended to other countries taking into consideration the local parameters.

It is worth mentioning that the results of mathematical models and simulators might not match a real-life situation due to the unlimited numbers of parameters that dictate those situations. However, such models attempt to give us means to explore possibilities. Here, we only wish to provide an approximation of the pandemic spreading in Saudi Arabia and to find out that, statistically, if the Saudi Government procedures are effective to stop the pandemic. In addition, we wish to raise the awareness of both public and officials in the speed of the pandemic spread. To this end, we developed a simulator from scratch using the C# language, by which we conducted several experiments based on different scenarios, attempting to understand what is happening on the Saudi soil in terms of the COVID-19 pandemic spread.

2. Related Works

Since the arrival of COVID-19, the new variant of the coronavirus, several scientific studies have proposed tracking, monitoring and prediction solutions analyzing the spread and behavior of this virus. In the following, we address the most recent studies, their operating methods, advantages and disadvantages.

2.1 Overview of methods based on mathematical modeling

Mathematical modeling is a powerful tool to design and evaluate the behavior of complex dynamic multi-parameter systems such as the spread of diseases. The study in [8] proposes to design mathematical models to trace the social process of propagation of the COVID-19 virus. Although this study presents a mathematical model to help understanding the social practices and the needed social actions in the case of pandemics, it is not followed by a real-case investigation supporting the results of application of the suggested models.

In [9], a framework for simulating the spread of the COVID-19 virus in different cities of China is proposed. This simulator combines the design of urban multi-layer public transport networks with the SEIR model proposing four states for people, namely, susceptible, exposed, infected and resistant. However, this system is specific to the Chinese case, and its extension to the Saudi context, for example, seems to be difficult since the proposed transport networks concern only some Chinese cities. Besides, this model supposes that the recovered persons are not subject to repeated infections, which is not true according to the latest declarations of the World Health Organization in April 2020.

Another study [10] proposes to simulate the spread of COVID-19 in Wuhan to make daily predictions based on a Kalman filter ensemble. Regarding the long-term forecasts, the approach relies on network-based model with Markovian processes. However, since the epidemiological parameters of COVID-19 do not follow the exponential distribution assumed by Markovian models, future works must use non-Markovian systems to better predict the COVID-19 trajectories of spread. Based on the results of the Chinese data making the transition from 15 thousand cases on February 12 to zero case on March15, the authors suggest "aggressive and continued measures" to control the epidemy.

In [11], the authors suggest a multidimensional graphical real-time COVID-19 simulator to monitor the propagation of the pandemic and its behavior. The simulator named MDNIDC-Mapping relies on mathematical modeling based on endogenous variables in different types of disks, nano-, micro-, sub-, mega- and general, to compute the infected cases in China and the world. Despite its efficiency, the proposed system does not rely on real data and assumptions which leave unknown several dimensions of the tackled disease.

2.2 Overview of methods based on machine learning and deep learning

Different artificial intelligence-based approaches have been used to solve various real-world problems [12–16]. In the case of an emergent and fast-spreading epidemy such as COVID-19, machine learning (ML) and deep learning (DL) techniques seem to be useful in helping the authorities to make right decisions. Indeed, ML and DL can consider different parameters such as medical resources, travel mobility and population density as features to predict the virus propagation.

In [17], the authors suggest a composite Monte-Carlo forecasting method combined with a fuzzy rule induction and a deep learning system. This combination allows considering an additional dimension of insights using the decision rules for helping the decision maker. However, the proposed method is not compared with other systems (proposed for the SARS in 2003, for example) to assess its real advantages.

In [18], the authors propose a deep learning algorithm for predicting the risks and the potential possibility that novel viruses infect humans. This research paper can be used for future studies on other viruses. However, since it requires a large amount of data, it seems to be inefficient when applied on viruses with rapid spread such as the SARS-CoV2.

2.3 Other methods

The study [19] proposes a stochastic agent-based and discrete-time model to simulate complex scenarios of infection for the Australian context. The proposed model takes into considerations real assumptions such as the school closing, the degree of respect of the social distancing and the international air traffic. A relevant finding of this study is that the closing of schools is effective only when coupled with high respect of the social distancing. Another interesting deduction concerns the needed duration to control the disease. The study suggests that 14 weeks are enough to control the disease if the level of respect of the social distancing exceeds 90%, while any duration is not sufficient if this level is below 70% of the population.

In [20], a SEIR model is suggested to predict the evolution of COVID-19 in different Chinese cities during the three first months of 2020. This system is used then to predict the evolution in other counties such as Italy and the USA. However, due to the different rates of infection in China and other countries, the results of this system are highly uncertain for other countries.

The study [21] presents a COVID-19 simulator for the case of Pakistan taking into account the mobility data and the population. The results indicate that the order of the infection in the Pakistani case will be tens of thousands despite the respect of the social distance for a long duration. In [22], the authors estimate the mortality and recovery ratios of confirmed cases per day. The proposed system relies on a SDIR (Susceptible, Dead, Infected, Recovered) model applied to data from China in order to forecast the situation three weeks ahead as the outbreak evolves. However, as reported by the authors, the system lacks certainty regarding the dynamics of the pandemic due to the inaccuracy of the available data and the unknown real number of infected cases.

In [23], a method is proposed to predict the evolution of COVID-19 based on quarantined surveillance Chinese data and a modified SEIR model. Although the study represents a real clinical data-based investigation and is more important than a simple mathematical model, it forecasts that the COVID-19 will be a local epidemy in Hubei and its global cases will be reduced sooner. This statement is contradicted in less than one month by the statistics of the cases of March 2020.

The investigation of these different systems proposed in the period from December 2019 to March 2020 indicates several limits and drawbacks such as non-use of real clinical big data and non-existence of a system dedicated to the Saudi context. Therefore, our objective is to propose a simulation to resolve some of the previous limitations, particularly, a simulation dedicated to the Saudi context.

3. The Proposed Model

In this work, a simulation model is built to investigate and monitor the spread speed of COVID-19 in Saudi Arabia. The spread rate of the virus within a society (city or country) depends upon many factors, such as the movements of the people, following the health care rules/recommendations, infection rate, etc. Each of these factors plays a role, which contributes to the spread rate with a certain probability. In the proposed system, we focus on the major factors that affect the spread rate significantly. Figure 1 shows the main screen of the proposed simulation model.

Each factor used in the simulator has its own effect on the results of the simulation. For example, if the hygiene rate is not high in a community, the people there are more likely to be infected with the virus, especially if the population is high. In this work, it is assumed that the hygiene rate is inversely proportional to the infection rate, and is equal to the complement of it.

The simulation factors have a cumulative effect on the pandemic vastness. However, some of the factors strongly affect the pandemic spread in one community and have less effect on another one; e.g., the restrictions degree on the people movements in a very crowded community should not be in the same degree in a less crowded community. Therefore, we need to find the most non-pharmaceutical measures that slow COVID-19 from spreading and keep the health care system under control.

The simulator is implemented using C# programming language to provide a graphical user interface. The main screen of the simulator (Figure 1) is divided into three main parts:

1. The canvas presents the visual part of the simulation. Here, the created population appears as points initialized at random locations in the canvas. The points perform a uniform random walk in the 4 directions, and each direction has the probability p=0.25. The color of the points (persons) presents their conditions: blue, red, green and black for

susceptible, infected, recovered and deceased subjects, respectively.

2. The control panel contains the factors that control the simulation procedure.

3. The output screen shows the results that help one to perform statistical analysis for each simulation run.

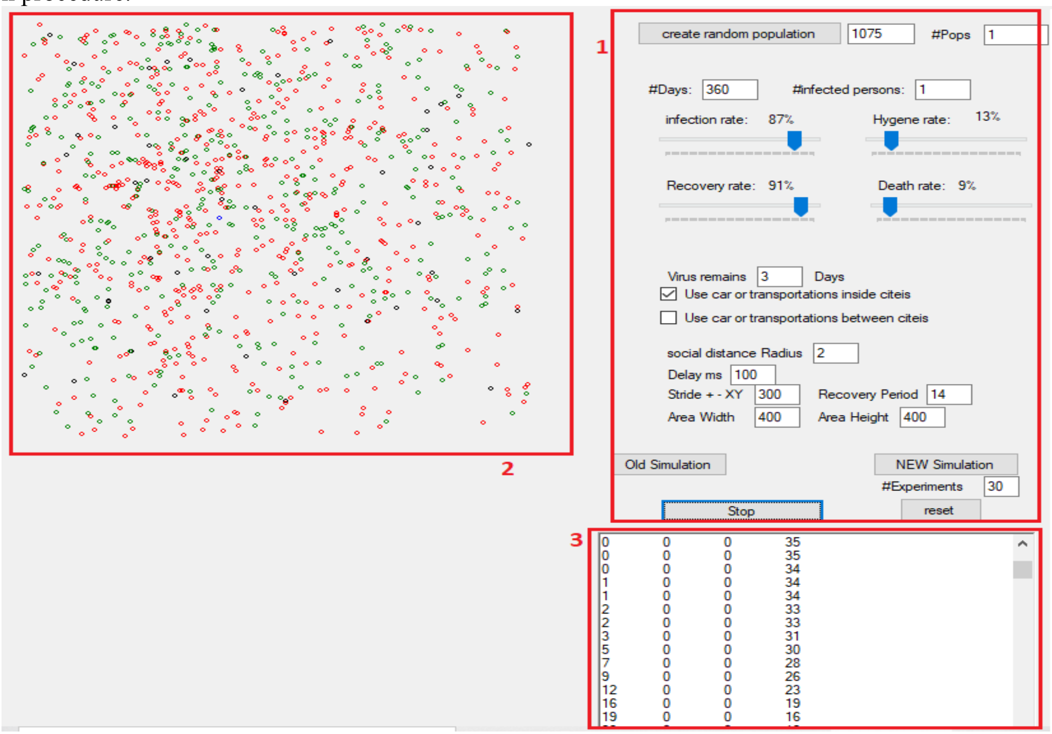


Fig. 1. The main screen of the COVID-19 simulator

The control panel of the simulator shown in Figure 2 contains the most effective factors that control the simulation procedure, as follows:

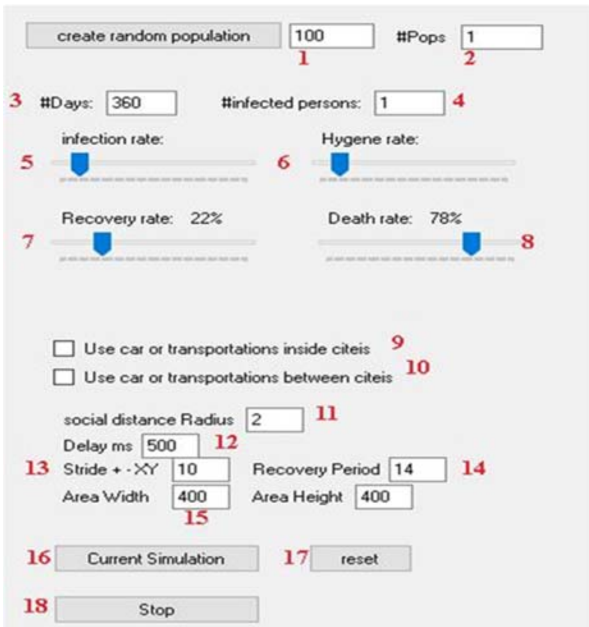


Fig. 2. The control panel of the simulator

1. The number of samples in the population, where 100 means that hundred persons are initialized at random locations in the canvas.

2. The number of populations that represents the number of cities to be simulated using given factors. In this work, we opt for 1 population only: this parameter is reserved for future use.
3. The duration of the simulation in days.
4. The initial number of infected people in each population started with one (Patient zero) in all experiments.
5. The infection rate that presents the probability of spreading the virus among susceptible subjects: if a susceptible subject approaches an infected person within the "social distance", then the probability of being infected equals this rate.
6. The hygiene rate that is the complement of the infection rate. If the hygiene rate is p, the infection rate is 1-p and vice versa. Here, hygiene does not only mean washing hands and using hygiene solutions, but also considering all the recommended healthy precautions such as wearing gloves, masks and other precautions that decrease the probability of infection even when approaching an infected subject. The rate is assumed to be the same for all the community.
7. The recovery rate that is the probability of an infected person to recover.
8. The death rate that is the probability of the infected person to die. The death rate is the complement of the recovery rate.
9. Activate/deactivate transportation within the population (city). We did not use this parameter in this study since it is reserved for future use.
10. Activate/deactivate the transportation between the populations (cities). We did not use this parameter in this study since it is reserved for future use.
11. The "social distance" based on the recommended social distancing of 2 meters between persons [24]. In this simulation, we assume that if the Euclidean distance between a susceptible subject and an infected subject is equal or less than the "social distance", this subject will be infected with a probability equal to the infection rate. We used 2 meters for all experiments; however, the end-user of this system can choose any distance.
12. The delay that is the time between the current frame and the next frame. It is used for visualization purposes, in particular, for small populations.
13. The "stride" that is the distance the subject (point) can move each time frame in any of the four directions. We opt for two random numbers in the range of [-Stride, +Stride], one added to the X-coordinate, and the other one added to the Y-coordinate of each subject. Therefore, the more movement restriction is forced on a community the less the stride is.
14. The recovery period that is the time required for an infected person to heal. Here, we assume it is 14 days, according to [25], and it can be changed for other scenarios if this application is used by practitioners.
15. The canvas area dimensions, that is, width and height of the populated area measured in meters. Each pixel in the drawing is equal to 1 meter, calculated based on the real density of the investigated population. Here, we use 500 x 500 meters, where the number of the population per 1 square kilometer is divided by 4.
16. Starts the simulation using the given factors (parameters).
17. Sets the simulation to the default parameters.
18. Stops the current (in progress) simulation.

There might be other important parameters to be considered, such as the variations in people's health history, different movement speed for everyone in the population, smoker rates in different populations, age distribution, etc. However, we currently have no available information about these factors, so we opt for the aforementioned parameters since information about them is available.

In addition to the random test data based on the population densities of the two major cities in Saudi Arabia, we also use real data about the infected subjects obtained from the Saudi Ministry of health, on which we made regression analysis to provide short-time forecasting.

4. Analysis of Results

We used the proposed simulator to investigate the spread of COVID-19 in two Saudi cities and compute the number of infected, deceased, susceptible and recovered subjects based on different real parameters and scenarios.

Until April 11, 2020, the Saudi government did not yet enforce a complete quarantine, and people were still walking in the streets. Therefore, our proposed simulation focuses on the walking distance of individuals.

The average person takes 4,961 steps per day. Step counts were the highest in Hong Kong, where people took an average of 6,880 steps a day, followed by China, with 6,189 steps, and Ukraine, with 6,107 steps. The countries with the fewest average daily steps were Malaysia, with 3,963 steps; Saudi Arabia, with 3,807 steps; and Indonesia, with 3,513 steps [26].

Focusing on Saudi Arabia, one step is normally estimated to be 0.762m, and therefore the average Saudi walks about 2,901 meter per day.

Our simulation allows the selected population (infected and uninfected subjects) to move randomly in the range of [0, 2091] m, or based on the investigated scenario, when the movement is restricted, where the "Stride" parameter is chosen for each scenario.

The World Health Organization advised people to stay separated for at least one meter [27], and later, they advised them to stay separated for at least two meters [24]. Therefore, the proposed simulator virtually transmits infection (based on infection rate) to susceptible subjects if they are approached by an infected subjected within 2m distance. However, this parameter can be altered by the end-user. Moreover, the infection might be transferred due to passing through the same infected point, and not necessarily

a face-to-face meeting, since the virus stays alive for a specific time (days). This time is another parameter we need to consider. By using the proposed model, one can statistically predict the number of infected, recovered, and deceased cases using several scenarios focusing on:
1. No quarantine, large movements of the subjects, large stride.
2. Quarantine, small movements of the subjects, small stride.
3. Partial quarantine (the current Saudi governmental procedures), intermediate movements of the subjects.
4. Low hygiene and, therefore, high infection rate.
5. High hygiene and, therefore, low infection rate.

In addition, the proposed simulator provides visual imitation of the spread of the pandemic. It uses real information as much as possible to be able to draw significant conclusions when only possible. We use the real population density of each simulated city. Each simulation run is repeated 30 times, and the results are averaged over these 30 trials.

4.1 Simulation scenarios

The simulations are run based on the information of two major Saudi cities, Riyadh and Jeddah. In order to calculate the simulation area needed for each city, its population density is computed, then scaled down to an area of 500 x 500 meters so as to visualize the results and get a faster simulation.

The city of Riyadh has a population of 7,731 million and covers a total surface area of 1,798 km^2 (square kilometers). Its population density is approximately 4,300 persons per km^2. Hence, the width of the simulator area is set to 500 x 500 m, and the population will be 4300 / 4 = 1075 subjects since we reduce the area to 500 x 500 meters (pixels). Regarding the city of Jeddah, the population is about 4,471,000 persons, and its area is 1600 km^2, which gives a population density of 2794 persons per km^2. Then the population used in the simulation should be 2794 / 4 = 699 occupying an area of 500 x 500m. Such tuning based on real situation gives more significance to the simulation results. For each of the two considered cities, two scenarios are considered: low versus high hygiene and low versus high movement. The results are obtained by two methods:
- Normal Random Walk, day-by-day infection.
- Random Walk, but taking into consideration polluted locations in the previous 3 days, assuming that the virus remains 4 days on some surfaces such as plastic, wood and aluminum. This parameter can be altered for different scenarios.

The factors (parameters) involved in our simulations are detailed in Table 1.

Table 1. Used parameters relying on real assumptions and values

Property	Value
Size of the random population	Riyadh: 1075 Jeddah: 699
Number of initial infected persons	1
Number of days for prediction	365
Infection rate	Variable (see scenarios)
Recovery rate	97%
Remaining period of virus	1 day (scenario 1) 4 days (scenario 2)
Social distance radius	2 meters
Delay	100 ms
Stride +-xy (movement)	Variable, in [0,500]
Recovery period	14 days [25]
Area	500 x 500
Simulation repeated (times)	30

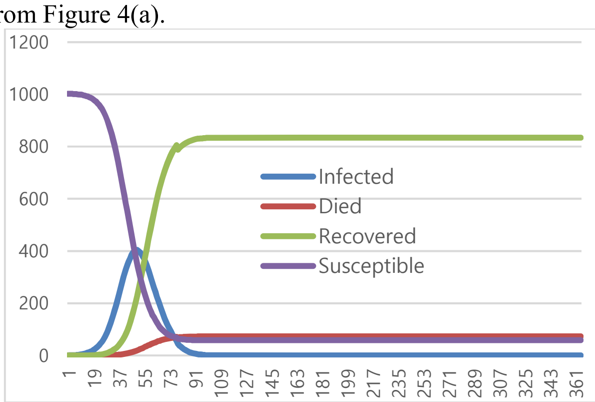
Scenario 1: Locations from previous day(s) are not infectious. When we run the simulator on both cities using small infection rate (10%) and small movement of subjects (+-10 m), starting with one infected subject, the subject remains infected for 14 days, then recovers without infecting any other subject. This is expected since the probability of meeting within the social distance is very low, in addition to the very low infection rate, which comes with considering high hygienic standards if used by any community. This expected result complies with the medical and official advises concerning the maintenance of social distancing and high hygiene. However, when both infection rate and random movement of the subject are increased, we obtain catastrophic results.

From Figure 3(a), it can be seen that the used parameters cause the virus to infect more than 400 subjects out of 1075, and if we scale this number up to match the real population of Riyadh, then we have about hundreds of thousands of infected cases, and this happens around the 50th day, when the deceased cases were more than 150 subjects. The slope of the curve of the infected cases shows the speed of the virus spread. When increasing the movement (stride) from 50m to 300m in the four directions, the results become more catastrophic, as it can be seen from Figure 3(b) where the infected subjects are more than 400. This is also expected, since not only the infection rate influences the spread of the disease, but also the increase of the movements, even without increasing the infection rate, leads to a higher number of infected cases. We have already used these expected results to raise the awareness of the Saudi public and officials through publishing a video (in Arabic)[1] showing these preliminary results, so as to foster the official advises concerning the "stay at home" slogan

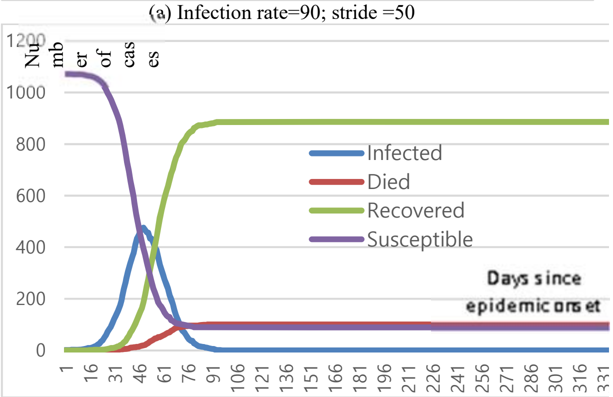
[1]https://www.youtube.com/watch?v=TJftq6shQuw&feature=youtu.be

and maintain high hygienic standards. By doing so, the simulation tells us that the pandemic will be stopped and people will return to the normal life that we know.

The same last observation is applied to the stride influencing the spread of the disease. Moreover, the density of the area seems to influence the propagation speed of the disease. Jeddah is less populated than Riyadh, therefore, it is found to be less infected by the pandemic as can be seen from Figure 4(a).



(a) Infection rate=90; stride =50

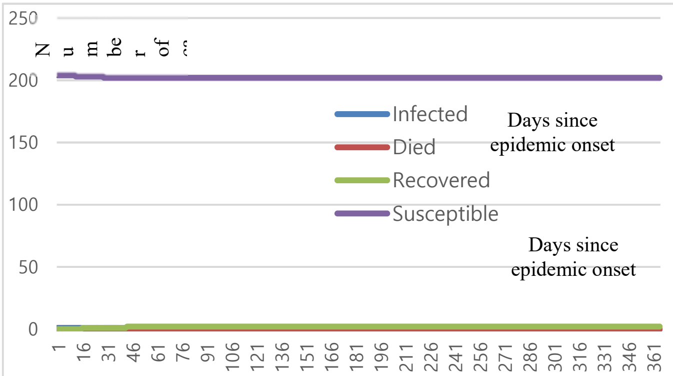


(b) Infection rate=90; stride =300

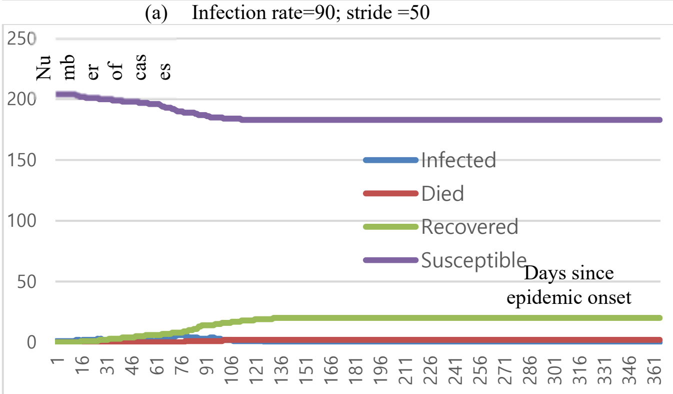
Fig. 3. Prediction of the infected, dead and susceptible cases of COVID-19 under different infection rates and strides for Riyadh city with 1-day infection

This is confirmed by the real infected cases reported by the Saudi officials. However, even in less dense communities, large movement of the population contributes to the increase of the number of infected cases, see Figure 4(b). The chart indicates that this factor, i.e., movement, can also affect the results.

Scenario 2: Locations from the previous 3 days are infectious. This scenario assumes that the virus remains on the location visited by any infected subject for 3 days, therefore, any susceptible subject who moves to any of these infectious locations becomes infected with the probability of the infection rate. This of course is expected to increase the number of infected subjects dramatically. Interestingly, even with all these polluted spots in the inhabitant area, if the infection rate is low (10%), which means high hygiene, and the movement is also low, e.g., below 10 m, the obtained results are good in terms of the number of infected subjects, which is almost zero, after healing the patient zero. However, given less hygiene, e.g., 90% infection rate, and higher movement. e.g., +-50m or even higher (+-300m), we have much more catastrophic situation when more than half, or even most, of the population is infected in a short period of time, particularly, in a dense city like Riyadh, see Figure 5. Regarding the less dense city of Jeddah, the result of scenario 2 is much worse than that of scenario 1 due to the many infectious spots in the inhabitant area. However, the pandemic is much less catastrophic than that of Riyadh due to the lower population density, as can be seen in Figure 6.



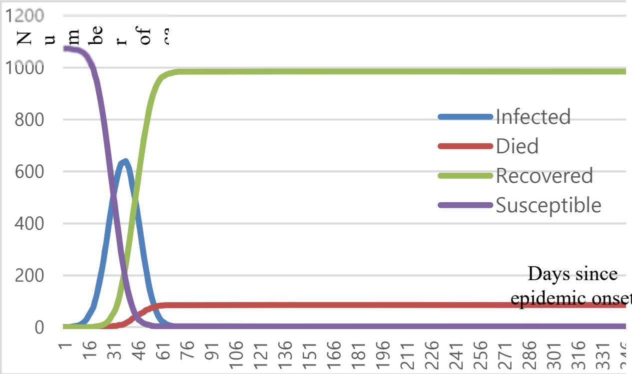
(a) Infection rate=90; stride =50



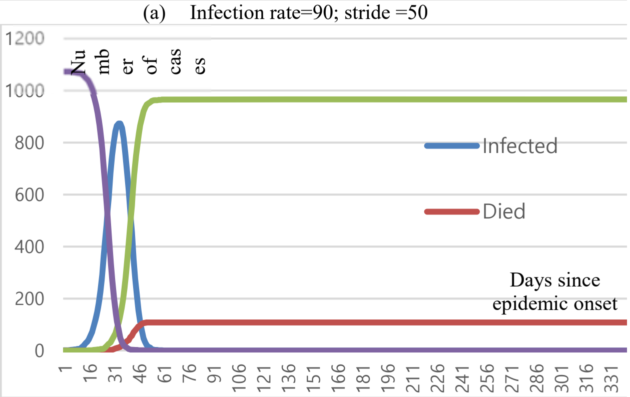
(b) Infection rate=90; stride =300

Fig. 4. Prediction of the infected, dead and susceptible cases of COVID-19 under different infection rates and strides for Jeddah city with 1-day infection

In the aforementioned scenarios, we presented the main application and discussed how the proposed work will be beneficial for the people, in order to educate them about the spread rate the COVID-19 pandemic and urge them to take all measures necessary to stop the pandemic spread, which are mainly decreasing the movement as possible (stay at home) and maintaining the highest hygienic standards, such as wearing gloves, masks, washing hands, faces, maintaining social distances, etc. Moreover, the proposed simulator might be beneficial for the decision-makers, as well, which might enable them to decide which parameters or regulations should they use, particularly, in terms of decreasing the movement of the persons to minimize the number of infected cases and to save human lives.



(a) Infection rate=90; stride =50



(b) Infection rate=90; stride =300

Fig. 5. Prediction of the infected, dead and susceptible cases of COVID-19 under different infection rates and strides for Riyadh city with 4-days infection

4.2 Regression analysis of data

In order to further understand the real situation of Saudi Arabia in terms of the officially recorded infected cases, and to better assess the behavior of the proposed simulator, we collected the real data of the COVID-19 infected cases from the same two major Saudi cities (Riyadh and Jeddah) for 14 days, starting from March 29th 2020 to April 10th 2020, in addition to the total cases in the Kingdom of Saudi Arabia (KSA), as can be seen on the left side of Table 2 and Figure 7 that shows the trends of the infected cases. Regression analysis based on real data can be used for short-term forecasting [28,29], therefore, we opt for linear and exponential regressions to forecast the obtained real data of the infected cases, in order to compare their results with the proposed simulator's results. As can be seen from Figure 7, the exponential and linear curves were fitted to be used for short-period forecasting. The best exponential curve fit for Riyadh is

$$y = \mathbf{22.644}e^{0.0722x} \quad (1)$$

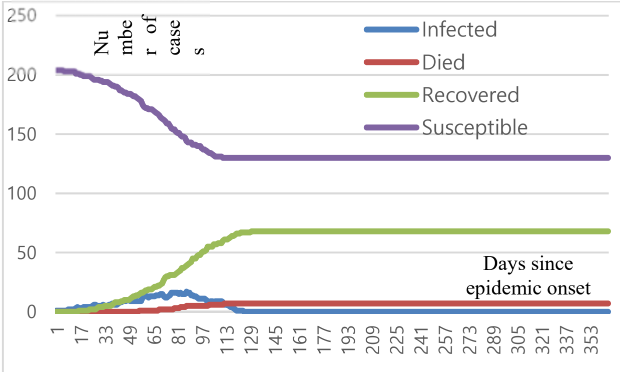
while for Jeddah it is $y = \mathbf{15.635}e^{0.0891x}$ (2)

The best linear curve fit for Riyadh is

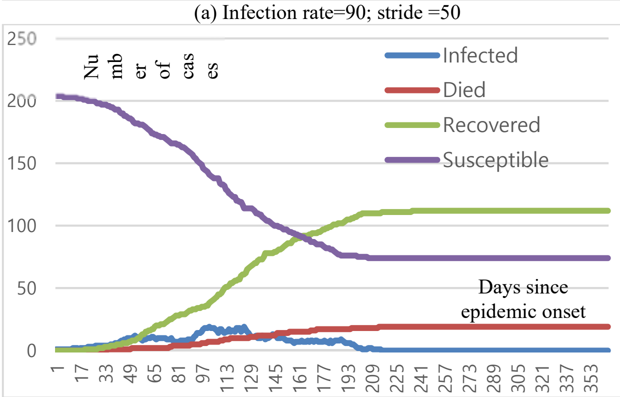
$$y = \mathbf{6.2171}x + 8.0667 \quad (3)$$

while for Jeddah it is

$$y = \mathbf{7.5803}x - 15.652 \quad (4)$$

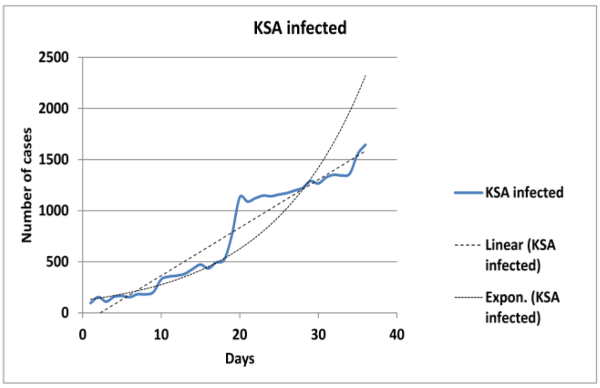


(a) Infection rate=90; stride =50

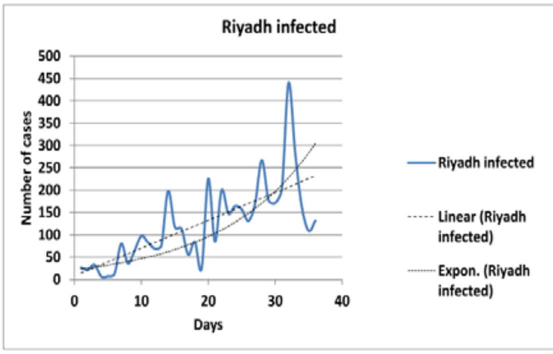


(b) Infection rate=90; stride =300

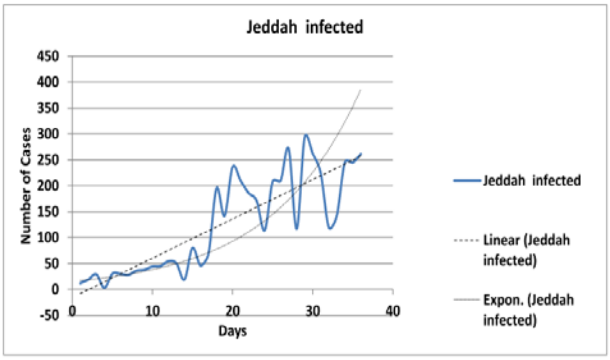
Fig. 6. Prediction of the infected, dead and susceptible cases of COVID-19 under different infection rates and strides for Jeddah city with 4-days infection



(a) All infected cases in KSA



(b) Infected cases in Riyadh



(c) Infected cases in Jeddah

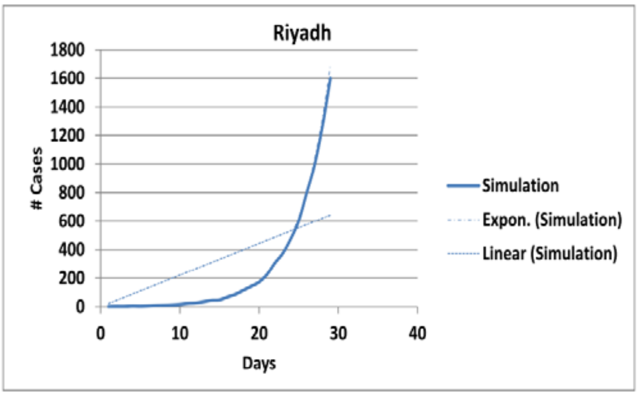
Fig. 7. Curve fitting for the number of infected cases in KSA, Jeddah and Riyadh, respectively

Table 2. Real and predicted values of COVID-19 cases using linear and exponential regression

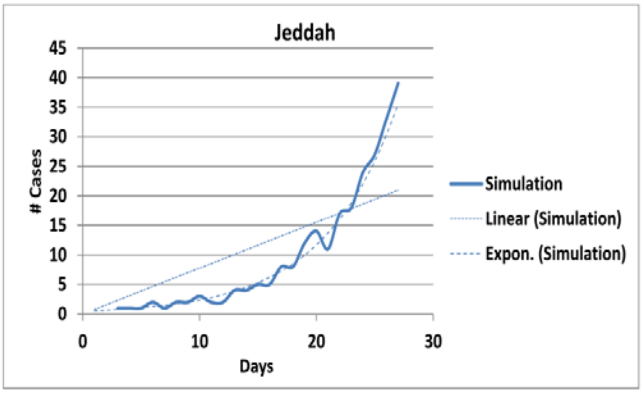
Real Values				Predicted Values						
					Linear			Exponential		
Day	KSA	Riyadh	Jeddah	Day	KSA	Riyadh	Jeddah	KSA	Riyadh	Jeddah
1 (20/04/2020)	1147	148	171	**15**	1631	238.0994	265	2515	327	423
2 (21/04/2020)	1141	164	114	**16**	1678	244.3165	272	2729	352	462
3 (22/04/2020)	1158	157	208	**17**	1725	250.5336	280	2962	378	505
4 (23/04/2020)	1172	131	210	**18**	1771	256.7507	288	3214	407	552
5 (24/04/2020)	1197	170	271	**19**	1818	262.9678	295	3488	437	603
6 (25/04/2020)	1223	267	117	**20**	1865	269.1849	303	3785	470	660
7 (26/04/2020)	1289	178	294	**21**	1912	275.402	310	4108	505	721
8 (27/04/2020)	1266	171	262	**22**	1959	281.6191	318	4458	543	788
9 (28/04/2020)	1325	203	224	**23**	2006	287.8362	325	4838	583	862
10 (29/04/2020)	1351	440	120	**24**	2053	294.0533	333	5251	627	942
11 (30/04/2020)	1344	282	142	**25**	2100	300.2704	341	5698	674	1030
12 (01/05/2020)	1362	161	245	**26**	2147	306.4875	348	6184	725	1126
13 (02/05/2020)	1552	109	245	**27**	2194	312.7046	356	6711	779	1231
14 (03/05/2020)	1645	131	261	**28**	2241	318.9217	363	7283	837	1345

By compensating for x in equations (1), (2), (3) and (4), which represents the day that we wish to forecast for, we get part of the forecast data in the right side of Table 2 (in our analysis, we use data of infected cases starting from 29th of March until 3rd of May, which is collected from [30,31]). The curve tends to be exponential in the dense city (Riyadh) and linear in the less dense city (Jeddah), and also linear in all KSA infected cases as the R^2 value suggests, and there is no indication that the curves will flatten soon. This means that the pandemic is spreading and is not reaching its peak. These results are compatible the previous simulation results in terms of the increase in the infected cases for the first 30 days, particularly, in some scenarios. However, we cannot compare the previous simulator's results with the regression results, because we used a small sample population inhabiting a small sample area, which was a quarter of 1 km^2, and if we want to scale up the results to meet the true number of population, we will end up with a large number of patients zero, which is not the case in Saudi Arabia.

Therefore, we opt for simulating the whole population (7,731,500 and 4,471,000) inhabiting the areas of 1798 and 1600 km2 for Riyadh and Jeddah, respectively. Hence, the Width and height of the simulator are 42403 and 40000 m for Riyadh and Jeddah, respectively. Since the comparison is to be with short-term forecasting, we run the simulator for 30 days starting with 7 infected subjects for Riyadh and 3 for Jeddah. The results are shown in Figure 8. We set the parameters as follows: number of infectious days = 3, infection rate = 90%, stride = 2901m, Area width and height are the same as each city's area, and population is the same as each city's population.



(a) Results on Riyadh



(b) Results on Jeddah

Fig. 8. Simulation results using the whole population of Riyadh and Jeddah, respectively

Table 3. R2 values for each method

Regression	Reported data		Simulation data	
	Riyadh	Jeddah	Riyadh	Jeddah
Linear	0.5148	0.7242	0.6156	0.775
Exponential	0.579	0.7369	0.9952	0.9679

We do not expect our simulation results to match the regression methods results because both methods depend on different factors: the regression methods depend on the historical data only, while the simulation results depend on a number of stochastic parameters. However, we are interested in the trend of the infected cases curve, i.e., whether the growth of the curve is linear or exponential. Table 3 shows the R^2 values for each method: the higher the R^2 value the more variation of the "independent" variable (day) can be explained by the variation of the "dependent" variable (the number of infected cases) [32]. As can be seen from Table 3, the trend is exponential in the dense city of Riyadh, while is more likely to be linear in the less dense city of Jeddah. This is in terms of the reported data, while we find that the simulated data tends to be more exponential rather than linear in both cities. However, it is more exponential in the denser city of Riyadh.

One-day prediction

The proposed simulator forecasts for a period of time (number of days), however, for one-day prediction, we opt for using a number of some of the well-known time series forecasting algorithms as used by [33], these include Simple Moving Average (SMA) [34], Weighted MA, and Exponential smoothing (ES) [35,36], to perform one-day forecasting on. We use Pinball loss function (PBL) to evaluate our prediction results. PBL can be calculated as follows:

$$L\mu(y, z) = \begin{cases} (y-z)\mu & , y \geq z \\ (z-y)(1-\mu) & , z > y \end{cases} \qquad (5)$$

where y is the real data, z is the predicted value and μ is the target quantile ($\mu = 0.95$).

Table 4 presents the forecasting results using SMA and WMA, based on the real infection data for Riyadh, Jeddah and KSA. Table 5 shows the forecasting results using various values of *a*.

The average values indicate that WMA and ES have close forecasting results. ES results are slightly better than WMA while both ES and WMA record better results than SMA method.

Table 4. SMA and WMA average PLF for 36 days in Riyadh, Jeddah and KSA

Period	SMA			WMA		
	Riyadh	Jeddah	KSA	Riyadh	Jeddah	KSA
1	29.327	23.237	45.681	29.327	23.237	45.681
2	25.131	24.047	61.129	26.132	22.842	55.257
3	26.452	26.271	77.788	26.340	24.787	69.167
4	26.934	27.598	94.506	26.103	31.475	83.531
5	28.862	28.217	110.98	26.993	27.235	98.231
6	31.054	29.852	127.19	28.548	28.633	112.98
7	32.942	31.569	143.02	30.224	30.093	127.61
8	34.645	33.544	158.66	32.320	31.635	142.30
9	35.953	35.079	173.98	34.293	33.436	156.96
10	37.619	37.060	188.98	36.441	35.386	171.66
Avg.	**30.892**	**29.647**	**118.19**	**29.672**	**28.876**	**106.34**

Table 5. ES average PLF for 36 days

	ES (Riyadh)	ES (Jeddah)	ES (KSA)
0.1	45.969	49.925	284.92
0.2	33.815	35.143	170.54
0.3	28.682	29.024	120.54
0.4	26.429	26.251	93.670
0.5	26.014	24.716	77.012
0.6	26.112	23.479	65.990
0.7	26.583	22.776	58.091
0.8	27.385	22.451	52.311
0.9	28.363	22.796	48.084
Avg.	**29.928**	**28.507**	**107.90**

5. Conclusion

In this paper, we investigated the problem of COVID-19 pandemic spread in Saudi Arabia under different practical scenarios. For this purpose, we created a new forecasting simulator model that follows the random and actual data trends of COVID-19 spread in two major big cities in Saudi Arabia, namely, Riyadh and Jeddah. Based on the simulation results for the different scenarios, this work showed that controlling some parameters, such as the infection rate and the movement of the population during the day, can play a crucial role in reducing the number of infected persons. Hence, the obtained results will contribute to raising the awareness of the public and officials of the seriousness of the pandemic spread.

Based on the simulation results and the regression analysis of the real data collected, it seems that the infection curve is not going to flatten soon, therefore, we recommend the officials to take more strict steps towards reducing the movement of the public and increasing the hygienic standards. The proposed simulator can be used by the Saudi officials to help control and fight the pandemic spread of the COVID-19, and to provide a tool to help in managing the emerging crises caused by the spread.

Besides, the developed software can be used by other countries for the same purposes provided it is customized to the local parameters. While the challenge involves forecasting the confirmed cases and fatalities, another advantage of this simulation is that it can be used for education. Using the provided graphical interface, the students at schools and universities and public, in general, will learn about the factors that cause the spread of the virus grow faster.

Polluting one point with a specific radius every day is not enough to spread the virus in the tendency of the real-life, and according to the regression results that we analyzed based on the reported cases. Therefore, we are thinking of polluting the whole track used by an infected subject each day and keeping it polluted for the next n days. This will increase the number of infected cases and allows our simulator to follow the same pace of the virus spread. This modification will be left for our future work, in addition to the investigation of more parameters, such as the use of cars and transportation within and between cities, and the consideration of multi-populations.

Funding: This work was supported by the University of Tabuk, KSA, via a grant for the project number 56-1441.
Conflicts of Interest: No conflict of interest to declare.
Sample Availability: Samples of the compounds are available from the authors.

References

[1] Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K.S.; Lau, E.H.; Wong, J.Y.; others. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England Journal of Medicine 2020.

[2] Koenig, K.L.; Bey, C.K.; McDonald, E.C. 2019-nCoV: The Identify-Isolate-Inform (3I) Tool Applied to a Novel Emerging Coronavirus. Western Journal of Emergency Medicine 2020, 21, 184.

[3] Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; others. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020, 395, 497–506.

[4] Shi, Y.; Wang, Y.; Shao, C.; Huang, J.; Gan, J.; Huang, X.; Bucci, E.; Piacentini, M.; Ippolito, G.; Melino, G. COVID-19 infection: the perspectives on immune responses, 2020.

[5] Xiong, H.; Yan, H. Simulating the infected population and spread trend of 2019-nCov under different policy by EIR model. Available at SSRN 3537083 2020.

[6] Wu, J.T.; Leung, K.; Leung, G.M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 2020, 395, 689–697.

[7] Read, J.M.; Bridgen, J.R.; Cummings, D.A.; Ho, A.; Jewell, C.P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. MedRxiv 2020.

[8] Rhodes, T.; Lancaster, K.; Rosengarten, M. A model society: maths, models and expertise in viral outbreaks, 2020.

[9] Li, T. Simulating the Spread of Epidemics in China on the Multi-layer Transportation Network: Beyond the Coronavirus in Wuhan. arXiv preprint arXiv:2002.12280 2020.

[10] Yang, Q.; Yi, C.; Vajdi, A.; Cohnstaedt, L.W.; Wu, H.; Guo, X.; Scoglio, C.M. Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei Province, China. medRxiv 2020.

[11] Ruiz Estrada, M.A. The Monitoring of COVID-19 Infected Cases by Continent under the Application of MDNIDC-Mapping Simulator. Available at SSRN 3545233 2020.

[12] Tarawneh, A.S.; Chetverikov, D.; Hassanat, A.B. Pilot comparative study of different deep features for palmprint identification in low-quality images. arXiv preprint arXiv:1804.04602 2018.

[13] Mnasri, S.; Nasri, N.; van den Bossche, A.; Val, T. Improved many-objective optimization algorithms for the 3D indoor deployment problem. Arabian Journal for Science and Engineering 2019, 44, 3883–3904.

[14] Tarawneh, A.S.; Hassanat, A.B.; Celik, C.; Chetverikov, D.; Rahman, M.S.; Verma, C. Deep Face Image Retrieval: a Comparative Study with Dictionary Learning. 2019 10th International Conference on Information and Communication Systems (ICICS). IEEE, 2019, pp. 185 192.

[15] Mnasri, S.; Van Den Bossche, A. ; Nasri, N.; Val, T. The 3D Deployment Multi-objective Problem in Mobile WSN: Optimizing Coverage and Localization. International Research Journal of Innovative Engineering. ISSN 2395-0560, Vol. 1 N. issue 5, May 2015.
[16] Mnasri, S.; Nasri, N.; Van Den Bossche, A.; Val, T. A Comparative analysis with validation of NSGA-III and MOEA/D in resolving the 3D indoor redeployment problem in DL-IoT. 2017 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC). IEEE, 2017, pp. 15–20.
[17] Fong, S.J.; Li, G.; Dey, N.; Crespo, R.G.; Herrera-Viedma, E. Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Applied Soft Computing 2020, p. 106282.
[18] Zhu, H.; Guo, Q.; Li, M.; Wang, C.; Fang, Z.; Wang, P.; Tan, J.; Wu, S.; Xiao, Y. Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm. BioRxiv 2020.
[19] Chang, S.L.; Harding, N.; Zachreson, C.; Cliff, O.M.; Prokopenko, M. Modelling transmission and control of the COVID-19 pandemic in Australia. arXiv preprint arXiv:2003.10218 2020
[20] Zhang, Y.; Yu, X.; Sun, H.; Tick, G.R.; Wei, W.; Jin, B. COVID-19 infection and recovery in various countries: Modeling the dynamics and evaluating the non-pharmaceutical mitigation scenarios. arXiv preprint arXiv:2003.13901 2020.
[21] Zia, K.; Farooq, U. Covid-19 outbreak in pakistan: Model-driven impact analysis and guidelines. arXiv preprint arXiv:2004.00056 2020.
[22] Anastassopoulou, C.; Russo, L.; Tsakris, A.; Siettos, C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PloS one 2020, 15, e0230405.
[23] Li, R.; Lu, W.; Yang, X.; Feng, P.; Muqimova, O.; Chen, X.; Wei, G. Prediction of the Epidemic of COVID-19 Based on Quarantined Surveillance in China. medRxiv 2020.
[24] Organization, W.H.; others. Management of ill travellers at points of entry–international airports, seaports and ground crossings–in the context of COVID-19 outbreak: interim guidance, 16 February 2020. Technical report, World Health Organization, 2020.
[25] Q&A on coronaviruses (COVID-19). https://www.who.int/news-room/q-a-detail/q-a-coronaviruses. Accessed: April 12th, 2020.
[26] Althoff, T.; Hicks, J.L.; King, A.C.; Delp, S.L.; Leskovec, J.; others. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017, 547, 336–339.
[27] WHO. (2020). COVID-2019 Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. Accessed: April 12th, 2020.
[28] Song, K.B.; Baek, Y.S.; Hong, D.H.; Jang, G. Short-term load forecasting for the holidays using fuzzy linear regression method. IEEE transactions on power systems 2005, 20, 96–101.
[29] Johannesen, N.J.; Kolhe, M.; Goodwin, M. Relative evaluation of regression tools for urban area electrical energy demand forecasting. Journal of cleaner production 2019, 218, 555–564.
[30] COVID 19 Dashboard: Saudi Arabia. https://covid19.moh.gov.sa/en/. Accessed: May 6th, 2020.
[31] COVID 19 interactive Dashboard. https://covid19.cdc.gov.sa/daily-updates/. Accessed: May 6th, 2020.
[32] Brown, A.M. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet. Computer methods and programs in biomedicine 2001, 65, 191–200.
[33] Al-tarawneh, G.A. PREDICTION OF STOCK PRICE USING A HYBRID TECHNICAL ANALYSIS METHOD. Science International Lahore 2019, 31, 391–396.
[34] Hooker, R.H. Correlation of the marriage-rate with trade. Journal of the Royal Statistical Society 1901, 64, 485–492.
[35] Brown, R.G. Exponential smoothing for predicting demand. Operations Research. INST OPERATIONS RESEARCH MANAGEMENT SCIENCES 901 ELKRIDGE LANDING RD, 1957, Vol. 5, pp. 145–145.
[36] Gardner Jr, E.S. Exponential smoothing: The state of the art. Journal of forecasting 1985, 4, 1–28.