Visual Data Analysis and Simulation Prediction for COVID-19 in Saudi Arabia Using SEIR Prediction Model

https://doi.org/10.3991/ijoe.v17i08.20099

Shakir Khan
Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
sgkhan@imamu.edu.sa

Abstract—The World Health Organization (WHO) reported the covid-19 epidemic a global health emergency on January 30 and confirmed its transformation into a pandemic on March 11. China has been the hardest hit since the virus's outbreak, which may date back to late November. Saudi Arabia realized the danger of the Coronavirus in March 2020, took the initiative to take a set of pre-emptive decisions that preceded many countries of the world, and worked to harness all capabilities to confront the outbreak of the epidemic. Several researchers are currently using various mathematical and machine learning-based prediction models to estimate this pandemic's future trend. In this work, the SEIR model was applied to predict the epidemic situation in Saudi Arabia and evaluate the effectiveness of some epidemic control measures, and finally, providing some advice on preventive measures.

Keywords—covid-19, coronavirus, SEIR model, pandemic, Saudi Arabia

1 Introduction

In December 2019, the world began facing the global health crisis, the new Coronavirus disease (COVID-19). The number of infected cases is continuously increasing. Most recently, the Middle East respiratory syndrome coronavirus (MERS-CoV) was first identified in Saudi Arabia in 2012 [1]. Since the pandemic's inception, the total number of infections in Saudi Arabia reached 333,193 infections, statistics of the Saudi Ministry of Health (September 28, 2020). The COVID-19 pandemic and the subsequent social distancing procedures have had significant impacts on limiting the number of cases and limiting the outbreak. Collecting and analysing data related to COVID-19 and learning from the data collected is critical for decision-makers. Data analysis and visualization are of particular importance for monitoring and evaluating control measures. It allows for the presentation of critical indicators and essential information about the dynamics of disease spread and the data's spatial and temporal visualization. It also allows display the geographical area and time interval related to these indicators allowing a more straightforward comparison of the effects of control measures and the possibility of predicting the future state of the case.

Mathematical models and simulations are essential tools for predicting an outbreak's possibility and severity and providing critical information to determine a pathological intervention's type and severity [15]. They resulted in reduced disease transmission and a more objective approach to managing the epidemic. This study aims to provide a local prediction for the peak of the epidemic for COVID-19 in Saudi Arabia using real-time data from March 2, 2020, to March 3, 2021[17]. There are several standard epidemiological models for modelling epidemics, such as SIR [2] [3]. However, since the study aims to analyse daily data series for COVID-19, it seeks to use a more accurate and reliable method, the common Susceptible-exposed-infected-recovered (SEIR), model. SEIR model takes into account various interventions to control the extent of the epidemic [4]. It classifies the population into sections and uses several mathematical equations based on predefined assumptions to predict infection spread peaks.

2 Visualization of spatio-temporal pandemic data

As of December 31st, 2019, 27 cases of idiopathic pneumonia were identified in Wuhan, Hubei Province, China [5]. The causative agent was identified from throat swab samples conducted by the Chinese Centre for Disease Control and Prevention (CCDC) on January 7th, 2020, and was subsequently named Coronavirus 2 (SARS-CoV-2). The World Health Organization WHO) has called the disease caused by the emerging coronavirus "COVID-19". Although Wuhan city practiced a severe quarantine measure for the entire city by closing all transportation to and from Wuhan, the virus turned into a pandemic that spread worldwide. On January 30th, 2020, WHO declared the COVID-19 outbreak in China a public health emergency of international concern and that posed a significant risk to countries with health systems [6].

The pandemic began in its early stage in China, where the first confirmed case of COVID-19 was recorded on January 22, 2020, until it reached the first Arab country in the United Arab Emirates at the end of January. The first infection was confirmed in Saudi Arabia on March 2nd (Figure 1).

Fig. 1. Global spread of coronavirus

3 Overview of the pandemic transmission

WHO has declared the (COVID-19) pandemic a global health emergency with the emergence of cases in countries other than China? In late November, China was the hardest hit since the virus outbreak. The Wuhan government did not realize the seriousness of the epidemic until January 23, which led to an outbreak of the virus abroad [7], as the COVID-19 epidemic spread suddenly. Simply 30 days taken to expand from Wuhan to mainland China and then to the outside world [8]. (Figure 2) shows the rapid increase in cases in endemic countries. On March 02, 2020, the Saudi Ministry of Health announced the registration of the first case of the emerging Coronavirus (COVID-19) for a citizen coming from Iran. Although the case was dealt with immediately and isolated, the Coronavirus takes an incubation period before the symptoms appear on patients, leading to the disease's rapid spread (Figure 3).

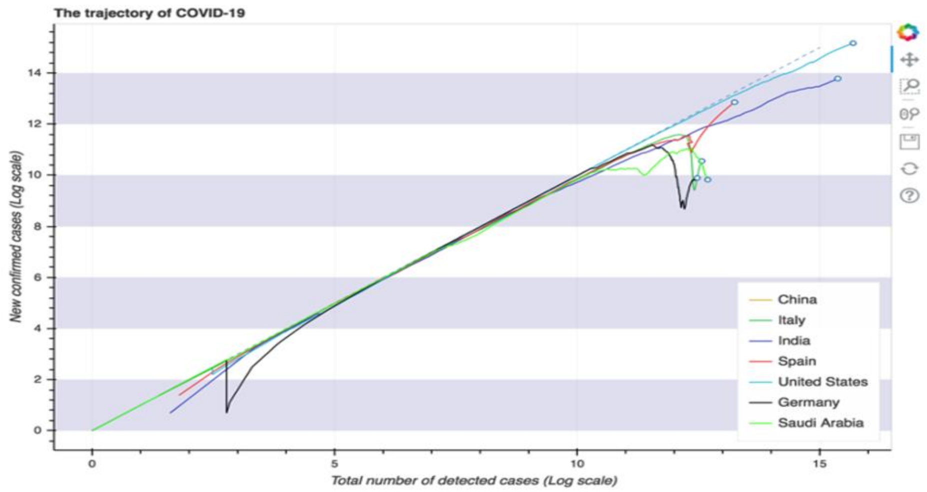

Fig. 2. The trajectory of covid-19

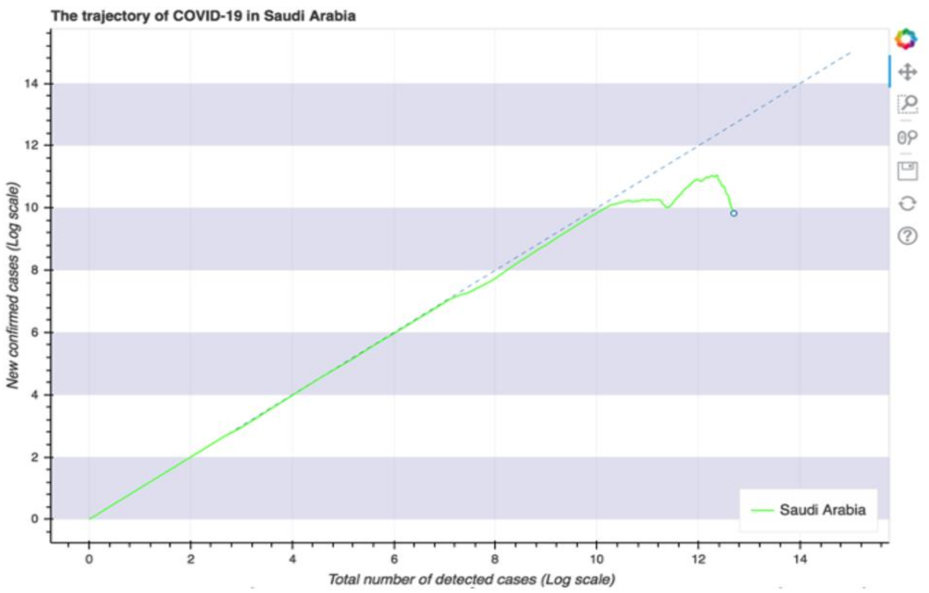

Fig. 3. The trajectory of covid-19 in Saudi Arabia

4 Overview of the pandemic situation in Saudi Arabia

In epidemiology, the semantic number *(R0)* measures the rate of infection spread [9]. This rate is calculated every five days, which is the incubation time for the virus inside the body. Although the incubation period for the virus ranges from 2 to 14 days, the incubation period's exact rate is five days. Every five days is a new epidemic cycle of the virus. The global spread of the Coronavirus *(R0 = 3)* means that one patient can transmit the infection to three people every five days [10]. The rate *(R0)* renders an accurate indication for measuring countries' efficiency measures in preventing the spread of infection. Saudi Arabia revealed early on its high capacity to manage the crisis. Where took a set of strict measures to make the rate decline from the number (R0 = 7), five days after the first case was recorded in Saudi Arabia on March 2, to the rate of infection spread (R0 = 1.04) in the virus cycle Sixth, on April 1 this year. According to the Saudi Health Statistics (September 5, 2020), the total number of cases has been cumulative since the first case in Saudi Arabia has reached 319,932 cases, of which 20,041 active cases are still receiving the necessary health care, and 1470 are critical. Simultaneously, the total number of recoveries reached 295,842 cases, while the death toll rose to 4,049 deaths. Thus, the recovery rate from the total Coronavirus infections in Saudi Arabia reaches 92.4%, while the percentage of active cases is 6.2%, while the percentage of critical cases is 0.459%. For the first time in 142 days, the daily statistics recorded the new infections of the Coronavirus (Covid-19), down to less than 800. As the Ministry of Health had monitored on April 17, 2020, 762 cases were recorded at that time. While on September 5, it announced that 791 had been recorded. New infection, 779 cases recovered and 34 deaths. Saudi Arabi is moving in a positive direction towards receding (Figure 4).

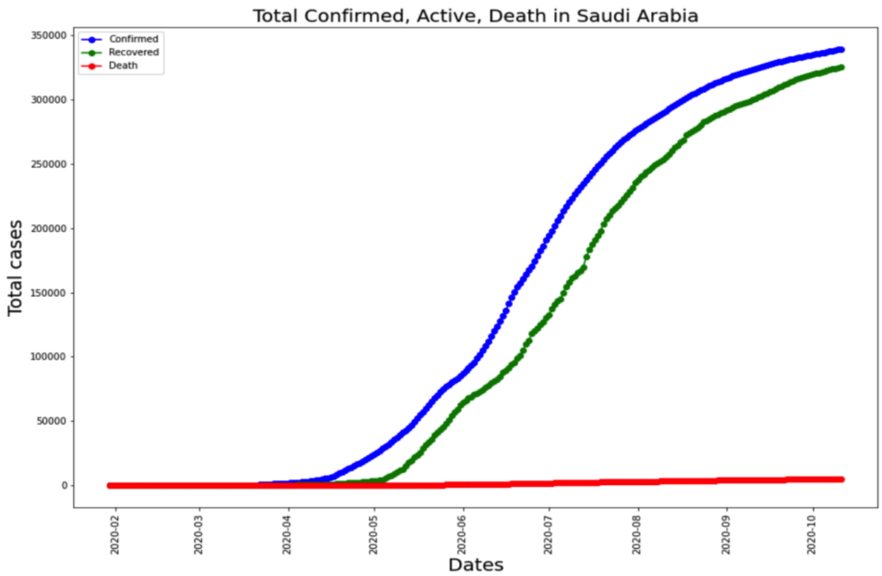

Fig. 4. Epidemiological curve in Saudi Arabia

5 Modelling of epidemiological data for predicting the situation of Covid-19 in Saudi Arabia

Several modeling studies have already been conducted for the COVID-19 epidemic. Wu et al. [11] proposed a Susceptible Exposed Infectious Recovered (SEIR) model to describe the spreading dynamics and predict the disease's national and global extent. It is constructive from the SIR model but adds the latency period (Exposed) compartment as a variable. Model development stages are as follows [12]:

1. SI model:
   Susceptible->Infectible
2. SIS model:
   Susceptible->Infectible-> Susceptible
3. SIR model:
   Susceptible->Infectible-> Recovery/Removed
4. SEIR model:
   Susceptible->Infectible-> Recovery/Removed-> Exposed

5.1 Model formulation

The SEIR model based on the parceling the society into four groups:

- **S**usceptible refers to individuals who can catch the infection and may become hosts if exposed.
- **E**xposed refers to individuals who are already infected but are asymptomatic.
- **I**nfectious refers to individuals who are showing signs of infection and can transmit the virus.
- **R**ecovered refers to previously infected individuals but are no longer infectious and already immune to the virus [13].

The SEIR model's peculiarity is the exposed precinct, characterized by infected individuals who cannot yet communicate with the virus. These individuals are in what is called the latent period. For the COVID-19 virus, this stage makes sense because it takes a specific time for the susceptible individual to be infected at time t, denoted by *S (t),* to enter the infection precinct *I (t).* Since the recovered individuals *R (t)* have immunity to infection, they do not influence transmission dynamics in any way upon contact with other individuals. (Figure 5) shows a graphical representation of Coronavirus progression in an individual. Where infectious occurs at *tL*, latency to infectious transition at *tlt*, symptoms appear at *tsy*, first transmission to another susceptible at *ttr*, and individual is recovered at *tR*.

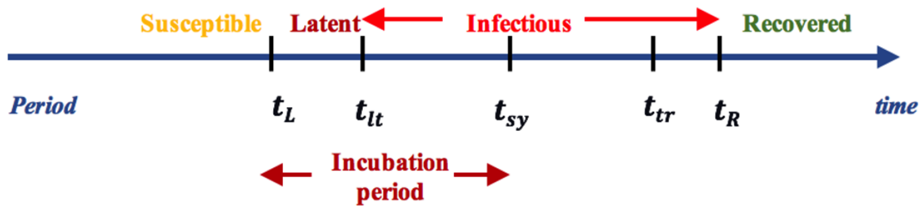

Fig. 5. Individual's coronavirus state

Coronavirus progression is described by the following system of nonlinear ordinary differential equations:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t), \quad (1)$$

$$\frac{dE(t)}{dt} = \beta S(t)I(t) - \gamma E(t) \quad (2)$$

$$\frac{dI(t)}{dt} = \gamma E(t) - \mu I(t),$$

$$\frac{dR(t)}{dt} = \mu I(t),$$

where $\beta \geq 0$ is the transmission rate; $\gamma \geq 0$ is the infectious rate; and $\mu \geq 0$ is the recovery rate. The initial conditions are given:

$S(0) = S0 > 0, E(0) = E0 \geq 0, I(0) = I0 > 0, R(0) = 0$

From (1 & 2),

$$\frac{d}{dt}[S(t) + E(t) + I(t) + R(t)] = 0$$

the population N is constant along time:

$S(t) + E(t) + I(t) + R(t) = N$, for any $t \geq 0$

The dynamics of the SEIR model are characterized by a set of four ordinary differential equations that correspond to the stages of the disease's progression:

$$\frac{dS}{dt} = -\frac{R_t}{T_{inf}} \cdot IS \quad (3)$$

$$\frac{dE}{dt} = \frac{R_t}{T_{inf}} \cdot IS - T_{inc}^{-1} E \quad (4)$$

$$\frac{dI}{dt} = T_{inc}^{-1} E - T_{inf}^{-1} I) \quad (5)$$

$$\frac{dR}{dt} = T_{inf}^{-1} I \quad (6)$$

Where T_{inf} is the average duration of the infection; and T_{inc} is the average incubation period, $T_{inc} = 5.1$ [14]. There is some effectiveness intervention will cause reproduction number *(R0)* reduce, such as inoculation, isolation. (Figure 6) shows the situation without intervention; the confirmed cases in Saudi Arabia are on the increase, which is not very reassured and deserves caution. Several interventions, such as bed nets and vaccines, decay over time (at the population and possibly the individual level). Hill function is one of these interventions that were applied in this study.

$$\frac{1}{(1+\frac{t}{L})^{k}} \qquad \text{Hill Function}$$

Where *L* is a description of the decay rate, either the time until half decay or the time until full decay in years, *k* is a shape parameter (no dimension). In Hill, the decay functions never reach zero and have half their original efficacies at time *L*. (Figure 7) shows the predicted confirmed cases after the intervention. The interventions that took place in Saudi Arabia to combat the spread of the epidemic. It includes preventive measures and precautions that were manifested in stopping foreign flights, stopping domestic flights, stopping Umrah, and this was one of the courageous decisions that saved the Islamic world, as well as closing schools, implementing partial prevention, and then total prevention, thousands were housed in isolation homes, and thousands of citizens were evacuated from abroad.

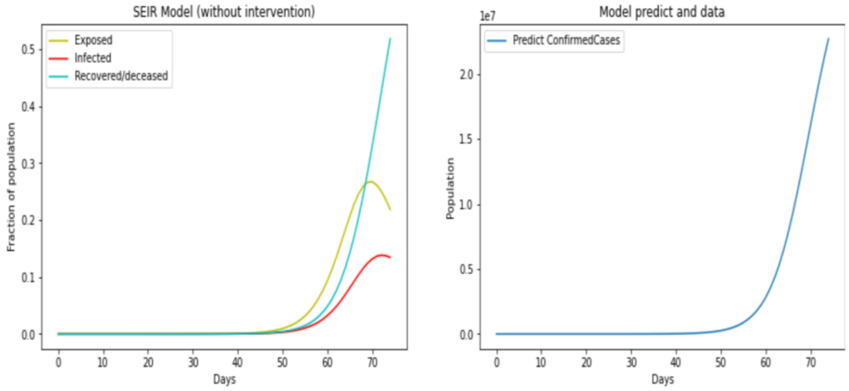

Fig. 6. SEIR model (without intervention)

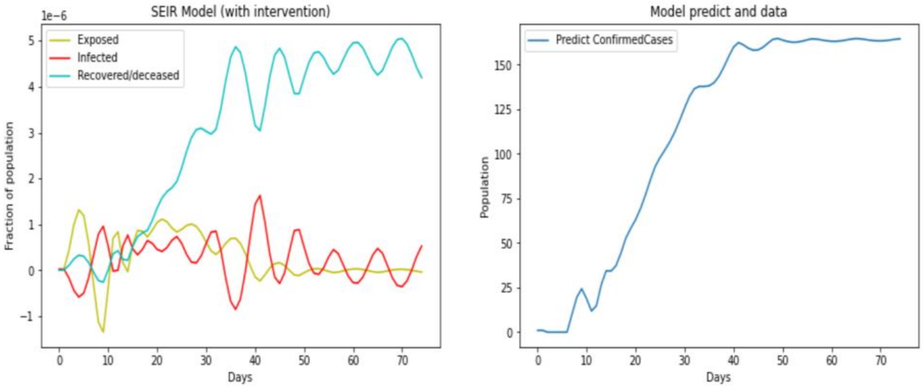

Fig. 7. SEIR Model (with intervention)

5.2 Simulation prediction

Simulation calculates the model retort using input data and preliminary surroundings. Forecast calculates the model retort at around stated extent of time in the future by means of the existing and previous values of dignified input and output values in addition to preliminary conditions. The phase mock-ups of the model response contest the time models of the input data applied for simulation. The furthermost mutual methodology is design identification, the groundwork of much-hyped machine learning and artificial intelligence. Simulation is another substitute in accepting business complications, forecasting upcoming trends, and endorsing best decisions. They explained the basics of simulation models and point out three of its advantages [19, 20]. Table 1 shows the dataset applied for the outcomes.

Table 1. Saudi Arabia Region: Date, Confirmed cases (C.C) and number of deaths (D)

Date	C.C	D	Date	C.C	D	Date	C.C	D
02/03/20	1	0	27/03/20	1104	3	21/04/20	11631	109
03/03/20	1	0	28/03/20	1203	4	22/04/20	12772	114
04/03/20	1	0	29/03/20	1299	8	23/04/20	13930	121
05/03/20	5	0	30/03/20	1453	8	24/04/20	15102	127
06/03/20	5	0	31/03/20	1563	10	25/04/20	16299	136
07/03/20	5	0	01/04/20	1720	16	26/04/20	17522	139
08/03/20	11	0	02/04/20	1885	21	27/04/20	18811	144
09/03/20	15	0	03/04/20	2039	25	28/04/20	20077	152
10/03/20	20	0	04/04/20	2179	29	29/04/20	21402	157
11/03/20	21	0	05/04/20	2402	34	30/04/20	22753	162
12/03/20	45	0	06/04/20	2605	38	01/05/20	24097	169
13/03/20	86	0	07/04/20	2795	41	02/05/20	25459	176
14/03/20	103	0	08/04/20	2932	41	03/05/20	27011	184
15/03/20	103	0	09/04/20	3287	44	04/05/20	28656	191
16/03/20	118	0	10/04/20	3651	47	05/05/20	30251	200

Date	C.C	D	Date	C.C	D	Date	C.C	D
17/03/20	171	0	11/04/20	4033	52	06/05/20	31938	209
18/03/20	171	0	12/04/20	4462	59	07/05/20	33731	219
19/03/20	274	0	13/04/20	4934	65	08/05/20	35432	229
20/03/20	344	0	14/04/20	5369	73	09/05/20	37136	239
21/03/20	392	0	15/04/20	5862	79	10/05/20	39048	246
22/03/20	511	0	16/04/20	6380	83	11/05/20	41014	255
23/03/20	562	0	17/04/20	7142	87	12/05/20	42925	264
24/03/20	767	1	18/04/20	8274	92	13/05/20	44830	273
25/03/20	900	2	19/04/20	9362	97	14/05/20	46869	283
26/03/20	1012	3	20/04/20	10484	103	15/05/20	49176	292

Simulation benefits in predictive analytics:

1. Simulation integrates signals omitted in the data [21]
2. Simulation has comparatively low data attainment and dispensation costs [21]
3. The accuracy of simulation predictions is extremely trustworthy [19]

To get a prediction on the future of epidemic development model's parameters were estimated and simulation activates, aiming to fit the SEIR model with the daily-confirmed cases data reported from Saudi Arabia. Figure 8 shows two curves for the cumulative model infections and the number of confirmed cases. The prediction curve fits with the actual number of cases. From that, it is possible to notice the increase in cases in the first third of March, and this is very close to the official totals announced, as the number of confirmed cases arrived in October 1,2020 to 10.557 the cumulative amount reached 335.097 (Figure 9).

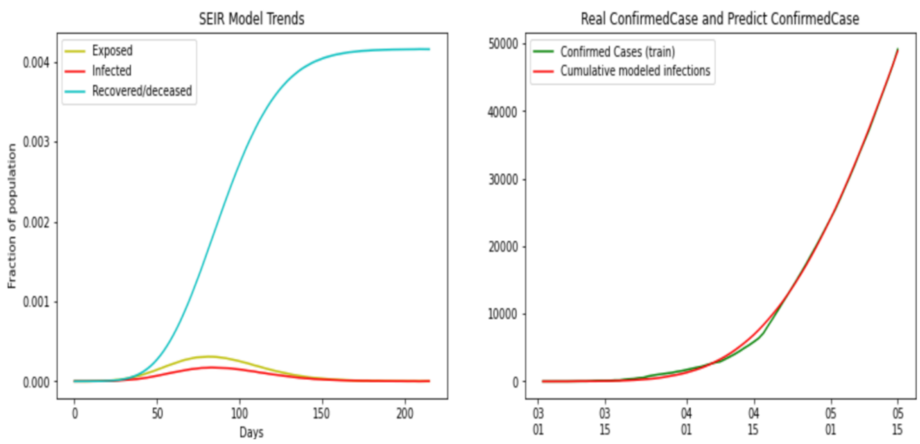

Fig. 8. Simulation prediction

The following parameters are searched and used different values as mentioned here

- T_inc = 5.2 # average incubation period

- T_inf = 2.9 # average infectious period
- R_0 = 3.954 # reproduction number
- k : for predict Fatalities= 3.95469597
- Hill decay. Initial values: k=3, L=15.32328881

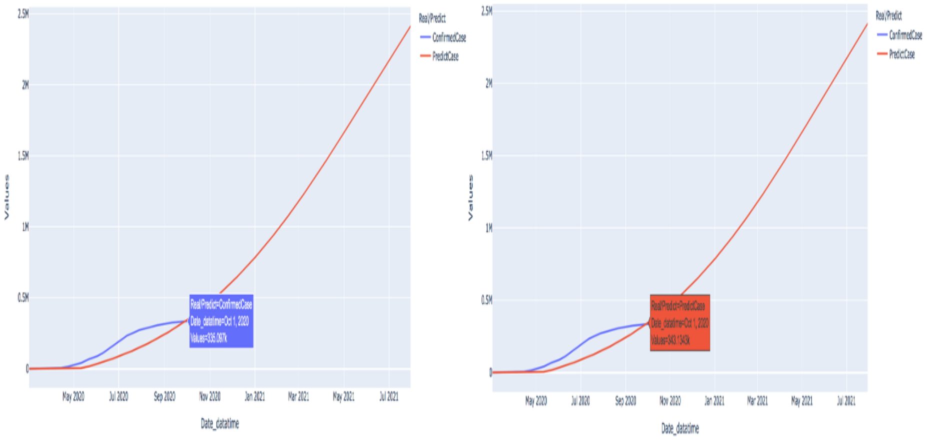

Fig. 9. Real confirmed cases and predict confirmed cases

5.3 Evaluation of Control Measures

The simulation confirms the effectiveness of the various control measures taken by Saudi Arabia early during the arrival of the virus in Saudi Arabia on March 2, 2020, where the first precautionary measure was taken directly on March 3, 2020. Proper quarantine is the most important, especially in the early stage of the epidemic's development. The occurrence of COVID-19 is believed to have started as early as mid-November. However, Wuhan's city did not alert the public until January 20, and effective quarantine measures were practiced until January 23, which led to the virus becoming a pandemic. As evidenced by the simulation in (Figure 10) the effect of early quarantine in Saudi Arabia, but if it is delayed by two days, it may lead to a cumulative weakness or increase in the number of infection. It is critical to perceive that too many affected people may exhaust or cripple available medical resources, as happened in Wuhan, which was then saved by an injection of medical resources (medical staff and supplies) from the country's rest. Although the prediction indicates the continuation of the situation in a stable manner compared to the situation of other countries such as US and France [16], covid-19 is still an unclear contagious disease, which means that we can only get an accurate prediction for SEIR after the occurrence has ended. Nevertheless, even though the SEIR model is a numerical simulation, the numbers give us an insight into how high the number of covid-19 cases can arise. These pathways could serve as a way for governments, companies, and individuals to plan and mitigate such a spike in affected cases. The spread of pandemic is greatly influenced by each country's policy and social responsibility. Therefore, everyone should work to reduce the curve and stop the spread by following the instructions laid down by the government, control measures, and refrain from mass gatherings.

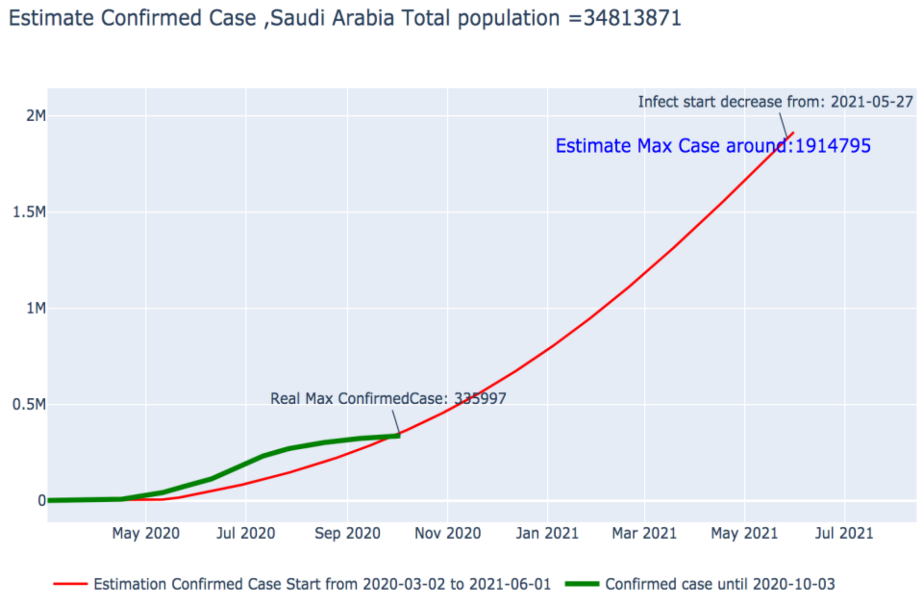

Fig. 10. Estimate confirmed cases in Saudi Arabia from March 02,2020 to May 31, 2021

6 Conclusion

In this study, the trend pattern of the covid-19 outbreak in Saudi Arabia was studied using the SEIR model. The most confirmed cases estimated from March 02, 2020 until May 31, 2021. Confirmed cases could reach 1914795 by May 31, 2021. These findings could help drive policies to tackle the covid-19 pandemic and formulate effective strategies to contain new restrictions on daily social activities and proactive mass testing of potential COVID-19 cases across the country. Finally, some containment recommendations were presented, such as avoid overcrowding and adhering to the precautionary measures. This work has great potential to produce the accurate results and controlling pattern for Covid-19 precautions. The research can be further improved in future to more accurately prediction the spread of the COVID-19 pandemic by training the model with recent data to lead to a more accurate prediction.

7 Glossary of terms

- *WHO*: World Health Organization.
- *COVID-19*: Corona virus Disease 2019.
- *SI*: Susceptible and Infectible.

- *SIS* model: Susceptible, Infectible and Susceptible.
- *SIR*: Susceptible, Infected and Recovered
- *SEIR*: Susceptible, Exposed, Infectious and Recovered.
- *MERS-CoV*: Middle East respiratory syndrome coronavirus
- *CCDC*: Chinese Centre for Disease Control and Prevention.
- *SARS-CoV-2*: Severe acute respiratory syndrome coronavirus 2.
- *R0*: Basic reproduction number, which refers to the average number of secondary infections produced by each new case of infection in a population in which everyone is susceptible.

8 References

[1] Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., Di Napoli, R. (2020). Features, evaluation and treatment coronavirus (COVID-19). In Statpearls [internet]. StatPearls Publishing.

[2] Van Mieghemy, P., Sahnehz, F. D., Scoglioz, C. (2014). An upper bound for the epidemic threshold in exact Markovian SIR and SIS epidemics on networks. In 53rd IEEE Conference on Decision and Control (pp. 6228-6233). IEEE. https://doi.org/10.1109/cdc.2014.7040365

[3] Shakir K, Amani A. (2020). Modeling of Coronavirus Behavior to Predict it's Spread. International Journal of Advanced Computer Science and Applications (IJACSA), 11(5). https://doi.org/10.14569/ijacsa.2020.0110552

[4] Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., ... & Davies, N. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The lancet infectious diseases. https://doi.org/10.1101/2020.01.31.20019901

[5] Lu, H., Stratton, C. W., & Tang, Y. W. (2020). Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. Journal of medical virology, 92(4), 401-402. https://doi.org/10.1002/jmv.25678

[6] Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., ... & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery. https://doi.org/10.1016/j.ijsu.2020.03.036

[7] Chen, B., Shi, M., Ni, X., Ruan, L., Jiang, H., Yao, H., ... & Ge, T. (2020). Visual data analysis and simulation prediction for COVID-19. arXiv preprint arXiv:2002.07096.

[8] Novel, C. P. E. R. E. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi, 41(2), 145.

[9] Liu, Y., Gayle, A. A., Wilder-Smith, A., Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of travel medicine. https://doi.org/10.1093/jtm/taaa021

[10] Yang, C., Wang, J. (2020). A mathematical model for the novel coronavirus epidemic in Wuhan, China. Mathematical Biosciences and Engineering, 17(3), 2708-2724.

[11] Wu, J. T., Leung, K., Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet, 395(10225), 689-697. https://doi.org/10.1016/s0140-6736(20)30260-9

[12] Bagal, D. K., Rath, A., Barua, A., Patnaik, D. (2020). Estimating the parameters of SIR model of COVID-19 cases in India during lock down periods. medRxiv. https://doi.org/10.1101/2020.06.03.20120899
[13] Rachah, A., Torres, D. F. (2017). Analysis, simulation and optimal control of a SEIR model for Ebola virus with demographic effects. arXiv preprint arXiv:1705.01079.
[14] Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine, 172(9), 577-582. https://doi.org/10.7326/m20-0504
[15] Khan, A., & Usman, M. (2019). Alzheimer's Disease Prediction Model Using Demographics and Categorical Data. International Journal of Online and Biomedical Engineering (iJOE), 15(15), 96-109. https://doi.org/10.3991/ijoe.v15i15.11472
[16] Khan, S. & Alqahtani, S. (2020). Big Data Application and its Impact on Education. International Journal of Emerging Technologies in Learning (iJET), 15(17), 36-46. Kassel, Germany: International Journal of Emerging Technology in Learning. https://doi.org/10.3991/ijet.v15i17.14459
[17] Karaçuha, E., Önal, N. Ö., Ergün, E., Tabatadze, V., Alkaş, H., Karaçuha, K., ... & Nu, N. V. N. (2020). Modeling and Prediction of the Covid-19 Cases With Deep Assessment Methodology and Fractional Calculus. IEEE Access, 8, 164012-164034. https://doi.org/10.1109/access.2020.3021952
[18] https://www.mathworks.com/help/ident/ug/definition-simulation-and-prediction.html
[19] Thompson, Jennifer & Hassan, O. & Rolland, Samuel & Sienz, J. (2015). The identification of an accurate simulation approach to predict the effect of operational parameters on the particle size distribution (PSD) of powders produced by an industrial close-coupled gas atomiser. Powder Technology. 291. https://doi.org/10.1016/j.powtec.2015.12.001
[20] Norman R. Swanson, Richard Urbach, (2015), Prediction and simulation using simple models characterized by nonstationarity and seasonality, International Review of Economics & Finance, Volume 40, 2015, Pages 312-323. https://doi.org/10.1016/j.iref.2015.02.027
[21] Alfredo Carter Ramirez, Maria Ludvigsson, (2014), Data Acquisition Methods for Discrete Event Simulation A Case Study at Volvo Cars Torslanda Master's thesis in Production Engineering https://publications.lib.chalmers.se/records/fulltext/202370/202370.pdf.

9 Author

Dr. Shakir Khan received his BSc, MSc and PhD in computer science in 1999, 2005 and 2011 respectively. He is member of the International Association of Online Engineering (IAOE) and IEEE; He is currently working as Associate Professor at College of Computer and Information Sciences in Imam Mohammad Ibn Saud Islamic University, Riyadh (Saudi Arabia). His research interest is Big Data, Data Science, Data Mining, Machine Learning, Internet of Things (IoT), and eLearning, Artificial Intelligence, Emerging Technology, Open Source Software, Library Automation and Mobile / Web Application. He published many research papers in international journals and conferences in his research domain. He has around 15 years of teaching, research and IT experience in India and Saudi Arabia. Dr. Khan is teaching bachelor and master degree courses in the college of computer at Imam University. He is reviewer for many international journals.

Article submitted 2020-11-27. Resubmitted 2021-04-11. Final acceptance 2021-04-11. Final version published as submitted by the authors.