DOI: 10.12962/j20882033.v32i1.7227 Received 27 July, 2020; Accepted 10 February, 2021

ORIGINAL RESEARCH

PREDICT THE SPREAD OF COVID-19 IN IRAN WITH A SEIR MODEL

Shirin Kordnoori*¹ | Mahboobe Sadat Kobari¹ | Hamidreza Mostafaei²

¹Dept. of Computer Engineering-Artificial Intelligenc, Islamic Azad University, Tehran, Iran

²Dept. of Statistics, Islamic Azad University, Tehran, Iran

Correspondence

*Shirin Kordnoori, Department of Computer Engineering-Artificial Intelligence, North Tehran Branch, Islamic Azad University, Tehran, Iran Email: sh.kordnoori@iau-tnb.ac.ir

Present Address

Department of Computer Engineering-Artificial Intelligence, North Tehran Branch, Islamic Azad University, Vafadar Blvd., Shahid Sadoughi St., Hakimiyeh Exit, Shahid Babaee Highway, Tehran, Iran

Abstract

The current coronavirus disease 2019 (COVID-19) outbreak has recently been declared a pandemic and spread over 200 countries and territories. Forecasting the long-term trend of the COVID-19 epidemic can help health authorities determine the transmission characteristics of the virus and take appropriate prevention and control strategies beforehand. Previous studies that solely applied traditional epidemic models or machine learning models were subject to underfitting or overfitting problems. This paper designed a predictive model based on the mathematical model Susceptible-Exposed-Infective-Recovered (SEIR). SEIR is represented by a set of differential-algebraic equations incorporated with machine learning techniques to fit the data reported to estimate the spread of the COVID-19 epidemic in long-term in the Islamic Republic of Iran up to the end of July 0f 2020. This paper reduced R0 after a certain amount of days to account for containment measures and used delays to allow for lagging official data. Two evaluation criteria, R2 and RMSE, had used in this research which estimates the model on officially reported confirmed cases from different regions in Iran. The results proved the model's effectiveness in simulating and predicting the trend of the COVID-19 outbreak. Results showed the integrated approach of epidemic and machine learning models could accurately forecast the long-term trend of the COVID-19 outbreak.

KEYWORDS:

COVID-19, Epidemic Peak, Generalized Additive Models, SEIR Model

1 | INTRODUCTION

A group of related RNA viruses that cause diseases in mammals and birds are named Coronaviruses. These viruses cause recessional extent infections that can range from mild to lethal in humans. Some cases of the common cold, which caused also

certain other viruses, predominantly rhinoviruses, are included in mild illnesses group. On the other hand, lethal manifold can cause SARS, MERS, and COVID-19.

In January 2020, the World Health Organization (WHO) declared the novel coronavirus spread in Public. In February 2020, WHO selected an official name, COVID-19 (stands for Coronavirus Disease 2019), for the infectious disease caused by the novel coronavirus, and later in March 2020 declared a COVID-19 Pandemic.

The answer to the immense pandemic consisted of a merger of containment and reduction strategies (primarily based upon social isolation), aimed to compensate for the clinical unpreparedness, and diminish or control the load on the overburdened health care systems by "flattening the infection curve." Gradually over a few weeks, as more countries and territories became affected, global travel was shut down, universities and schools were closed, followed by bars, restaurants and other entertainment venues, and finally by most churches and other religious or spiritual gatherings. The general population was asked to either quarantine or observe social distancing, depending upon suspicion of active symptoms, and upon the gravity of the local situation. These measures have been directly influenced in most places by (1) limitations in testing abilities; (2) the unpreparedness of the clinical field, of medical resources and suppliers to cater to an outbreak of this magnitude; (3) the difficulty in providing an immediate prevention plan.

One of the fundamental ideas within the sub-specialty of mathematical epidemiology is to model the outbreak of an infectious disease through a population. For predicting the outbreak of viruses there are various models. Along with vaccines or diagnostic tests, if not handled effectively, mathematical modeling can be a useful tool for drafting plans for prediction and designing appropriate intermediacies for rapid control of infectious diseases.

Compartment models for years have used by epidemiologists to make vital predictions about epidemics like measles and polio. This is how can estimate number of people will be infected, how quickly the infections will spread through a population, how long the pandemic will last, what measures will effect on the disease: vaccination, hospitalization, etc., and more. Understanding the disease through such data and models helps clarify the best way can use always limited resources to fight the disease.

2 | PREVIOUS RESEARCHES

In this case several approaches have been proposed to model the pandemic in the literature, including Susceptible-Infected-Removed (SIR) model^[1, 2], Susceptible-Exposed-Infected-Removed (SEIR) model^[3], Susceptible-Infected-Recovered-Dead (SIRD) model^[4, 5], and fractional-derivative SEIR^[5], and SEIRD^[6]. While some recent studies are addressing this epidemic using the aforementioned models^[4, 7], compartmental models simplify the mathematical modelling of infectious diseases. The population is assigned to sections with labels - for example, S, I, or R, (Susceptible, Infectious, or Recovered). The order of the labels usually shows the flow patterns between the sections; for example, SEIS means susceptible, exposed, infectious, then susceptible again. The origin of such models is the early 20th century, with an important work being done by Kermack and McKendrick in Bacaër^[8].

These models are most often run with usual and deterministic differential equations and used with a stochastic framework. These frameworks are more realistic but much more complicated for analyzing. Models try to predict things such as how a disease spreads, or the total number infected, or the duration of an epidemic, and to estimate various epidemiological parameters such as the reproductive number. Such models can show how different public health interventions may affect the outcome of the epidemic, e.g., what the most efficient technique is for issuing a limited number of vaccines in a given population.

The classical Susceptible Exposed Infectious Recovered (SEIR) model represents one of the most adopted mathematical models for characterizing and forecasting the epidemic diseases. In this model it is possible to include key epidemic parameters for COVID-19, such as the latent time, infected time, protection rate and time evolution of the recovery. This allows us to estimate the inflection point, ending time and total infected cases.

The purpose of this study was to estimate the COVID-19 epidemic in Iran based on the SEIR model. Several authors have worked on mathematical modeling of the novel coronavirus. In late 2019, a novel form of Coronavirus, named SARS- CoV-2 (stands for Severe Acute Respiratory Syndrome Coronavirus 2), started spreading in the province of Hubei in China, and claimed numerous human lives^[9, 10].

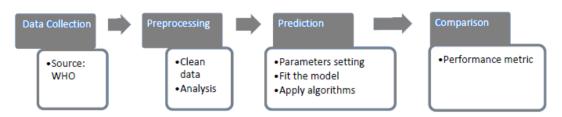


FIGURE 1 Flowchart of the current research process.

In 2019, Putra et al.^[11] used Particle Swarm Optimization (PSO) algorithm to estimate parameters (Susceptible, Infected, Recovered) in the SIR model. The results indicate that the suggested 38 methods are precise enough with low error compared to analytical methods. Mbuvha and Marwala^[12] calibrated the SIR model to South Africa after considering different scenarios for R0 (reproduction number) for reporting infections and healthcare resource estimation for the next few days. In 2020 Qi, Xiao et al. proposed that both daily temperature and relative humidity influenced the occurrence of COVID-19 in Hubei province and in some other provinces^[13]. In 2020, Salgotra et al.^[14] developed two COVID-19 prediction models based on genetic programming and applied this model in India and showed genetic evolutionary programming models are highly reliable for COVID-19 cases in India.

Zareie et al.^[15] applied the SIR model to the prediction of coronavirus spread in Iran based on China parameters [15]. In 2019, Lijuan et al.^[16] proposed the SEIR model which illustrates the relation among susceptible, exposed, infectious, and recovered individuals. It is the widely used model that predicted the outbreak of coronavirus in China as well as in other countries^[17].

Yang et al.^[18] proposed the modified SEIR model by introducing the two new parameters move-in and move-out for the inflow and outflow of susceptible individuals respectively. Lin et al.^[19] discussed the conceptual SEIR model by incorporating the factors government action and public perception.

3 | MATERIAL AND METHOD

3.1 | Material

The study was carried out in several phases:

- 1. Data were collected from World Health Organization (WHO) and John Hopkins University.
- 2. In order to avoid any duplicated and missing values data were analyzed and preprocessed.
- 3. Apply machine learning algorithms in prediction phase.
- 4. Numerical tests were performed using Python and R and executed.

The flowchart of the research methodology is provided in Figure 1 .

3.2 | SEIR Model

SEIR model which is inspired from fundamental model, SIR model, is a representation which used widely to describe a disease spread. SEIR model consists of four-compartment levels: Susceptible, Infectious, and Removed which are similar to SIR model, and one additional compartment is added between the Susceptible and Infectious compartment called Exposed. This can be achieved via Susceptible–Exposed–Infected–Removed/Recovered (SEIR) models. A brief description of these compartments is given as follows.

Susceptible individuals are those people who have no immunity to the disease but they are not infectious. Since there is no vaccine yet developed for this disease, we can say that the entire community is exposed to get infected by this disease and hence, the "Susceptible" compartment can be represented by the entire population. An individual in the "Susceptible" level can move into the next level of the model (Infectious) through contact with an infectious person. By this single transmission, the number

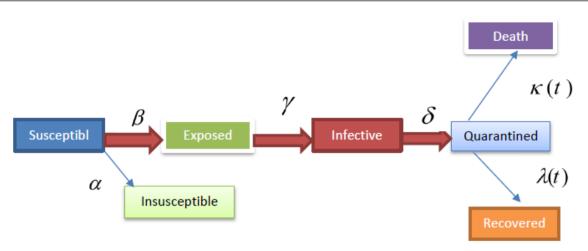


FIGURE 2 A generic SEIR schema.

of susceptible/infectious people reduces/increases by one, respectively. **Exposed** compartment is dedicated to those people who are infectious but they do not infect others for a period of time namely incubation or latent period. **The infectious** people are the next group of people who have the disease and can spread it to susceptible people. Infectious people can move to the "Removed" compartment by recovering from the disease. **The removed** compartment includes those who are no longer infectious and the ones who have dies from the disease (closed cases).

Figure 2 illustrates a typical SEIR scenario in which It models the interaction of people between different conditions: the susceptible (S), exposed (E), infective (I), and recovered (R).

- S is the number of individual susceptible at time t.
- I is the number of infected individuals at time t.
- R is the number of recovered individuals at time t.
- E presents the fraction of individuals that have been infected but does not show any signs.

This diagram can be converted into a mathematical format via the set of differential equations which represented as follows.

$$\frac{ds}{dt}(t) = -\beta \frac{S(t) I(t)}{N} - \alpha S(t)$$
(1)

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

$$\frac{dI}{dt}(t) = \gamma E(t) - \delta I(t)$$
(3)

$$\frac{dQ}{dt}(t) = \delta I(t) - \lambda y(t) Q(t) - K(t) Q(t)$$
(4)

$$\frac{dR}{dt}(t) = \lambda \ y(t) \ Q(t) \tag{5}$$

$$\frac{dD}{dt}(t) = K(t) Q(t) \tag{6}$$

$$\frac{dP}{dt}(t) = \alpha \ S(t) \tag{7}$$

 α presents the protection rate, β shows the infection rate, δ illustrates the inverse of the average latent time, λ_0 and λ_1 displays the inverse of the average quarantine time, coefficients used in time dependent cure rate, K_0 and K_1 are coefficients used in the time dependent mortality rate^[7].

3.3 | Prediction

In the present section, some machine learning techniques were used for COVID-19 case predictions in Iran. Machine learning is a branch of computer science in which data could teach algorithms. The learning process could be done as supervised, unsupervised, and/or semi-supervised learning forms. In this section, some approaches that are used for prediction of cases (confirmed and deaths) of COVID-19 Pandemic are provided.

3.3.1 | Generalized Additive Models

Many data in the environments do not fit with simple linear models and are best described generalized additive models (GAMs). GAMs are simply a class of statistical Models in which the usual Linear relationship between the Response and Predictors are replaced by several Nonlinear smooth functions to model and capture the Non linarites in the data. These are also a flexible and smooth technique which helps us to fit Linear Models which can be either linearly or non-linearly dependent on several Predictors X_i to capture Nonlinear relationships between Response and Predictors.

The concept of additive models with GLM can be combined to procure the idea of GAM, as shown in Eq. 8.

$$gi(\mu Y) = \sum_{i} (f_i(X_i)) \tag{8}$$

The purpose of GAMs is to maximize the quality of the prediction of a dependent variable Y from various distributions, by specific non-parametric functions of the predictor variable which are connected to dependent variable via a link function. The structure of GAM can be written as Eq. 9.

$$g(E(Y)) = \beta + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m)$$
(9)

Where: g(E(Y)) is the link function that links the expected value to the predictor variables $x_1, ..., x_m$, and $f_1(x_1) + f_2(x_2) + ... + f_m(x_m)$ is the functional form with an additive series. Response variable *Y*.

Alternatively, the advantage of GAM is to limit the error in prediction of a dependent variable Y from various distributions by assessing unspecific functions which are connected by means of link function with the dependent variable. The GAM allows a broad range of distributions for the response variable to be adopted, and link functions for measuring the effects of the predictor variables on the dependent repressors.

In this paper, GAMs used to estimate β , μ_Y , I(t) fit by Restricted Maximum likelihood (REML) and log distribution. The uncertainty of data on the first days of the epidemic in Iran is one of the limitations in this paper.

3.3.2 | Times Series Forecasting With the Prophet Algorithm

The Prophet algorithm is an open-source tool developed by Facebook' s Data Science team, and its main goal is business forecasting. The Prophet algorithm works well with time-series data that have seasonal effects and are robust in dealing with missing data. In the Prophet algorithm, the forecast could be written as shown in Eq. 10.

$$\hat{y}_{T+h|t} = \bar{y} = \frac{y_1 + y_2 \dots + y_T}{T}$$
(10)

in which $y_1 + y_2 + ... + y_T$ are denoted as historical data, and $\hat{y}_{T+h|t}$ is a short-hand to forecast $y_{T+h|t}$ based on available data.

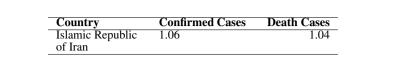


TABLE 1 R2 score for confirmed and death cases.

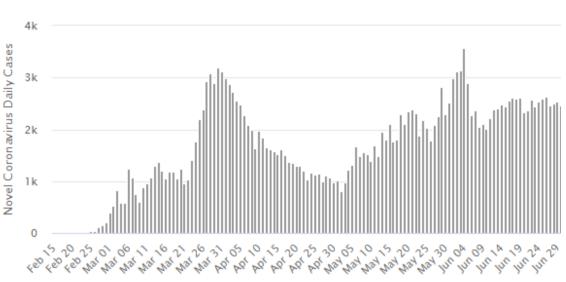


FIGURE 3 Confirmed/deaths/recovered cases in Iran.

4 | RESULTS AND DISCUSSION

4.1 | Evaluation criteria

With the aim of forecasting, the GAM function defined in Equation and was applied to collected data and results have been illustrated in Figures. As it is shown in Figures, the function is fitted until the trend of cases is increases and to evaluate the performance of metric R2 scores used for confirmed and death cases. Another metric that has been used in experiments is the root mean square error (RMSE), as shown in Eq. 11.

$$R^2 = 1 - \frac{RSST}{SS} \tag{11}$$

Where RSS means residual sum of squares is defined by Eq. 12.

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$
(12)

Where y_i is residual sum of squares and $f(x_i)$ is predicted value of y_i .

4.2 | Results

We could not recognize a better model criterion to find distribution of data in the first week of the epidemic, because diffusion parameter of the model with Iranian data through non-linear least squares or polynomial regression models posed problems. The most important point is to emphasize the timing of the epidemic peak, hospital readiness, government measures and public readiness to reduce social contact. Confirmed cases in Iran up to 29 of June are present in Figure 3.

With the aim of forecasting, the function which defined in Eq. 9 was applied to collected data and results have been illustrated in Figure 4. The GAM function is fitted until the trend of cases is increases to evaluate the performance of metric R2 scores used for confirmed and death cases. Results are presented in Table 1.

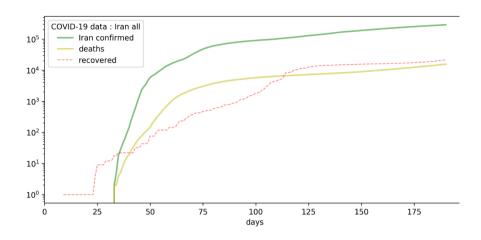


FIGURE 4 Confirmed/deaths/recovered cases in Iran.

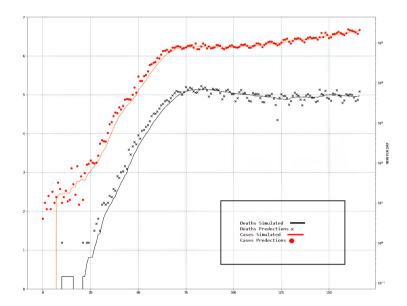


FIGURE 5 Cases/Deaths in per day.

Figure 4 shows confirmed/deaths/recovered cases in Iran with 175 data points from 1st of January of 2020 up to end of the July of 2020. This estimation shows that the COVID-19 Iran which raised from March 25, 2020 and peak during March 25, 2020 - March 27, 2020, and a decreased from March 28, 2020 - April 15, 2020, but unfortunately, in July month the epidemic peak will occur again and confirmed cases will increase up to more than 250000 cases. On the other hand, by this plot it can be understanding that deaths and recovered cases will have no change with constant change.

Another metric that has been used in experiments is the root mean square error (RMSE), and the results of RMSE depicted in Figure 4 which reached 1.41, in this case the prophet algorithm used for fitting data. Figure 5 presents the prediction cases/deaths in per day which referred to end of the July of 2020 and simulated.

At the time of the spread of COVID-19, the best and most important steps must be taken to overcome the coronavirus epidemic. It can only be overcome with active collaboration of different businesses and industries such as the medical industry, transportation, government, producer of technology products, etc.

Multiple essential tasks must be done to control the spread of the virus; 1) Controlling the origin of the virus. In the face of the virus, the most effective strategy is to keep any influenced patient in a relatively limited space, hospital or home, to prevent

the virus from spreading. Patients who influenced by the virus should be treated at the hospital immediately. 2) Controlling the infectious of the virus. Restrictive measures must be taken against the general population or the population that may spread the disease to prevent a major epidemic from continuing. 3) Virus exploring. In addition, the main source of the virus must be traced to understand the source of the virus and effective measures must be taken to completely eliminate the source of the virus.

For an entire contaminated or potentially contaminated city, continuous air detection must be carried out to effectively detect or trace the virus in the air, and masks must be worn by all citizens, especially in public places. New mobile hospitals should be built that can handle suspected cases in a centralized manner. This is important to diminish the pressure of curing the populace of infected patients in large hospitals. For example, during this outbreak, in only 10 days for people who infected with the coronavirus, the Huoshenshan Hospital in Wuhan was built. During the whole process, the contagious situation should be reported in a timely and clear manner. It also excludes abrupt public pressure and reduces inessential psychic health problems.

Artificial intelligence technology can play a leader role in almost every aspect, including traffic handling, infectious disease, logistics supply chain, etc. What said are very important features of a modern data-driven smart city. All can be accurately tracked and any population can be accommodated, if the status of each citizen is listed. Therefore, the flow of population can be controlled in a more regular way. Artificial intelligence technologies can be used to employ smart devices to support diagnosis and treatment, and can be used in telecommunications, online training and intelligent manufacturing to ensure minimal disruption to people's lives. Some hospitals use intelligence systems. Train stations and airports can install powerful thermal imagers to measure the passengers' body temperature. During the whole control process, efficiency and speed is extremely important and cross-disciplinary research should be managed. There are several determinations for governments to make. However, people respond to opinion on preventing transmission is more important than government measures.

5 | CONCLUSION

This paper designed and developed a predictive model of COVID-19 disease based on an artificial intelligence model and mathematical SEIR model. Because, artificial intelligence technology can play a key role in almost every aspect, including traffic management, infection detection, logistics supply chain, etc. SEIR model represented by estimating the four parameters of basic reproductive number, time-dependent transmission rate, time-dependent recovery rate, and time-dependent death rate from Covid-19 outbreak in world, and used the number of Covid-19 infections rate in Iran. In this paper used:

- Mathematical approaches based on SEIR model were proposed to predict the epidemiology in Iran
- Two machine learning approaches, GAMs and prophet algorithm, described and the data fitted based on them to represent results.
- Evaluation criteria was R2 and RMSE; R2 score by using GAMs model reached that confirmed cases would be 1.06 and death cases would be 1.04. RMSE score reached 1.41.

Unfortunately, in July month the epidemic peak will occur again and confirmed cases will increase up to more than 250000 cases. Several essential tasks must be done while controlling the outbreak proposed in this paper.

In this paper, all forecasting was addressed without considering of scenario of social distancing and quarantine that makes it valuable as a future direction. This paper presented SEIR as epidemiology model; it would be interesting to test other epidemiology models. Moreover, it is worthwhile to combine the mathematical model with other observations such as Policy intervention, human behavior, and constraints.

CREDIT

Shirin Kordnoori: Investigation, Software, Conceptualization, Writing - Original Draft. **Mahboobe Sadat Kobari:** Data Curation, Validation, Writing - Review Editing. textbfHamidreza Mostafaei: Formal analysis, Supervision, Writing - Review Editing

References

- 1. Zhong L, Mu L, Li J, Wang J, Yin Z, Liu D. Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model. IEEE Access 2020;8:51761–51769.
- 2. Nesteruk IG. Statistics Based Models for the Dynamics of Chernivtsi Children Disease. Research Bulletin of the National Technical University of Ukraine "Kyiv Politechnic Institute" 2017;5:26–34.
- Neher RA, Dyrdak R, Druelle V, Hodcroft EB, Albert J. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. Swiss Medical Weekly 2020;150(w20224):1–8.
- 4. Caccavo D. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. medRxiv 2020;https://www.medrxiv.org/content/early/2020/04/21/2020.03.19.20039388.
- 5. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS ONE 2020;15(3):1–21.
- 6. Khan MA, Atangana A. Modeling the dynamics of novel coronavirus (2019-nCov) with fractional derivative. Alexandria Engineering Journal 2020;59(4):2379–2389.
- 7. Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. medRxiv 2020;https://www.medrxiv.org/content/early/2020/02/18/2020.02.16.20023465.
- Bacaër N. A short history of mathematical population dynamics. 1 ed. London: Springer; 2011. https://link.springer.com/ book/10.1007%2F978-0-85729-115-8.
- 9. Ellerin T, Farid H, Krakower D, LeWine HE, McCarthy C, Memon B, et al., Coronavirus Resource Center Harvard Health; 2021. https://www.health.harvard.edu/diseases-and-conditions/coronavirus-resource-center.
- 10. Xu B, Gutiérrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. Scientific Data 2020;7.
- 11. Putra S, Muktamar K, Zulkarnain. Estimation of Parameters in the SIR Epidemic Model Using Particle Swarm Optimization. American Journal of Mathematical and Computer Modelling 2019;4(4):83–93.
- 12. Mbuvha R, Marwala T. On Data-Driven Management of the COVID-19 Outbreak in South Africa. medRxiv 2020;https: //www.medrxiv.org/content/early/2020/04/15/2020.04.07.20057133.
- 13. Qi H, Xiao S, Shi R, Ward MP, Chen Y, Tu W, et al. COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. Science of The Total Environment 2020;728:138778. https://www.sciencedirect.com/science/article/pii/S0048969720322956.
- Salgotra R, Gandomi M, Gandomi AH. Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming. Chaos, Solitons & Fractals 2020;138:109945. https://www.sciencedirect.com/science/article/pii/ S0960077920303441.
- 15. Zareie B, Roshani A, Mansournia MA, Rasouli MA, Moradi G. A Model for COVID-19 Prediction in Iran Based on China Parameters. Arch Iran Med 2020;23(4):244–248. http://aimjournal.ir/Article/aim-15640.
- 16. Lijuan Z, Fuchang W, Xuqin Z, Zhitong J. Global Stability Analysis on One Type of SEIR Epidemic Model with Floating Population. Journal of Institute of Disaster-Prevention Science and Technology 2019;21(2):78–81.
- 17. Safi MA, Garba SM. Global stability analysis of SEIR model with holling type II incidence function. Computational and Mathematical Methods in Medicine 2012;2012:1–8.
- Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease 2020;12(3). https://jtd.amegroups. com/article/view/36385.

19. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. International Journal of Infectious Diseases 2020 apr;93:211–216.

How to cite this article: Kordnoori S., Kobari M.S., Mostafaei H., (2021), Predict The Spread of COVID-19 in Iran with A SEIR Model, *IPTEK The Journal of Technology and Science*, *32*(*1*):*31-40*.