Medical & Biological Engineering & Computing (2022) 60:969–990
https://doi.org/10.1007/s11517-022-02525-z

ORIGINAL ARTICLE

Simulation of the COVID-19 patient flow and investigation of the future patient arrival using a time-series prediction model: a real-case study

Mahdieh Tavakoli[1] · Reza Tavakkoli-Moghaddam[1] · Reza Mesbahi[1] · Mohssen Ghanavati-Nejad[1] · Amirreza Tajally[1]

Received: 26 August 2021 / Accepted: 1 February 2022 / Published online: 12 February 2022

Abstract

COVID-19 looks to be the worst pandemic disease in the last decades due to its number of infected people, deaths, and the staggering demand for healthcare services, especially hospitals. The first and most important step is to identify the patient flow through a certain process. For the second step, there is a crucial need for predicting the future patient arrivals for planning especially at the administrative level of a hospital. This study aims to first simulate the patient flow process and then predict the future entry of patients in a hospital as the case study. Also, according to the system status, this study suggests some policies based on different probable scenarios and assesses the outcome of each decision to improve the policies. The simulation model is conducted by Arena.15 software. The seasonal auto-regressive integrated moving average (SARIMA) model is used for patient's arrival prediction within 30 days. Different scenarios are evaluated through a data envelopment analysis (DEA) method. The simulation model runs for predicted patient's arrival for the least efficient scenario and the outputs compare the base run scenario. Results show that the system collapses after 14 days according to the predictions and simulation and the bottleneck of the ICU and CCU departments becomes problematic. Hospitals can use simulation and also prediction tools to avoid the crisis to plan for the future in the pandemic.

Keywords Simulation · COVID-19 pandemic · Patient flow · Scenario evaluation · Time-series prediction

1 Introduction

COVID-19 looks to be the worst pandemic disease in the last decades due to its number of infected people, deaths, and also the staggering demand for healthcare services, especially hospitals. Based on Worldometer's COVID-19 data reports, the number of confirmed infected people is growing every day. As reported on 7 May 2021, the total number of worldwide cases of coronavirus is 157,049,695. Of these, 3,274,689 (2%) people were dead and 134,416,166 (85%) were recovered, and the number of active patients is 19,358,840 (13%).[1] As COVID-19 is a global public health emergency, hospitals are the most important organizations that can help to avoid the destructive effects of this pandemic.

[1] https://www.worldometers.info/coronavirus/

✉ Reza Tavakkoli-Moghaddam
tavakoli@ut.ac.ir

[1] School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran

The first and most important step in facing this pandemic is to identify the suspected patients who enter the hospital whether they are infected. To do this, all patients go through a certain process. In this case, if they are not infected, they will not enter the treatment system, but if they are infected, they need to go through certain steps in the hospital, based on their symptom level. So, the COVID-19 patient journey in hospitals should be specified to find the bottlenecks in this process. According to circumstance happened during the pandemic, there is a critical need to change some resources (e.g., nurses, physicians, and facilities) or planning in different process of hospitals to serve the patient in a stable situation [40]. Usually, understanding the entire process for a patient's journey is difficult. In this case, researchers most use queue theory, Petri nets, and simulation for clarifying the structure of patient processes [8, 17]. In this study, we use the simulation tool for visualizing the suspected COVID-19 patient's journey process.

To improve the performance of hospitals as a system, it is necessary to have a dynamic understanding of it. To achieve such an understanding, simulation provides an ideal tool for determining and allocating the capacity needed to respond to demand in a timely manner and minimize delay. Also, simulation is a convenient alternative with less time and more cost than most traditional statistical methods [7]. In general, discrete event simulation models that have been used in various studies to analyze health care delivery systems mainly focus on two areas (a) optimizing the flow of patients in different departments and (b) allocating resources to improve services. In optimizing the flow of patients, the goal is to improve the output of patients and reduce waiting times, and the second area is to improve the use of resources and determine the amount of resources needed (physical and human) to provide quality services [5].

For the second step, there is a crucial need for predicting the future trend to be ready for facing the challenges and making preparations especially for the administrative levels [39]. Developing some accurate models for predicting the infected ones in the future can help decision-makers suggest appropriate policies. Besides, it is important to assess the effectiveness and impacts of every policy before running [35]. Predicting the ones who will be infected in the future can help to obtain the pressure in the treatment process and plan for avoiding the overloading through the number of nurses, physicians, beds, etc. However, statistics show a high degree of uncertainty in COVID-19 infected people's behavior [47]. Thus, advanced and accurate predictive models are essential [32]. Machine learning (ML) has recently been used for prediction models of COVID-19 with a high-level ability and reliability that most researchers have acclaimed [9, 11, 19, 33]. Although some researchers used ML for other former pandemics, such as H1N1 influenza, Ebola, dengue fever, and swine fever [2, 13, 21, 37], using ML for COVID-19 outbreak prediction is rare in researches and it is not saturated. Since there is no more information about the probable effective parameters on patient arrival, the seasonal auto-regressive integrated moving average (SARIMA) time-series model will be used to predict the number of COVID-19 patients who will arrive in the next few days [6, 18].

For this aim, in this study, we try to first predict the future entry of patients in the case study hospital. Then, according to the system status, we suggest some policies based on different probable scenarios and assess the outcome of each scenario to help the policy's decision-making. So, the current study has six steps as follows:

- Simulating the current process for COVID-19 suspected patient arrival in the case study hospital
- Investigating the outputs for the current process, such as total time, average waiting time, discharged patient ratio, and the cost
- Predicting the patient entry within two months later using machine learning algorithms
- Evaluating and comparing different scenarios
- Running the worst scenario by considering the predicted patient arrival to simulate the pessimistic situation in the future
- Analyzing the system in the pessimistic situation and suggesting some solutions

The rest of this paper is organized as follows. The related literature and studies are reviewed in Sect. 2. Methods and materials are explained in Sect. 3. The hospital case study is described completely in Sect. 4. Section 5 describes simulation models of the current process and scenarios as results. Evaluating the scenarios and simulation of predicted process are discussed in Sect. 6. Finally, the conclusion and future suggestions are discussed in Sect. 7.

2 Literature review

This study includes two research streams. First, the reviewed studies focus on simulation models for different processes especially in hospitals. Then, the prediction models in healthcare services in the literature are investigated. In the first stream, simulation techniques have been used in most papers in hospital processes due to their ease of inception and also less costly in the first stream since they can optimize the process while risks are decreasing (Diaz & Dawson, 2020a). Also, sometimes, simulation models can help the decision-makers to find the best scenario or policy for aiming the optimum outputs in different goals, such as resource allocation, process sequence, and order of activities [45, 46]. For example, Azadeh et al. [4] used a simulation approach for finding the best and optimum policy for maintenance. They first simulated the maintenance system based on historical data and then used a Taguchi method for evaluating different scenarios for the system and calculated output values for each scenario. They evaluated the efficiency of each scenario through a data envelopment analysis (DEA) method and selected an optimal scenario. Pan et al. [26] simulated an ophthalmic specialist outpatient clinic in Singapore. They focused on patient and information flow. Finally, they proposed several improving strategies to decrease turnaround time and analyzed the scenarios via the design of experiment (DOE) method. (Diaz & Dawson, 2020a) used simulation for a COVID-19 resuscitation process in a 47-bed pediatric emergency department over 2 weeks. They considered the arrival of patients, resuscitation, and disposition of patients besides the facilities and staff in their simulation model. They could understand which changes can lead to a more efficient process by comparing the outputs before and after each change. Finally, they concluded the optimal room layout, number of equipment, and staff.

Alban et al. [1] used stochastic process simulation for ICU capacity management during the COVID-19 pandemic.

They simulate an ICU for the patients who are COVI-19 and non-COVID-19. They assessed the increase in COVID-19 patient entry during the pandemic to help a hospital manager for a better management decision. They defined a stochastic queuing model for patient flow. They finally investigated the impact of each decision such as needing to transfer a patient to other hospitals and decreasing bed occupancy rate based on the service level output. Zeinalnezhad et al. [43] used simulation techniques of a heart clinic during the COVID-19 pandemic. According to bottlenecks available in the base process, three scenarios were proposed for improving the waiting time of the process as the target variable. They used timed colored Petri nets for workflow simulation. Finally, they compared the three scenarios waiting time output and chose the best strategy. Melman et al. [24] tried to balance the hospital resources during the COVID-19 pandemic while using the discrete-event simulation model. They used data of COVID-19 patient flow for a hospital in the UK. They proposed three different resource allocation scenarios and evaluated them with simulation run outputs.

Recently, the papers tried to combine some other techniques besides the simulation tools to improve the outputs and close them to reality. For example, Kovalchuk et al. [20] used simulation for patient flow in acute coronary syndrome (ACS) unit. They also combined some machine learning approaches to identify each patient path class. So, classifying the patient, they could improve the length of stay of patients. Ordu et al. [25] proposed the integrated forecasting-simulation–optimization approach in a hospital to help the managers for their resource allocation problem. So, they first predicted the professions demand in hospital, then in the second step, simulated the patient journey in a case study hospital, and finally, developed an optimization model for bed and staff allocation based on the outputs of the two previous steps. They expressed that their proposed model could be as a decision support system. Sasanfar et al. [36] simulated the emergency department (ED) of a hospital to find the best resource allocation policy. They simulated an ED in a case study hospital in Iran and could decrease waiting times 23.1% (3.31 min) and 81.7% (10.58 min) for internal and emergent patients, respectively. Pereira et al. [29] evaluated a public hospital efficiency while using a DEA with simulation. They used a Monte Carlo method to model the hospital supply chain and defined several providers and then evaluated them using the DEA to find the target area. Finally, they could justify the target scenario for their case study hospital. Teberga Campos et al. [10] tried to simulate the COVID-19 infected ones' pattern to inform hospitals and carried out their simulation results in a case study. So, they could improve patient waiting time, movement intensity, length of stay (LOS), and either adoption rate.

In the second stream, some studies used the prediction models such as the time-series model and regression models. In this stream, Tomar & Gupta [39] tried to predict the COVID-19 spread in India and used long short-term memory (LSTM) and curve fitting to predict the COVID-19 cases in India within 30 days based on the available data. The effectiveness and related results for solutions (e.g., isolation and lockdown) were investigated. Heo et al. [15] tried to predict and monitor military hospitals in South Korea. They aimed to select the important patients due to medical resource shortage. An application gathers some information about age,body temperature; pre-disease physical status; history of cardiovascular disease; hypertension; visit to a region with an outbreak; and symptoms of chills, feverishness, dyspnea, and lethargy. So, important patients were selected through prediction models. Ardabili et al. [3] used machine learning techniques to predict COVID-19 outbreak and used several machine learning and mathematical model (e.g., logistic, linear, logarithmic, quadratic, cubic, compound, power, and exponential) for predicting COVID-19 outbreak and compared them. Multi-layered perceptron (MLP), adaptive network-based fuzzy inference system (ANFIS), and finally, time series were employed for prediction as ML techniques. Their results showed that ML models have fewer errors for prediction and are more powerful.

Besides, some studies focused on new COVID-19 cases in the future by predicting using the risk factors or the arrivals history. They used different machine learning techniques to predict how many infected cases may occur in future days [16],G. [45, 46]. Several recent papers tried to predict new cases through time series. For this aim, Zeroual et al. [44] forecasted the COVID-19 patients based on time series. They tried to predict the new cases infected in short term to be used by resource managers. They used five methods for time-series forecasting considering recent historical data of infected and recovered cases in the USA, China, France, Spain, and Italy. Finally, they compared the error metrics of each model such as RMSE and MAE. Maleki et al. [22] also predicted the COVID-19 new confirmed and recovered cases through time-series modeling. To this aim, they used autoregressive models in time series as TP-SMN (two pieces-scale mixture normal distributions). Since the infected cases trend and then their entrance to a special hospital includes uncertainty, some researchers tried to forecast new cases using time-series models while uncertainty was considered. In this regard, Ye & Yang [42] predicted the future cases of COVID-19 in China through time series in an uncertain environment. Based on their result, the prediction accuracy was greater compared to the classical time series. In addition to the time-series models, other prediction models were used for the new case prediction, such as the linear regression model. For instance, Rath et al. [31] used a multiple linear regression model for the new active cases of COVID-19 prediction based on a WHO data set. They compared the results of their method with a simple linear regression. Roy et al. [34] used the additive regression model for infected cases all over the world based on global data for each country and compared them to help the economic evaluation for countries based on their future infected people. However, support vector machine (SVM) models were also used for forecasting. Parbat & Chakraborty [27] used an SVM model for COVID-19 case prediction based on the data of the recent 2 months. They could develop a prediction model with 97% accuracy for all cases of infection, deaths, and recoveries.

Based on the reviewed literature, no study has used the integrated simulation, DEA, and time-series models, especially in the COVID-19 patient process. For instance, Diaz & Dawson (14, Alban et al. [1], Melman et al. [24], and Teberga Campos et al. [10] only used simulation tools for optimal layout when they did not have any approach for deciding optimality. Besides, Zeinalnezhad et al. [43] and Pan et al. [26] first simulated their case study processes and then used Petri nets and DOE for analyzing the identified scenarios. In the literature reviewed, only Azadeh et al. [4] and Pereira et al. [29] tried to evaluate the scenarios by the DEA based on the simulation outputs that we carry out in our study. There was no study that used a time-series prediction model for the simulation model which predicted the input. So, this study has the novelty in its methodology and the case study process, which is the most complete patient flow for COVID-19-suspected patients. The main contributions of this research can be summarized and highlighted below:

- Developing a simulation model of the COVID-19 patient flow in a hospital as a real-case study
- Predicting the different categories of patient (i.e., outpatient, emergency, and inpatient) arrival using the time-series model
- Proposing various scenarios based on different levels of input variables using the Taguchi method
- Evaluating the proposed scenarios based on their input and simulation output using the DEA method
- Predicting the bottlenecks of the patient flow process by simulating the worst scenario with predicted patient arrivals
- Suggesting some public health policies according to different scenarios and assessing the outcome of each scenario for the policy's decision-makers

3 Materials and methods

A brief overview of the research methodology framework is shown in Fig. 1. In the first stage, the data about the COVID-19 patient flow are collected. Then, the simulation model is designed based on the distribution functions for each parameter. In the base run of the simulation model, the current bottlenecks of the process are identified.

In the second stage, since the COVID-19 patient journey in a hospital during a pandemic is a complex process and consists of different input parameters affecting the process output, different values can be considered for each of these parameters for each unit. Furthermore, various combinations of these parameters lead to several scenarios. For defining all possible combinations of the input variables, the Taguchi method is used. For each scenario proposed by this method, the simulation model will run and the output variables will measure.

In the third stage, the best scenario should be identified based on the output comparison. It is less probable that a scenario has the best value for all three outputs. If there is a scenario that has the best outputs comparing with the other scenarios, it will be considered as the best scenario. Otherwise, for evaluating the scenarios, the DEA method is considered. Finally, in the last stage, the patient entry in the future will predict and the predicted values will be the input of the best scenario simulation model. The future bottlenecks will be recognized to inform the hospital managers. The tools and techniques used in this study will be explained in detail.

3.1 Arena simulation

An Arena is an application software with high modeling capabilities and a powerful simulation tool that allow users to create and test a simulation model, while also having an easy-to-use interface. The Arena can simulate a discrete event system (DES) that accelerates the analysis of the behavior of a process or system over time. So, before we get into the practical implementation of a business process, it is best to first model and evaluate it so that we can better decide on some changes in that process and improve it [12]. Furthermore, before it becomes costly and productive, we realize the best of that business process and make the best decisions. The capabilities of Arena are as follows:

- Showing a graphical representation of process flows for even the most complex business processes
- Monitoring, analyzing, and better understanding the behavior of workflows
- Guessing more accurately the efficiency, response time, and bottlenecks of a new system or design
- Evaluating the impact of error rates
- Changing or improving how the system is configured and tasks are performed
- Testing different ways to find the best solution for a topic

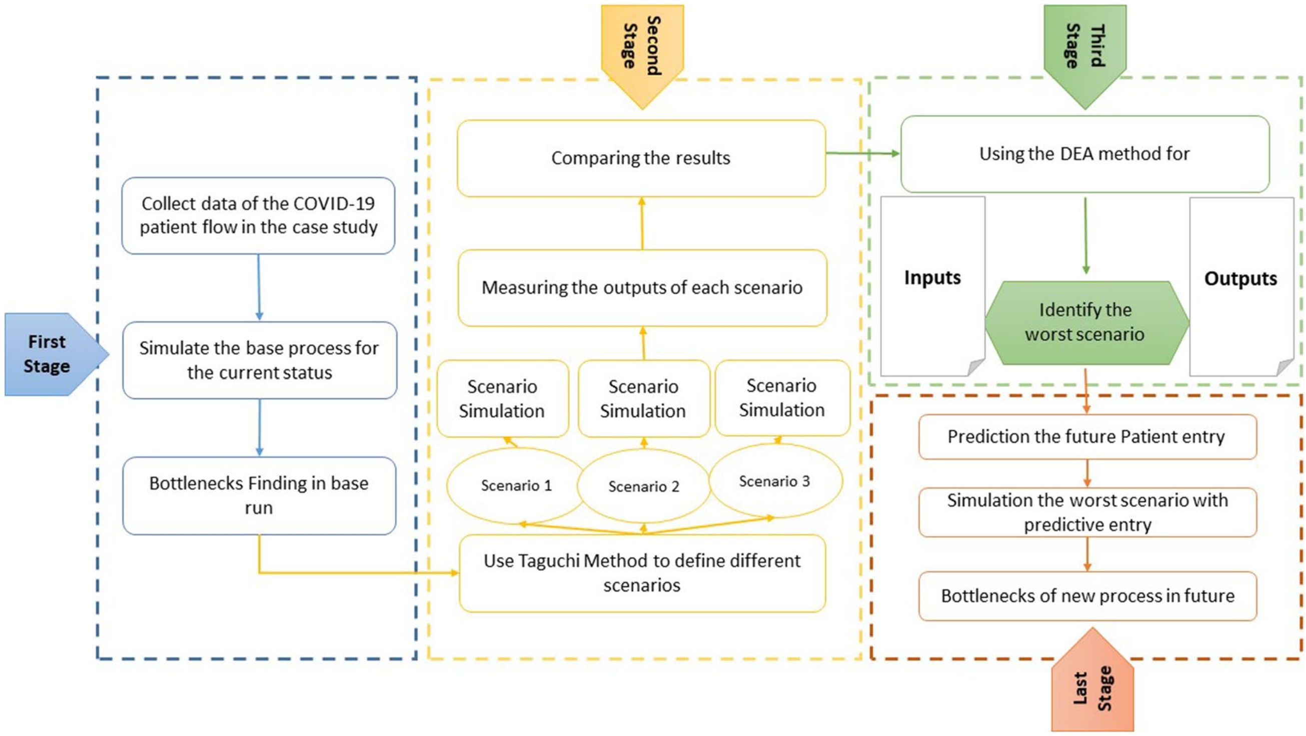

Fig. 1 Research methodology framework

- Showing the results graphically and numerically to increase the acceptance and understanding of decisions [23]

3.2 Data envelopment analysis (DEA)

Data envelopment analysis (DEA) is a mathematical planning model for evaluating the performance of decision-making units (DMUs) that have multiple inputs and multiple outputs. Charnes et al. (1978) proposed this method as a CCR model by the first letter of their names for calculating the efficiency of each DMU by solving a nonlinear mathematical model as Eqs. (1)-(3):

$$e0 = \max \sum_r u_r y_{ro} / \sum_i v_i x_{io} \tag{1}$$

s.t.

$$\sum_r u_r y_{rj} - \sum_i v_i x_{ij} \leq 0 \tag{2}$$

$$u_r, v_i \geq \varepsilon \tag{3}$$

where x_{io} means the value of input i, y_{ro}, the value of output r, v_i, the weight of input i, u_r, the weight of output r, number j of DMUs, and ε is a positive parameter.

However, this nonlinear programming can be transferred into linear as Eqs. (4)-(8) (Cook and Seiford,2009):

$$e0 = \max \sum_r \mu_r y_{ro} \tag{4}$$

$$\text{s.t.} \sum_i v_i x_{io} = 1 \tag{5}$$

$$\sum_r \mu_r y_{rj} - \sum_i v_i x_{ij} \leq 0 \tag{6}$$

$$\mu_r, v_i \geq \varepsilon \tag{7}$$

$$\text{where} \mu_r = t u_r, v_i = t v_i, \text{and} t = \left(\sum_i v_i x_{io}\right)^{-1}. \tag{8}$$

In this method, an efficient boundary curve is created from a series of points that are determined by linear programming. The linear programming method determines whether the DMU is on the edge of efficiency or outside it. Thus, efficient and inefficient units are separated from each other based on Fig. 2.

The parameters considered inputs are as follows:

- Number of physicians in the emergency COVID-19 special line
- Number of nurses in the emergency COVID-19 special line
- Number of physicians in ICU
- Number of nurses in ICU

- Number of physicians in CCU
- Number of nurses in CCU
- Number of radiologists in a CT-scan unit
- Number of service providers in the laboratory
- Number of beds in ICU
- Number of beds in CCU
- Number of beds in the emergency COVID-19 special line

Also, the outputs include as follows:

- Total time: the time through the entire patient flow
- Average patient waiting time: the average of the waiting times through all the processes in all activities of all departments
- Discharged patient ratio: the fraction number of discharged patients to the patients who entered the hospital
- Cost: the cost of inputs for the hospital as the service provider

3.3 Taguchi design

To examine the different values of input factors affecting the entire system, for every eleven input parameters, three levels were chosen and given in Table 1. In this table, each column indicated three levels and inputs are in rows. These levels are defined based on the minimum, medium, and maximum values that can be in the system based on the hospital capacity and budget. If we want to examine a full factorial experiment for eleven inputs having three levels, the number of required requirements will be 3^{11} (i.e., 177,147) experiments. However, the Taguchi method reduces the number of experiments to fewer experiments that can be investigated easier (Davim, 2003). The steps of this method include the following: (1) choose control factors, (2) choose suitable levels for factors, (3) choose an orthogonal array, which is appropriate for the control factors, (4) carry out the experiments, and (5) analyze the experiments and find the best combination for levels of factors. Based on orthogonal arrays of Taguchi, for eleven of three-level factors, 27 scenarios are suggested as given in Table 2. While the value for each factor is 1 it means the minimum value, 2 the medium value, and three the maximum value.

3.4 Seasonal auto-regressive integrated moving average (SARIMA)

For the prediction of patient arrival based on their historical arrival, the SARIMA(p, d, q),(P, D, Q)m model is used. This model is used because of its simplification and appropriation for predicting the patient arrival. Also, it is used when there are only two columns in the dataset (i.e., data and event frequency) in non-linear cases, which have seasonal behavior [28]. However, other time-series model which is more complicated (i.e., LSTM) is used one there are other exogenous features impact on the event frequency [41]. Since the data available in the case study includes the information of date and patient arrival and other information (e.g., the region population, the infection rate, the connection frequency, and some detailed features), we will use the SARIMA model for time-series prediction. This model includes several parameters that can be tuned to achieve optimal performance. These parameters are trend elements and seasonal elements as follows:

Trend elements:
pTrend auto-regression order.
dTrend difference order.
qTrend moving average order.
Seasonal elements:
PSeasonal autoregressive order.
DSeasonal difference order.
QSeasonal moving average order.
mNumber of time steps for a single seasonal period.

To get the best prediction, the values of SARIMA(p, d, q),(P, D, Q)m should be optimized. For this aim, we used "grid search" to iteratively explore different combinations of these parameters. The evaluation metric used for the grid search is the Akaike information criterion (AIC) value. The AIC measures how well a model fits the data while considering the overall complexity of the model [30].

The augmented Dickey-Fuller (ADF) test is used for checking stationary. The ADF approach is essentially a statistical significance test that compares the p-value with the critical values and does the hypothesis testing. Using this test, we can determine whether the processed data are stationary or not with different levels of confidence. If p-value > 0.05, then the zero hypothesis with the stationary will reject data [38].

The metric used for evaluation is the root mean squared error (RMSE) as Eq. (9), where y_t which is the actual patient entry on the date (t), $y_t^{predicted}$ is the predicted value of the patient entry on the date (t), and n is the number of test dates.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(y_t - y_t^{predicted}\right)^2}{n}} \tag{9}$$

4 Case study

4.1 Data collection

This study was performed at a hospital in Iran. This hospital provides services to COVID-19 patients in four normal

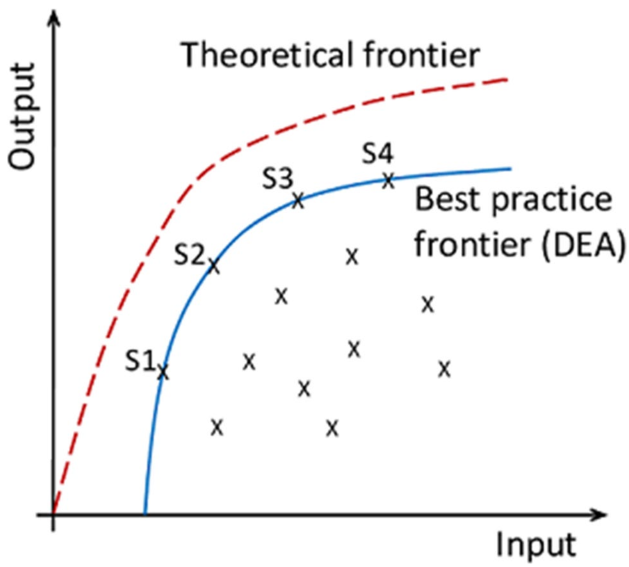

Fig. 2 DEA separation frontier

units with 104 beds and three intensive care units with 70 beds. The laboratory and CT scan units of this hospital are also active 24 h a day of outpatients, emergencies, and inpatients who have been hospitalized before for other reasons. The emergency unit of this hospital also has 10 beds for the severe symptom patients who stay in these beds until the inpatient unit bed becomes empty.

The data for patient entry (i.e., outpatients, emergencies, and inpatients), symptoms (i.e., no symptoms, moderate, and severe), laboratory test results and CT-scan reports, and length of stay in normal and intensive units for all entered cases are collected from 3 May to 5 October 2020. Figure 3 shows the data related to patients collected from the case study. However, the information about the available beds, the human resources in each unit, the waiting times of patient in every stage, and the average time of each activity is obtained from the hospital information system (HIS), and the managers for each category of patients (i.e., no symptoms, moderate, and severe) will be explained in five different flows in the next section as the "base process."

4.2 Base process

In the patient flow of the COVID-19-suspected ones in the case study hospital, three groups of patients enter the process. These three categories include outpatients, emergencies, and inpatients. The definition of each category has been identified:

- Outpatients are patients who go directly to a lab or CT scan based on their suspicion or some of the symptoms of COVID-19 disease visibility.
- Emergencies are patients who have called the emergency services due to the visibility of some symptoms and are delivered by ambulance to the emergency unit of the hospital.
- Inpatients include patients who have already been hospitalized for other reasons before and need to be tested for COVID-19 services due to the visibility of some symptoms or even before their surgery operations.

Outpatients take three steps based on the severity of their first symptoms. If they have no symptoms, they usually go to the lab. If they have moderate symptoms, they will have a CT scan for a chest x-ray. It is rare for outpatients to have severe symptoms, but if so, they go to the triage of the emergency unit of the COVID-19 line. Emergencies are also delivered by ambulance to the triage of the hospital's emergency COVID-19 line. In this case, with the diagnosis of emergency unit triage, if their symptoms are not severe, they are sent to the laboratory. If they have moderate symptoms, they will have a CT scan for a chest x-ray. Most patients referred to the emergency department have severe symptoms, in which case, while staying on one of the beds in the

Table 1 Input variable levels

Input number	Input variable	Level 1	Level 2	Level 3
1	Number of physicians in emergency COVID-19 special line	2	4	6
2	Number of nurses in emergency COVID-19 special line	5	6	7
3	Number of physicians in ICU	2	4	6
4	Number of nurses in ICU	16	20	24
5	Number of physicians in CCU	2	4	6
6	Number of nurses in CCU	6	7	8
7	Number of the service providers in CT-scan unit	4	8	12
8	Number of the service providers in Laboratory	14	18	22
9	Number of beds in ICU	32	52	72
10	Number of beds in CCU	10	18	38
11	Number of beds in emergency COVID-19 special line	5	10	15

Table 2 Taguchi OA L27

Scenario	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8	Input 9	Input 10	Input 11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2
3	1	1	1	1	3	3	3	3	3	3	3
4	1	2	2	2	1	1	1	2	2	2	3
5	1	2	2	2	2	2	2	3	3	3	1
6	1	2	2	2	3	3	3	1	1	1	2
7	1	3	3	3	1	1	1	3	3	3	2
8	1	3	3	3	2	2	2	1	1	1	3
9	1	3	3	3	3	3	3	2	2	2	1
10	2	1	2	3	1	2	3	1	2	3	1
11	2	1	2	3	2	3	1	2	3	1	2
12	2	1	2	3	3	1	2	3	1	2	3
13	2	2	3	1	1	2	3	2	3	1	3
14	2	2	3	1	2	3	1	3	1	2	1
15	2	2	3	1	3	1	2	1	2	3	2
16	2	3	1	2	1	2	3	3	1	2	2
17	2	3	1	2	2	3	1	1	2	3	3
18	2	3	1	2	3	1	2	2	3	1	1
19	3	1	3	2	1	3	2	1	3	2	1
20	3	1	3	2	2	1	3	2	1	3	2
21	3	1	3	2	3	2	1	3	2	1	3
22	3	2	1	3	1	3	2	2	1	3	3
23	3	2	1	3	2	1	3	3	2	1	1
24	3	2	1	3	3	2	1	1	3	2	2
25	3	3	2	1	1	3	2	3	2	1	2
26	3	3	2	1	2	1	3	1	3	2	3
27	3	3	2	1	3	2	1	2	1	3	1

hospital's emergency COVID-19 line, they go to the laboratory and CT-scan unit at the same time until their admission will be done. Inpatients, as needed or sometimes observing the symptoms of the disease, go to the laboratory and radiology to ensure that they are not infected with COVID-19 to continue their treatment in another disease that they have. In this case, they are sometimes confirmed to have COVID-19, which must be treated at the same time as their underlying disease and coronary heart disease. If inpatients have no symptoms but need a COVID-19 test, they are sent to a laboratory first. If they have moderate symptoms, they first go to the radiologist and then to the laboratory. In this case, these patients are in their non-coronary wards and are transferred for tests and returned to their wards. However, if they have severe symptoms, they are immediately isolated in their ward to prevent transmission to other patients in the non-coronary wards of the hospital and sent to the laboratory and radiology at the same time.

In general, in COVID-19 disease, observing the symptoms is the priority in making a patient decision. After that, the CT-scan results and finally the test result determine whether or not a patient is infected. There are five different flows of patients, which are explained as follows:

1. First flow: Outpatients and emergencies with no symptoms wait in the laboratory and will be tested after a while. The test result is also prepared after an average of 6 h. Ten percent of the time, it is necessary to repeat the test. If there is no need to repeat the test, based on the test results, then patients follow three ways: (I) If their test is negative, they leave the hospital, (II). If the test is positive, some patients leave the hospital on their own to go home, some patients go home based on physician's orders to be quarantined, and some patients prefer to go to another hospital, and (III) the rest of the patients go to the radiology for CT scan. In radiology, patients are waiting in a queue, and after an average of 25 min, while no waiting for emergency patients, a CT scan is performed. The CT-scan result is ready after 2 h on average. Based on the CT-scan result along with the test result, the following conditions occur:

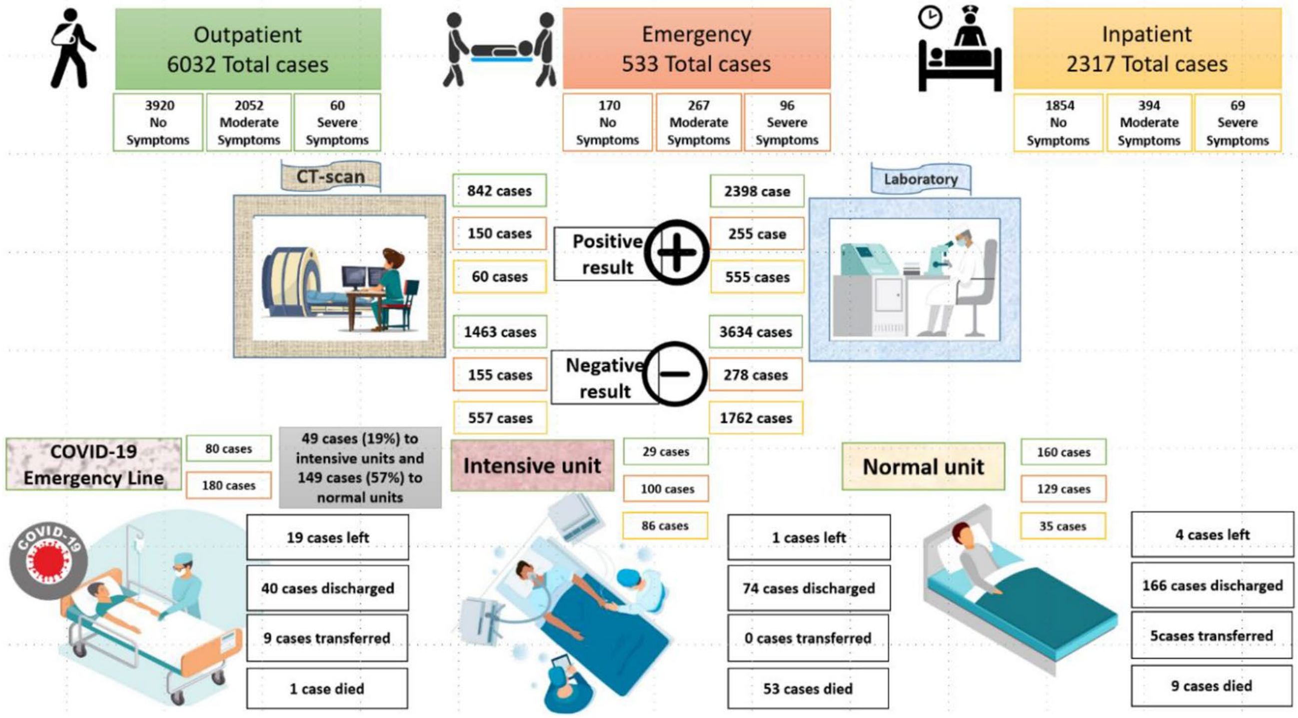

Fig. 3 Data description of the case study

• If the CT scan is positive, the test result is positive, and the patient has no symptoms, he/she goes home based on physician's orders; however, he/she must be quarantined at home and rest until complete recovery.

• If the CT-scan result is negative while the test result is positive and the patient has no specific symptoms, he/she goes home for quarantine based on the physician's orders.

2. Second flow: Outpatients and emergencies with moderate symptoms first go to the radiology and then go to the laboratory. Based on the CT-scan result along with the test result, in all four conditions, patient should wait for the COVID-19-unit admission.
3. Third flow: Outpatients and emergencies with severe symptoms should first go to the triage of the COVID-19 emergency department of the hospital. In this case, because the patient's symptoms are severe, they stay in the COVID-19 emergency line and go to the radiology and laboratory at the same time. After preparing the results, regardless of the results, they remain in the COVID-19 emergency line.
4. Fourth flow: Inpatients with no symptoms are transferred to a laboratory. The test result is also prepared after an average of 6 h. In 10% of the time, it is necessary to repeat the test. If there is no need to repeat the test, patients take three steps based on the test results: (I) If their test is negative, they return to their unit and continue their previous treatment, (II) If their test is positive, they will be taken to the radiology for CT scan. After an average of 5 min of waiting for their CT scan to be done, the CT-scan answer is ready after 2 h on average. Based on the CT scan result along with the test result, the following conditions occur:

• If the CT scan and the test are positive and the patient has no symptoms, the patient should be isolated in his/her unit.

• If the CT scan result is negative while the test result is positive and the patient has no specific symptoms, the patient should remain in his/her unit.

5. Fifth flow: Inpatients with moderate and severe symptoms are first referred to radiology and then transferred to the laboratory. Due to the moderate and severe symptoms, these patients are isolated in their unit, regardless of their results, and are waiting for COVID-29-unit admission.

In the hospital's COVID-19 units, patients are separated into normal units or intensive care units based on clinical diagnosis. So that if the patient has a clinical disease before, he/she will be admitted to intensive care units and otherwise to normal units. In intensive care units, after undergoing the relevant treatments based on the physician's order, patients are first transferred to normal units and complete their treatment there. Then, patients will

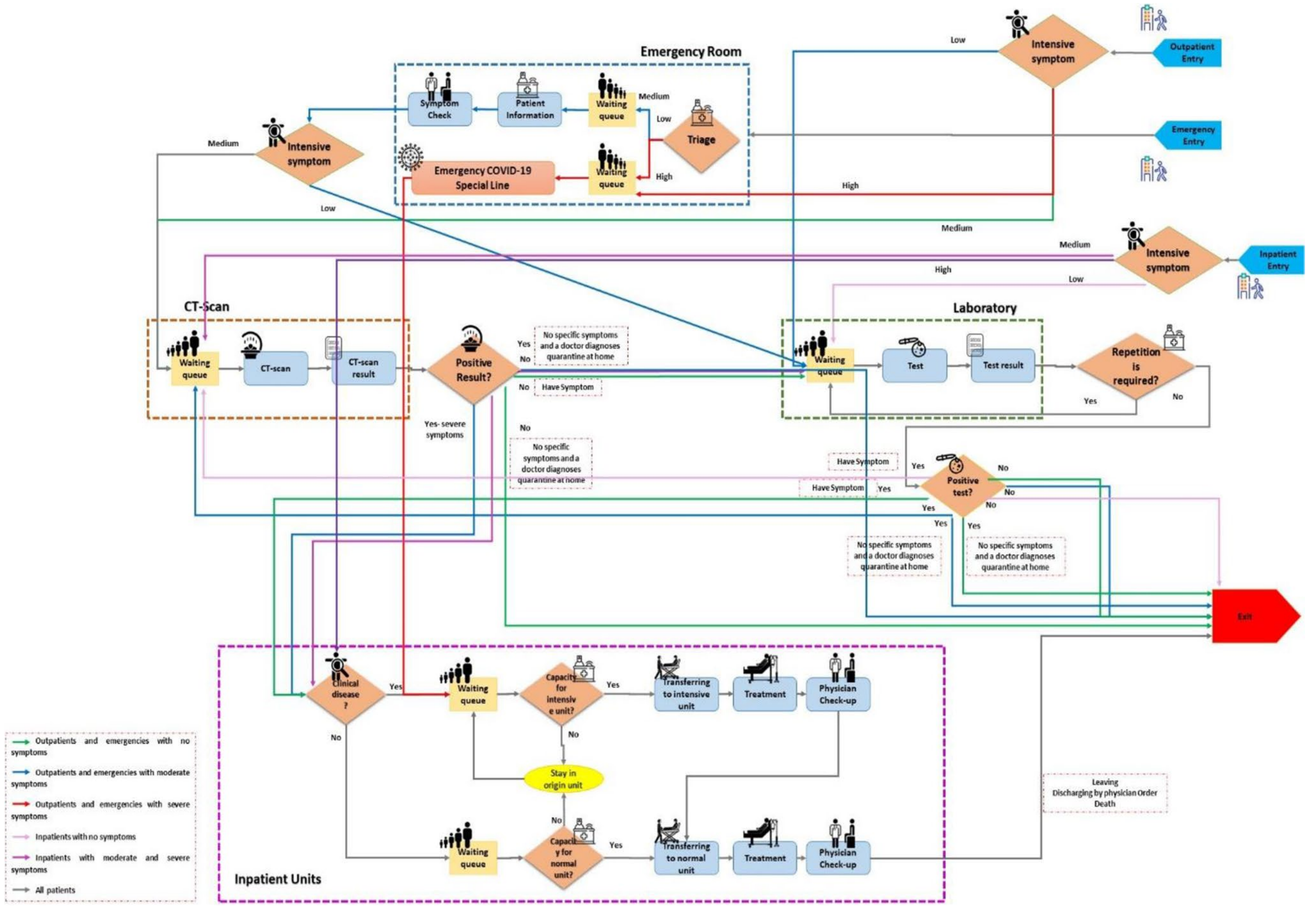

Fig. 4 COVID-19 patients monitoring process

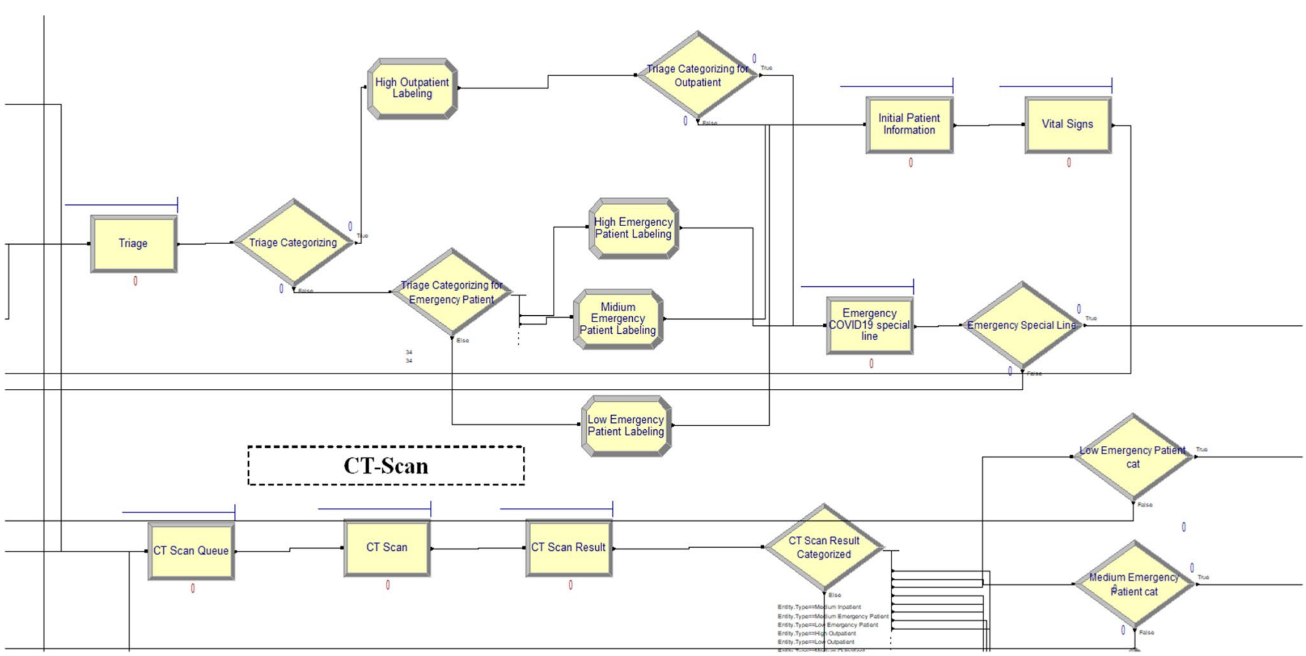

Fig. 5 Part of the simulation model of the COVID-19 patient flow

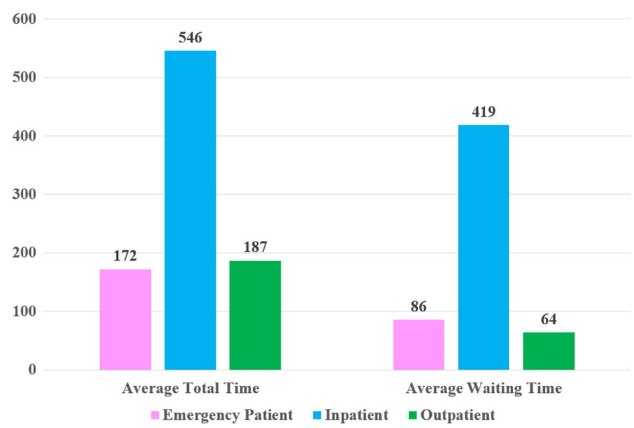

Fig. 6 Comparison of the average total time and waiting time of each category

leave the system in three modes. Either they are discharged based on the physician's order due to complete recovery or continuing treatment, or they unfortunately die, or are transferred to another hospital to continue their treatment. The process explained above is depicted in Fig. 4 in five flows described and a general flow, which is common for all categories of patients.

5 Results

5.1 Base patient flow simulation

In this section, the base patient flow is simulated. For simulating the model of the present study, the data collected from the case study hospital, and then in one of the Arena software add-ins called input analyzer, the input data are converted into distributed functions and these functions are used in the process. Also, the distribution functions of the variables are explained in Section A in the Supplementary Materials.

In the process of discrete event simulation performed with Arena software, different modules are used. In the simulation of this research, the "Create" module is used to create the entities of the studied process (i.e., patients). Patients are divided into three categories, namely, outpatients, emergencies, and inpatients when each of them has its flow. The "Process" module is used to perform various activities during the process, for example, activities (e.g., laboratory test, CT scan, and treatment). The "Hold" module is used by different entities to wait in the queue at different workstations. In some parts of the simulation, the "Decide" module is used to divide the different paths of the entities. Another module used in this model is the "Assign" module, which is used to separate entities in different sections for accurate monitoring. The "Record" module is used to record data of different parts of the process. All entities are also removed from the process by the "Dispose" module. Also, Fig. 5 is part of the simulation model of the patient flow in the hospital. However, the complete simulation model is shown in Section B in the Supplementary Materials.

According to the outputs of the base patient flow model, the total time of outpatients in the system is from 56 to 2998 min. These patients vary in the length of time and stay in the system according to their various conditions. However, these patients are in the system for 187 min on average. Also, the waiting time of outpatients in the system is from 42 to 258 min, with an average of 64 min. Emergencies are in the system from 85 to 12,129 min and an average of 172 min. Also, their average

Table 3 Simulation outputs of scenarios

Scenario	Total time (minutes)	Average patient waiting time (minutes)	Discharged patient ratio	Cost (,million toman)	Scenario	Total time (minutes)	Average patient waiting time (minutes)	Discharged patient ratio	Cost (million toman)
1 (base)	172	86	0.919	21,034	15	166	81	0.918	39,550
2	165	81	0.922	35,270	16	168	85	0.919	25,346
3	124	59	0.918	51,906	17	149	75	0.919	41,838
4	164	81	0.921	37,774	18	170	86	0.922	41,350
5	148	76	0.919	46,910	19	168	84	0.920	42,990
6	171	87	0.920	25,486	20	167	85	0.921	29,602
7	131	64	0.915	49,414	21	167	84	0.919	36,554
8	171	86	0.920	26,390	22	164	79	0.921	31,890
9	170	87	0.921	33,126	23	169	86	0.922	31,402
10	167	84	0.923	36,902	24	135	68	0.918	45,494
11	165	81	0.917	43,854	25	168	86	0.919	43,834
12	165	82	0.918	28,066	26	129	62	0.919	47,986
13	168	85	0.923	46,346	27	170	87	0.920	27,038
14	172	87	0.922	22,998					

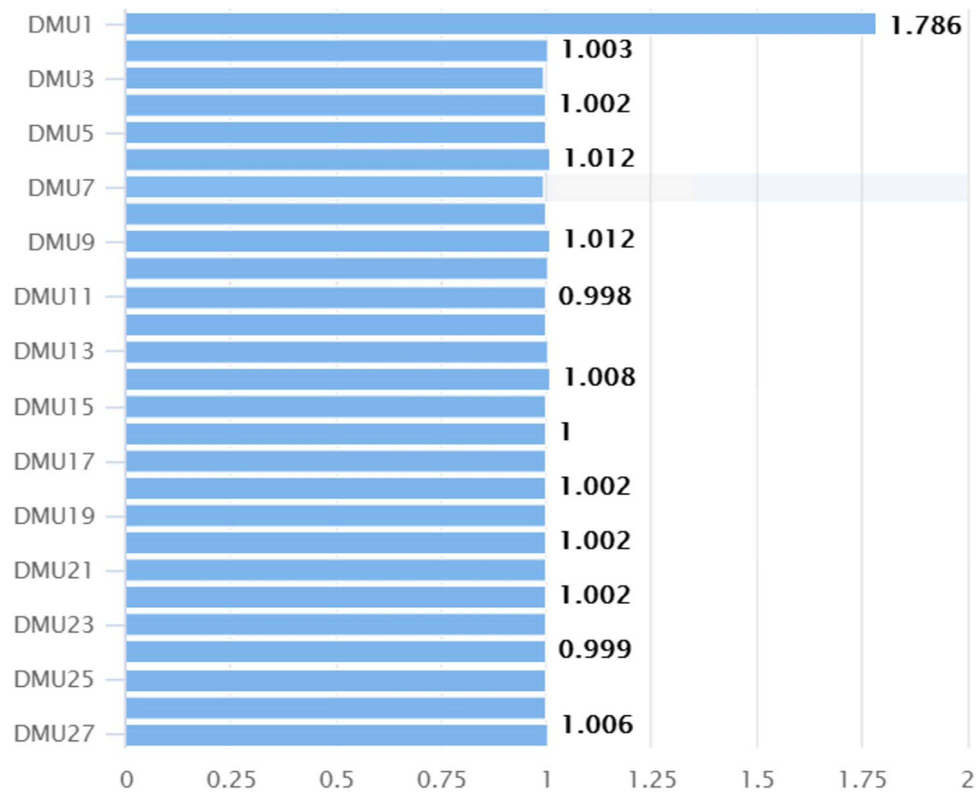

Fig. 7 DEA results of the scenarios (i.e., DMUs)

waiting time is 86 min in the range of 61 to 364 min. Inpatients also stay in the system from 25 to 8753 min, with an average of 546 min remaining in the system. The average waiting time for inpatients is 419 when the range of it was between 13 and 293 min. The comparison of the average total time and waiting time of each category is shown in Fig. 6.

6 Scenarios patient flow simulation

As different 27 scenarios are defined in Table 2 with three levels described in Table 1, based on the simulation model running, the outputs of the model are obtained. The outputs of the model include the following:

- Total time: the time through all the processes in all activities of all departments
- Average waiting time: the average of the waiting times through all the processes in all activities of all departments
- Discharged patient ratio: the fraction number of discharged patients to all patients who arrived in the hospital
- Cost: the cost of each scenario for the hospital provider

Total time, average patient waiting time, and discharged patient ratio are extracted from the simulation run in each scenario. The last output, which is the cost, is calculated based on the average cost of all inputs needed for each scenario based on the expert opinion and wages. So, the outputs of 27 scenarios for all types of patients are shown in Table 3.

7 Discussion

7.1 Scenario evaluation with the DEA

Now, when all the inputs and outputs are specified for all scenarios, the DEA can be done with eleven inputs and

Table 4 DEA results of scenarios

Scenario no	DEA result	Rank	Scenario no	DEA result	Rank
1	1.786	1	15	0.999	10
2	1.003	6	16	1	9
3	0.996	12	17	1	9
4	1.002	7	18	1.002	7
5	0.999	10	19	1	9
6	1.012	2	20	1.002	7
7	0.995	13	21	1	9
8	1.001	8	22	1.002	7
9	1.012	2	23	1.002	7
10	1.004	5	24	0.999	10
11	0.998	11	25	1	9
12	0.998	11	26	1	9
13	1.004	5	27	1.006	4
14	1.008	3			

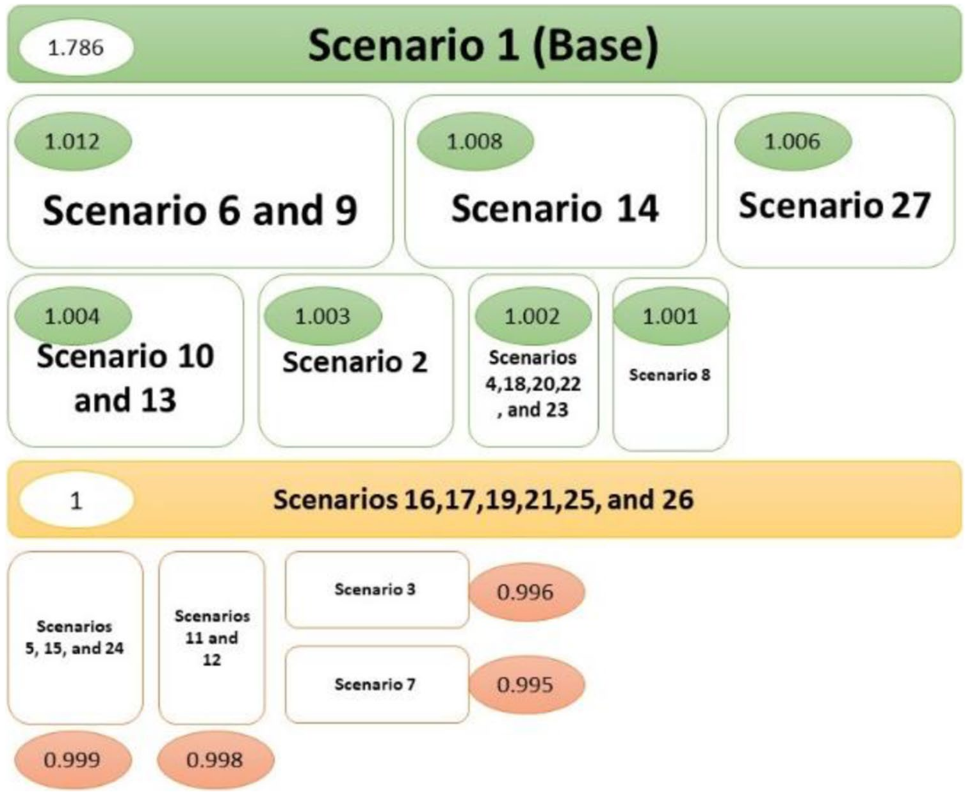

Fig. 8 DEA results of the same scenarios (i.e., DMUs)

four outputs to evaluate the scenarios and compare them as shown in Fig. 7. Also, Table 4 is depicted the efficiency score of each scenario and its rank.

As can be seen in Fig. 7, the least-efficient scenarios are scenarios of 3, 7, 11, and 12, which are less or equal to 0.998 efficient scores. Also, the scenarios of 5, 15, and 24 are 0.999 efficient. The base model has the best efficient score. Other scenarios, which have more input variables and sources, are not more efficient. The reason for this is because of the time need for the result of the CT scan and laboratory. So that the added resources (e.g., physicians, nurses, and beds) do not affect the total time and the waiting time amazingly. Consequently, the discharged patient ratio is not changed really. So, with more resources that lead to more

Table 5 Output effects of scenarios

Scenario	Total time (minutes)	Average patient waiting time (minutes)	Discharged patient ratio	Cost (million toman)	Scenario	Total time (minutes)	Average patient waiting time (minutes)	Discharged patient ratio	Cost (million toman)
1 (base)	172	86	0.919	21,034	15	6 ↓	5 ↓	0.001 ↓	18,516 ↑
2	7 ↓	5 ↓	0.003 ↑	14,236 ↑	16	4 ↓	1 ↓	=	4312 ↑
3	48 ↓	27 ↓	0.001 ↓	30,872 ↑	17	23 ↓	11 ↓	=	20,804 ↑
4	8 ↓	5 ↓	0.002 ↑	16,740 ↑	18	2 ↓	=	0.003 ↑	20,316 ↑
5	24 ↓	10 ↓	=	25,876 ↑	19	4 ↓	2 ↓	0.001 ↑	21,956 ↑
6	1 ↓	1 ↑	0.001 ↑	4452 ↑	20	5 ↓	1 ↓	0.002 ↑	8568 ↑
7	41 ↓	22 ↓	0.004 ↓	28,380 ↑	21	5 ↓	2 ↓	=	15,520 ↑
8	1 ↓	=	0.001 ↑	5356 ↑	22	8 ↓	7 ↓	0002 ↑	10,856 ↑
9	2 ↓	1 ↑	0.002 ↑	12,092 ↑	23	3 ↓	=	0.003 ↑	10,368 ↑
10	5 ↓	2 ↓	0.004 ↑	15,868 ↑	24	37 ↓	18 ↓	0.001 ↓	24,460 ↑
11	7 ↓	5 ↓	0.002 ↓	22,820 ↑	25	4 ↓	=	=	22,800 ↑
12	7 ↓	4 ↓	0.001 ↓	7032 ↑	26	43 ↓	24 ↓	=	26,952 ↑
13	4 ↓	1 ↓	0.004 ↑	25,312 ↑	27	2 ↓	1 ↑	0.001 ↑	6004 ↑
14	=	1 ↑	0.003 ↑	1964 ↑					

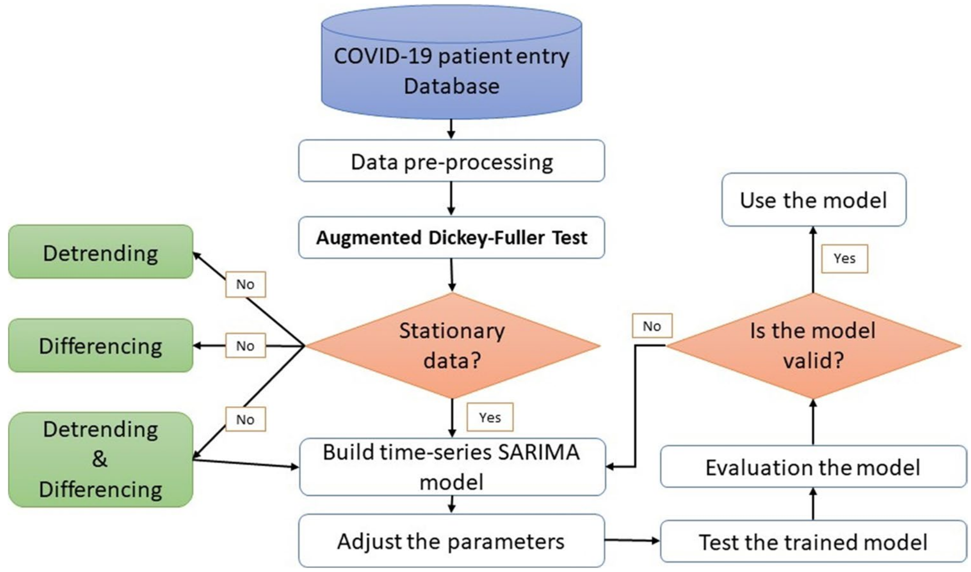

Fig. 9 Time-series prediction steps

costs in all scenarios and not better outputs, the efficient scores of all scenarios are less than the base scenario.

However, in Fig. 8, the scenarios with the same efficiency score are shown. Although they alter the outputs with different values, entirely, their performance is the same. The decision-makers should select among them considering their priority of total time, waiting time, and discharge ratio. Besides, the increase or decrease amount of each output based on all scenarios in comparison to the base scenario is calculated in Table 5.

Based on Tables 4 and 5, we have to find the worst scenario. For finding the worst scenario, we do not focus on the DEA efficiency score and consider the output values, too. So, the scenario of number 14 can be considered as the worst based on its effects on outputs although it was efficient. However, it did not change total time, increased the average waiting time with more costs while improved the discharged patient ratio.

7.2 Time-series prediction

This study proposes a time-series framework for three categories of patient entries (i.e., outpatients, emergencies, and inpatients). The framework of the prediction models is illustrated in Fig. 9. First, the raw data are preprocessed and checked for the stationary test. Then, the data are divided into train and test sets using the train-test-split module

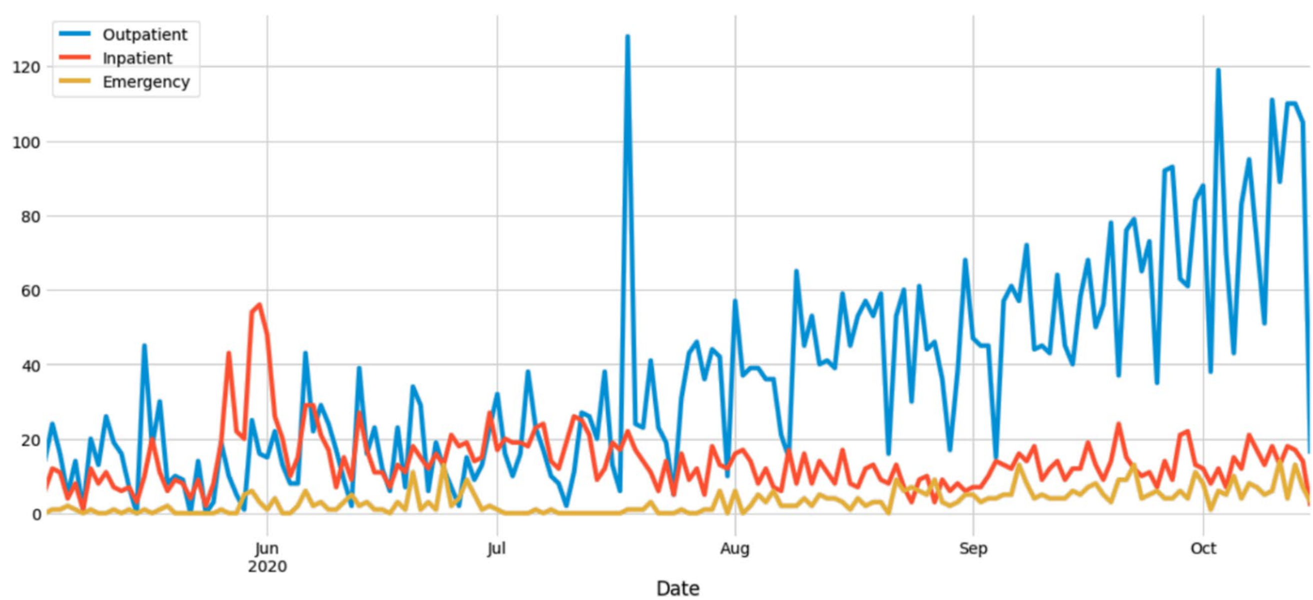

Fig. 10 Daily patient entry of emergency patients, outpatients, and inpatients

on "Sklearn. Model_selection" package in Python programming language (70% train and 30% test). Finally, the SARIMA model is constructed. Then, the patient entry is predicted for 30 days later. The accuracy of the model will be verified by comparing the measured data with the real data via the RMSE.

For data preprocessing, the missing values for someday patient arrivals and the average values are considered. For checking the stationary of data, the rolling mean and standard deviation of each column are calculated within 6 days for mean and 24 days in standard deviation. Daily patient entry of emergencies, outpatients, and inpatients are depicted in Fig. 10 based on collected data. As can be seen in Fig. 11, we see that the rolling mean itself has a trend component even though the rolling standard deviation is fairly constant with time. For our time series to be stationary, we need to ensure that both the rolling statistics (mean and standard deviation) remain time-invariant or constant with time. Thus, the curves for both of them have to be parallel to the x-axis, which in the outpatient arrival is not so. Table 6 shows the ADF test results for three types of patients.

To help data to be stationary, detrending is done as Eq. (10) and the detrending patient's arrivals are shown in Fig. 12.

$$\text{y.detrend} = (\text{y} - \text{y.rolling(window} = 6\text{).mean())}/\text{y.rolling(window} = 24\text{).std()} \quad (10)$$

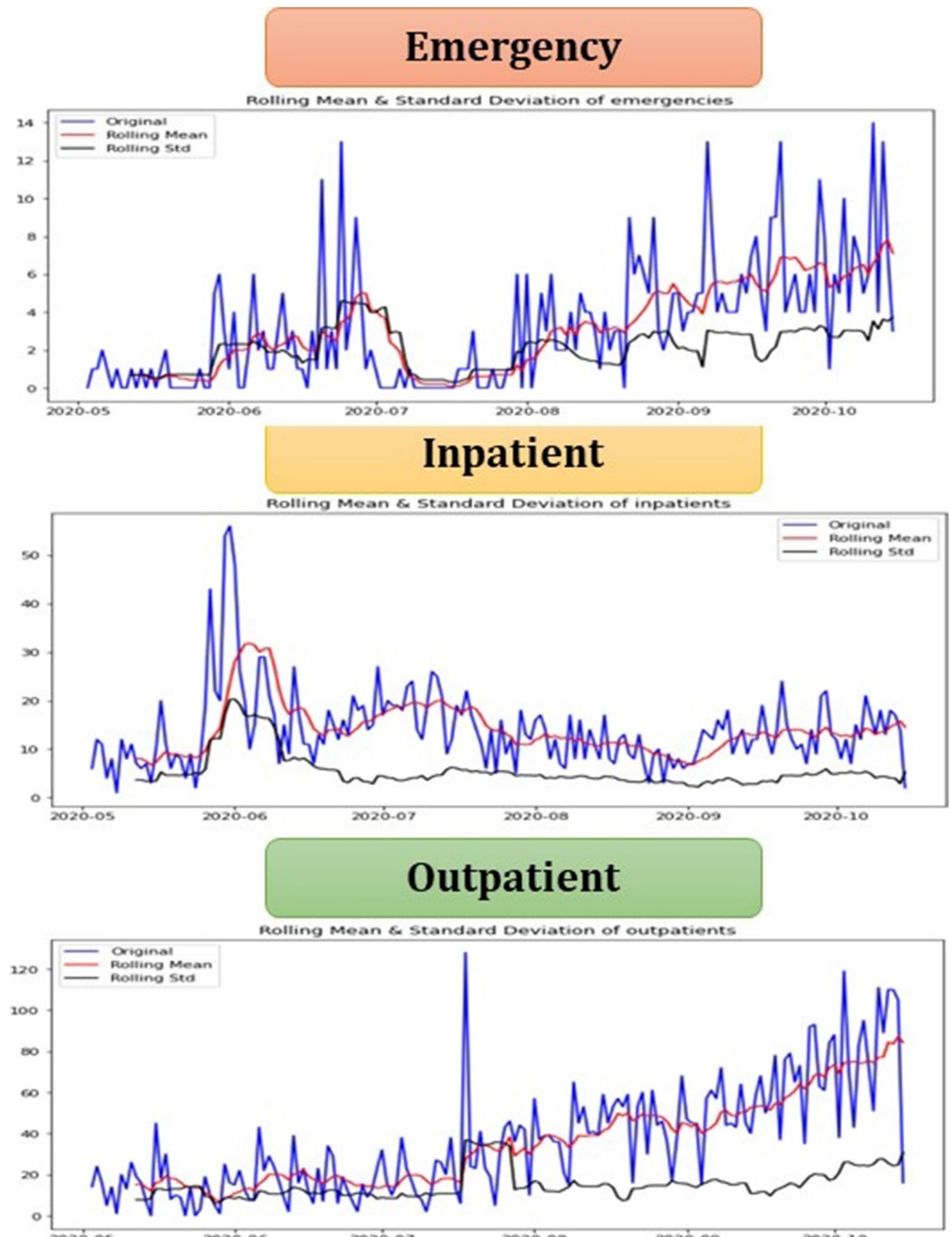

Fig. 11 Rolling mean and standard deviation of patient entry of emergency patients, outpatients, and inpatients

Table 6 ADF test results

Type of patients	*P*-value	99% level of confidence	95% level of confidence	90% level of confidence	
Emergency	0.071	−3.471	−2.879	−2.576	Data are not stationary with all confidence levels
Inpatient	0.005	−3.472	−2.880	−2.576	Data are stationary with all confidence levels
Outpatient	0.996	−3.474	−2.880	−2.577	Data are not stationary with all confidence levels

Finally, Table 7 shows the ADF test results for outpatients and emergencies after detrending. Now, three types of patients are stationary and the prediction model can be developed.

For tuning the parameters of the SARIMA model, the "gridsearch" method is used. This is a parameter-tuning solution. The key point about the performance of this method is that, for each possible combination of parameters in the grid, the model is constructed and evaluated. Hence, it can be said that this algorithm has a search nature. We define different ranges for the parameters and selected the AIC as evaluation metric. As mentioned before, the best prediction model is the model with the lowest AIC value. Based on results, the model of $SARIMA(1, 1, 1) \times (0, 1, 1, 12)$ has the lowest AIC value for inpatient entry, $SARIMA(0, 1, 1) \times (0, 1, 1, 12)$ for outpatient and also emergency entry.

Results for the next 30 days of patient arrival prediction for all patient categories are shown in Fig. 13. The RMSE of the SARIMA with a season length of 12 for inpatients, outpatients, and emergencies are 4.87, 27.54, and 3.06, respectively. The gray area above and below the orange line in this figure represents the 95% confidence interval and as

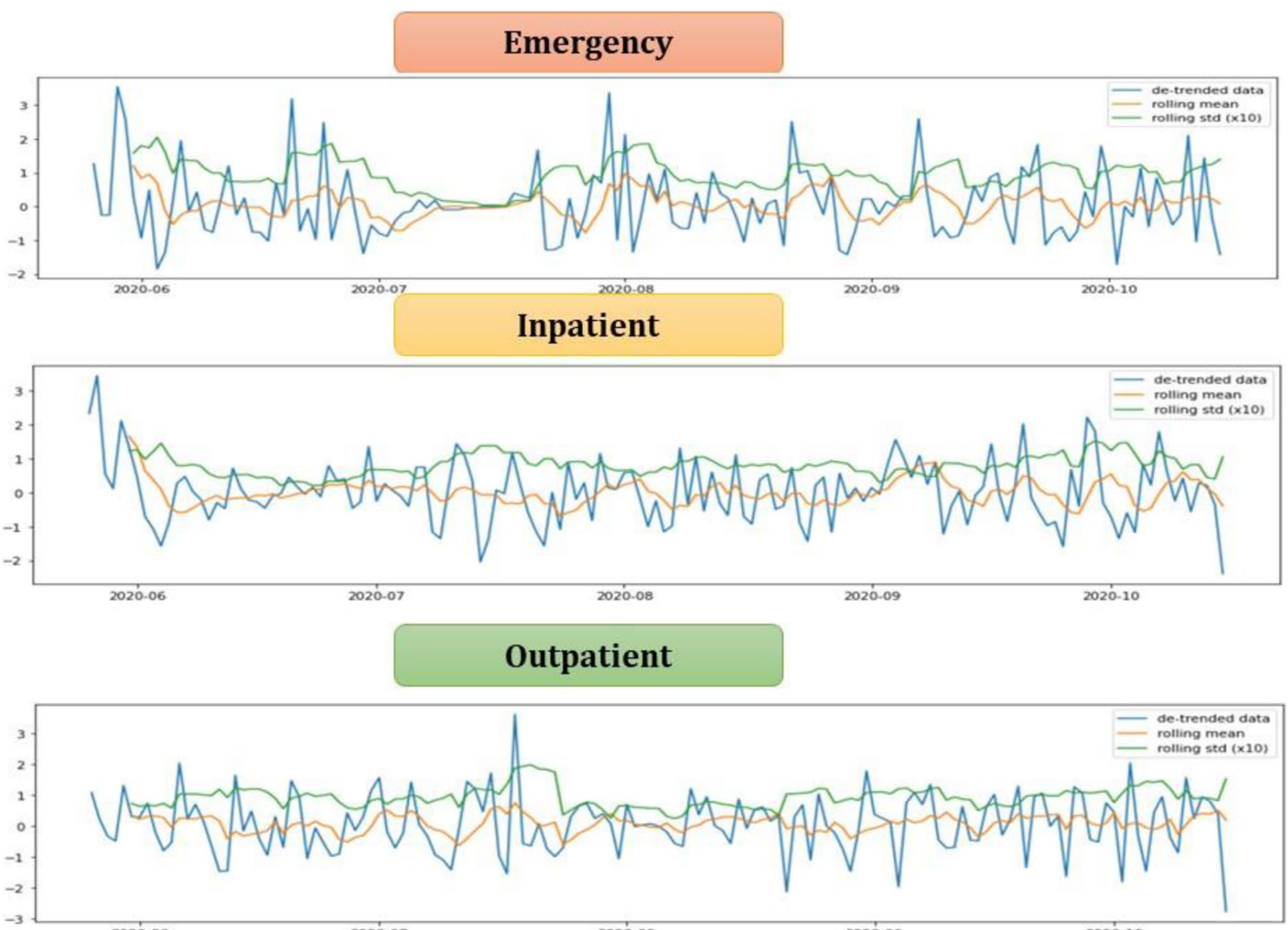

Fig. 12 Rolling mean and standard deviation of detrending patient arrivals

Table 7 ADF test results after detrending

Type of patients	*P*-value	99% level of confidence	95% level of confidence	90% level of confidence	
Emergency	0.000	−3.478	−2.882	−2.587	Data are stationary with all confidence levels
Outpatient	0.000	−3.482	2.884	−2.578	Data are stationary with all confidence levels

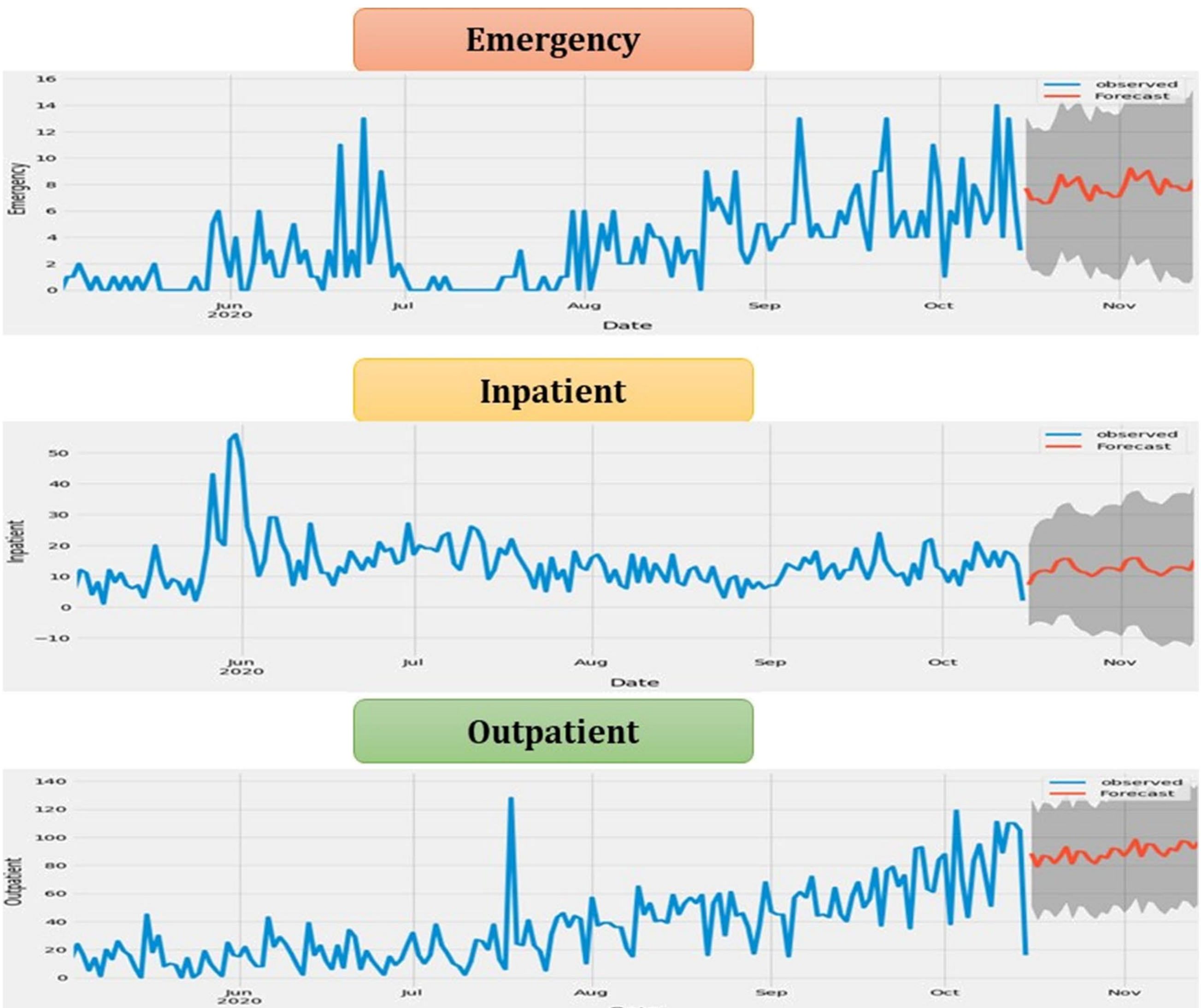

Fig. 13 SARIMA patient arrivals prediction within 30 days

with virtually all forecasting models, as the predictions go further into the future, the less confidence we have in our values. In this case, we are 95% confident that the actual patient arrivals will fall inside this range as shown in Table 8 for each patient category.

8 Managerial implications

Assume that the worst situation will be happening. So, the simulation model of scenario 14 is run again with the upper bound of time-series prediction of patient arrivals. This scenario had the worst outputs among other scenarios and could be considered the worst situation. However, all the scenarios can be run again, we obtain whether there are some bottlenecks in the patient flow in a pessimistic situation. In the simulated system used in this study, the most bottlenecks are related to the ICU, CCU, and corona special beds. After these three bottlenecks, the number of nurses in CCU and ICU wards and laboratories is the biggest challenge in the system. In this regard, to evaluate the stability of the system against the number of patients and their deterioration and the ability to respond to existing needs, the future of the system should be simulated. Therefore, scenario 14, which has the least increase in resources in the expressed bottlenecks and also does not improve significantly compared to the baseline, is selected and the number of patients admitted to the system in all three types of outpatients, inpatients, and emergency

Table 8 Predicted patient arrival based on the SARIMA time-series model

Date	Emergency			Inpatient			Outpatient		
	Predicted mean	Lower bound	Upper bound	Predicted mean	Lower bound	Upper bound	Predicted mean	Lower bound	Upper bound
2020–10-16	7.714765	2.380448	13.049082	7.134354	−5.983584	20.252292	88.456091	50.784333	126.127850
2020–10-17	6.833679	1.484595	12.182763	10.583554	−4.733418	25.900526	79.107164	41.431561	116.782767
2020–10-18	6.892752	1.479368	12.306135	11.636985	−4.618131	27.892101	86.825900	48.962829	124.688971
2020–10-19	6.567111	1.102572	12.031651	11.834827	−4.962317	28.631972	85.994904	48.049741	123.940067
2020–10-20	6.603010	1.081454	12.124566	11.281688	−5.925693	28.489070	81.500242	43.416682	119.583802
2020–10-21	7.447026	1.868506	13.025545	14.715041	−2.839410	32.269492	86.285286	48.058568	124.512003
2020–10-22	8.719867	3.084997	14.354738	15.536804	−2.332997	33.406605	92.999393	54.630500	131.368285
2020–10-23	7.832468	2.141802	13.523135	15.553853	−2.614186	33.721891	81.214134	42.703539	119.724730
2020–10-24	8.189716	2.443796	13.935636	12.704779	−5.751255	31.160813	89.928065	51.276291	128.579838
2020–10-25	8.544715	2.744067	14.345363	11.422879	−7.314197	30.159955	89.499555	50.707116	128.291993
2020–10-26	7.496343	1.641479	13.351207	11.067539	−7.945247	30.080324	84.071060	45.138466	123.003655
2020–10-27	6.763107	0.854525	12.671690	9.925537	−9.358467	29.209541	81.285414	42.213163	120.357665
2020–10-28	7.909421	1.877557	13.941284	11.006963	−8.721438	30.735365	86.817813	47.286145	126.349482
2020–10-29	7.342166	1.254010	13.430323	12.539233	−7.549615	32.628081	85.150103	45.484902	124.815305
2020–10-30	7.373639	1.225027	13.522251	12.686006	−7.724744	33.096757	91.987222	52.161673	131.812771
2020–10-31	7.050426	0.845532	13.255320	12.455023	−8.259257	33.169303	91.257414	51.314495	131.200334
2020–11-01	7.086111	0.823717	13.348506	11.699061	−9.311660	32.709781	86.751139	46.661027	126.841251
2020–11-02	7.930145	1.609969	14.250322	15.036483	−6.260328	36.333293	91.537515	51.293059	131.781971
2020–11-03	9.202985	2.825613	15.580357	15.812873	−5.762232	37.387979	98.251469	57.853992	138.648946
2020–11-04	8.315586	1.881522	14.749651	15.808462	−6.039282	37.656206	86.466228	45.916222	127.016234
2020–11-05	8.672834	2.182572	15.163096	12.949238	−9.166743	35.065220	95.180157	54.478202	135.882111
2020–11-06	9.027833	2.481857	15.573810	11.662538	−10.71796	34.043044	94.751647	53.898308	135.604985
2020–11-07	7.979461	1.378240	14.580682	11.304927	−11.33677	33.946633	89.323152	48.318989	130.327315
2020–11-08	7.246225	0.590219	13.902232	10.161851	−12.73796	33.061667	86.537506	45.383068	127.691945
2020–11-09	8.392539	1.615944	15.169134	11.242769	−12.07175	34.557293	92.069905	50.444961	133.694850
2020–11-10	7.825284	0.991351	14.659218	12.774799	−0.880407	36.430006	90.402196	48.633432	132.170959
2020–11-11	7.856757	0.961474	14.752041	12.921458	−11.04040	36.883318	97.239314	55.299390	139.179239
2020–11-12	7.533544	0.578227	14.488861	12.690422	−11.57018	36.951032	96.509507	54.424582	138.594431
2020–11-13	7.569229	0.555947	14.582511	11.934434	−12.61316	36.482033	92.003231	49.763246	134.243216
2020–11-14	8.413263	1.341502	15.485025	15.271844	−9.551368	40.095056	96.789607	54.385763	139.193451

patients using the time-series machine learning method is predicted. Based on the simulation findings of scenario 14 in the future of the system, it is concluded that the system will collapse after 14 days according to the predictions made. This means that the bottleneck of the ICU and CCU becomes problematic. In this regard, the following solutions must be taken for the system to continue:

- Creating more capacities for hospitalization of coronary patients in the studied hospital
- Creating the capacity to hospitalize coronary patients referred to the studied hospital in other hospitals
- Establishment of temporary capacities (i.e., hospitals) to transfer patients required to be admitted to those places
- Transferring more patients to their homes and providing services remotely and in patients' homes

It is mentioned that, for practical use and exploitation, managers of the hospital can analyze each decision they are going to make using this proposed model as what has been carried out above. So, it can be a decision support tool for evaluating every policy before implementation.

9 Conclusion

In this study, the COVID-19 patient flow in the hospital of a case study was first investigated. Then, the process with detailed data was simulated and the outputs were obtained. Consequently, 27 scenarios based on 11 inputs were defined based on the Taguchi method and simulated all scenarios. Then, the DEA method was used to calculate the efficiency score of scenarios. Finally, the worst scenario was simulated with predicted patient arrivals, which was the output of the SARIMA time-series model and the bottlenecks were identified. Moreover, we tried to highlight the simulation tools as decision support systems for hospital managers, who are willing to be more efficient and rely on data as the data-driven decision-makers. Since simulation can visualize the future and help the managers in human resource planning, facilities procurement, and other strategic and tactical decisions, we demonstrate the proposed approach in the case study for helping the managers in decision-making.

Our study presents some limitations. First, we only considered the patient arrival rate and not the other features which could be the impact on being infected. Second, we did not consider the impact of workload on the physician and nurse capability or even their infection as well as their specialty level. Third, we did not consider the beds can be transferred from other units to ICU and their quality. As these are the limitations of our study, we should highlight that, although our proposed approach could be as a decision support system, it does not guarantee optimal results that could be continued by future studies.

Furthermore, future studies can focus on different prediction models of patient arrivals based on other exogenous features for other time-series prediction models such as LSTM and machine learning regression models. Also, using other methods of investigating different inputs' effects on the outputs (e.g., system dynamic approach) can analyze different scenario results. Besides, other studies can focus on the same problem using process mining tools.

Author contribution All authors (Mahdieh Tavakoli, Reza Tavakkoli-Moghaddam, Reza Mesbahi, Mohssen Ghanavati-Nejad, Amirreza Tajally) contributed to all parts of this research including conceptualization; formal analysis; resources; methodology; supervision; data collection and investigation; software; validation; and writing — review & editing and Reza Tavakkoli-Moghaddam has the role of project administration.

Availability of data and material All data generated or analyzed during this research are included in this published article.

Code availability Not applicable.

Declarations

Ethics approval The authors certify that they have no affiliation with or involvement with human participants or animals performed by any of the authors in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this paper.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

1. Alban A, Chick SE, Dongelmans DA, Vlaar APJJ, Sent D, van der Sluijs AF, Wiersinga WJ, Vlaar APJJ, Sent D (2020) ICU capacity management during the COVID-19 pandemic using a stochastic process simulation. Intensive Care Med 46(8):1624–1626. https://doi.org/10.1007/s00134-020-06066-7
2. Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, Mizukami Y, Tadono T (2019) Spatiotemporal dengue fever hotspots associated with climatic factors in taiwan including outbreak predictions based on machine-learning. Geospat Health 14(2):183–194. https://doi.org/10.4081/gh.2019.771
3. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM (2020) COVID-19 outbreak prediction with machine learning. Algorithms 13(10):249
4. Azadeh A, Sheikhalishahi M, Firoozi M, Khalili SM (2013) An integrated multi-criteria taguchi computer simulation-DEA approach for optimum maintenance policy and planning by

incorporating learning effects. Int J Prod Res 51(18):5374–5385. https://doi.org/10.1080/00207543.2013.774496
5. Bhosekar A, Ekşioğlu S, Işık T, Allen R (2021) A discrete event simulation model for coordinating inventory management and material handling in hospitals. Annals of Operations Research 1–28
6. Bonakdari H, Pelletier J, Martel-pelletier J (2020) A reliable time-series method for predicting arthritic disease outcomes: new step from regression toward a nonlinear artificial intelligence method. Comput Methods Programs Biomed 189:105315. https://doi.org/10.1016/j.cmpb.2020.105315
7. Bosbach WA, Heinrich M, Kolisch R, Heiss C (2021) Maximization of open hospital capacity under shortage of sars-cov-2 vaccines—an open access, stochastic simulation tool. Vaccines 9(6): https://doi.org/10.3390/vaccines9060546
8. Brenner S, Zeng Z, Liu Y, Wang J, Li J, Howard PK (2010) Modeling and analysis of the emergency department at university of Kentucky Chandler Hospital using simulations. J Emerg Nurs 36(4):303–310. https://doi.org/10.1016/j.jen.2009.07.018
9. Burke RM, Shah MP, Wikswo ME, Barclay L, Kambhampati A, Marsh Z, Cannon JL, Parashar UD, Vinjé J, Hall AJ (2019) The norovirus epidemiologic triad: predictors of severe outcomes in US norovirus outbreaks, 2009–2016. J Infect Dis 219(9):1364–1372. https://doi.org/10.1093/infdis/jiy569
10. Campos AT, Dos Santos CH, Gabriel GT, Montevechi JAB (2021) Safety assessment for temporary hospitals during the COVID-19 pandemic: a simulation approach. Safety Science, 105642
11. Carlson CJ, Dougherty E, Boots M, Getz W, Ryan SJ (2018) Consensus and conflict among ecological forecasts of Zika virus outbreaks in the United States. Sci Rep 8(1):1–15. https://doi.org/10.1038/s41598-018-22989-0
12. Coelli FC, Ferreira RB, Almeida RMVRR, Pereira WCA (2007) Computer simulation and discrete-event models in the analysis of a mammography clinic patient flow. Comput Methods Programs Biomed 87(3):201–207. https://doi.org/10.1016/j.cmpb.2007.05.006
13. Darwish A, Rahhal Y, Jafar A (2020) A comparative study on predicting influenza outbreaks using different feature spaces: application of influenza-like illness data from Early Warning Alert and Response System in Syria. BMC Res Notes 13(1):1–8
14. Diaz MCG, Dawson K (2020) Use of simulation to develop a COVID-19 resuscitation process in a pediatric emergency department. Am J Infect Control 48(10):1244–1247. https://doi.org/10.1016/j.ajic.2020.07.032
15. Heo JN, Park JA, Han D, Kim H-JJ, Ahn D, Ha B, Seog W, Park YR (2020) COVID-19 outcome prediction and monitoring solution for military hospitals in South Korea: development and evaluation of an application. J Med Internet Res 22(11):e22131. https://doi.org/10.2196/22131
16. Idris NM, Chiam YK, Varathan KD, Ahmad WAW, Chee KH, Liew YM, Md Idris N, Chiam YK, Varathan KD, Wan Ahmad WA, Chee KH, Liew YM (2020) Feature selection and risk prediction for patients with coronary artery disease using data mining. Med Biol Eng Compu 58(12):3123–3140. https://doi.org/10.1007/s11517-020-02268-9
17. Karan NB, Kahraman S (2019) Evaluation of posterior airway space after setback surgery by simulation. Med Biol Eng Compu 57(5):1145–1150. https://doi.org/10.1007/s11517-018-1943-8
18. Khan YA, Abbas SZ, Truong B-CC (2020) Machine learning-based mortality rate prediction using optimized hyper-parameter. Comput Methods Programs Biomed 197:105704. https://doi.org/10.1016/j.cmpb.2020.105704
19. Kleiven EF, Henden J-AA, Ims RA, Yoccoz NG (2018) Seasonal difference in temporal transferability of an ecological model: near-term predictions of lemming outbreak abundances. Sci Rep 8(1):1–6. https://doi.org/10.1038/s41598-018-33443-6
20. Kovalchuk SV, Funkner AA, Metsker OG, Yakovlev AN (2018) Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification. J Biomed Inform 82:128–142
21. Liang R, Lu Y, Qu X, Su Q, Li C, Xia S, Liu Y, Zhang Q, Cao X, Chen Q (2020) Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. Transbound Emerg Dis 67(2):935–946
22. Maleki M, Mahmoudi MR, Wraith D, Pho K-HH (2020) Time series modelling to forecast the confirmed and recovered cases of COVID-19. Travel Med Infect Dis 37(March):101742. https://doi.org/10.1016/j.tmaid.2020.101742
23. Markovitch NA, Profozich DM (1996) Arena software tutorial. Proceedings Winter Simulation Conference 437–440
24. Melman GJ, Parlikad AK, Cameron EAB (2021) Balancing scarce hospital resources during the COVID-19 pandemic using discrete-event simulation. Health Care Management Science 1–19
25. Ordu M, Demir E, Tofallis C, Gunal MM (2021) A novel healthcare resource allocation decision support tool: a forecasting-simulation-optimization approach. J Oper Res Soc 72(3):485–500
26. Pan C, Zhang D, Kon AWM, Wai CSL, Ang WB (2015) Patient flow improvement for an ophthalmic specialist outpatient clinic with aid of discrete event simulation and design of experiment. Health Care Manag Sci 18(2):137–155. https://doi.org/10.1007/s10729-014-9291-1
27. Parbat D, Chakraborty M (2020) A python based support vector regression model for prediction of COVID19 cases in India. Chaos Solit Fractals 138:109942. https://doi.org/10.1016/j.chaos.2020.109942
28. Parviz L (2020) Comparative evaluation of hybrid sarima and machine learning techniques based on time varying and decomposition of precipitation time series. J Agric Sci Technol 22(2):563–578
29. Pereira MA, Ferreira DC, Figueira JR, Marques RC (2021) Measuring the efficiency of the Portuguese public hospitals: a value modelled network data envelopment analysis with simulation. Expert Syst Appl 181:115169
30. Rahmanian V, Bokaie S, Haghdoost A, Barouni M (2021) Predicting cutaneous leishmaniasis using SARIMA and Markov switching models in Isfahan, Iran: a time-series study. Asian Pac J Trop Med 14(2):83–93. https://doi.org/10.4103/1995-7645.306739
31. Rath S, Tripathy A, Tripathy AR (2020) Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. Diabetes Metab Syndr 14(5):1467–1474. https://doi.org/10.1016/j.dsx.2020.07.045
32. Reis J, Yamana T, Kandula S, Shaman J (2019) Superensemble forecast of respiratory syncytial virus outbreaks at national, regional, and state levels in the United States. Epidemics 26(July 2018):1–8. https://doi.org/10.1016/j.epidem.2018.07.001
33. Rivers-Moore NA, Hill TR (2018) A predictive management tool for blackfly outbreaks on the Orange River, South Africa. River Res Appl 34(9):1197–1207. https://doi.org/10.1002/rra.3357

34. Roy AN, Jose J, Gautam N, Nathalia D, Suresh A, Pradesh U, Sarasvati H, Development H, Sunil A, Gautam N, Nathalia D, Suresh A (2020) Prediction and spread visualization of COVID-19 pandemic using machine learning. Preprints.Org, May, 1–17. https://doi.org/10.20944/preprints202005.0147.v1
35. Rypdal M, Sugihara G (2019) Inter-outbreak stability reflects the size of the susceptible pool and forecasts magnitudes of seasonal epidemics. Nature Communications 10(1). https://doi.org/10.1038/s41467-019-10099-y
36. Sasanfar S, Bagherpour M, Moatari-Kazerouni A (2021) Improving emergency departments: simulation-based optimization of patients waiting time and staff allocation in an Iranian hospital. Int J Healthc Manag 14(4):1449–1456
37. Tapak L, Hamidi O, Fathian M, Karami M (2019) Comparative evaluation of time series models for predicting influenza outbreaks: application of influenza-like illness data from sentinel sites of healthcare centers in Iran. BMC Res Notes 12(1):1–6. https://doi.org/10.1186/s13104-019-4393-y
38. Tohidinik HR, Mohebali M, Mansournia MA, Kalhori SRN, Niakan Kalhori SR, Ali-Akbarpour M, Yazdani K (2018) Forecasting zoonotic cutaneous leishmaniasis using meteorological factors in eastern Fars province, Iran: a SARIMA analysis. Tropical Med Int Health 23(8):860–869. https://doi.org/10.1111/tmi.13079
39. Tomar A, Gupta N (2020) Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Sci Total Environ 728:138762
40. Virani SA, Clarke B, Ducharme A, Ezekowitz JA, Heckman GA, McDonald M, Mielniczuk LM, Swiggum E, Van Spall HGCC, Zieroth S (2020) Optimizing access to heart failure care in Canada during the COVID-19 pandemic. Can J Cardiol 36(7):1148–1151. https://doi.org/10.1016/j.cjca.2020.05.009
41. Wu DCW, Ji L, He K, Tso KFG (2021) Forecasting tourist daily arrivals with a hybrid Sarima-Lstm approach. J Hosp Tour Res 45(1):52–67. https://doi.org/10.1177/1096348020934046
42. Ye T, Yang X (2020) Analysis and prediction of confirmed COVID-19 cases in China with uncertain time series. Fuzzy Optimization and Decision Making, 1–20. https://doi.org/10.1007/s10700-020-09339-4
43. Zeinalnezhad M, Chofreh AG, Goni FA, Klemeš JJ, Sari E (2020) Simulation and improvement of patients' workflow in heart clinics during COVID-19 pandemic using timed coloured petri nets. Int J Environ Res Public Health 17(22):8577. https://doi.org/10.3390/ijerph17228577
44. Zeroual A, Harrou F, Dairi A, Sun Y (2020) Deep learning methods for forecasting COVID-19 time-series data: a comparative study. Chaos, Solitons and Fractals 140. https://doi.org/10.1016/j.chaos.2020.110121
45. Zhang G, Xu J, Yu M, Yuan J, Chen F (2020) A machine learning approach for mortality prediction only using non-invasive parameters. Med Biol Eng Comput 1–44
46. Zhang H, Best TJ, Chivu A, Meltzer DO (2020) Simulation-based optimization to improve hospital patient assignment to physicians and clinical units. Health Care Manag Sci 23(1):117–141
47. Zhong L, Mu L, Li J, Wang J, Yin Z, Liu D (2020) Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model. Ieee Access 8:51761–51769. https://doi.org/10.1109/ACCESS.2020.2979599

Mahdieh Tavakoli is currently the PhD candidate in School of Industrial Engineering at the University of Tehran, Iran. She obtained her BSc and MSc degrees from the Alzahra University and Tarbiat Modares University in Tehran in 2015 and 2017, respectively. She has started using industrial engineering functions in healthcare systems for 5 years and experienced different projects in hospitals in fields of process mining, simulation, data analysis, and system dynamic. Her interest areas are optimization, data-driven decision-making, and process mining.

Reza Tavakkoli-Moghaddam is a distinguished Professor of Industrial Engineering at College of Engineering, University of Tehran in Iran. He obtained his PhD in Industrial Engineering from Swinburne University of Technology in Melbourne (1998), his MSc in Industrial Engineering from the University of Melbourne in Melbourne (1994) and his BSc in Industrial Engineering from the Iran University of Science and Technology in Tehran (1989). He serves as Editorial Board of five reputable academic journals. He is the recipient of the 2009 and 2011 Distinguished Researcher Awards and the 2010 and 2014 Distinguished Applied Research Awards at the University of Tehran. He was selected as National Iranian Distinguished Researcher in 2008 and 2010. He has obtained the outstanding rank as the top 1% scientist and researcher in the world elite group since 2014. He has published 4 books, 37 book chapters, and more than 1000 journal and conference papers.

Reza Mesbahi is currently the PhD candidate in School of Industrial Engineering at the University of Tehran, Iran. He obtained his BSc degree in Biomedical Engineering from Science and Research Branch at Islamic Azad University in 2006 and MSc degree in System Engineering from the University of Science and Technology in Tehran in 2008. He trained by Top Brands Company, such as Getinge Group and St. Jude Medical in the field of biomedical and healthcare technology in Germany and Belgium, and then started working in health care systems as a professional specialist in health data and development projects in healthcare systems since 2010 under the supervision of Emergency Medical services in Iranian MOH.

Mohssen Ghanavati-Nejad is currently the PhD candidate in School of Industrial Engineering at the University of Tehran, Iran. He obtained his BSc and MSc degrees in Industrial Engineering from Qom University of Technology and Tarbiat Modares University in Tehran in 2015 and 2017, respectively. He is very interested in digital transformation and lean approaches, business startup, business model development, and data-driven marketing.

Amirreza Tajally is currently the MSc student in School of Industrial Engineering at the University of Tehran. He obtained his BSc degree in Industrial Engineering from North Tehran Branch at Islamic Azad University. He has strong background in data-driven optimization and machine learning and has more than 6 years experienced in the instructing of data science. Also, he implemented data-driven tools and techniques for organizational problem-solving in different fields, such as automotive industries and healthcare systems. His main research interests are focused on data-driven decision-making, customer behavior analytics, and statistical learning with applications ranging from manufacturing, transportation, and retail to health management and cognitive computing.