



Care process optimization in a cardiovascular hospital: an integration of simulation–optimization and data mining

Masoumeh Vali¹ · Khodakaram Salimifard¹ · Amir H. Gandomi² · Thierry J. Chausalet³

Accepted: 10 June 2022
© The Author(s) 2022

Abstract

To provide health services, hospitals consume electrical power and contribute to the CO₂ emission. This paper aims to develop a modelling approach to optimize hospital services while reducing CO₂ emissions. To capture treatment processes and the production of carbon dioxide, a hybrid method of data mining and simulation–optimization techniques is proposed. Different clustering algorithms are used to categorize patients. Using quality indicators, clustering methods are evaluated to find the best cluster sets, and then patients are categorized accordingly. Discrete-event simulation is applied to each patient category to estimate performance measures such as number of patients being served, waiting times, and length of stay, as well as the amount of CO₂ emission. To optimize performance measures of patient flow, metaheuristic searches have been used. The dataset of Bushehr Heart Hospital is considered as a case study. Based on K-means, K-medoid, Hierarchical clustering, and Fuzzy C-means clustering methods, patients are categorized into two groups of high-risk and low-risk patients. The number of patients being served, total waiting time, length of stay, and CO₂ emitted during care processes are improved for both groups. The proposed hybrid method is an effective method for hospitals to categorize patients based on care processes. The problems and the proposed solution approach reported in this study could be applicable to other hospitals, worldwide to help both optimize the patient flow and minimize the environmental consequences of care services.

✉ Thierry J. Chausalet
chausst@westminster.ac.uk

Masoumeh Vali
m.vali@mehr.pgu.ac.ir

Khodakaram Salimifard
salimifard@pgu.ac.ir

Amir H. Gandomi
gandomi@uts.edu.au

¹ Computational Intelligence & Intelligent Optimization Research Group, Business School, Persian Gulf University, Bushehr 75169, Iran

² Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo NSW 2007, Australia

³ Health and Social Care Modelling Group, School of Computer Science and Engineering, University of Westminster, 115 New Cavendish St, London W1W 6UW, UK

Keywords Patient flow · Healthcare · Carbon footprint · Discrete-event simulation · Data mining · Simulation–optimization

1 Introduction

In recent years, healthcare has become an important part of our daily life and it has been challenging to deliver high-quality care with limited resources (Strome et al., 2013). In many countries, healthcare has become a thriving sector of the economy (Yang et al., 2015) and has gone through technological advancements. For example, information technology is currently being used as part of healthcare management systems (Prokosch & Ganslandt, 2009).

Data mining can be defined as the analysis of data that discovers relationships or identifies patterns between various elements of a data set. It has been applied to extract hidden patterns within patient data in the healthcare system, including clinical medicine (Iavindrasana et al., 2009), adverse drug reaction signal detection (Karimi et al., 2015), big data analytics (Ghassemi et al., 2015), diabetes (Sigurdardottir et al., 2007), and skin diseases.

Discrete event simulation (DES) models a system by simulating its sequence of events or processes over time. Due to the complexity associated with healthcare processes, DES is the most widely used decision support tool for assessing trade-offs between the multiple objectives of healthcare systems. Simulation-based optimization could be used to find solutions to problems with a large number of conceivable scenarios (Fetter & Thompson, 1965).

Climate change is occurring because of the accumulation of greenhouse emissions in the atmosphere due to the combustion of fossil fuels. The major greenhouse gases responsible for climate change and global warming are carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O). Healthcare infrastructures have a large footprint in climate change and hospitals as a major part of this system have a high demand for electricity, lighting, heating, and the energy for ventilation, electric and electronic equipment, and air conditioning (Bi & Hansen, 2018a).

From an operations research perspective, carbon footprint (CFP) reduction has been tackled at strategic (Badri et al., 2013), tactical, and operational (Absi et al., 2013) levels. Planning and advanced scheduling techniques can play key roles in supporting CFP reduction (Liu, 2014). Intense climate events can have a direct or indirect effect on human health by disrupting ecosystems, agriculture, food, water quality, air quality, and by damaging infrastructure (Organization, 2014). All these effects place a great burden on health systems. Given the importance of healthcare and environmental threats due to increasing levels of greenhouse gases (GHGs), health authorities need to use effective ways to decrease the carbon footprint in healthcare systems. Pollard et al. (Pollard et al., 2013) used data from a case study to propose a bottom-up modeling framework to help with decisions regarding both cost and carbon in healthcare. Research results have confirmed that a bottom-up approach is effective for estimating and modeling the carbon footprint in healthcare.

The study in this paper was conducted at the Bushehr Heart Hospital in Iran. Negative effects of waiting on patients are well studied (e.g., Sigurdardottir et al., 2007). However, clinical staff also experience negative effects (Viccellio, 2017). Growing queues of patients put staff under significant work pressure and often require them to deal with frustrated patients. In the long run, such pressures can create morale problems and likely contributes to absenteeism. We were brought in to help the clinic staff diagnose the causes of poor patient flow and to identify effective solutions. We used simulation modeling as the main tool to help in our diagnostic and improvement efforts.

One major contribution of our work is to show how a simulation analysis of patient flow can significantly improve waiting time in a specialized hospital. To the best of our knowledge, our study is the first of its kind for a heart hospital. Another major contribution is to show how simulation modeling and the “hard” quantitative analysis it provides can assist in convincing involved parties to implement improvements. The clinic previously attempted to improve patient waiting-time performance by testing localized initiatives using standard Plan-Do-Study-Act (PDSA) methods (Fetter & Thompson, 1965). While this approach likely helped in creating a culture more accepting of change, the modeling we performed provided a systems perspective in addition to the quantitative evidence that showed that the improvements should work. With the preponderance of scientific staff in healthcare settings, a quantitative evidence-based approach can be important for a successful implementation.

The proposed approach used in this research is made up of two parts. In the first part, data mining is used to investigate data to discover relevant relationships among them. In the second part, a simulation–optimization model is developed to find an optimized patient flow while minimizing CO₂ emissions. The remainder of this paper is organized as follows. In Sect. 2, related works are reviewed. Section 3 describes the research methodology. Section 4 presents the results of patients’ clustering and optimization of the care process. Section 5 discusses the findings of this research and presents conclusions based on the outputs of this research.

2 Literature review

Air pollution is a leading cause of global mortality and morbidity in the twenty-first century (Eckelman et al., 2018). A large number of studies on the relationship between climate change and death rate have been conducted (Sheridan et al., 2011) and have shown that for example blood pressure and cardiovascular diseases are related to diurnal temperature range and that heart disease incidence increases with air pollution. The healthcare sector like other industries requires improving its environmental performance thus many new facilities have been built based with this in mind in the past decades (Pinzone et al., 2012). Appropriate policies are needed to cope with the demand for climate change management within the health sector (Frumkin et al., 2008). Healthcare infrastructures have a large footprint in climate change and hospitals as a major part of this system have a high demand for electricity, lighting, heating, and the energy for ventilation, electric and electronic equipment, and air conditioning (Bi & Hansen, 2018b); in fact, all medical equipment needs the energy to function (Chevalier et al., 2009). Considering the importance of patient flow in hospitals and the usage of medical equipment in care processes, this work integrates data mining and simulation–optimization modeling to find an optimized patient flow while minimizing CO₂ emission.

The merging of simulation and optimization methods has seen remarkable growth in recent years (Sheridan et al., 2011). Klassen and Yoogalingam (2009) proposed a simulation–optimization approach that was used to determine optimal rules for outpatient healthcare service scheduling problems. Their approach uses more variables and factors for system modeling as compared to previous studies. Kasaie and Kelton (2013) proposed a simulation–optimization framework for resource allocation in the control of epidemics interventions, and analyzed the behavior of RA outcomes concerning different investment strategies and sought optimal allocations. Cabrera et al. (2012) presented an Agent-Based modeling (ABM) approach to design a decision support system for Healthcare Emergency Department (ED). Osorio et al. (2017) presented an integrated simulation–optimization model to support both strategic

and operational decisions for production planning in the blood supply chain. This method improved key indicators such as shortages, outdated units, donors required, and cost.

Healthcare encompasses many processes dealing with the treatment, diagnosis, and prevention of disease, injury, and other mental and physical impairments. Data mining has been used in previous studies to extract hidden patterns in patient data (Sun & Reddy, 2013; Yoo et al., 2012). Bruno (Bruno et al., 2014) proposed an explorative data mining approach to identify examinations followed by patients with a given disease. Their results showed the effectiveness of the proposed approach for discovering interesting groups of patients based on disease severity and similar examination history.

Xu et al. (2016) proposed an alternating optimization approach that was used to discover clusters in the positive class and to optimize the classifiers that separate each positive cluster from the negative samples. Mahoto et al. (2014) used clustering techniques to transform patient diagnostic exam data into patient vectors based on three clustering algorithms including DBSCAN, K-means, and Hierarchical algorithms and showed that DBSCAN performed better than the other algorithms.

Several studies have been published regarding the combination of simulation/optimization and data mining. Ng et al. (2011) proposed the integration of DES and data mining techniques for the analysis of general systems that are particularly suitable for production systems. Codrington-Virtue et al. (2006) developed an intelligent patient management system for use in the Accident and Emergency (A&E) setting based on DES and clustering techniques to calculate the maximum number of treatment places and nurse units required to service A&E ambulance arrivals. Their study also demonstrated how A&E ambulance arrivals can be categorized into diagnosis sub-groups according to length of stay quantiles. Ceglowski et al. (2016) combined data mining and DES to identify bottlenecks at the interface between the ED and hospital wards. Their model provided a value-added view of a hospital emergency department, treatment and disposal, and the occurrence of queues for treatment.

Amaran et al. (2016) compared and contrasted simulation optimization (SO) to algebraic model-based mathematical programming. The capacity problem of perinatal networks in the United Kingdom was considered by Asaduzzaman et al. (2010), while bed occupancy levels in an intensive care unit were assessed using simulation–optimization by Mallor and Azcárate (2014).

The above literature review shows that the patient flow, in general, and emergency departments have been considered with great concern. It also reveals that the combination of data mining, discrete event simulation, and optimization in improving the patient flow has been rarely considered. In addition to optimizing the length of stay, the number of patients discharged from the hospital, and waiting time, the amount of carbon produced by medical equipment in the hospital has been investigated. Reviewing the literature revealed that carbon footprint has been neglected in the context of healthcare in many middle and low income countries, and has not been considered in heart hospitals, thus motivating our research.

3 Research methodology

This research uses clustering algorithms to cluster patients and DES to capture the complexity of the patient flow. Then, the clustered patient flow is optimized based on waiting time, length of stay, patient throughput, and CO₂ emission, using OptQuest (Eckelman & Sherman, 2016). The three stages of the methodology are described in Fig. 1.

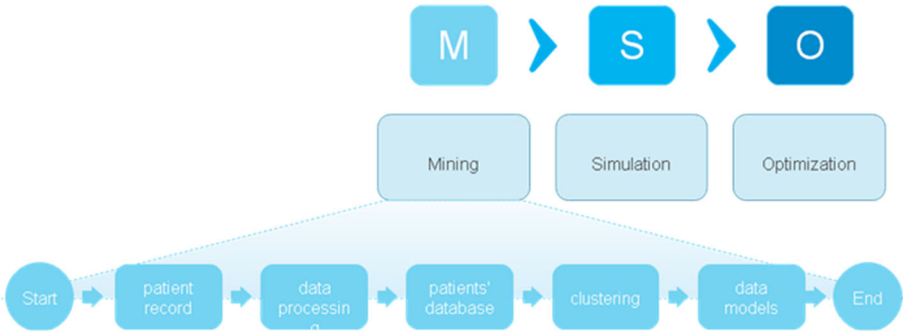


Fig. 1 Stages of the proposed methodology

Data mining (M), the first stage is composed of five steps: patient recording, data processing, retrieving patients' database, data clustering, and data modeling. Following data collection, is data preprocessing, consisting of data cleaning, data integration, data selection, and data transformation. Data are cleaned because real-world data are sometimes noisy, inconsistent, and incomplete. Then, the data are stored in a database. Next, data relevant to the analysis are retrieved from the database. Finally, data are transformed and consolidated into different forms that are suitable for the mining procedure.

For the second stage simulation (S), the general framework of the flexible job-shop scheduling problem (FJSP) is used for patients' flow modeling. The care units that a patient must go to during treatment and the average electricity consumption of equipment in each unit per patient are explained in Sect. 4. The last stage of the methodology is optimization (O), where OptQuest is used to optimize the objective function. According to the relationship between environment and health (Schulz et al., 2016), and the role of the health sector against climate change (Frumkin et al., 2008), in addition to throughput, waiting times and length of stay, reduction of carbon dioxide emissions due to the use of electrical equipment in the treatment process is considered.

3.1 Data mining: clustering methods and internal validation

Data mining emerged in the middle of the 1990s as a new approach to data analysis and knowledge discovery. The term "Data Mining" was first registered for the 2010 Medical Subject Headings (Yoo et al., 2012). Data mining has been used for pattern recognition (Kaya & Schoop, 2019), database design (Chaudhuri, 1998), artificial intelligence (Navale et al., 2016), visualization, and applications in healthcare (Tomar & Agarwal, 2013). One of the definitions mostly used states that "data mining is the analysis of observational data sets to summarize the data in novel ways and to find unsuspected relationships that are both useful and understandable to the data owner" (Hand et al., 2001).

Clustering forms one of the major classes of data mining algorithms. Clustering is an approach in which data are categorized into different groups or clusters in such a way that each group contains similar data points (Ibrahim et al., 2013). In healthcare systems, data points represent clinical profiles. Since patients with similar diseases need fairly similar types of care, the system should be able to design diagnostic patterns for treatment. Clear and tested clusters based on comorbidities can help clinicians select treatments for specific patients. In turn, this can assist with resource planning and system performance. In this research, four

clustering methods are used to categorize patients: K-means (Duda et al., 2012; Jain et al., 1999), K-medoid (Na et al., 2010), hierarchical clustering (Jain et al., 1999), and fuzzy C-means (Mannila, 1996).

To ensure that a technique produces reliable results, validation is vital. Clustering validation is recognized as essential to the success of clustering applications (Jain & Dubes, 1988) and evaluates the goodness of clustering results (Liu, et al., 2010). Internal validation measures include Root-mean-square error, R-squared, Dunn's index, Silhouette index, among many others (Liu et al., 2010). In this research we evaluated the clustering performance of each method using two of the most commonly used measures the Dunn index (Azuaje, 2002) and the Silhouette score (Wang et al., 2003).

Silhouette analysis is used to study the separation distance between the resulting clusters and measure how close each object in one cluster is close to another object in another cluster. Silhouette score values lie between -1 to $+1$. The value of $+1$ indicates the correct clustering of data points while the value of -1 shows that data points are not properly clustered. *Dunn's validation index* is characterized as the ratio of the minimum distance between two clusters and the size of the biggest cluster (Azuaje, 2002).

3.2 Simulation optimization method

Discrete event simulation (DES) is a computer-based methodology utilized in modeling complex dynamic and stochastic systems, including health care delivery, and characterized by its speed and high flexibility. Nowadays, DES software is often embedded with robust tools to support optimization in a variety of applications, including manufacturing (Rivera-Gómez et al., 2016), and operations scheduling (Cadi, et al., 2015).

DES is useful in hospitals where patient demand outstrips medical system capacity, and low-cost approaches to improve health care delivery are essential. It allows users to estimate the impact of operational changes before expanding resources (Abo-Hamad & Arisha, 2013).

Simulation optimization is an important enhancement of the simulation methodology because optimization is often desired in the design of systems. For instance, Li and Wang (2012) modeled and compared the impact of different ordering policies utilizing OptQuest simulation. Zhang et al. (2020) developed an ED model to evaluate different assignment strategies for expected patient waiting time, care quality, physician, and hospital profit. Lin et al. (2013) presented a system for multi-objective simulation optimization that combines the power of genetic algorithm with data envelopment analysis to evaluate the simulation results and guide the search process.

4 Case study and results

Our case study is based at Bushehr Heart Hospital (BHH), which is a hospital in southern Iran specializing in Cardiovascular disease (CVD), one of the most prevalent causes of death throughout the world (Sufi & Khalil, 2010). The Bushehr's Heart Hospital has eight care units including triage, cardiopulmonary resuscitation (CPR), emergency department (ED), coronary care unit I and II (CCUI and CCU II), post coronary care unit (PCCU), intensive care units I and II (ICU I and ICU II), Catheterization Laboratory (Cath Lab), and operating rooms (ORs). Also, it has two administration units including reception and discharge units which have been conceptualized as workstations in this research.

The conceptual model of the patient flow is illustrated in Fig. 2. Patients arrive either as

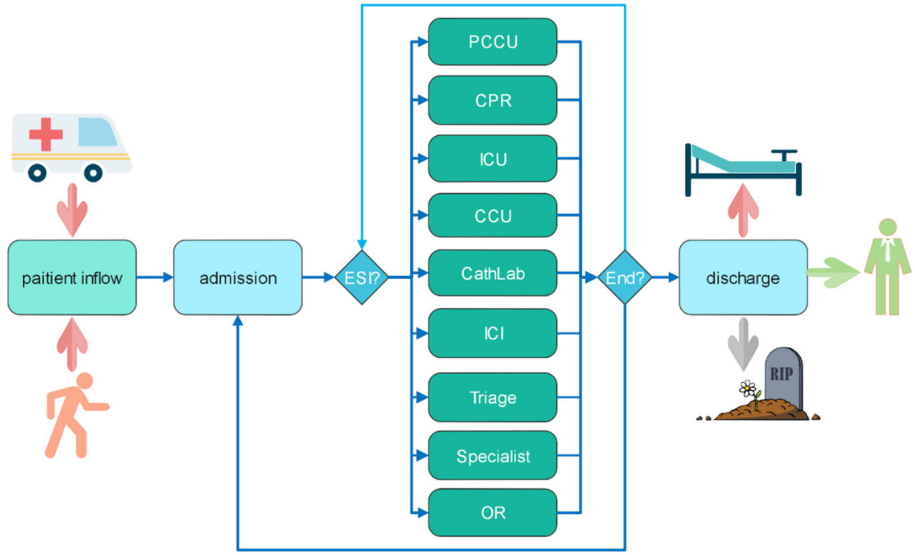


Fig. 2 Conceptual model of patient flow simulation

walk-in or by ambulance. On arrival, they are registered at the admission desk and based on their conditions, they receive the required treatment. Patients will be discharged when the treatment is successfully completed, or they are transferred to an inpatient ward or another hospital. Unfortunately, sometimes, the treatment is not successful, and the patient passes away.

Patient flow is defined as the movement of patients through a set of care units in the hospital. Based on interviews with the head nurse and the supervisor, ten common pathways have been discovered. To validate the discovered pathways, a database of patients has been investigated using a mix of descriptive and advanced data analytics techniques. Data extracted from the repository have been transformed and structured as excel files. Using the Emergency Severity Index, and data analytics (Bachhety et al., 2021), the experimental pathways defined by staff were confirmed.

Upon arriving and based on their condition, patients can be categorized in one of the five levels of severity using the Emergency Severity Index. Depending on their ESI, patients follow a different sequence of treatment and care. The care units visited by a patient during treatment processes are illustrated in Fig. 3. As shown, patients categorized as ESI 1 (*resuscitation*) follow either Route 11, Route 12, or Route 13. Patients with ESI 2 (*emergent*) are categorized as acute cardiovascular disease (ACS); follow either Route 21, Route 22, or Route 23. Patients with ESI 3 (*urgent*) follow either Route 31 or Route 32, while patients with ESI 4 (*nonurgent*) follow Route 4. Patient categorized as ESI 5 (*referred*) follow Route 5.

The collected data cover those patients who have visited the hospital within one year, from August 2017 to July 2018. That is, 11,700 patients were referred to the hospital in total, of which 5% were in the ESI 1 category and another 10% were in the ESI 2. The ESI for the others was found to be 30%, 30%, and 25% for ESI 3, ESI 4, and ESI 5, respectively. In order to capture the patient flow, four clustering algorithms have been applied. The results are reported in the following sub-section.

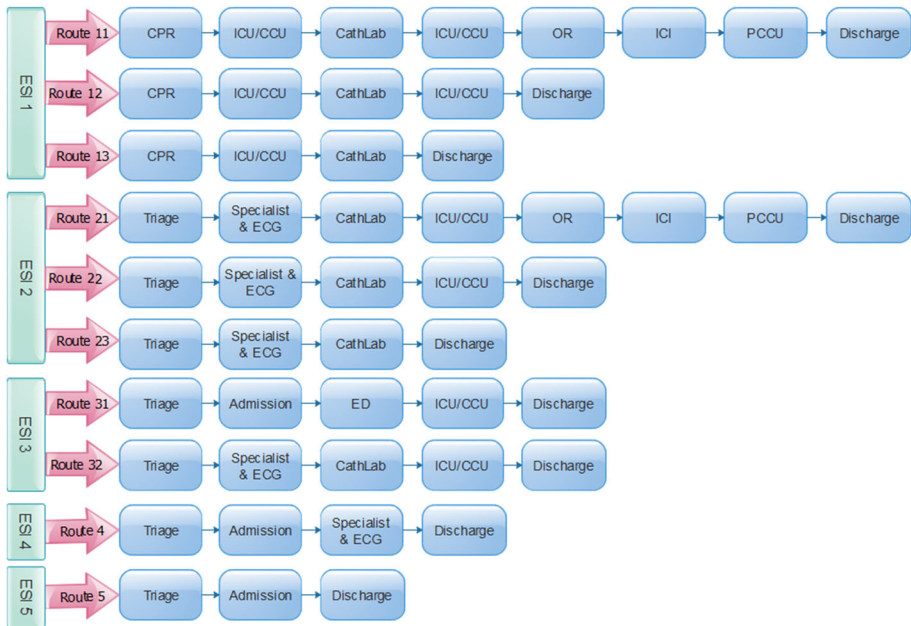


Fig. 3 Treatment routes of patients based on the ESI

4.1 Patients data clustering

According to Nyman (2007), LOS is an important performance measure for a hospital. Since patients must pay for the cost of care services they receive, the cost is an important performance measure for a hospital. Based on the gathered information of BHH, patients who underwent surgery, coronary artery bypass grafting (CABG), and primary percutaneous coronary intervention (pPCI)/PCI had longer stay and cost than other patients. Furthermore, age, gender, and blood cholesterol are important and influential factors in heart disease. By selecting age, gender, cost, LOS, CABG, pPCI/PCI, and blood cholesterol features, patients in the BHH dataset were categorized into two groups using clustering algorithms (K-means, K-medoid, Hierarchical clustering, and fuzzy C-means).

As can be seen in Table 1, hierarchical clustering with two clusters outperformed the other methods, based on both the Silhouette score (0.8520) and the Dunn index (0.4548). This is also confirmed in Fig. 4, where Silhouette and Dunn's index have the highest values (shown in boldface) for the hierarchical algorithm with two clusters.

The resulting clusters are justified in two ways. First, based on the BHH, it is very important for the hospital authorities to classify patients based on the so-called "cost class". The low-cost patients and the high-cost patients are the two classes from this point of view. Secondly, as it is shown in Table 1 and illustrated in Fig. 4, these two classes are technically confirmed using machine learning techniques.

According to Fig. 5a, cost is an appropriate feature in comparison to other features to separate the observations into two clusters. As seen in Fig. 5b, a cost boundary line of provides a separation of observations into two clusters with an accuracy of 0.99.

Table 1 Internal validation of clustering algorithms

Algorithm	Validation method	Number of clusters				
		2	3	4	5	6
Hierarchical	Dunn	0.4548	0.0678	0.0387	0.0387	0.0331
	Silhouette	0.8520	0.7030	0.6455	0.6428	0.7091
K-means	Dunn	0.0061	0.0120	0.0176	0.0188	0.0241
	Silhouette	0.7119	0.7115	0.7007	0.7069	0.6903
K-medoid	Dunn	0.0056	0.0001	0.0007	0.0008	0.0003
	Silhouette	0.7109	0.3895	0.3977	0.4164	0.3715
Fuzzy	Dunn	0.0089	0.0004	0.0001	0.0004	0.0003
	Silhouette	0.7079	0.4124	0.3662	0.2994	0.2094

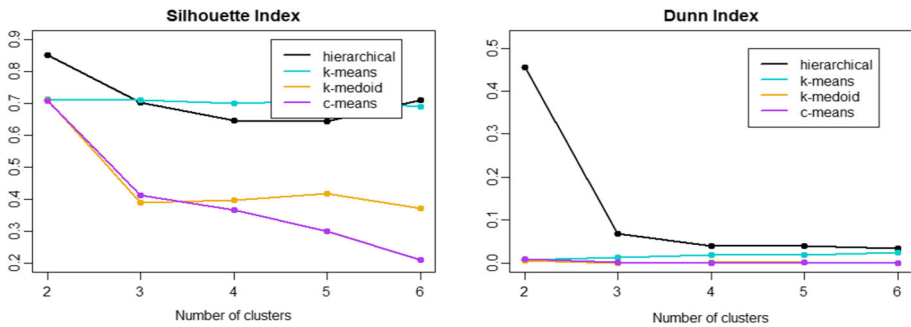


Fig. 4 Comparison of the results of clustering algorithms

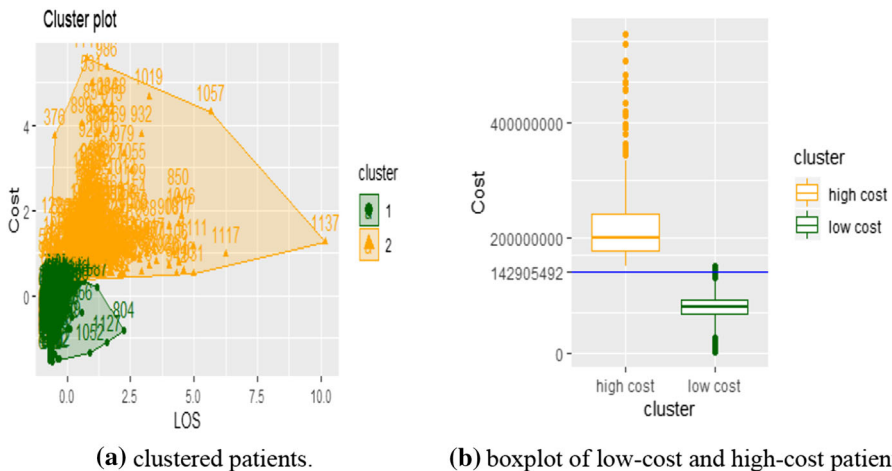
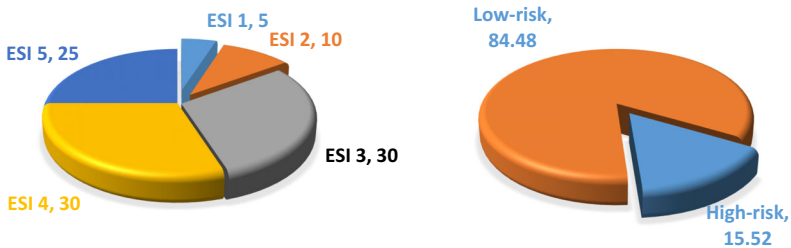


Fig. 5 Results of hierarchical clustering



Fig. 6 Medical versus modeling view of data



(a) Percentage of patients based on their ESI

(b) Percentage of high-risk and low-risk patients

Fig. 7 Percentage of patients based on ESI and categories

As shown in Fig. 6, based on the information obtained from BHH data and for modeling purpose, patients from ESI 1, ESI 2, and ESI 3 who have received services CABG or PCI / (PPCI) (or both) are regrouped as high-risk / high-cost patients (cluster 1), while the other patients are labeled as low-risk / low-cost patients (cluster 2). The higher cost of treating a patient is a consequence of being in the high-risk category. Hence, in this study, this fact has been considered as one of the useful features in differentiating high-risk and low-risk patients.

The clustering results indicate that approximately 90% of ESI 1, 70% of ESI2, and 13% of ESI3 patients are in the high-risk cluster. Figure 7a shows the percentage of patients based on their ESI as reported at the end of the previous section, while Fig. 7b shows the percentage of high-risk and low-risk patients based on the clustering approach.

4.2 Simulation input model

To create a correct simulation model, it is necessary to determine the right probability distribution function for those inputs of the model which follow random behavior. Based on historical data and graphical representations, probability distributions of the time of essential procedures were determined using classical Kolmogorov–Smirnov test. Triangular distribution provided good fit for the time of most activities, while an exponential distribution was adequate for modelling patient inter-arrival time. Table 2 shows the time distributions of activities that are essential for patients. These distributions were also verified by clinical staff.

Table 2 Input distributions for simulation model

Purpose	Distribution	<i>p</i> -value
Ambulance patient arrivals	Exponential (700)	< 0.001
Walk-in patient arrivals	Exponential (57)	< 0.001
Reception admission	Triangular (5, 10, 15)	< 0.001
Triage	Triangular (1,2,3)	< 0.005
ED	Triangular (240,300,360)	< 0.005
Electrocardiography	Triangular (5,10,25)	< 0.005
Radiology	Triangular (45,50,60)	< 0.001
Doctor / Specialist	Triangular (5,12,23)	< 0.006
Angiography	Triangular (10,20,30)	< 0.005
CPR	Triangular (45,82,120)	< 0.001
Cath Lab	Triangular (180,360,600)	< 0.001
CCU I	Triangular (4320,5040,5760)	< 0.001
CCU II	Triangular (1440,2160,2880)	< 0.001
PCCU	Triangular (1440,2880,4320)	< 0.005
ICU I	Triangular (4320, 5760, 7200)	< 0.001
Angioplasty	Triangular (15,60,90)	< 0.001
ICU II	Triangular (1440,2880,4320)	< 0.001
OR	Triangular (100,180,360)	< 0.001

4.3 Carbon emission calculation

To calculate the CFP, electricity consumption by equipment in different care units was investigated. Electricity consumption depends, almost linearly, on the amount of time the equipment is used during the treatment process. Statistical results show that the time usage of each piece of equipment follows a triangular distribution with parameters (*min*, *mode*, *max*) (see Table 8 in Appendix). Electricity consumption of equipment is taken from technical specifications, shown in Table 9 in the Appendix.

The emission factor is conventionally expressed in terms of carbon dioxide emitted for every unit of energy delivered, e.g., kilograms of carbon dioxide per kilowatt-hour ($kgCO_2/kWh$). The amount of produced $kgCO_2$ in the hospital is calculated using Eqs. (1) and (2).

$$C_i = \sum_{k=1}^K \sum_{j=1}^J (EF * T_{ijk} * W_{jk} * Z_{ijk}) \quad \forall i = 1, 2, \dots, I \quad (1)$$

$$TCO_2 = \sum_{i=1}^I C_i \quad (2)$$

where TCO_2 is the total amount of carbon dioxide produced in the hospital, C_i is the total amount of carbon dioxide produced per patient, EF is the emission factor, T_{ijk} indicates the usage time (hours) of equipment j in care unit k for patient i , W_{jk} is the rate of power consumption (kW) of equipment j in the care unit k , and Z_{ijk} equals one if equipment j is used in care unit k for patient i ; otherwise, it is zero.

To calculate the total CO₂ emitted, it is required to have the emission factor. Migone et al. (2010) estimated Greenhouse Gas (GHG) emission of the electricity generation sector for Iranian power plants and showed that Iran's national grid emission factor (EF) was 0.58, 0.62, 0.61 and 0.62 kgCO₂/kWh for years 2007, 2008, 2009 and 2010 respectively. Despite the development of the hydropower and renewable energy power plants and their shares in generated power, Iran's grid EF has not changed dramatically, mostly because of the simultaneous development of fossil fuel power plants that counterbalances this positive effect. Therefore, we used the four years weighted *EF* average of 0.61 in our models.

4.4 Patient flow simulation model

The focal point of patient flow analysis is how patients are moved throughout the treatment process and from activity to activity. The flow of a patient could vary from a simple sequence of some care services to a very complex flow with lots of decisions, branching, repetitions, and reworks. The complexity of the flow depends on the patient's conditions and uncertainties.

Patients who arrived by ambulance are categorized as ESI 1 and transferred to CPR immediately, whilst those who walk into hospital first go to the admission desk to be routed to the appropriate treatment activity. The routing is based on probabilities that were determined from historical data and observation. In the simulation model, each arriving patient is routed according to the clustering group and the ten routing schemes. Upon generating incoming patients, their associated ESI label are also generated using a probability distribution. The corresponding probability distribution is based on the number of patients of each five ESI categories within the population of patients who have visited the hospital in the last two years. Service time for each activity was randomly generated using the probability distributions presented in Table 2. In simulating each activity, the amount of CO₂ emitted is calculated using the duration that an equipment contributes to patient treatment, and the electricity consumption of the device (see Table 9 in the Appendix).

The simulation model was run for a year (365 days) and replicated 50 times to ensure that model outputs are accurate enough. To further validate the model, a t-Test was used to see whether the mean value of the simulation results was statistically different from the actual values for the year from August 2017 to July 2018. As seen in Table 3, there is no significant difference between the simulation output and the actual data.

Table 4 shows average simulation results for each patient group (low risk and high risk). According to Table 4, the average waiting time is higher for low-risk (10 min) than for high-risk (7 Minutes) while the LOS is lower (4,273 versus 2,595 min). The amount of CO₂ produced per high-risk patient and low-risk patient is 19.18 and 13.26, respectively, resulting in a total amount of carbon dioxide of 14,615 for high-risk patients and of 145,224 for low-risk patients.

Table 3 The t-test results of comparing the mean of simulation output and the actual data

Patients	Simulation output	Actual data	t-Statistics	Sig
High-risk	762	754	1.194	0.25
Low-risk	10,955	10,946	-0.635	0.54

Table 4 Average simulation results for one year

Measures	High-risk	Low-risk
Total number of patients	762	10,955
Waiting time (min.) per patient	7	10
LOS (min.) per patient	4,273	2,595
CO ₂ emitted (kg) per patient	19.18	13.26
Total CO ₂ emitted (kg)	14,615	145,224

Simulation enables us to find the best configuration among a set of predetermined scenarios. Optimization is then applied to search for an optimal configuration among many (infinite) scenarios subject to specified constraints.

4.5 Simulation-based optimization

OptQuest is a generic optimization package that treats the simulation model as a black box by considering inputs and outputs of the simulation model and combines the metaheuristics of Neural Networks (NNs), Scatter Search (SS), and Tabu Search (TS) into a single search heuristic. To optimize the hospital performance criteria (including number of patients being served, waiting time, length of stay, and amount of CO₂ produced), a mathematical model is proposed with one objective and ten constraints.

In the following optimization problem, Eq. (3) is a single-objective function f_i representing the hospital performance criterion to be optimized with $i = 1$ for number of patients served (to be maximized), $i = 2$ for waiting time, $i = 3$ for total length of stay, and $i = 4$ for total amount of CO₂ produced. For f_2, f_3 , and f_4 , the model is a minimization problem. Variables x_1, x_2, \dots , and x_8 represent the number of beds in ED, Cath lab, PCCU, CCU I, CCU II, ICU I, ICU II, and operating rooms, respectively. The value of α_i is calculated as the average value of simulation outputs over the 50 runs. For example, for $i = 2$, i.e., waiting time, α_2 is the average waiting time calculated using the output of all simulation runs. Equation (4) represents four different constraints of the optimization model. For instance, $f_2(x_j) \leq \alpha_2$ forces the optimization model to choose a solution in which the optimized total waiting time is at least as good as the simulation results. Equation (5) provides bounds on the number of beds in each care unit $x_j; j = 1, \dots, 8$ as defined by hospital authorities based on operational requirements and financial conditions and shown in Table 5. Equation (6) indicates that the mathematical model is an integer programming problem.

$$\max_{1 \leq i \leq 4} f_i(x_j) \quad (3)$$

$$f_i(x_j) \leq \alpha_i \quad j = 1, \dots, 8; \quad \forall i = 1, \dots, 4 \quad (4)$$

$$x_{jL} \leq x_j \leq x_{jU} \quad j = 1, 2, \dots, 8 \quad (5)$$

$$x_j \in Z^+ \quad j = 1, 2, \dots, 8 \quad (6)$$

The purpose of using the simulation–optimization model is to determine the number of beds in the hospital wards so that number of patients discharged from the hospital is maximized, and length of stay, waiting time, and amount of carbon produced due to the use

Table 5 Bounds on the number of beds defined by hospital

Care unit	Minimum	Average	Maximum
ED	6	7	9
ICU I	7	8	10
ICU II	4	5	7
CCU I	11	12	14
CCU II	3	4	6
PCCU	13	14	16
Cath lab	1	2	3
OR	1	2	3

of medical equipment during the treatment process are minimized. Optimizing the number of hospital beds plays an important role in improving hospital performance.

Optimized simulation outputs for both high-risk and low-risk patients are shown in Table 6. For example, the optimal value of the objective function f_1 is 11,125 for low-risk patients, while it is 776 for high-risk patients. Considering all four objective functions, and the willingness to keep a conservative approach to the number of beds, the hospital authorities have decided to set the number of beds in each unit as the maximum value suggested by the four objective functions.

Table 7 shows the percentage improvement of the objective functions $f_1, f_2, f_3,$ and f_4 for

Table 6 Optimized values of objective functions and number of beds

Patient group	f_i	Optimum	Cath lab	PCCU	CCU I	CCUII	ED	ICU I	ICU II	ORs
Low	f_1	11,125	2	14	12	4	7	8	5	2
	f_2	9	1	16	14	4	9	10	7	1
	f_3	2554	3	16	11	4	9	7	6	2
	f_4	144,436	1	14	12	5	7	10	6	1
High	f_1	776	1	13	11	3	7	7	5	1
	f_2	6	2	15	13	3	9	7	7	1
	f_3	4245	2	15	13	5	8	9	6	2
	f_4	14,429	2	13	13	3	9	7	6	1
Number of beds decided by hospital			3	16	14	5	9	10	7	2

Table 7 Percentage improvement obtained compared to current status

Patient group	(f_1)	(f_2)	(f_3)	(f_4)
low-risk patients	1.5%	11%	1.6%	0.55%
high-risk patients	1.8%	17%	0.66%	1.3%

both low-risk and high-risk patients after optimization. As it is seen, the highest improvement is in f_2 for both low-risk and high-risk patients.

5 Findings and conclusions

This study reports the successful improvement in patient flow achieved at a heart hospital in Iran. It proposed a hybrid method combining data mining and simulation–optimization approach to improve care delivery in a cardiovascular hospital.

In the data mining part, four clustering algorithms (K-means, K-medoid, hierarchical clustering, and fuzzy C-means) were applied to cluster patients based on age, gender, cost, LOS, CABG, and pPCI/PCI, features. The clustering results were evaluated using Dunn’s index and Silhouette index and showed that hierarchical clustering with two clusters performed better than the other clustering algorithms. Hence, patients were classified into two categories, namely high-risk and low-risk patients.

Then, a simulation-based methodology was applied to each cluster of patients to track performance measures of the treatment process. The OptQuest package was used to optimize number of patients being served, total waiting time, LOS, and the amount of CO₂ produced during the process. The use of simulation–optimization models was shown to be particularly valuable for identifying process improvement and quantifying the resulting improvements in hospital performance.

Considering the environmental impact of hospitals is a great challenge while maintaining a good level of care services. The proposed approach in this study helped a hospital to resolve this challenge. Although our research was applied to a specific hospital in Iran, the results are applicable to most other hospitals. It appears that other hospitals and healthcare, in general, have comparable performance measures and environmental concerns. Therefore, the problems and potential solutions described in this study would be applicable to many hospitals, worldwide.

The proposed approach could be extended from different angles. Time-dependent flows of patients could help to bridge environmental concerns with other crucial challenges such as scheduling and resource management. We could then use timed colored Petri nets to model different flow branching and resources.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Tables 8 and 9.

Table 8 Medical equipment in each unit

Department	Medical equipment	Number of electrical equipment	Demanding Time to use electrical equipment per patient (minutes)		
			Min	Ave	Max
Emergency Department	Monitoring of Vital Signs	1	1	2	3
	CPR Beds	1	45	82	120
	Monitoring of Vital Signs	1	45	82	90
	Syringe Pump	3	30	50	75
	General Motorized Suction	2	20	30	40
	Electro Shock	2	0.5	1	1.5
	Portable Ventilator	1	10	50	80
	Blood Gas Analyzer	1	3	5	7
	ED Beds	7	240	300	360
	Monitoring of Vital Signs	7	240	300	360
	Syringe Pump	7	240	300	360
	Electrocardiograph	2	3	4	5
Cath lab	Echocardiograph	1	15	17	20
	Angiography bed	1	10	20	30
	Angiographic Injector	1	1	2	3
	CARM radiography	1	10	20	30
	Angiographic specialized monitors	6	10	20	30
	Electro shock*	2	0.5	1	1.5

Table 8 (continued)

Department	Medical equipment	Number of electrical equipment	Demanding Time to use electrical equipment per patient (minutes)		
			Min	Ave	Max
Recovery	Pacemaker*	2	10	20	30
	Recovery Beds	8	180	360	420
	Syringe pump	10	180	360	420
	Monitoring of vital signs	8	180	360	420
	Angiography bed	1	15	60	90
	Angiographic Injector	1	1	2	3
	CARM radiography	1	15	60	90
	Angiographic specialized monitors	3	15	60	90
	Electro shock*	2	0.5	1	1.5
	Pacemaker*	2	40	60	90
ICU I	ICU Beds	8	4320	5760	7200
	Portable Ventilator	8	1440	2160	2880
	Motorized Suction Surgery	4	1220	1330	1440
	Monitoring of Vital Signs	8	4320	5760	7200
	Syringe pump	10	4320	5760	7200
	Mattress with air pump	8	4320	5760	7200
	Multichannel electrocardiograph	1	3	4	5
	Electro Shock*	2	0.5	1	1.5
	Balloon Pump*	1	1440	2880	7200

Table 8 (continued)

Department	Medical equipment	Number of electrical equipment	Demanding Time to use electrical equipment per patient (minutes)		
			Min	Ave	Max
ICU II	Blood Gas Analyzer	1	3	5	7
	Dialysis machine*	1	360	540	720
	Echocardiograph	3	15	17	20
	Warm touch*	8	4320	5760	7200
	Electrocardiograph	1	3	4	5
	ICU Beds	5	1440	2880	4320
	Portable Ventilator	5	1440	2880	4320
	Motorized Suction Surgery	3	5	10	20
	Monitoring of Vital Signs	5	1440	2880	4320
	Syringe pump	7	1440	2880	4320
	Mattress with air pump*	5	1440	2880	4320
	Multichannel electrocardiograph	1	3	4	5
	Electro Shock*	2	0.5	1	1.5
	Balloon Pump*	2	1440	2880	4320
	Blood Gas Analyzer	1	3	5	7
	Dialysis machine*	1	360	540	720
	Electrocardiograph	1	3	4	5
Echocardiograph	1	15	17	20	
Carotid Sono*	1	3	4	5	
Chest X ray*	1	15	18	20	

Table 8 (continued)

Department	Medical equipment	Number of electrical equipment	Demanding Time to use electrical equipment per patient (minutes)		
			Min	Ave	Max
CCU	CCU Beds	12	4320	5040	5760
	Monitoring of vital signs	12	4320	5040	5760
	Portable Ventilator*	1	1220	1330	1440
	Flowmeter O ₂ *	12	1220	1330	1440
	Multichannel electrocardiograph	1	3	4	5
	Syringe pump	12	4320	5040	5760
	Motorized Suction Surgery*	4	1	2	3
	Pacemaker*	3	4320	5040	5760
	Electro Shock*	4	0.5	1	1.5
	Mattress with air pump*	2	4320	5040	5760
	Electrocardiograph	1	3	4	5
	Echocardiograph	1	15	17	20
	Balloon Pump*	2	1440	2880	5760
	Dialysis machine*	1	360	540	720
	CCU Beds	4	1440	2160	2880
CCU II	Monitoring of vital signs	4	1440	2160	2880
	Portable Ventilator*	2	1220	1330	1440
	Flowmeter O ₂ *	2	1220	1330	1440

Table 8 (continued)

Department	Medical equipment	Number of electrical equipment	Demanding Time to use electrical equipment per patient (minutes)		
			Min	Ave	Max
PCCU	Multichannel electrocardiograph	1	3	4	5
	Syringe pump	4	1440	2160	2880
	Motorized Suction Surgery*	1	1	2	3
	Pacemaker*	2	1440	2160	2880
	Electro Shock*	2	0.5	1	1.5
	Mattress with air pump*	2	1440	2160	2880
	Electrocardiograph	1	3	4	5
	Echocardiograph	1	15	17	20
	Balloon Pump*	2	1440	2160	2880
	Dialysis machine*	1	360	540	720
	Carotid Sono*	1	3	4	5
	Chest X ray*	1	15	18	20
	PCCU Beds	14	1440	2880	4320
	Syringe pump	14	1440	2880	4320
	Motorized Suction Surgery*	2	10	15	20
Monitoring of vital signs	14	1440	2880	4320	
Multichannel electrocardiograph	1	3	4	5	
Pacemaker*	2	1440	2880	4320	

Table 8 (continued)

Department	Medical equipment	Number of electrical equipment	Demanding Time to use electrical equipment per patient (minutes)		
			Min	Ave	Max
OR	Electro shock*	1	1	1.5	2
	Mattress with air pump*	2	1440	2880	4320
	Echocardiograph*	1	15	17	20
	Dialysis machine*	1	360	540	720
	Hydraulic operating room bed	2	240	300	360
	Autoclaves	1	45	100	135
	Electric Sternum Saw	2	2	3	4
	Monitoring of Vital Signs	3	240	300	360
	Medical monitor	4	240	300	360
	Anesthesia Machine	2	240	300	360
	Syringe Pump	10	240	300	360
	Motorized Suction Surgery	4	20	30	40
	Blood Shaker	1	20	30	40
	Two-cavity ceiling flashing light	2	240	300	360
	Cerebral cortex	1	240	300	360
	Automatic coagulation timer	3	5	10	15
	Trans esophageal Echocardiogram (TEE)	1	30	60	120
	Electro Counter	1	30	40	60

Table 8 (continued)

Department	Medical equipment	Number of electrical equipment	Demanding Time to use electrical equipment per patient (minutes)		
			Min	Ave	Max
	Heart-lung machine	2	130	150	180
	Blood warmer	1	10	15	20
	Salt Set	1	20	30	40
	laryngoscope	2	1	2	3
	Pacemaker*	1	60	80	120
	Electro Shock*	3	0.5	1	1.5
	Balloon Pump*	1	240	300	360
Reception	Portable Monitor	1	5	10	15
Discharge	Portable Monitor	1	10	15	20

Symbol *: The medical equipment used for patients with special conditions

According to head nurses of each unit, of the total number of patients entering ICU I, ICU II, CCU I, CCU II, PCCU I, PCCU II, OR and Cath lab units, approximately 0.3, 0.2, 0.15, 0.1, 0.05, 0.01 and 0.01 of them need special equipment, respectively

The Cath lab is consisting of two parts, one part for coronary angiography and angioplasty and another for recovery. Time for coronary angiography and angioplasty follow triangular distribution with parameters (10, 20, 30), and (15, 60, 90), respectively

Table 9 Electricity Consumption (KWh) of medical equipment in the BHH

Department		Medical equipment	Electricity consumption (KWh)
Emergency Department	Triage	Monitoring of Vital Signs	0.14
		Monitoring of Vital Signs	0.14
		Syringe Pump	0.03
		General Motorized Suction	0.15
		Electro Shock	0.22
		Portable Ventilator	0.53
		Blood Gas Analyzer	0.02
		Monitoring of Vital Signs	0.14
		Syringe Pump	0.03
		Electrocardiograph	0.14
		Echocardiograph*	1
Cath lab	Angiography	Angiographic Injector	0.26
		CARM radiography	0.50
		Angiographic specialized monitors	0.24
		Electro shock*	0.22
		Pacemaker*	0.01
		Recovery	Syringe pump
		Monitoring of vital signs	0.14
	Angioplasty	Angiographic Injector	0.26
		CARM radiography	0.50
		Angiographic specialized monitors	0.24
		Electro shock*	0.22
Pacemaker*		0.01	
ICU	ICU I	Portable Ventilator	0.53
		Motorized Suction Surgery	0.15
		Monitoring of Vital Signs	0.14
		Syringe pump	0.03
		Mattress with air pump	0.02
		Multichannel electrocardiograph	0.02
		Electro Shock*	0.22
		Balloon Pump*	0.50
		Blood Gas Analyzer	0.02
		Dialysis machine*	2
		Echocardiograph	1
	Warm touch*	1.2	
	Electrocardiograph	0.14	
	ICU II	Portable Ventilator	0.53
		Motorized Suction Surgery	0.15

Table 9 (continued)

Department	Medical equipment	Electricity consumption (KWh)	
CCU	CCU I	Monitoring of Vital Signs	0.14
		Syringe pump	0.03
		Mattress with air pump*	0.02
		Multichannel electrocardiograph	0.02
		Electro Shock*	0.22
		Balloon Pump*	0.50
		Blood Gas Analyzer	0.02
		Dialysis machine*	2
		Electrocardiograph	0.14
		Echocardiograph	1
		Carotid Sono*	1
		Chest X ray*	1.43
		Monitoring of vital signs	0.14
		Portable Ventilator*	0.53
		Flowmeter O ₂ *	2.20
		Multichannel electrocardiograph	0.14
	Syringe pump	0.03	
	Motorized Suction Surgery*	0.15	
	Pacemaker*	0.01	
	Electro Shock*	0.22	
	Mattress with air pump*	0.02	
	Electrocardiograph	0.14	
	Echocardiograph	1	
	Balloon Pump*	0.50	
	Dialysis machine*	2	
	CCU II	Monitoring of vital signs	0.14
		Portable Ventilator*	0.53
		Flowmeter O ₂ *	13.2
		Multichannel electrocardiograph	0.02
		Syringe pump	0.03
		Motorized Suction Surgery*	0.15
		Pacemaker*	0.01
Electro Shock*		0.22	
Mattress with air pump*		0.02	
Electrocardiograph		0.14	
Echocardiograph		1	
Balloon Pump*		0.50	
Dialysis machine*		2	

Table 9 (continued)

Department	Medical equipment	Electricity consumption (KWh)
PCCU	Carotid Sono*	1
	Chest X ray*	1.43
	Syringe pump	0.03
	Motorized Suction Surgery*	0.15
	Monitoring of vital signs	0.14
	Multichannel electrocardiograph	0.03
	Pacemaker*	0.01
	Electro shock*	0.22
	Mattress with air pump*	0.02
	Echocardiograph*	1
OR	Dialysis machine*	2
	Autoclaves	1
	Electric Sternum Saw	1
	Monitoring of Vital Signs	0.14
	Medical monitor	0.22
	Anesthesia Machine	0.02
	Syringe Pump	0.03
	Motorized Suction Surgery	0.15
	Blood Shaker	0.02
	Two-cavity ceiling flashing light	0.17
	Cerebral cortex	0.12
	Automatic coagulation timer	6.50
	Trans esophageal Echocardiogram (TEE)	1
	Electro Counter	0.85
	Heart–lung machine	0.22
	Blood warmer	0.28
	Salt Set	3.52
	laryngoscope	2.20
	Pacemaker*	0.01
	Electro Shock*	0.22
Balloon Pump*	0.50	
Reception	Portable Monitor	0.21
Discharge	Portable Monitor	0.21

References

- Abo-Hamad, W., & Arisha, A. (2013). Simulation-based framework to improve patient experience in an emergency department. *European Journal of Operational Research*, 224(1), 154–166.
- Absi, N., et al. (2013). Lot sizing with carbon emission constraints. *European Journal of Operational Research*, 227(1), 55–61.
- Amaran, S., et al. (2016). Simulation optimization: A review of algorithms and applications. *Annals of Operations Research*, 240(1), 351–380.
- Asaduzzaman, M., Chausaulet, T. J., & Robertson, N. J. (2010). A loss network model with overflow for capacity planning of a neonatal unit. *Annals of Operations Research*, 178(1), 67–76.
- Azuaje, F. (2002). A cluster validity framework for genome expression data. *Bioinformatics*, 18(2), 319–320.
- Bachhety, S., Kapani, S., & Jain, R. (2021). Big Data Analytics for healthcare: Theory and applications. In A. K. D. G. N. Dey (Ed.), *Applications of big data in healthcare* (pp. 45–67). Academic Press.
- Badri, H., Bashiri, M., & Hejazi, T. H. (2013). Integrated strategic and tactical planning in a supply chain network design with a heuristic solution method. *Computers & Operations Research*, 40(4), 1143–1154.
- Bi, P., & Hansen, A. (2018a). Carbon emissions and public health: An inverse association? *The Lancet Planetary Health*, 2(1), e8–e9.
- Bi, P., & Hansen, A. (2018b). Carbon emissions and public health: An inverse association? *The Lancet Planetary Health*, 2(1), e8–9.
- Bruno, G., et al. (2014). A clustering-based approach to analyse examinations for diabetic patients. In *2014 IEEE International Conference on Healthcare Informatics*. IEEE.
- Cabrera, E., et al. (2012). Simulation optimization for healthcare emergency departments. *Procedia Computer Science*, 9, 1464–1473.
- Ceglowski, R., Churilov, L., & Wasserthiel, J. (2016). Combining data mining and discrete event simulation for a value-added view of a hospital emergency department. *Operational research for emergency planning in healthcare* (Vol. 1, pp. 119–138). Springer.
- Chaudhuri, S. (1998). Data mining and database systems: Where is the intersection? *IEEE Database Engineering Bulletin*, 21(1), 4–8.
- Chevalier, F., Gareil, P., & Levitan, J. J. D. P. (2009) *Hospitals in the 27 Member States of the European Union*. Codrington-Virtue, A., et al. (2006). A system for patient management based discrete-event simulation and hierarchical clustering. In *19th IEEE Symposium on computer-based medical systems (CBMS'06)*. IEEE.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. Wiley.
- Eckelman, M. J., & Sherman, J. (2016). Environmental impacts of the US health care system and effects on public health. *PLoS ONE*, 11(6), e0157014.
- Eckelman, M. J., Sherman, J. D., & MacNeill, A. J. J. P. M. (2018). Life cycle environmental emissions and health damages from the Canadian healthcare system: An economic-environmental-epidemiological analysis. *PLoS Medicine*, 15(7), 1002.
- El Cadi, A. A., et al. (2015). A joint optimization-simulation model to minimize the makespan on a repairable machine. In *2015 International conference on industrial engineering and systems management (IESM)*. IEEE.
- Fetter, R. B., & Thompson, J. D. (1965). The simulation of hospital systems. *Operations Research*, 13(5), 689–711.
- Frumkin, H., et al. (2008). Climate change: The public health response. *American Journal of Public Health*, 98(3), 435–445.
- Ghassemi, M., Celi, L. A., & Stone, D. J. (2015). State of the art review: The data revolution in critical care. *Critical Care*, 19(1), 1–9.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining (adaptive computation and machine learning)*. MIT Press.
- Iavindrasana, J., et al. (2009). Clinical data mining: A review. *Yearbook of Medical Informatics*, 18(01), 121–133.
- Ibrahim, N. H., et al. (2013). A hybrid model of hierarchical clustering and decision tree for rule-based classification of diabetic patients. *International Journal of Engineering and Technology*, 5, 2013.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323.
- Karimi, S., et al. (2015). Text and data mining techniques in adverse drug reaction detection. *ACM Computing Surveys (CSUR)*, 47(4), 1–39.
- Kasaie, P., & Kelton, W. D. (2013). Simulation optimization for allocation of epidemic-control resources. *IIE Transactions on Healthcare Systems Engineering*, 3(2), 78–93.

- Kaya, M.-F. & Schoop, M. (2019). *Application of data mining methods for pattern recognition in negotiation support systems*. In *International conference on group decision and negotiation*. Springer.
- Klassen, K. J., & Yoogalingam, R. (2009). Improving performance in outpatient appointment services with a simulation optimization approach. *Production and Operations Management*, 18(4), 447–458.
- Li, S. L. & Wang, C. H. (2012). *Analysis for quick response strategy using OptQuest simulation*. In *Applied Mechanics and Materials*. 2012. Trans Tech Publ.
- Lin, R.-C., Sir, M. Y., & Pasupathy, K. S. (2013). Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: Specific application to determining optimal resource levels in surgical services. *Omega*, 41(5), 881–892.
- Liu, Y., et al. (2010). *Understanding of internal clustering validation measures*. In *2010 IEEE international conference on data mining*. IEEE.
- Liu, C.-H. (2014). Approximate trade-off between minimisation of total weighted tardiness and minimisation of carbon dioxide (CO₂) emissions in bi-criteria batch scheduling problem. *International Journal of Computer Integrated Manufacturing*, 27(8), 759–771.
- Mahoto, N. A., Shaikh, F. K., & Ansari, A. Q. (2014). Exploitation of clustering techniques in transactional healthcare data. *Mehran University Research Journal of Engineering & Technology*, 33(1), 77–92.
- Mallor, F., & Azcárate, C. (2014). Combining optimization with simulation to obtain credible models for intensive care units. *Annals of Operations Research*, 221(1), 255–271.
- Mannila, H. (1996). *Data mining: Machine learning, statistics, and databases*. In *Proceedings of 8th international conference on scientific and statistical data base management*. IEEE.
- Migone, M. B., et al. (2010) *Emission factor calculation of Iran's grid connected power plants*. Rahbord Energy (REC): Tehran, Iran.
- Na, S., Xumin, L. & Yong, G. (2010). *Research on k-means clustering algorithm: An improved k-means clustering algorithm*. In *2010 Third International Symposium on intelligent information technology and security informatics*. 2010. IEEE.
- Navale, G., et al. (2016). Prediction of stock market using data mining and artificial intelligence. *International Journal of Computer Applications*, 134(12), 9–11.
- Ng, A. H., et al. (2011). Simulation-based innovation using data mining for production systems analysis. *Multi-objective evolutionary optimisation for product design and manufacturing* (pp. 401–429). Springer.
- Nyman, M.A. *Patient flow: Reducing delay in healthcare delivery*. In *Mayo Clinic Proceedings*. Elsevier.
- Organization, W. H. (2014) *Quantitative risk assessment of the effects of climate change on selected causes of death, 2030s and 2050s*.
- Osorio, A. F., et al. (2017). Simulation-optimization model for production planning in the blood supply chain. *Health Care Management Science*, 20(4), 548–564.
- Pinzone, M., Lettieri, E., & Masella, C. (2012). Sustainability in healthcare: Combining organizational and architectural levers. *International Journal of Engineering Business Management*, 4, 38.
- Pollard, A. S., et al. (2013). Mainstreaming carbon management in healthcare systems: A bottom-up modeling approach. *Environmental Science & Technology*, 47(2), 678–686.
- Prokosch, H.-U., & Ganslandt, T. (2009). Perspectives for medical informatics. *Methods of Information in Medicine*, 48(01), 38–44.
- Rivera-Gómez, H., et al. (2016). Production control problem integrating overhaul and subcontracting strategies for a quality deteriorating manufacturing system. *International Journal of Production Economics*, 171, 134–150.
- Schulz, M., Romppel, M., & Grande, G. (2016). Built environment and health: A systematic review of studies in Germany. *Journal of Public Health*, 40(1), 8–15.
- Sheridan, S., et al. (2011). Heat-related mortality and heat watch-warning systems in the United States: Recent developments. *Epidemiology*, 22(1), S13.
- Sigurdardottir, A. K., Jonsdottir, H., & Benediktsson, R. (2007). Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. *Patient Education and Counseling*, 67(1–2), 21–31.
- Strome, T. L., & Liefer, A. (2013). *Healthcare analytics for quality and performance improvement*. Wiley.
- Sufi, F., & Khalil, I. (2010). Diagnosis of cardiovascular abnormalities from compressed ECG: A data mining-based approach. *IEEE Transactions on Information Technology in Biomedicine*, 15(1), 33–39.
- Sun, J. & Reddy, C. K. (2013). *Big data analytics for healthcare*. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*.
- Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. *International Journal of Bio-Science and Bio-Technology*, 5(5), 241–266.
- Viccellio, R. S. R. V. J. S. W. M. A. (2017). Emergency department (ED) overcrowding: Evidence-based answers to frequently asked questions. *Revista Médica Clínica Las Condes*, 28(2), 213–219.
- Wang, L., et al. (2003). Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12), 1505–1518.

- Xu, T., et al. (2016). *A joint sparse clustering and classification approach with applications to hospitalization prediction*. In *2016 IEEE 55th conference on decision and control (CDC)*. IEEE.
- Yang, J.-J., et al. (2015). Emerging information technologies for enhanced healthcare. *Computers in Industry*, *69*, 3–11.
- Yoo, I., et al. (2012). Data mining in healthcare and biomedicine: A survey of the literature. *Journal of Medical Systems*, *36*(4), 2431–2448.
- Zhang, H., et al. (2020). Simulation-based optimization to improve hospital patient assignment to physicians and clinical units. *Health Care Management Science*, *23*(1), 117–141.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.