



Machine learning based integrated scheduling and rescheduling for elective and emergency patients in the operating theatre

Masoud Eshghali¹ · Devika Kannan^{2,3}  · Navid Salmanzadeh-Meydani^{2,4} · Amir Mohammad Esmaeeli Sikaroudi⁵

Accepted: 4 January 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

As the only largest source of revenue and cost in a hospital, the operation room (OR) scheduling problem is a hot research topic. Nonetheless, an integrated model is the missing key to managing and improving the efficiency of ORs. This paper presents a fully integrated model regarding three concepts: meditating elective patients and emergency patients together, considering ORs and downstream units, and proposing hierarchical weekly, daily, and rescheduling models. Due to the inherent randomness in emergency patient arrival, a random forest machine learning model and geographical information systems are used to obtain the emergency patient surgery duration and arrival time, respectively. According to the machine learning model in weekly and daily scheduling, initially, fixed capacity is reserved for emergency patients. When an emergency patient arrives, the surgery starts if a reserved OR is available. Otherwise, the first available OR will be dedicated to the patient due to an emergency patient's higher priority than an elective patient. In this case, it is needed to reschedule the OT schedule for the remaining patient. Moreover, the three-phase model guarantees that an emergency patient assigns to an OR within a specific time limit. To solve the models, genetic algorithm and particle swarm optimization are developed and compared. In addition, a real-world case study is undertaken at a hospital. The results of comparing the proposed approach to the hospital's current scheduling show that the three-phase model had a considerable positive effect on the ORs schedule.

Keywords Elective and emergency patients · Operating theater scheduling · Rescheduling · Operating room planning · Machine learning

✉ Devika Kannan
deka@iti.sdu.dk

¹ Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721, USA

² Centre for Sustainable Supply Chain Engineering, Department of Technology and Innovation, University of Southern Denmark, 5230 Odense M, Denmark

³ School of Business, Woxsen University, Sadasivpet, Telangana, India

⁴ Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

⁵ Department of Computer Science, University of Arizona, Tucson, AZ 85721, USA

1 Introduction

Health care systems constitute one of the most critical service sectors in any society. According to the World Health Organization, the average global health expenditure per capita has more than doubled (from US\$480 in 2000 to US\$1111 in 2018). Moreover, in March 2020, the World Health Organization proclaimed the coronavirus disease (COVID-19), a global pandemic (Ali & Kannan, 2022; Fattahi et al., 2023; Sohrabi et al., 2020; Ferreira et al., 2022). As a result, the need for hospital resource such as intensive care units (ICU) and wards has increased significantly over the world. In addition, aging phenomena put emphasis on the importance of health care systems, signaling its growing need for attention and planning. Clearly, hospitals play a critical role in health systems and operating rooms (ORs) are the most expensive services as they deploy a great number of precious resources such as surgeons, staff, and equipment (Macario, 2010; van Oostrum et al., 2008). In addition, ORs are the foremost source of annual income and have a direct effect on patient safety (Liu et al., 2011). Then, it is conceivable that any improvements in OR productivity could lead to a more productive hospital and better patient satisfaction (Durán et al., 2017). Accordingly, integrated planning and scheduling can be the capstone to managing and improving the efficiency of hospitals.

In the real world, patients, after being operated in ORs, frequently stay in downstream units, such as the recovery unit or post-anesthesia care unit (PACU), ICU, CCU, and wards, which are also limited (Fügenger et al., 2014). As a result, considering just ORs capacity does not result in practical scheduling for patients, since the lack of enough downstream units keeps patients from moving forward, significantly decreasing ORs utilization. For instance, when there are not enough available beds in recovery units for incoming patients after surgery, some of them must remain in ORs until one recovery bed becomes available.

Elective patients can stay on a waiting list and be scheduled in advance. However, emergency patients must be operated on as soon as possible. Therefore, the arrival of emergency patients may lead to delays in elective patients surgeries because of their priority. A great number of hospitals deal with emergency surgeries problem by reserving some operating theater (OT) capacity. There are three ways to do it (Van Riet & Demeulemeester, 2015): (1) reserving some ORs' capacity (reserved-OR), (2) use elective patients OR (shared-OR) (3) a combination of (1) and (2) (van Essen et al., 2012). In situation (1), emergency patients are operated on in the reserved OR, but they should wait for the reserved OR to become available if the reserved ORs are busy. In this situation, the utilization of ORs is chronically low. In situation (2), emergency patients are operated on in the first available OR. In this strategy, although the utilization of OR is reasonable, the high rate of elective patient surgery cancellation is undeniable. Situation (3) is a combination of situations (1) and (2), which means that the emergency patient is operated on at once if the reserved OR is available. Otherwise, they have to wait until the reserved ORs or one of the elective ORs becomes available. van Veen-Berkx et al. (2016) demonstrated that strategy 1 works better than strategy 2 in terms of utilization, overtime, and case cancellations.

Due to the inherent randomness in emergency patient arrival, historical data and geographical information systems are applied to obtain the emergency patient arrival time. Therefore, proper forecasting of this process can be used as a basis for management to better allocate OTs resources. Machine learning is an artificial intelligence subtype that employs algorithms, learning large quantities of data iteratively but not explicitly programmed to do so (Obermeyer & Emanuel, 2016). They can extract schematics from various data sources, explain

them, and create a predictive model (Rajkomar et al., 2019). Machine learning models can be a proper method for forecasting emergency patients' arrivals time and surgery duration.

The contributions of this paper are summarized as follows:

- The fully integrated model is developed for the OTs. The proposed model is integrated into three concepts:
 1. Considering elective patients and emergency patients,
 2. Regarding ORs and downstream units.
 3. Proposing hierarchical weekly, daily, and rescheduling mathematical models.
- The model ensures that the emergency patients will be assigned to an OR at most within the specified amount of time.
- OTs' capacity is reserved for emergency patients while multi-shared-OR scheduling is considered for emergency patients.
- Since cancelling patients' surgery may result in dissatisfaction and anxiety, the rescheduling model ensures that their surgery day will not change.
- Geographical information systems and random forest machine learning models are modified to predict the arrival time of emergency patients and patient surgery duration, respectively.

The remainder of this paper is organized as follows. Section 2 presents a detailed literature review. In Sect. 3, the problem definition and three-phase model are presented. In Sect. 4 solution methods are developed. Section 5 describes case study. Sections 6 and 7 provides computational results and discussion, respectively. Eventually, The conclusion is summarized in Sect. 8.

2 Literature review

The scheduling of OT is a broad topic. A detailed literature review can be found in (Cardoen et al., 2009; Guerriero & Guido, 2011; Rahimi & Gandomi, 2021; Van Riet & Demeulemeester, 2015; Zhu et al., 2019). The weekly scheduling (Di Martinelly & Meskens, 2017; Fei et al., 2009, 2010; Luo & Wang, 2019; Soudi et al., 2019), the daily scheduling (Aringhieri et al., 2015; Atighehchian et al., 2020; Gauthier & Legrain, 2016; Guido & Conforti, 2017; Guido et al., 2018; Riise et al., 2016; Varmazyar et al., 2020), and the rescheduling (Addis et al., 2014, 2016; Akbarzadeh et al., 2019; Gul et al., 2015) are well studied in the literature. Therefore, the literature focuses here that consider emergency patients and elective patients in OTs scheduling or applying machine learning models for OTs planning.

Soudi et al. (2019) benefitted from a hybrid flow workshop scheduling formulation for ORs weekly planning. Also, they applied a chance-constrained model for the arrival of emergency patients. Erdem et al. (2012) formulated a mixed-integer linear programming model to reschedule elective surgeries on the arrival of emergency patients. They considered ORs and recovery units to minimize the cost of overtime, postponing surgeries, and emergency patients' rejections. Essen et al., (2012) developed a new concept to minimize the emergency patient surgery waiting time. They introduced the break-in-moment (BIM), which is where an emergency surgery could start an operation, and the break-in-interval, which defines as the interval between two consecutive BIMs concepts. Zonderland et al. (2010) considered semi-urgent surgeries and elective patients simultaneously for balancing the OR idle time and canceling elective surgeries. Lamiri et al. (2007) studied a stochastic optimization model to minimize elective patients' and ORs' overtime costs. They tried to schedule elective patients

in OR and reserve a random capacity for emergency patients. Bargetto et al. (2018) provided a dynamic optimization model to balance emergency patients' waiting times, the delay of elective patient surgery, and the overtime of staff. They considered emergency patients with different emergency levels to arrive as a nonhomogeneous Poisson process. Meersman and Maenhout (2022) analyzed the capacity distribution between elective and emergency patients to optimize patient appointment waiting times. Kroer et al. (2018) examined elective and emergency surgeries with uncertain duration aimed at reducing OR and overtime costs.

A stochastic programming model was developed by Farzad and Mohammad (2016) to represent the problem, taking into account priority rules based on moral values, and aiming to reduce waiting time and overtime for elective surgeries. Arab Momeni et al. (2022) proposed a mixed-integer mathematical programming model to assign COVID-19 as well as non-COVID-19 patients and specialty teams to ORs. Wang et al. (2016) considered uncertainty in surgical durations as well as emergency patient arrivals time and utilized a discrete simulation model. Jebali and Diabat (2017) considered the duration of patient surgery and the length of stay in ICU as random variables and formed a two-stage chance-constrained model for emergency and elective patients in the advance scheduling problem. Bruni et al. (2015) proposed a stochastic programming framework with three recourse strategies for the uncertainty of arrivals of emergency patients and surgery duration. Silva and de Souza (2020) considered approximate dynamic programming with an integer model to monitor daily scheduling for emergency and elective patients.

Tuwatananurak et al. (2019) applied a proprietary machine learning algorithm, leap Rail, to estimate the duration of patient surgery. The algorithm learned how to predict based upon a data set of 15,000 surgery cases. Fairley et al. (2019) used the machine learning algorithm to forecast the time of the PACU. As a result, total PACU holdings have been reduced, leading to substantial cost savings. Schiele et al. (2021) considered a three-step model for master surgery scheduling, in which they integrated the ORs and downstream units. They predict ICU bed occupancy based on a neural network model. Shuvo et al. (2020) proposed a deep reinforcement learning-based policy for inadequate beds in the Bangladesh hospital. They applied the machine learning algorithm, Advantage Actor-Critic (A2C), to reduce the cost.

Surgery cancellations are a cause of ineffective use of OTs. Luo and Wang (2019) used machine learning models to identify the surgeries with high risks of cancellation. They applied random forest, support vector machine, and XGBoost models and found that the random forest model attained the greatest performance. These new techniques make it possible for OT managers to provide preventive measures to reduce cancellation rates in OTs. Luo et al. (2016) applied techniques of machine learning to predict the cancellation of surgery at the West China Hospital. Models such as bagging, boosting, Bayesian additive regression trees, and the random forest were used. Erekat et al. (2020) examined four different standard data mining techniques to predict surgical cancellations for successful schedules of surgical patients. The prediction of surgery cancellation leads to efficient scheduling that eventually leads to cost reductions. Another example is the use of these new technologies is in robotic surgery, which involves significant costs. Zhao et al. (2019) considered 500 robotic surgeries and examined 28 variables. They applied a number of supervised machine learning techniques, including neural networks and random forest. As mentioned, the machine learning model seems to be very helpful in the OT schedule.

According to the reviewed literature, many authors have ignored downstream units and regarded them as limitless resource, which is not conceivable in the real world. Although some resources have been addressed in prior papers, the lack of sufficient downstream units prevents patients from going forward and results in improper realistic scheduling. In addition, while many authors have focused on elective patients, it is vital to take all patients into

consideration for complete and effective planning and scheduling. In this study, elective and emergency patients are taken into account in ORs and downstream units simultaneously. Also, a three-phase model, weekly and daily scheduling as well as rescheduling, is proposed to integrate all the decisions in the OTs department. To the best of our knowledge, such an integrated model has not been discussed in the literature which can be completely useful for OT management. Also, geographical information systems are utilized for to forecast the emergency patient arrival time. In addition, a random forest machine learning model is used to predict emergency patients' surgery duration.

3 Problem definition and formulation

This paper develops an approach for scheduling and rescheduling for both elective and emergency patients in OTs. OR capacities are divided into some blocks, which are assigned in advance. For this purpose, a three-phase model is extended to schedule patients weekly and daily as well as to reschedule patients after arriving emergency patients if needed. The Three-phase model is presented hierarchically, where the results of each phase are the input for the subsequent phases. Figure 1 demonstrates the flowchart of the scheduling process.

Phase 1 is weekly scheduling in which the scheduler needs to allocate OTs to surgical groups for the next week. In the same way, he or she reserves some OT capacity for emergency patients according to their arrival rate. Phase 2 studies daily scheduling in which the exact time and sequencing of elective patient's surgery are determined within the assigned blocks for each patient. Finally, phase 3 implies rescheduling. In this phase, when an emergency patient arrives, he or she needs to be scheduled within a limited specific time. This may disrupt a subset of daily scheduled elective surgeries, and then it is needed to be rescheduled.

The assumptions are as follows:

- Elective and emergency patients are considered together in OT, composed of multiple ORs, recovery, ICU, CCU, and ward units.
- Some capacity in OTs is reserved for emergency patients, and they can also use elective patients ORs if needed.
- All patients are transferred to the recovery unit after surgery, and they are transferred to one of the ICU, CCU, or ward. Patients from the ICU and CCU are also transferred to the ward unit after they have passed the specific conditions.
- The duration of the patient's stay in the OR and the recovery unit are measured in the sum of the number of blocks and sub-block. The duration of the patient's stay in other downstream parts is measured in the sum of the number of days.
- The ORs are opened at 8 AM and closed at 4 PM.
- Each day is divided into four blocks, and each block is divided into ten sub-blocks.

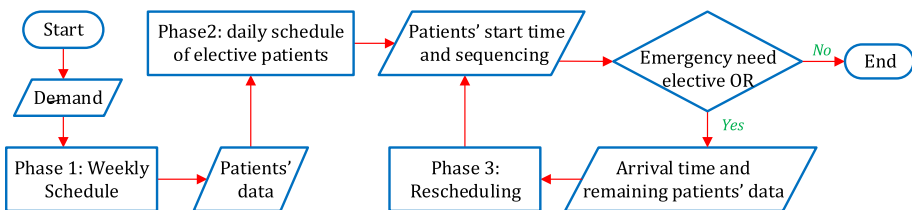


Fig. 1 Flowchart of the scheduling process

- The start time of each surgery can be changed, yet the day of operation is unchangeable.

3.1 Weekly scheduling

Before developing weekly scheduling, the arrival patterns of emergency patients must be forecasted to reserve some OT capacities. The number of reserved ORs can be different from one block to another since their arrival patterns can be different.

Indices and sets:

O : Set of ORs indexed by o

I : Set of surgeries days indexed by i

J : Set of surgeries blocks indexed by j

S : Set of surgeries groups indexed by s

Parameters:

L_s : The length of average surgery duration of the surgical group s

LE : The length of average surgery duration of the emergency patient

LR_s : The length of stay in the recovery room of the surgical groups

LI_s : The length of stay in the ICU of the surgical group s

LC_s : The length of stay in the CCU of the surgical group s

LW_s : The length of stay in the ward unit of the surgical group s

BR_i : The number of available beds in the recovery room on day i

BI_i : The number of available beds in ICU on day i

BC_i : The number of available beds in CCU on day i

BW_i : The number of available beds in the ward unit on day i

S_{ijs} : The number of surgeons from the surgical group s on day i and block j

D_s : Demand for surgical group s

DE_{ij} : Demand for emergency patients on day i and block j

Decision variables:

X_{ijos}^w : A binary variable, equal to 1, if block j on the day i from the OR o assigns to the surgical group s ; 0, otherwise

Y_{ijs}^w : An integer variable equal to the number of assigned blocks to the surgical group s on the day i on block j

Z_{ijo}^w : A binary variable, equal to 1, if block j on the day i from the OR o assigns to the emergency patients; 0, otherwise

Q_{is}^w : An integer variable, equal to the number of assigned blocks to the surgical group s on day i

The mathematical formulation of the weekwngly scheduling is as follows:

$$\min \sum_i \sum_j \sum_o \sum_s j \times X_{ijos}^w \quad (1)$$

subject to:

$$\sum_o X_{ijos}^w \leq S_{ijs} \quad \forall i, j, s \quad (2)$$

$$\sum_s X_{ijos}^w + Z_{ijo}^w \leq 1 \quad \forall i, j, s \quad (3)$$

$$\sum_i \sum_j \sum_o X_{ijos}^w \geq D_s \times L_s \quad \forall s \tag{4}$$

$$\sum_o Z_{ijo}^w \geq DE_{ij} \times LE \quad \forall i, j \tag{5}$$

$$\sum_o X_{ijos}^w \leq Y_{ijs}^w \quad \forall i, j, s \tag{6}$$

$$\sum_j Y_{ijs}^w \leq Q_{is}^w \quad \forall i, s \tag{7}$$

$$\sum_s \left(Y_{ijs}^w + Y_{i(j-1)s}^w + \dots + Y_{i(j-LR_s+1)s}^w \right) \times \left(1/L_s \right) \leq BR_i \quad \forall i, j \tag{8}$$

$$\sum_s \left(Q_{is}^w + Q_{(i-1)s}^w + \dots + Q_{(i-LI_s+1)s}^w \right) \leq BI_i \quad \forall i \tag{9}$$

$$\sum_s \left(Q_{is}^w + Q_{(i-1)s}^w + \dots + Q_{(i-LC_s+1)s}^w \right) \leq BC_i \quad \forall i \tag{10}$$

$$\sum_s \left(Q_{(i-LI_s-LC_s)s}^w + Q_{(i-LI_s-LC_s-1)s}^w + \dots + Q_{(i-LI_s-LC_s-LW_s+1)s}^w \right) Q_{(i-LI_s+1)s}^w \leq Bw_i \quad \forall i \tag{11}$$

$$X_{ijos}^w, Z_{ijo}^w \in \{0, 1\}, Y_{ijs}^w, Q_{is}^w \in int \tag{12}$$

The objective function (1) strives to schedule a surgical group each day as soon as possible. Constraint (2) allows the assigning of OR to the surgical groups when a needed surgeon is available. Constraint (3) assures that each block is assigned to one patient or left empty. Constraints (4) and (5) ensure that each surgical group’s total assigned time must be more than the needed time. Constraints (6) and (7) show the number of surgery in each block and day. Constraints (8–11) consider the integration of downstream units to ORs. Constraint (12) shows the domain of variables.

3.2 Daily scheduling

By the weekly scheduling, it is determined that each block from each OR is assigned to which surgical group. A daily scheduling model is developed to find the optimal sequence of patients’ surgery. When an emergency patient arrives, the patient will be supported in the emergency department with initial treatments before transferring to an OR. This transfer must be done within a limited amount of time. The following notations are used for the daily scheduling mathematical modeling in addition to the explained notations of the weekly scheduling:

Indices and sets:

K: Set of sub-blocks indexed by *k*

P: Set of the patients from the surgical groups indexed by *p*

Parameters:

X_{ijos}^w : The output of the weekly scheduling indicating block *j* of operation room *o* on the day *i* assigns to surgical group *s*

Z_{ijok}^w : The output of the weekly scheduling indicating block *j* of operation room *o* on the day *i* assigns to emergency group *k*

LO_{sp} : The surgery duration of patient p from the surgical group s

W_{sp} : Priority of patient p from the surgical group s

C_{ijk} : A factor in the objective function whose purpose is to put the operations in the earliest possible time ($C_{ijk} = c_i \times i + c_j \times j + c_k \times k$ where $c_i > c_j > c_k$)

R_{ijk} : The number of elective patients ORs that could potentially assign to the emergency patient if needed at this time (day i , block j , and sub-block k)

Decision variables:

X_{ijkosp}^d : A binary variable, equal to 1, if sub-block k from block j on the day i from the OR o assigns to patient p from the surgical group s ; 0, otherwise

ST_{ijkosp}^d : A binary variable, equal to 1, if sub-block k from block j on the day i from the OR o assigns to patient p from the surgical group s and the patient's surgery starts in this sub-block; 0, otherwise

EM_{ijkosp}^d : A binary variable, equal to 1, if sub-block k from block j on the day i from the OR o from the surgical group s is empty; 0, otherwise

The mathematical formulation of the daily scheduling is as follows:

$$\min \sum_i \sum_j \sum_k \sum_o \sum_s \sum_p C_{ijk} \times W_{sp} \times ST_{ijkosp}^d \quad (13)$$

subject to :

$$X_{ijkosp}^d + EM_{ijkosp}^d \leq X_{ijos}^w \forall i, j, k, o, s, p \quad (14)$$

$$\sum_p X_{ijkosp}^d + EM_{ijkosp}^d \leq 1 \forall i, j, k, o, s \quad (15)$$

$$ST_{ijkosp}^d \leq X_{ijkosp}^d \forall i, j, k, o, s, p \quad (16)$$

$$\sum_i \sum_j \sum_k \sum_o ST_{ijkosp}^d = 1 \forall s, p \quad (17)$$

$$\sum_o \left(Z_{ijok}^w + \sum_s \left(\sum_p ST_{ijkosp}^d + EM_{ijkosp}^d \right) \right) \geq R_{ijk} \forall \{i, j, k\} \quad (18)$$

$$ST_{ijkosp}^d \times LO_{sp} \leq \left(ST_{i(jk)osp}^d + ST_{i(jk+1)osp}^d + \dots + ST_{i(jk+LO_{sp}-1)osp}^d \right) \forall i, j, k, s, p, o \quad (19)$$

$$\sum_o \sum_s \sum_p \left(ST_{i(jk-LO_{sp})osp}^d + ST_{i(jk-LO_{sp}-1)osp}^d + \dots + ST_{i(jk-LO_{sp}-LR_{sp}+1)osp}^d \right) \leq BR_i \forall i, j, k \quad (20)$$

$$\sum_j \sum_k \sum_o \sum_s \sum_p \left(ST_{ijkosp}^d + ST_{(i-1)jkosp}^d + \dots + ST_{(i-LI_{sp}+1)jkosp}^d \right) \leq BI_i \forall i \quad (21)$$

$$\sum_j \sum_k \sum_o \sum_s \sum_p \left(ST_{ijkosp}^d + ST_{(i-1)jkosp}^d + \dots + ST_{(i-LC_{sp}+1)jkosp}^d \right) \leq BC_i \forall i \quad (22)$$

$$\sum_j \sum_k \sum_o \sum_s \sum_p \left(ST_{(i-LI_{sp}-LC_{sp})jkosp}^d + ST_{(i-LI_{sp}-LC_{sp}-1)jkosp}^d + ST_{(i-LI_{sp}-LC_{sp}-LW_{sp}+1)jkosp}^d \right) \leq BW_i \forall i \quad (23)$$

$$X_{ijkosp}^d, ST_{ijkosp}^d, EM_{ijkos}^d \in \{0, 1\} \quad (24)$$

In objective function (13), function C_{ijk} is an increasing function. Therefore, it strives to start patients' operations as soon as possible with regarding the priority of each patient. In constraint (14), using the weekly scheduling output which assigns each block to a surgical group, first, each block is divided into ten sub-blocks which is represented by k , and eventually, each sub-block is assigned to elective patients or left empty. Constraint (15) ensures that at most one surgery is conducted in each OR each time. Constraints (16) and (17) indicate the start time for each elective patient. Thus, among the blocks that are assigned to each patient, the first sub-block is the start time of each patient. Constraint (18) dedicates to emergency patients and ensures that any emergency patient could operate in elective OR in a maximum limited time if reserved ORs are not available. This guarantees the ability to assign at least R_{ijk} of the elective patients OR to the emergency patient. There will be empty ORs or an operation begins in one of the OR, where the operation can be delayed and given to emergency patients. Constraint (19) notifies the completion time of the surgery as well as indicates that an operation from the beginning to the end should be done without interruption in one OR. Constraint (20) states the number of occupied recovery beds which in each period should not be more than the available number. Constraints (21–23) consider the integration of other sections to ORs. Constraint (24) demonstrates the domain of variables.

3.3 Rescheduling

In phase 1, OTs are reserved for emergency patients, Z_{ijok}^w is the reserved ORs. In phase 2, R of the ORs in each time potentially are available to assign to emergency patients. Whenever an emergency patient arrives randomly, the patient must transfer to an OR within a limited time. This limitation is crucial and completely different from one patient to another. One emergency patient must be operated on in a maximum of 15 min, while another one could wait until tomorrow. The patient is operated on at once if the reserved OR is empty. Otherwise, he or she must wait until the reserved ORs or one of the elective ORs becomes available. In the proposed model, the maximum waiting time for emergency patients is controllable for elective ORs, which can be from 1 sub-block to whatever required. Suppose an emergency patient arrives at T_a , the earliest time when a reserved OR becomes available is T_r , and his or her time limit is T_l , if $T_r - T_a \leq T_l$, he or she can wait until the reserved OR become available. In contrast, if $T_r - T_a > T_l$, the patient must operate in the first available ORs. Therefore, the starting times of surgeries that begun earlier than T_a remain fixed and the remaining surgeries must be rescheduled to take the emergency surgery into account.

This reschedule model is referred to as the online model, in which the daily remaining elective patients are rescheduled again. In this model, the number of patients reduces to the remaining patients (p') of the day, who should operate in this day according to daily scheduling. Also, overtime is allowed, and the objective function tries to minimize the total number of overtimes. In the weekly and daily scheduling, the number of needed surgeons is known in advance. Therefore, it is possible to schedule patients without considering the availability of surgeons. Yet in rescheduling, patients with the available surgeons must schedule. Therefore let S_{js} as a parameter to indicate the number of available surgeons who can operate in the surgical group s in block j . In addition, integrity constricts for ICU, CCU, and the ward units are not necessary since they are considered in the daily schedule, yet the rescheduling process cannot change the day of the surgery. In the rescheduling model, all the parameters

and decision variables are the same as the daily and weekly scheduling, yet the difference is the rescheduling program for the remaining patients in one day.

The following notations are used in addition to the explained notations in the mathematical model.

Indices and sets:

P' : Set of remaining patients of the day indexed by p'

Q : Set of blocks in the overtime

Parameters:

$LO_{sp'}$: The length of surgery duration of patient p' from the surgical group s

$LR_{sp'}$: The length of recovery duration of patient p' from the surgical group s

S_{jks} : The number of available surgeons from the surgical group s in block j and sub-block k

C_{jk} : A factor in the objective function whose purpose is to put the operations in the earliest possible time ($C_{jk} = c_j \times j + c_k \times k$ where $c_j > c_k$)

R_{jk} : The number of ORs, which could assign to emergency patients in block j and sub-block k

Decision variables:

$X^r_{jkosp'}$: A binary variable, equal to 1, if sub-block k from block j of the OR o assigns to patient p' from the surgical group s ; 0, otherwise

$ST^r_{jkosp'}$: A binary variable, equal to 1, if sub-block k from block j of the OR o assigns to patient p' from the surgical group s and the patient's surgery starts in this sub-block; 0, otherwise

EM^r_{jkos} : A binary variable, equal to 1, if sub-block k from block j of the OR o from the surgical group s is empty; 0, otherwise

XS^r_{jkos} : A binary variable, equal to 1, if a surgeon from the surgical group s in sub-block k from block j assigns to the OR o ; 0, otherwise

The mathematical formulation of the rescheduling is as follows:

$$\min \sum_{j \in Q} \sum_k \sum_o \sum_s \sum_{p'} C_{jk} \times X^r_{jkosp'} \quad (25)$$

subject to:

$$\sum_o XS^r_{jkos} \leq S_{jks} \forall j, k, s \quad (26)$$

$$\sum_s XS^r_{jkos} \leq 1 \forall j, k, o \quad (27)$$

$$X^r_{jkosp'} + EM^r_{jkos} \leq XS^r_{jkos} \forall j, k, o, s, p' \quad (28)$$

$$\sum_{p'} X^r_{jkosp'} + EM^r_{jkos} \leq 1 \forall j, k, o, s \quad (29)$$

$$ST^r_{jkosp'} \leq X^r_{jkosp'} \forall j, k, o, s \quad (30)$$

$$\sum_j \sum_o \sum_k ST^r_{jkosp'} = 1 \forall s, p' \quad (31)$$

$$\sum_o \left(Z_{ijok}^w + \sum_s \left(\sum_{p'} X_{jkosp'}^r + EM_{jkos}^r \right) \right) \geq R_{jk} \forall \{j, k\} \in \Phi \tag{32}$$

$$ST_{jkosp'}^r \times LO_{sp'} \leq \left(X_{(jk)osp'}^r + X_{(jk+1)osp'}^r + \dots + X_{(jk+LO_{sp}-1)osp'}^r \right) \forall j, k, s, p', o \tag{33}$$

$$\sum_o \sum_s \sum_{p'} \left(ST_{(jk-LO_{sp'})kosp'}^r + ST_{(jk-LO_{sp'}-1)kosp'}^r + \dots + ST_{(jk-LO_{sp'}-LR_{sp'}+1)kosp'}^r \right) \leq BR \forall j, k \tag{34}$$

$$X_{jkosp'}^r, X_{jkos}^r, ST_{jkosp'}^r, EM_{jkos}^r \in \{0, 1\} \tag{35}$$

In the objective function (25), function C_{jk} is an increasing function. Therefore, it strives to start patients' operations as soon as possible to minimize the total number of overtime sub-blocks. Constraints (26) and (27) assign surgeons to ORs. In the constraint (28) and (29), each sub-block assigns to elective patients or leave empty. Constraints (30) and (31) indicate the start time for each elective patient. Thus, from among the blocks that are assigned to each patient, the first sub-block is the start time of each patient. Constraint (32) is for emergency patients and ensures that any emergency patient could operate in elective OR in a maximum limited time if reserved ORs are not available. This Constraint guarantees that at least one OR will be able to assign to the emergency patient in each sub-block. In these sub-blocks, there will be empty ORs or an operation begins in one of the OR, where the operation can delay and give to emergency patients. Constraint (33) demonstrates the completion time of the surgery, and also indicates that an operation from the beginning to the end should be done without interruption in one OR. Constraint (34) states the number of occupied recovery beds in each period should not be more than the available number. Constraint (35) shows the domain of variables.

4 Solution method

The proposed mathematical models of this paper can be regarded as a flexible job shop scheduling problem that has been proved to be NP-hard (Fattahi et al., 2007). Therefore, for large-size models, it is required to find a near-optimal solution within a reasonable time. To do so, GA and PSO are modified and applied for solving the models.

4.1 Genetic algorithm

GA is a well-known meta-heuristic method, which is based on the evolutionary concept and has been extensively applied to tackle different optimization problems. For instance, Marques et al. (2014), Hooshmand et al. (2018), and Wang et al. (2022) used GA to solve OR scheduling problems.

In the proposed GA, the chromosome is a $4 \times N(\sum n_s)$ matrix as shown in Fig. 2, where n_s is the number of patients in each surgical group and N is the total number of patients in the entire surgical groups. Columns 1 to n_1 are the first surgical group patients. Columns

Fig. 2 Chromosome structure of the problem

	1	2	...	n_1	n_1+1	...	n_1+n_2	...	N
<i>O</i>	2	1	...	2	8	...	7	...	1
<i>I</i>	3	2	...	5	2	...	4	...	5
<i>J</i>	2	2	...	4	2	...	2	...	1
<i>K</i>	5	9	...	1	7	...	5	...	8

Fig. 3 The pseudo-code of GA steps

```

Iter = 0
Initialize population P
Evaluate(P(Iter))
While not converged || maxIterates do
    Pp(Iter) = selectParents(P(Iter))
    Pc(Iter) = reproduction(Pp)
    Mutate(Pc(Iter))
    Evaluate(Pc(Iter))
    P(Iter+1) = nextGeneration(Pc(Iter),
P(Iter))
    Iter = Iter + 1
End

```

$n_1 + 1$ to $n_1 + n_2$ are the second surgical group patients, etc. This matrix has 4 rows. The first row shows ORs in which the patient will operate. The second row indicates the day of operation. The third, as well as fourth rows, show the exact start time of each operation (block and sub-blocks). According to this, for example, the first patient will operate on the third day of the week and in the second block and the fifth sub-block in the second OR. Figure 3 demonstrates the pseudo-code of the proposed GA.

4.2 Particle swarm optimization

PSO initially has been introduced for simulating swarm behavior. Swarm intelligence inherently represents optimum behavior in nature. Therefore, swarm simulation can lead to optimization (Kennedy & Eberhart, 1995). PSO reported being a successful algorithm in several fields such as antennas, biomedical, etc. (Poli, 2007, 2008; Devika et al., 2016). The velocity of each bird in a swarm is related to adjacent birds and stochastic factors. Furthermore, swarming birds move around the whole swarm central position and tend to land where food can be found (Heppner & Grenander, 1990). In PSO, the central position of the swarm is the best solution found in the entire swarm and each particle has its own best solution. Two mentioned rules on birds' swarm lead to the velocity formulation of PSO shown in Eq. (36).

$$newV_p = \omega V_p + \varphi_p rand(Pbest, x_p) + \varphi_g rand(Gbest, x_p) \quad (36)$$

where V_p is the velocity of particle p . $Pbest$ and $Gbest$ notify the best-known solution of each particle and the swarm, respectively. x_p displays the current particle solution. Parameter ω is the inertia factor. Parameters φ_p and φ_g indicate that how much each particle should rely on the best swarm known solution or its own local solution.

In this research, PSO is implemented with the adaptive inertia parameter of Taherkhani and Safabakhsh (2016). PSO with adaptive inertia parameter is shown in Eqs. (37–39). In

Fig. 4 The pseudo-code of PSO steps

```

P = Particle initialization
pBest = first particle
For i = 1 to maxIterations
  For each particle p in P do
    Obj = f(p)
    If Obj is better than f(pBest)
      pBest = p
    End
  End
gBest = best p in P
For each particle p in P do
  Adaptive update of velocity v
  p = p + v
End
End
    
```

the PSO solution structure, each row represents a particle, which all variables are serialized in columns. Index i is the particle number and j is the dimension number. Equation (39) is the velocity updating. The pseudo-code of PSO can be found in Fig. 4.

$$\delta_i(t) = \begin{cases} 1 & \text{if } f(x_i(t-1)) < f(Pbest(t)) \\ -1 & \text{else} \end{cases} \tag{37}$$

$$w_{ij}(t+1) = \begin{cases} \min(w_{max}, w_{ij}(t) + (1 - w_0) * \exp\left(\frac{(x_{ij}(t+1) - Pbest_{ij}(t))^2}{-2\sigma^2}\right)) + \varepsilon \text{if } \delta_i(t) > 0 \text{ and } \delta_i(t-1) > 0 \\ \min\left(w_{min}, w_{ij}(t) - w_0 * (1 - \exp\left(\frac{(x_{ij}(t+1) - Pbest_{ij}(t))^2}{-2\sigma^2}\right))\right) + \varepsilon \text{if } \delta_i(t) < 0 \text{ and } \delta_i(t-1) < 0 \\ w_{ij}(t) \text{ else} \end{cases} \tag{38}$$

$$V_{ij} = \omega_{ij} V_{ij} + \varphi_p rand(Pbest_{ij}, x_{ij}) + \varphi_g rand(Gbest_{ij}, x_{ij}) \tag{39}$$

Initial particle values and velocity are set randomly with respect to minimum and maximum expected values of the variables of the problem. Except for parameter ω which is adaptive to runtime, φ_p is set to -0.6 and φ_g is 2.6 (Pedersen, 2010). Furthermore, the maximum velocity of the particles is considered to avoid divergence of particles. To validate PSO, the problem is solved by GA. Discovering the first feasible solution is tested for each algorithm. The maximum iteration number of PSO is set to 1000.

5 Case study

The case study of this paper is a hospital located in the sixth district of Tehran. The sixth district is a major commercial and administrative center and is ranked 17th populated among 22 districts (Iran, 2016). Despite having a low population, a large number of people travel to this district during the day (Atlas, 2020). It has the third rank in car accidents and the first rank in casualties. Therefore, most of the emergency patients arriving at the hospital are due to car accidents. The investigated hospital has 10 ORs with capacities of more than 200

patients per week and eight elective surgical groups. Elective surgeries are operated on five days a week from Saturday to Wednesday. The usual working hour is from 8 AM to 4 PM and may take up to 4 h of overtime if needed. It is supposed the length of each block and sub-block to be 2 h and 12 min, respectively. Therefore, a week consists of 200 blocks and 2000 sub-blocks. To estimate the emergency arrivals to the hospital, historical time-related data about accidents and geographical information related to various types of accidents are required. Figure 5 demonstrates the absorbed travels in Tehran districts and summarizes the data used for emergency patients' arrivals.

The data discussed so far represent the distribution of average accidents on daily basis, however, the arrivals during a day are required for modeling. To achieve sub-block arrivals, the traffic load is used to represent the activity level in the city during the day. Traffic load is obtained from Google map data which is provided by the Traffic Index website (index, 2021), and is represented by weighted colors (green, orange, red, and dark red) used by Google Maps to express traffic congestion index (TCI). The traffic data is distributed by TCI and interpolated by cubic smoothing spline interpolation for emergency patients' arrivals. The cubic spline is a parametric model whose parameters can be optimized to obtain well-shaped curvature (Kouibia et al., 2011). Figure 6 shows the interpolated average TCI index during a day.

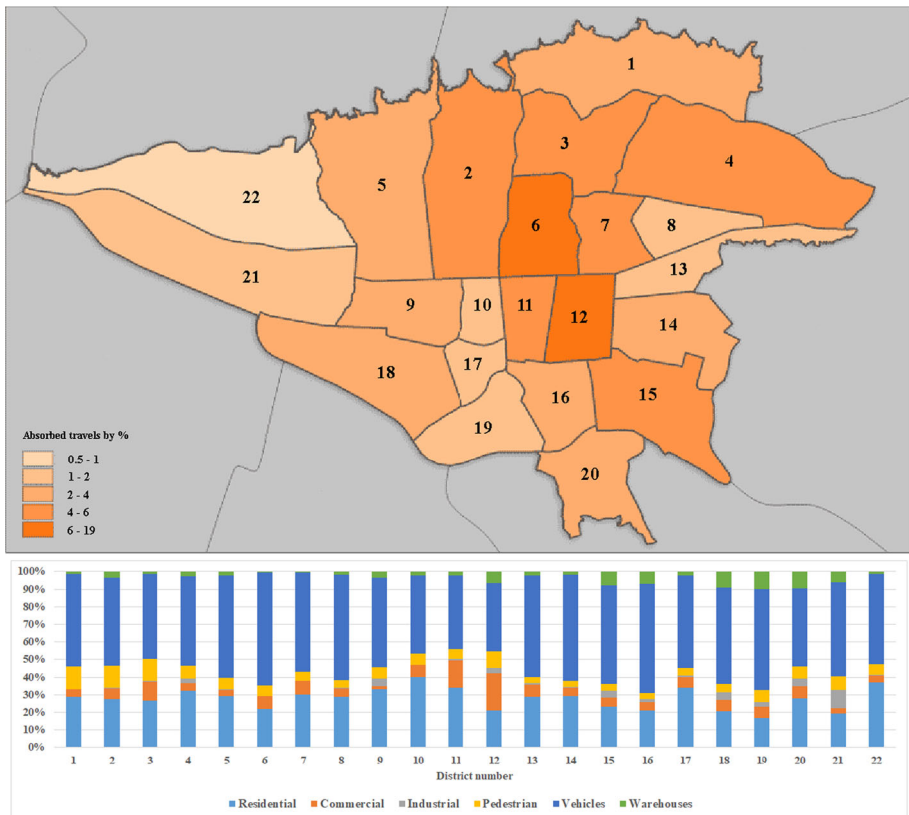


Fig. 5 Travel absorbed by different districts and different types of accidents used for emergency surgery arrivals

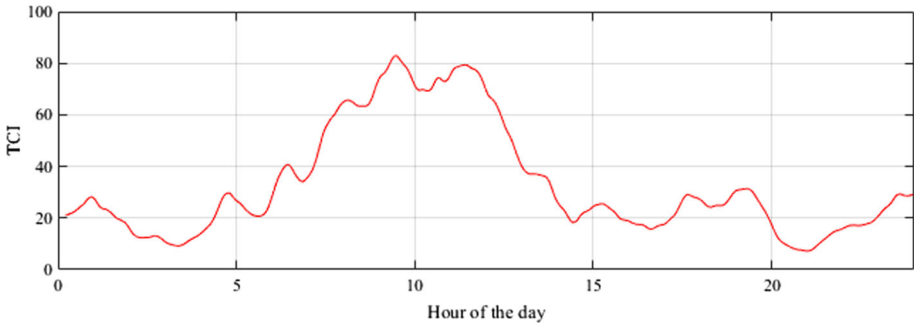


Fig. 6 Average TCI index for Tehran interpolated by cubic spline

Surgical duration is a continuous variable, and it is difficult to be determined directly. Machine learning techniques can help us to predict surgical duration more accurately from various parameters of a surgery. Although the surgical duration is a continuous variable and can be handled as a regression problem, the prediction needs to be fed into the mathematical model and it should be binned first to fit the model. Surgical duration is categorized into four scenarios based on unsupervised learning. Expected maximization is used to obtain the clusters whose result can be seen in Fig. 7. Different cluster numbers are tested and analyzed by Silhouette clustering validity measure (Lovino et al., 2022). The number of clusters which maximizes the Silhouette measure is selected. A random forest algorithm is applied to predict the duration scenario based on the surgery characteristics. Random forest is a well-known

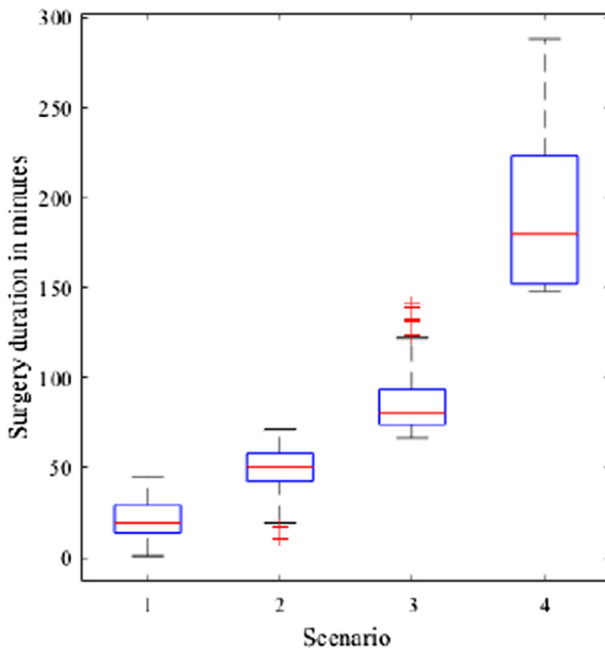


Fig. 7 Different scenarios generated by expected maximizing clustering on surgery duration

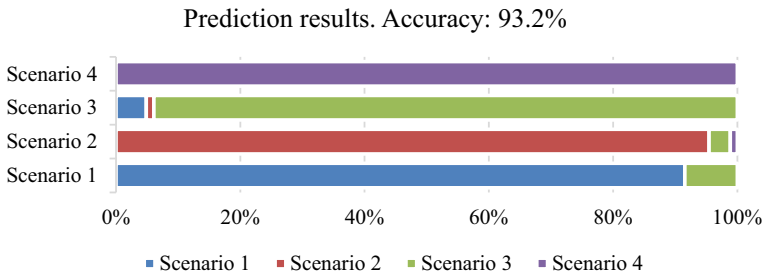


Fig. 8 Prediction accuracy of the random forest model

machine learning model used in medical prediction studies which has been proven to be an effective classification model (Kong et al., 2020; Lin et al., 2019). It is an ensemble learning algorithm that generates multiple decision trees to enhance prediction accuracy. The features which are used to build a prediction model are the age of the patient, surgery type, type of anesthesia, preconditions of the patient, operation room number, number of staff, month, and day of the week. The operation room number in addition to the location of the surgery represents the operation team experience and the physical equipment of the room. Figure 8 shows the prediction performance by applying tenfold cross-validation. The average duration of the predicted scenario is used for the optimization model. As shown in Fig. 8 the longer surgeries are predicted with higher accuracy and the longest surgery duration doesn't have any misclassification.

6 Computational results

This section explains the results of numerical experiments that examine the performance of the solution methods. All computations are performed on Intel Core i7-3632QM equipped with an 8 GB RAM computer. Customized GA and PSO are coded in the MATLAB programming language. Furthermore, the ILOG IBM CPLEX is applied to solve small-size models.

6.1 Performance of solution methods

In weekly scheduling, surgery blocks are assigned to surgery groups. Therefore, the dimensionality is not high and it is possible to achieve exact solutions in a reasonable time for this model. Since the number of patients is high in daily scheduling, the dimension of the daily scheduling problem is high. On the other hand, the rescheduling model is considered only for the remaining patients in a day. Thus, the dimensionality of this model is not high. Consequently, the weekly scheduling and rescheduling models are solved with CPLEX, and GA and PSO are applied to find the daily scheduling. To evaluate the performance of the proposed algorithms, 20 small-size test problems are generated. Table 1 provides the characteristics of these problems.

For each test problem, first, CPLEX results are computed. Then, the same problem is solved 5 times using the proposed GA and PSO. To tune the parameters of GA and PSO, the Taguchi method is applied but the corresponding results are not reported because of brevity. The average objective function of 5 runs is examined for evaluation. The gap between the exact solutions of test problems from both GA and PSO is considered as a criterion of the

Table 1 Characteristics of generated test problems

Test problem	S	K	I	J	O
1	3	2	2	2	2
2	3	3	3	3	2
3	3	4	3	3	3
4	3	6	3	5	3
5	4	4	4	3	3
6	4	4	4	3	2
7	4	4	4	4	2
8	5	3	4	4	3
9	5	3	4	3	4
10	6	3	4	4	3
11	6	3	4	4	4
12	5	4	5	4	3
13	5	4	5	4	3
14	7	3	5	4	3
15	6	4	5	4	4
16	6	5	5	4	4
17	6	5	5	4	5
18	7	5	5	4	5
19	7	5	5	5	5
20	7	6	5	5	5

accuracy of the algorithms' response and is calculated as Eq. (40).

$$Gap_a^t(\%) = \frac{Z_a^t - Z_{CPLEX}^t}{Z_{CPLEX}^t} \times 100 \quad \forall a \in \{1, 2, \dots, 20\} \text{ and } t \in \{GA, PSO\} \quad (40)$$

where Z_a^t implies the objective function value of the proposed methods for each test problem.

Table 2 reports the results of CPLEX, GA, and PSO for daily scheduling. As seen in this table, by growing the dimensions of the problems, the difference between the solution times for the CPLEX, and proposed GA, and PSO is increasing. Given the small gap between the exact solution of the problem and the algorithm's, it seems that the proposed GA gap is around 2%, and the PSO gap is around 4%. Therefore, the GA has a better function and can reliably use to solve real-world problems that the CPLEX cannot solve at the logical time.

6.2 Results

After assessing the performance of the proposed solution methods, we concentrate on a real case study that inspired this research. Patients' surgeries performed at the hospital during the first 20 weeks of 2020 are taken into account to assess the performances of the proposed three-phase model in the real-world problem. This is accomplished by comparing the current patient schedule for these weeks to the optimum schedule of the three-phase model.

Table 2 Results of test problems

Test problem	Time (Sec.)			Solution			GAP (%)	
	CPLEX	GA	PSO	CPLEX	GA	PSO	GA	PSO
1	60	45	51	1657	1692	1770	2.1	6.8
2	70	57	59	2154	2154	2211	0.0	2.6
3	206	68	79	2703	2768	2873	2.4	6.3
4	1023	218	229	5432	5560	5703	2.4	5.0
5	359	103	111	4662	4755	4781	2.0	2.6
6	466	115	119	4925	4987	5103	1.3	3.6
7	420	137	145	4639	4776	4799	3.0	3.4
8	291	94	100	3418	3477	3591	1.7	5.1
9	269	90	105	3173	3245	3310	2.3	4.3
10	725	185	203	5401	5482	5482	1.5	1.5
11	789	127	151	4441	4523	4539	1.8	2.2
12	913	174	202	5935	6068	6102	2.2	2.8
13	948	189	193	5935	6037	6202	1.7	4.5
14	1000	202	214	6427	6549	6582	1.9	2.4
15	1302	240	242	5602	5722	5823	2.1	3.9
16	1639	273	280	6965	7111	7145	2.1	2.6
17	1403	244	256	6192	6303	6422	1.8	3.7
18	2945	320	332	9013	9147	9368	1.5	3.9
19	3012	312	327	7251	7414	7533	2.2	3.9
20	3241	398	406	9882	10,056	10,261	1.8	3.8

Doctors' suggestions for elective patients who had surgery in these twenty weeks are considered for surgery duration and length of stay in downstream units. Also, geographical information systems and random forests model are applied for years 2019 and 2020 to predict emergency patient's arrival time and surgery durations. Regarding the 20 weeks data, the three-phase model is applied. The weekly scheduling and rescheduling models are solved with CPLEX. In addition, GA is applied to find the daily scheduling. Table 3 shows the results of the real-world case. Equation (41) calculates the Improvement percentage of the three-phase model compare to the current patient scheduling.

$$Improvement_w(\%) = \frac{Remaining_w - Overtime_w}{Total\ number\ of\ sub - blocks} \times 100 \quad \forall w \in \{1, 2, \dots, 20\} \quad (41)$$

where $Remaining_w$ and $Overtime_w$ indicate the number of remaining sub-blocks after the last elective patient surgery and the total number of overtime sub-blocks after rescheduling, respectively. The results show that by applying the proposed model, the performance of OT can improve by approximately 10.5% on average.

Figure 9 illustrates the elective patients scheduling on the first day that the rescheduling model is needed, in which horizontal and vertical axes indicate ORs and sub-blocks, respectively. For example, in OR1, S4-P3 means that the surgery of the third patient from the fourth surgical group begins at sub-block 8 and finishes at sub-block 18. If an emergency

Table 3 Real-word case results

Week no	Elective patients	Emergency patients	Remaining sub-blocks	Rescheduling	Overtime sub-block	Improvement percentage (%)
1	151	60	222	2	16	10.3
2	148	64	270	2	17	12.7
3	159	61	272	5	20	12.6
4	157	60	180	1	3	8.9
5	154	41	97	0	0	4.9
6	153	62	228	2	23	10.3
7	146	50	243	3	8	11.8
8	150	56	263	3	17	12.3
9	147	61	270	2	11	13.0
10	153	47	270	1	4	13.3
11	145	60	198	2	23	8.8
12	158	55	121	2	10	5.6
13	147	47	285	0	0	14.3
14	157	50	221	2	9	10.6
15	148	59	243	4	24	11.0
16	151	65	185	3	13	8.6%
17	149	45	198	2	14	9.2%
18	145	62	249	3	17	11.6%
19	155	58	234	2	18	10.8%
20	153	67	222	3	21	10.1%

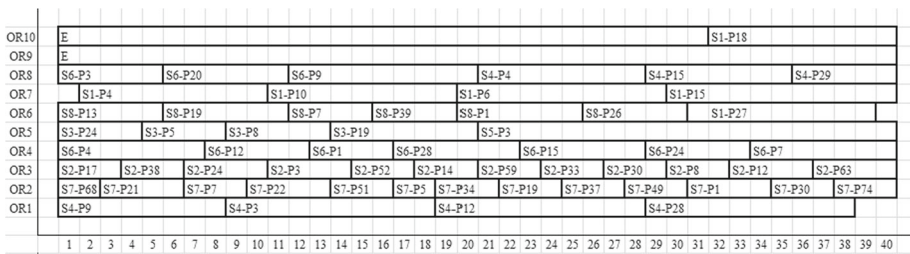


Fig. 9 Patients' daily scheduling

patient arrives at the time of implementing the program and the reserved ORs are full, and the emergency patient cannot wait until the reserved ORs become available, the elective ORs should assign to the emergency and reschedule elective patients. EMERGENCY1 is the first emergency patient that needs to use the elective patient ORs and it is needed to run the rescheduling model. EMERGENCY1 arrives in the hospital at 12:30 PM and cannot wait until reserved ORs become available. Therefore, the surgeries that began before this time are unchanged which are determined in Fig. 10. The first available OR (OR2) will assign to EMERGENCY1, and also the remaining patient will be rescheduled as shown in Fig. 11.

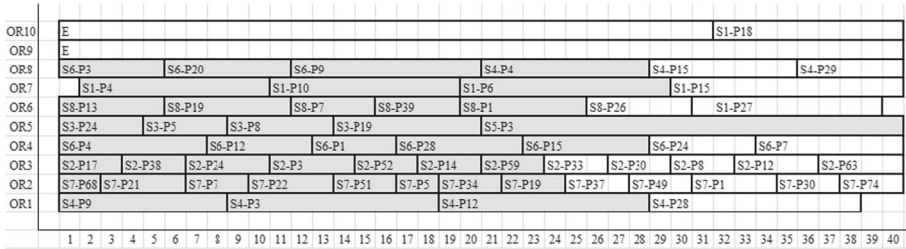


Fig. 10 Remaining patients' scheduling when EMERGENCY1 arrives

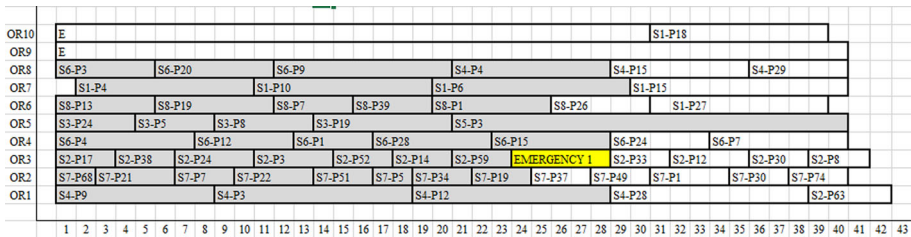


Fig. 11 Patients' rescheduling. (Color figure online)

7 Discussion

Limited surgical capacities of hospitals which stems from various factors such as pandemics, and aging make effective OT management critical. Although the ORs planning and scheduling problem is popular research that has been widely studied in the literature, there is still a considerable research gap in this area. It is shown that integration is a missing key in this field of research. Considering the integrated planning is of great importance because many surgeries are canceled due to the lack of enough downstream units which keeps patients from moving forward, and emergency patients unexpectedly arrive at the OT. Therefore, the proposed model takes into account the elective and emergency patients scheduling for OT in a hospital instead of ORs. This OT is composed of multiple ORs, PACU, ICU, CCU, and wards. Also, our three-phase model integrates all the decisions in the OT department.

The results showed that near-optimal solutions can be found for large instances with more than 200 surgeries in a reasonable time. In addition, since the lack of enough downstream units keeps patients from moving forward, significantly decreasing ORs' utilization, the integrated model considers the needed downstream units for patients to ensure the practicality of the proposed model. Moreover, the proposed model guarantees that an emergency patient assigns to an OR within a specific limit time and the rescheduling model ensures that the patient's surgery day will not change, which may result in patient dissatisfaction and anxiety.

By increasing the presence of pandemics, the lack of hospital resources, and the excessive workload of medical staff, in general, are aggravated. Hospital resources must expand as quickly as possible to restore their previous performance level and enable them to continue providing quality care to patients. As a result of the pandemic, health systems are unable to supply hospitals appropriately and rapidly with the necessary sources. Therefore, operations research techniques such as the three-phase model described in this paper can help to ensure that hospital resources are correctly allocated, and patients are treated fairly in such a situation.

It is shown that machine learning is an effective way for predicting the duration of a surgery, which is a nondeterministic variable that depends on various parameters. Surgery durations are clustered based on maximum Silhouette measure for different number of clusters. Four duration scenarios are generated by the clusters and the random forest classifier could successfully predict the scenario based on the surgery characteristics.

8 Conclusion

This paper addresses the integrated elective and emergency patient scheduling and rescheduling model in operating theater (OT), which consists of multiple operating rooms (ORs), recovery, ICU, CCU, and ward units. Elective surgeries are scheduled in the OTs in advance on which some OT capacity must be reserved for emergency patients who may arrive randomly. Once an emergency patient arrives, it is ensured that ORs will assign to the patient in a limited time, if needed.

For developing schedules for OTs, a three-phase model was presented. Phase 1 allocates the surgical groups to OR blocks on each day of a week. Phase 2 develops a daily detailed patient schedule and sequence. Although some OR capacity is reserved for emergency patients in the weekly schedule, they could use elective patients OR whenever they need. Since the arrival of emergency patients is completely random, geographical information systems and random forest machine learning models were applied to forecast the arrival time and surgery duration of emergency patients. Phase 3 is an online model to reschedule elective patients' daily scheduling when an emergency patient needs to use elective patients OR. Phase 3 is a small sample of Phase 2, which can easily be solved in a reasonable time.

To solve the proposed NP-hard problem, genetic algorithm (GA), as well as particle swarm optimization (PSO), was modified and proposed. In addition, the Taguchi method is utilized for tuning the parameters of the proposed algorithms. The results showed the high accuracy of GA in comparison with PSO, and it is able to find the solution for the real-world problem in an appropriate time with a negligible gap. The proposed three-phase models were applied for a hospital in Tehran as a real-world case. The comparison between the results of the proposed models with the current scheduling for the hospital was performed to identify the performance of the proposed models. Comparing the proposed method to the hospital's present scheduling reveals that the three-phase model has a significant positive impact on the OT schedule. The results indicated that within the scheduling provided by the proposed models, the performance of the OT department will go up by about 10.5% on average.

Eventually, some suggestions are provided as to future research directions. OTs scheduling can be performed under uncertainty or probabilistic conditions. In addition, applying machine learning models to forecast elective patient's surgery duration and/or the length of stay in downstream units may help researchers to make improvements in OT management.

Funding There is no funding for this research work.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Addis, B., Carello, G. & Tanfani, E. A robust optimization approach for the operating room planning problem with uncertain surgery duration. In *Proceedings of the international conference on health care systems engineering, 2014*. (pp. 175–189) Springer.
- Addis, B., Carello, G., Grosso, A., & Tanfani, E. (2016). Operating room scheduling and rescheduling: A rolling horizon approach. *Flexible Services and Manufacturing Journal*, 28, 206–232.
- Akbarzadeh, B., Moslehi, G., Reisi-Nafchi, M., & Maenhout, B. (2019). The re-planning and scheduling of surgical cases in the operating room department after block release time with resource rescheduling. *European Journal of Operational Research*, 278, 596–614.
- Ali, I., & Kannan, D. (2022). Mapping research on healthcare operations and supply chain management: a topic modelling-based literature review. *Annals of Operations Research*, 315, 29–55.
- Arab Momeni, M., Mostofi, A., Jain, V., & Soni, G. (2022). COVID19 epidemic outbreak: operating rooms scheduling, specialty teams timetabling and emergency patients' assignment using the robust optimization approach. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04667-7>
- Aringhieri, R., Landa, P., Soriano, P., Tanfani, E., & Testi, A. (2015). A two level metaheuristic for the operating room scheduling and assignment problem. *Computers and Operations Research*, 54, 21–34.
- Atighehchian, A., Sepehri, M. M., Shadpour, P., & Kianfar, K. (2020). A two-step stochastic approach for operating rooms scheduling in multi-resource environment. *Annals of Operations Research*, 292, 191–214.
- Atlas, T. 2020. *Atlas of Tehran Metropolis* [Online]. Available: <https://atlas.tehran.ir/Default.aspx?tabid=227> [Accessed].
- Bargetto, R., Garaix, T., Xie, X., & ENGINEERING,. (2018). Dynamic Insertion of Emergency Surgeries With Different Waiting Time Targets. *IEEE Transactions on Automation Science*, 16, 87–99.
- Bruni, M., Beraldi, P., & Conforti, D. (2015). A stochastic programming approach for operating theatre scheduling under uncertainty. *IMA Journal of Management Mathematics*, 26, 99–119.
- Cardoen, B., Demeulemeester, E., & Beliën, J. (2009). Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics*, 119, 354–366.
- Devika, K., Jafarian, A., Hassanzadeh, A., & Khodaverdi, R. (2016). Optimizing of bullwhip effect and net stock amplification in three-echelon supply chains using evolutionary multi-objective metaheuristics. *Annals of Operations Research*, 242(2), 457–487.
- di Martinelly, C., & Meskens, N. (2017). A bi-objective integrated approach to building surgical teams and nurse schedule rosters to maximise surgical team affinities and minimise nurses' idle time. *International Journal of Production Economics*, 191, 323–334.
- Durán, G., Rey, P. A., & Wolff, P. (2017). Solving the operating room scheduling problem with prioritized lists of patients. *Annals of Operations Research*, 258, 395–414.
- Erdem, E., Qu, X., & Shi, J. (2012). Rescheduling of elective patients upon the arrival of emergency patients. *Decision Support Systems*, 54, 551–563.
- Erekat, A., Servis, G., Madathil, S. C., & Khasawneh, M. T. (2020). Efficient operating room planning using an ensemble learning approach to predict surgery cancellations. *IJSE Transactions on Healthcare Systems Engineering*, 10, 18–32.
- Essen, J. T., Hans, E. W., Hurink, J. L., & Oversberg, A. (2012a). Minimizing the waiting time for emergency surgery. *Operations Research for Health Care*, 1, 34–44.
- Fairley, M., Scheinker, D., & Brandeau, M. L. (2019). Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health Care Management Science*, 22, 756–767.
- Farzad, G., & Mohammad, S. M. (2016). A stochastic surgery sequencing model considering the moral and human virtues. *Modern Applied Science*, 10, 68.
- Fattahi, M., Keyvanshokoo, E., Kannan, D., & Govindan, K. (2023). Resource planning strategies for healthcare systems during a pandemic. *European Journal of Operational Research*, 304(1), 192–206.
- Fattahi, P., Mehrabad, M. S., & Jolai, F. (2007). Mathematical modeling and heuristic approaches to flexible job shop scheduling problems. *Journal of Intelligent Manufacturing*, 18, 331–342.
- Fei, H., Chu, C., & Meskens, N. (2009). Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research*, 166, 91–108.
- Fei, H., Meskens, N., & Chu, C. (2010). A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers and Industrial Engineering*, 58, 221–230.
- Ferreira, F. A., Kannan, D., Meidutė-kavaliauskienė, I., & Vale, I. M. (2022). A sociotechnical approach to vaccine manufacturer selection as part of a global immunization strategy against epidemics and pandemics. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04347-y>
- Fügener, A., Hans, E. W., Kolisch, R., Kortbeek, N., & Vanberkel, P. T. (2014). Master surgery scheduling with consideration of multiple downstream units. *European Journal of Operational Research*, 239, 227–236.

- Gauthier, J. B., & Legrain, A. (2016). Operating room management under uncertainty. *Constraints*, 21, 577–596.
- Guerrero, F., & Guido, R. (2011). Operational research in the management of the operating theatre: A survey. *Health Care Management Science*, 14, 89–114.
- Guido, R., & Conforti, D. (2017). A hybrid genetic approach for solving an integrated multi-objective operating room planning and scheduling problem. *Computers and Operations Research*, 87, 270–282.
- Guido, R., Groccia, M. C., & Conforti, D. (2018). An efficient matheuristic for offline patient-to-bed assignment problems. *European Journal of Operational Research*, 268, 486–503.
- Gul, S., Denton, B. T., & Fowler, J. W. (2015). A progressive hedging approach for surgery planning under uncertainty. *INFORMS Journal on Computing*, 27, 755–772.
- Heppner, F., & Grenander, U. 1990. A stochastic nonlinear model for coordinated bird flocks. In *American Association For The Advancement Of Science, Washington, DC(USA). 1990*.
- Hooshmand, F., Mirhassani, S., & Akhavan, A. (2018). Adapting GA to solve a novel model for operating room scheduling problem with endogenous uncertainty. *Operations Research for Health Care*, 19, 26–43.
- Index, T. 2021. In *GoogleMap traffic layer query* [Online]. Available: <http://trafficindex.org/> [Accessed].
- Iran, S. C. O. 2016. In *8th National Population and Housing Census* [Online]. Available: <https://www.amar.org.ir/english/Population-and-Housing-Censuses> [Accessed].
- Jebali, A., & Diabat, A. (2017). A Chance-constrained operating room planning with elective and emergency cases under downstream capacity constraints. *Computers and Industrial Engineering*, 114, 329–344.
- Kennedy, J., & Eberhart, R. 1995. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, (pp. 1942–1948).
- Kong, G., Lin, K., Hu, Y., & Making, D. (2020). Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Medical Informatics*, 20, 1–10.
- Kouibia, A., Pasadas, M., & Rodríguez, M. L. (2011). Optimization of parameters for curve interpolation by cubic splines. *Journal of Computational Applied Mathematics*, 235, 4187–4198.
- Kroer, L. R., Foverskov, K., Vilhelmsen, C., Hansen, A. S., & Larsen, J. (2018). Planning and scheduling operating rooms for elective and emergency surgeries with uncertain duration. *Operations Research for Health Care*, 19, 107–119.
- Lamiri, M., Dreou, J., & Xie, X. Operating room planning with random surgery times. In *2007 IEEE International Conference on Automation Science and Engineering, 2007*. (pp. 521–526) IEEE.
- Lin, K., Hu, Y., & Kong, G. (2019). Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *International Journal of Medical Informatics*, 125, 55–61.
- Liu, Y., Chu, C., & Wang, K. (2011). A new heuristic algorithm for the operating room scheduling problem. *Computers and Industrial Engineering*, 61, 865–871.
- Lovino, M., Randazzo, V., Ciravegna, G., Barbiero, P., Ficarra, E., & Cirrincione, G. (2022). A survey on data integration for multi-omics sample clustering. *Neurocomputing*, 488, 494–508.
- Luo, L., Liu, H., Hou, X., & Shi, Y. Machine learning methods for surgery cancellation. In *2016 13th International Conference on Service Systems and Service Management (ICSSSM), 2016*. (pp. 1–4) IEEE.
- Luo, Y. Y., & Wang, B. (2019). A new method of block allocation used in two-stage operating rooms scheduling. *IEEE Access*, 7, 102820–102831.
- Macario, A. (2010). What does one minute of operating room time cost? *Journal of Clinical Anesthesia*, 4, 233–236.
- Marques, I., Captivo, M. E., & Pato, M. V. (2014). Scheduling elective surgeries in a Portuguese hospital using a genetic heuristic. *Operations Research for Health Care*, 3, 59–72.
- Meersman, T., & Maenhout, B. (2022). Multi-objective optimisation for constructing cyclic appointment schedules for elective and urgent patients. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04628-0>
- Obermeyer, Z., & Emanuel, E. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375, 1216.
- Pedersen, M. E. H. 2010. Good parameters for particle swarm optimization. *Hvass Lab., Copenhagen, Denmark, Tech. Rep. HL1001*.
- Poli, R. (2007). *An analysis of publications on particle swarm optimization applications*. Department of Computer Science, University of Essex.
- Poli, R. (2008). Analysis of the publications on the applications of particle swarm optimisation. *Journal of Artificial Evolution and Applications*, 2008, 3.
- Rahimi, I., & Gandomi, A. H. (2021). A comprehensive review and analysis of operating room and surgery scheduling. *Archives of Computational Methods in Engineering*, 28, 1667–1688.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380, 1347–1358.

- Riise, A., Mannino, C., & Burke, E. (2016). Modelling and solving generalised operational surgery scheduling problems. *Computers and Operations Research*, *66*, 1–11.
- Schiele, J., Koperna, T., & Brunner, J. O. (2021). Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. *Naval Research Logistics*, *68*, 65–88.
- Shuvo, S. S., Ahmed, M. R., Kabir, S. B. & Shetu, S. A. Application of machine learning based hospital up-gradation policy for Bangladesh. In *7th international conference on networking, systems and security, 2020*. (pp. 18–24)
- Silva, T. A., & de Souza, M. C. (2020). Surgical scheduling under uncertainty by approximate dynamic programming. *Omega*, *95*, 102066.
- Sohrabi, C., Alsafi, Z., O'neil, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, *76*, 71–76.
- Soudi, A., Heydari, M., & Mazdeh, M. M. (2019). A new approach for integrated surgical procedure scheduling with arrival uncertainty. *International Journal of Operational Research*, *34*, 430–449.
- Taherkhani, M., & Safabakhsh, R. (2016). A novel stability-based adaptive inertia weight for particle swarm optimization. *Applied Soft Computing*, *38*, 281–295.
- Tuwatananurak, J. P., Zadeh, S., Xu, X., Vacanti, J. A., Fulton, W. R., Ehrenfeld, J. M., & Urman, R. D. (2019). Machine learning can improve estimation of surgical case duration: A pilot study. *Journal of Medical Systems*, *43*, 44.
- Van Oostrum, J. M., Van Houdenhoven, M., Hurink, J. L., Hans, E. W., Wullink, G., & Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *Or Spectrum*, *30*, 355–374.
- Van Riet, C., & Demeulemeester, E. (2015). Trade-offs in operating room planning for electives and emergencies: A review. *Operations Research for Health Care*, *7*, 52–69.
- Van Veen-Berkx, E., Elkhuizen, S. G., Kuijper, B., & Kazemier, G. (2016). Dedicated operating room for emergency surgery generates more utilization, less overtime, and less cancellations. *The American Journal of Surgery*, *211*, 122–128.
- Varmazyar, M., Akhavan-Tabatabaei, R., Salmasi, N., & Modarres, M. (2020). Operating room scheduling problem under uncertainty: Application of continuous phase-type distributions. *IIE Transactions*, *52*, 216–235.
- Wang, J.-J., Dai, Z., Chang, A.-C., & Shi, J. J. (2022). Surgical scheduling by Fuzzy model considering inpatient beds shortage under uncertain surgery durations. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04645-z>
- Wang, S., Roshanaei, V., Aleman, D., & Urbach, D. (2016). A discrete event simulation evaluation of distributed operating room scheduling. *IIE Transactions on Healthcare Systems Engineering*, *6*, 236–245.
- Zhao, B., Waterman, R. S., Urman, R. D., & Gabriel, R. A. (2019). A machine learning approach to predicting case duration for robot-assisted surgery. *Journal of Medical Systems*, *43*, 32.
- Zhu, S., Fan, W., Yang, S., Pei, J., & Pardalos, P. M. (2019). Operating room planning and surgical case scheduling: A review of literature. *Journal of Combinatorial Optimization*, *37*, 757–805.
- Zonderland, M. E., Boucherie, R. J., Litvak, N., & Vleggeert-Lankamp, C. L. (2010). Planning and scheduling of semi-urgent surgeries. *Health Care Management Science*, *13*, 256–267.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.