INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 0 (2024) 1–26
DOI: 10.1111/itor.13584

Robust solutions via optimisation and predictive process monitoring for the scheduling of the interventional radiology procedures

Matteo Di Cunzolo[a], Massimiliano Ronzani[b,*], Roberto Aringhieri[a], Chiara Di Francescomarino[c], Chiara Ghidini[b], Alberto Guastalla[a] and Emilio Sulis[a]

[a] *Computer Science Department of the University of Torino, Corso Svizzera 185, Torino 10149, Italy*
[b] *Fondazione Bruno Kessler, Via Sommarive 18, Povo 38123, Italy*
[c] *Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, 38123, Povo, Italy*
E-mail: matteo.dicunzolo@edu.unito.it[Cunzolo]; mronzani@fbk.eu[Ronzani]; roberto.aringhieri@unito.it[Aringhieri]; c.difrancescomarino@unitn.it[Francescomarino]; ghidini@fbk.eu[Ghidini]; alberto.guastalla@unito.it[Guastalla]; emilio.sulis@unito.it[Sulis]

Received 3 January 2024; received in revised form 14 November 2024; accepted 17 November 2024

Abstract

Interventional radiology (IR) is an increasingly used medical specialty relying on the possibilities offered by medical imaging guidance technologies to perform minimally invasive procedures (both diagnostic and therapeutic) through very small incisions or body orifices. Although the operative context is quite similar to that of the classical operating room (OR) literature, to the best of our knowledge management problems arising in the IR operative context never appeared in the healthcare management literature. This is even more true for studies that combine the OR approach with automatic extraction of information from real hospital health record data as in the present study. Two specific features characterise our case study with respect to the traditional OR literature: due to the Italian legislation, the anaesthetist (usually in a very limited number) must be present for the entire duration of the procedure ($C1$), and the IR does not have its own ward but receives inpatients from different wards ($C2$). The aim of this paper is to introduce a novel approach to determine a robust solution for our case study problem addressing both features $C1$ and $C2$. Our approach is based on the interplay between optimisation and predictive process monitoring (PPM) models. The obtained results show that the proposed approach produces schedules that achieve higher usage rate, lower overtime and more patients operated on than the original schedule. We also show that the integration of PPM models within the optimisation workflow improves the quality of the output schedule with respect to the standard one-shot optimisation.

Keywords: optimisation; robustness; process mining; predictive process monitoring; interventional radiology; case study

*Corresponding author.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. Introduction

Interventional radiology (IR) is an increasingly used medical specialty relying on the possibilities offered by medical imaging guidance technologies to perform minimally invasive procedures (both diagnostic and therapeutic) through very small incisions or body orifices. Compared with traditional methods, the advantage is to decrease risks, pain and recovery time. Real-time visualisation also allows precise guidance to the abnormality. Although the operative context is quite similar to that of the classical operating room (OR) literature, to the best of our knowledge management problems arising in the IR operative context never appeared in the healthcare management literature (see, e.g., the reviews by Samudra et al., 2016, Hof et al., 2017, and Zhu et al., 2019).

This is even more true for studies that combine the OR approach with automatic extraction of information from real hospital health record data, as in the present study. In particular, the case study considered in this paper arises from the 'Molinette' hospital in Turin, one of the largest Italian cities, and a partner of the Circular Health for Industry (CH4I) research project (https://ch4i.di.unito.it/). Compared to traditional OR settings, procedures in IR are generally shorter and rely on specialised medical equipment, meaning that procedural variety has less impact on optimisation performance. However, two features of our case study characterise it with respect to traditional OR literature: first, the anaesthetist (usually in a very limited number) must be present for the entire duration of the procedure ($C1$); second, the IR unit does not have its own ward and instead receives inpatients from different wards ($C2$).

The feature $C1$ implies the need of considering, at the same time, (i) the problem of selecting patients from the waiting list and assigning their procedure to an IR room on a certain day over the planning horizon and (ii) the problem of determining the precise sequence of the IR procedures in such a way to avoid the overlap among different procedures assigned to the same anaesthetist. The resulting joint approach is not common also in the OR literature (see, e.g., Siqueira et al., 2018).

The feature $C2$ introduces a novel source of uncertainty not yet considered in the OR literature (see, e.g., Beaulieu et al., 2012): actually, the fact that the inpatients arrive from different wards inside the hospital means that they are not under control of the IR team introducing an exogenous factor affecting the IR activities, as reported by the medical doctors during the CH4I project.

To address this uncertainty, we leverage process mining (PM) in combination with optimisation. PM (van der Aalst et al., 2012) is a discipline at the intersection of data science and process management, focusing on analysing and improving operational processes based on event logs. Event logs capture the sequence of activities in operational processes, and PM techniques are used to discover process models, assess conformance and identify inefficiencies or deviations from expected behaviour. Extensive applications of these techniques in the healthcare domain exist (Munoz-Gama et al., 2022).

In particular, in this work, we employ predictive process monitoring (PPM) (Maggi et al., 2014; van der Aalst and Carmona, 2022), a subset of PM that uses predictive models, often based on machine learning (ML), to forecast future events or outcomes within a process. PPM is particularly valuable for predicting the remaining time of an ongoing process, detecting potential deviations and enabling proactive decision-making. In this paper, we employ PPM models to predict specific aspects of IR procedures, which are then used to inform the optimisation process.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The aim of this paper is to introduce a novel approach to determine a robust solution for our case study problem, addressing both features $C1$ and $C2$, that integrates optimisation and PPM models.

In particular, we employ two PPM models. A first PPM model predicts how long an IR procedure occupies the IR operating room in accordance with the procedure type and the patient's ward of origin. Such a time is then used to generate an initial optimised solution. An iterative procedure validates such a solution by exploiting a second PPM model – capable of predicting the overall daily running time of the IR operating room – in such a way as to determine a final and more robust solution.

The approach proposed here strongly differs from that previously discussed in Di Cunzolo et al., 2022: first, we consider only the problem of selecting patients from the waiting list and assigning their procedure to an IR operating room in a certain day over the planning horizon; then, there the PM techniques adopted serve only to identify delays and lagging cases in an analysis prior to the optimisation, conversely here we adopt ML-based PPM models in interplay with the optimisation.

The paper is organised as follows. Section 2 presents some related works. Section 3 introduces the reader to the PM and PPM. Section 4 introduces the case study considered in the paper. Section 5 describes the methodology developed while the evaluation, and the results are discussed in Section 6. Section 8 contains the final discussion and presents some future research directions.

2. Related work

Operational researchers provided in the last three decades several decision models and optimisation tools aimed at addressing the complexity of the OR management as evidenced by the large number of reviews in the last 15 years (Cardoen et al., 2010a; Cardoen et al., 2010b; Guerriero and Guido, 2011; Van Riet and Demeulemeester, 2015; Samudra et al., 2016; Hof et al., 2017; Zhu et al., 2019).

OR planning and scheduling are defined by three main decision levels: strategical, tactical and operational that are, respectively, based on long-, medium- and short-term objectives (Hulsof et al., 2012). Apart from recent integrated approaches (see, e.g. Aringhieri et al., 2015; Siqueira et al., 2018), usually these levels are studied or analysed separately, given by the complexity of the problem and the time horizon (e.g., the larger the horizon, the harder is the problem). Such studies can be further challenged by the inherent stochasticity of their main parameters, such as the surgery duration, the length of stay, the arrival of non-elective patients and often combined with staffing and scheduling decisions (Lamiri et al., 2008; Duma and Aringhieri, 2015; Landa et al., 2016; Duma and Aringhieri, 2018; Wang et al., 2018; Guido, 2024) as well as the limited availability of historical data (Coban et al., 2023).

The availability and the management of resources is one of the key factors in the development of the planning and scheduling strategies (Duma and Aringhieri, 2019; Landa et al., 2018). More generally, the problem of patient scheduling considering downstream resources is receiving increasing attention (see, e.g., van den Broek d'Obrenan et al., 2020; Schneider et al., 2020; Zhu et al., 2022). On the contrary, the problem of integrating OR planning and personnel scheduling (such as medical doctors, anaesthetists, etc.) is less considered (Breuer et al., 2020).

PM is a relatively recent discipline that focuses on extracting knowledge from data collected in enterprise information system databases (van der Aalst, 2016). In healthcare, they are called health information systems. Such data can be processed and organised into event logs, which is a

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

structured way of collecting a set of traces, each of which contains the activities performed for any particular process instance (van der Aalst, 2011). Process models can be represented using different modelling paradigms ranging from procedural (van der Aalst, 2016) to declarative (Pesic, 2008) and hybrid languages (van der Aalst et al., 2017).

Hospital Information System database can be the basis for applying PM and optimisation techniques in healthcare. An example application involves the discovery of patient paths (Prodel et al., 2015) by building process models from event logs (Prodel et al., 2018). Integer linear programming has been used to discover a Petri net model (Van der Werf et al., 2008), or to reproduce meaningful patient paths, and finally to build simulations that can consider behavioural and operational aspects (Halawa et al., 2021).

The combination of PM, optimisation, and simulation methods can be the basis for building decision support systems in healthcare (Moreira et al., 2019). PM techniques can be used in conjunction with optimisation techniques to improve planning, including through online optimisation algorithms (Duma and Aringhieri, 2020, 2023).

3. Process mining and predictive process monitoring

In this section, we introduce the main concepts of PM and its sub-field PPM.

3.1. Process mining

PM (van der Aalst et al., 2012) is a discipline that combines data science and business process science. Its focus is to improve operational processes by extracting and exploiting knowledge in a systematic way from event data.

PM techniques use a combination of event data and process models to perform several tasks: provide insights on the process, identify inefficiencies and bottlenecks, discover deviations with respect to the expected behaviour of the process, assess process performance and compliance with respect to a given model as well as support improvements of the process itself. These tasks are grouped into three main categories:

- *Process discovery*, which uses event data to discover a model that effectively describes the process;
- *Conformance checking*, which verifies if event data are in compliance with a given process model;
- *Process enhancement*, which exploits event data to contextually improve the existing process model.

Event data are collected in enterprise information systems, as in the case of hospital information systems in the healthcare scenario in which we are interested. To be suitable for the PM analysis, the data extracted from an information system have to be properly preprocessed and transformed in the so-called *event log* format. An event log is a collection of cases, and each case is in turn a sequence of events. Each case represents an execution of the process under study. In the following, we describe the main components of an event log:

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- *Event*. An event e_i refers to the execution of an activity (or event class). It is usually identified by three pieces of information: the activity that has been executed, the case ID which identifies the process instance in which the event has happened, and the timestamp related to the time in which the event has occurred. Furthermore, an event can carry additional information, also known as data payload (or *event attributes*), such as the resource involved in the execution of the activity, and other data recorded with the event.
- *Trace*. A trace (or case) corresponds to a single instance of the process and consists of a time-ordered sequence of events $\sigma = \langle e_1, e_2, \ldots, e_n \rangle$. In addition to the event sequence, also known as *control-flow* information, traces can include attributes that do not change from event to event in the trace execution (*trace attributes*), for example the birth place of a patient. For some, PM tasks can be useful to standardise the length of the traces in the dataset. We define a prefix σ^k of length k as a truncation of the trace σ if $k \leq n$: $\sigma^k = \langle e_1, e_2, \ldots, e_k \rangle$; or its extension if $k > n$: $\sigma^k = \langle e_1, e_2, \ldots, e_n, \varepsilon, \ldots, \varepsilon \rangle$, where ε is a padding symbol.
- *Event log*. An event log is a collection of traces. In an event log, traces can be temporally ordered based on their initial timestamp (the timestamp of the first event).

An important remark is that the same dataset can be looked at from different perspectives, and this conditions the way in which traces and events are defined. This is particularly relevant in the case of the distinguishing characteristics of PM in healthcare (Munoz-Gama et al., 2022). For example, a healthcare dataset collecting patients' examinations from different medical clinics, can be inspected by looking at the history of the patient through all the medical clinics, or by looking at the history of each clinic through all its patients. These are two different ways to analyse the same data. In the first case, the patient care pathway represents the trace and the medical clinics are the events. In the second case, each clinic represents a trace and the events correspond to the patients visiting the clinic.

PM techniques can be classified as backward-looking (e.g., finding the causes of inefficiencies in a process) or forward-looking (e.g., predicting the remaining processing time of a running trace). In this work, we focus on forward-looking techniques and specifically on PPM which is described in the next section.

3.2. *Predictive process monitoring*

PPM (Maggi et al., 2014; van der Aalst and Carmona, 2022) is a branch of PM that aims to predict some aspects of the future evolution of an ongoing or yet-to-start process execution. The ability to forecast relevant traits of the process execution can be really valuable in several domains and scenarios, allowing organisations to prevent undesired outcomes, issues and delays.

PPM techniques can be classified using three main features: the type of prediction, the technique or approach used to obtain the prediction and the information exploited in order to obtain the prediction. In the following, we separately analyse these three dimensions, with a special focus on the techniques used in this work.

Focusing on the prediction type, we can identify three main types of prediction:

- *Outcome prediction*, which aims to predict the possible outcome of a process (e.g., positive/negative);

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by Cochraneltalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- *Next activity prediction*, which aims to predict the continuation of a process execution;
- *Time-related prediction*, which aims to predict time-related aspects, as, for instance, the cycle time of a process execution or the time that a process instance will require to complete.

In this work, we are interested in time-related predictions (and more in-detail predictions on the deviances from the expected durations), since we aim to use the predicted execution time of surgical procedures to improve the scheduling.

Techniques and approaches used in PPM are grouped into two main classes: *model-based methods*, which rely on and exploit explicit process models; and *supervised learning* approaches, which rely on ML and statistical techniques, such as classification and regression models, as well as neural networks, by encoding event log information in terms of features. In this work, we consider the latter class, indeed we will build two classical ML regressors to predict the execution time of single and multiple surgical procedures, respectively.

There are many pieces of information contained in event logs that can be exploited in order to obtain a prediction. They can be classified into four classes: the *control flow*, that is the sequence of events of the trace or trace prefix; the *data payloads*, that is event and trace attributes associated with the event and the trace, respectively, *unstructured textual information* available with or included within the log; and *contextual information*, such as workload or resource availability or inter-dependencies between concurrent traces. In this work, we consider the first three types of information.

The information corresponding to every trace or prefix has to be properly encoded to be used by the predictive ML model. In this work, we used two popular encodings:

- The *frequency-based encoding* is a way of encoding control-flow information by keeping track of how many times each activity appears in the given trace. Note that this type of encoding does not keep track of the sequential order of the events in the trace. In this work, we will consider a simple variation of this encoding that contains also trace attributes (structured and unstructured) together with the control-flow information. The resulting feature vector is $(\vec{t}, f_{a_1}, \dots, f_{a_N})$, where $\vec{t}$ are trace attributes, N is the total number of different activities (the number of event classes) in the log and f_a is the number of times in which the activity a appears in the trace.
- The *complex index-based encoding* includes, besides the trace attributes, the sequence of events appearing in the trace together with their data payloads. The resulting feature vector is $(\vec{t}, e_1, \vec{p}_1, \dots, e_k, \vec{p}_k)$, where $\vec{t}$ are trace attributes, e_i is the ith event in the trace (according to the temporal order of the events in the trace) an $\vec{p}_i$ is the data payload of the ith event that includes both structured and unstructured information.

In ML-based PPM, encoded traces are then used to train the ML model. In this work, we adopt Random Forest regression models for time-related predictions. Random Forests are ensemble learning methods that build multiple decision trees and aggregate their results to improve predictive accuracy and robustness against overfitting.

The training methodology follows these steps: the dataset is split chronologically into three subsets – training, validation and test sets. The training set is used to train the model, the validation set is employed for hyperparameter optimisation and the test set is reserved for the final evaluation of the model's performance. Chronological splitting unsures a fair assessment by using the most

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

recent traces in the dataset for evaluation, reflecting a real-world setting where predictions are made on future process instances.

For model evaluation, standard metrics are used depending on the nature of the prediction task. In the case of regression models, a common choice is the mean absolute error (MAE), which measures the average magnitude of prediction errors. This metric provides a straightforward assessment of how closely the model's predictions align with the actual outcomes.

4. Case study

In this work, we focus on the real-world scenario based on the Department of Diagnostic Imaging and Interventional Radiology of the 'Molinette', which is part of the City of Health and Science (CHS) of Torino (https://www.cittadellasalute.to.it/), one of the main hospitals in Italy. IR is a medical speciality where minimally invasive procedures are performed using medical imaging guidance. The procedures are performed through very small incisions or body orifices, achieving the therapeutic goal with the minimum trauma for the patients.

In recent years, thanks to the availability of new materials, IR has replaced some traditional surgical techniques. Common interventions are radiological methods, computed tomography scans, magnetic resonance imaging and ultrasound. For example, interventional neuroradiology, which focuses on the diagnosis and treatment of diseases of the head, neck and spine, has proven to be of crucial importance in preventing ischaemic stroke.

IR procedures can be diagnostic or therapeutic. Diagnostic procedures support diagnosis making or guide medical treatments: they include image-guided biopsy of tumours and injection of an imaging contrast agent into blood vessels or ducts. Therapeutic procedures provide direct treatment: they include administration of drugs via catheter, placement of medical devices (e.g., stents) and angioplasty of narrowed structures.

In this paper, we use the term *procedure* to indicate the diagnostic or therapeutic intervention a patient undergoes and the term *service* to indicate the specific test or treatment a procedure is composed of. For instance, a procedure can be composed of two types of arteriography and a chemoembolization treatment. Procedures and services can also be classified as clean, dirty and infectious (e.g., Covid-19 cases). Dirty and infectious procedures and services, indeed, are the ones requiring that the OR to be cleaned more in-depth.

The operating theatre is composed of four different IR operating rooms (or IR rooms) denoted by S_1, S_2, S_3 and S_4. They have different equipment for different types of procedures: S_1 and S_2 are devoted to angioplasty, while S_3 and S_4 are used for procedures requiring echography and/or computed axial tomography.

4.1. Dataset

The dataset provided by the CHS hospital contains anonymised data of 400 procedures that have been performed in the IR room S_1 of the Department of Diagnostic Imaging and Interventional Radiology, in the period April 2022–September 2022, for a total of 27 weeks and 119 days. In Table 1, we report additional details on the daily number of procedures, procedure durations and IR session durations and overtime.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Details on procedures and IR sessions

	Min	Max	Mean	Median
Number of procedures per session	1	8	3.4	3
Procedures duration	4 minutes	2 hours 35 minutes	31 minutes	23 minutes
IR session total duration	20 minutes	7 hours 40 minutes	4 hours 20 minutes	4 hours 40 minutes
IR session overtime	0 minutes	3 hours 10 minutes	49 minutes	30 minutes

The dataset contains, for each procedure carried out, the patient's unique identifier (i.e., a medical record number) and the date of the procedure, demographic information related to the patient (i.e., age and gender), the clinical question (in natural language) that the procedure aims at addressing, the specific services provided when carrying out the procedure, the technique used and the hospital ward the patient comes from.

Moreover, the dataset contains detailed temporal data related to the time in which (i) the patient entered the IR department (the information regarding the patient exiting the IR department is not available in the dataset); (ii) the patient entered and exited the OR; (iii) the anaesthesia starts and ends; (iv) the patient is ready for the operation after receiving the anaesthesia; (v) the surgical procedure is started and ended; and (vi) the OR is cleaned and restored.

In addition, for every service provided by the IR department, we received two time estimations made by the doctors: the estimated execution time of the surgical procedure and the estimated OR occupancy, which also included the time for room restoration and cleaning. Finally, we received also the list of clean and dirty services.

While the dataset provides a robust foundation for evaluating our system with real-world data, it is important to acknowledge its limitations. The dataset is based on a single OR, which, while providing valuable insights, restricts the ability to explore optimisation across multiple rooms. Despite this, the real-world context of the data offers a meaningful evaluation of the system's performance, demonstrating its effectiveness in a realistic setting and offering valuable guidance for future improvements and adaptations in more complex scenarios.

5. Methodology

In this section, we describe the proposed methodology that combines PM and optimisation techniques. The main idea is to exploit the information contained in historical data about the utilisation of the IR operating rooms to improve the optimisation of the weekly IR management. Namely, we use PPM models to predict the deviations from the estimated execution time of procedures (PPM phase) and we then leverage this information – in place of the time estimations provided by the doctors – in the optimisation phase.

In Section 5.1, we detail the two PPM models employed to predict the temporal deviation of the single procedure OR occupancy and the daily OR session overtime, respectively, from doctors' estimates. In Section 5.2, we introduce a general mathematical model dealing with the problems determined by the feature $C1$, that is (i) the problem of selecting patients from the waiting list and assigning their procedure to an IR room on a certain day over the planning horizon and (ii)

© 2024 The Author(s).

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the problem of determining the precise sequence of the IR procedures in such a way to avoid the overlap among different procedures assigned to the same anaesthetist. In Section 5.3, we finally describe the whole approach that exploits the interplay between optimisation and PPM models.

5.1. *Predictive process monitoring*

In this section, we describe the PPM models that have been developed in order to support the optimisation phase. The aim of these models is to predict the execution time of the single procedures and the daily IR room occupancy and to use this information in place of the estimated time suggested by the doctors.

We trained two different regression models. The first model, *PPM1*, predicts the delay or advance of a procedure with respect to the time estimated by the doctors, independently of its position in the schedule of the IR room. The second model, referred to as *PPM2*, predicts the overall daily running time of the IR operating room taking into account the delay of all the procedures that occurred.

In the rest of this section, we describe in detail the two models.

5.1.1. *Data pre-processing*

First of all we pre-process the dataset in the following way. For each service, we add the execution time estimates suggested by the doctors: the duration of the surgical procedure and the occupancy of the OR including room restoration and cleaning time. In addition, based on the findings of Ronzani et al. (2022), we opt for not using the field 'clinical question' in its natural language form, since it has been shown that the direct use of the textual field in supervised learning PPM hardly improves the performance due to its high variability. Instead, we extract two binary flags from it, which correspond to the appearance of the words 'TARE' and 'TACE' in the text, respectively. These two words denote procedures which are commonly performed in IR. *TARE* is an abbreviation for Transarterial radioembolisation. It is a procedure made of two arteriographies: an arteriography of the celiac tripod and a liver arteriography. *TACE* is an abbreviation for transarterial chemoembolisation. It is a procedure composed of three services, the same two arteriographies as in TARE and loco-regional chemoembolisation.

5.1.2. *PPM1*

With the first PPM model, we aim to predict possible deviations (delays or advances) of a procedure with respect to the time estimate given by the doctors, independently of the position of the procedure within the daily schedule of the IR operating room (*IR session*). For this aim, we adopt the following perspective in which a trace is represented by the single procedure and the events correspond to the services that compose the procedure. We encode each trace in the following way. Since in the data, there is no information on the order in which the services are performed inside the procedure, we adopt the frequency encoding (see Section 3) to encode events. In addition, we consider the following trace attributes: patient's gender and age; TARE and TACE flags (see Data pre-processing); the patient's originating hospital ward and hospitalisation status; the technique used during the procedure; and the surgical time and OR occupancy estimated by the doctors. The surgical time and OR occupancy are given for every service; to obtain the estimate for the complete

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

procedure, we sum the values of all the services preformed during the procedure. Nominal features are encoded with one-hot encoding.

The label is defined as the difference between the real OR occupancy and the one estimated by the doctors.

The PPM model is obtained by training a Random Forest regressor (https://scikit-learn.org) The training procedure includes the optimisation of the hyperparameters. Further details on the training and the evaluation of the model are reported in Section 6.1.

5.1.3. PPM2

Differently from *PPM1*, which considers the single procedure independently of its position in the scheduling of the IR operating room, the second PPM model aims to predict the effective daily occupancy time (or *total IR time*) of the IR room taking into account the sequence of the scheduled procedures. In this way, we wish to leverage

- possible correlations between procedures occurring in the same *IR session* and deviations from the estimated time;
- possible correlations between the order of the scheduled procedures in the same *IR session* and deviations from the estimated time.

To this aim, we consider a different perspective on the data, in which the *IR session* defines a trace, and the events in the trace are the scheduled procedures. We encode every trace in the following way. Since we want to leverage the sequence of procedures, we opt for the complex-index encoding (see Section 3) that keeps track of the order of the events. We consider the following event attributes: patient's gender and age; TARE and TACE flags (see Data pre-processing); the patient's originating hospital ward and hospitalisation status and the technique used during the procedure. Clearly, we do not include the timestamps of the events in the encoding, since we want to predict the room occupancy starting only from the procedure schedule before any of the events are actually performed. In addition, we consider the following trace attributes: the weekday, the total surgical time estimated by the doctors and an estimate for the *total IR time* that can be either the initial estimation by the doctors or the prediction of *PPM1*. The surgical time and IR room occupancy estimates are summed over all the procedures scheduled in the *IR session*.

The label is defined as an upper bound of the *total IR time*, which represents the occupancy time of the OR in an *IR session*. To be more precise the label is defined as follows:

$$\text{label} = \sum_{\text{procedure}} \max(\text{procedure actual duration, procedure estimated duration}), \tag{1}$$

where the sum is over all the procedures of the *IR session* and the estimated duration of a procedure can be either the initial estimate by the doctors or the prediction of *PPM1*.

The definition of the label is related to the evaluation methodology reported in Section 6.2. Indeed one of the metrics used therein is the *total IR time*, and (1) defines an upper limit for it, since it considers all the delays but does not consider potential in advance. Indeed, a procedure in advance will not, in general, anticipate the next procedure that will anyway start at the scheduled time, so it will not reduce in general the *total IR time*.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The PPM model is obtained by training a Random Forest regressor. The training procedure includes the optimisation of the hyper-parameters. Further details on the training and the evaluation of the model are reported in Section 6.1.

5.2. Optimisation

We introduce an integer linear programming model that jointly addresses the two problems implied by the feature $C1$, that is (i) the problem of selecting patients from the waiting list and assigning their procedure to an IR room on a certain day over the planning horizon and (ii) the problem of determining the precise sequence of the IR procedures in such a way to avoid the overlap among different procedures assigned to the same anaesthetist. In our modelling approach, we would like to be as general as possible and in accordance with the OR literature. As a consequence, we refer to the usual terminology adopted in the literature and consider the whole operative context depicted in Section 4 made of four IR rooms and two types of procedures (or specialties) assuming that all the IR teams are always available.

Let I, J and K be, respectively, the set of patients, specialities and IR rooms, which are indexed by i, j, k. We would remark that set J is composed of two specialties, that is the procedures operated on S_1 and S_2 and the ones operated on S_3 and S_4, respectively. Let I_j be the set of patients belonging to the speciality j. Let T be the set of the days belonging to the planning horizon, indexed by t. Let (k, t) be the *IR session* on the IR room k during the day t.

Let p_i and s_{kt} be, respectively, the duration of the procedure of the patient i, and the maximum time allowed for procedures in the IR session (k, t). The priority of the patient i is denoted with r_i and is defined in accordance with the realistic patient priority model introduced by Valente et al. (2009) and deeply tested at S. Martino University Hospital of Genoa, Italy. The value r_i is a function of L_i, that is the time (in days) spent by the patient i waiting for the treatment, and m_i, that is the maximum time (in days) before treatment of the patient i, that is the *ideal* time within which a patient should receive their treatment.

The binary parameter τ_{kt}^{j} is equal to 1 if the OR k is assigned to speciality j on day t, 0 otherwise. The binary parameter $u_{i\ell}$ is equal to 1 if the patient i should be scheduled before the patient ℓ, 0 otherwise. This parameter is used to represent the sequence among clean, dirty and Covid-19 procedures.

Let A be the set of the available anaesthetists indexed by α. To represent the need and the availability of the anaesthetists, we introduce a binary parameter a_i, which is equal to 1 when the procedure of the patient i requires an anaesthetist, 0 otherwise, and the parameter $\bar{A}_{\alpha t}$, which represents the time availability of the anaesthetist α on day t, in minutes.

Before stating the optimisation model, some decision variables should be introduced. The binary variable x_{ikt} is equal to 1 if the patient i is selected for a procedure and assigned to the IR session (k, t), 0 otherwise. The binary variable $y_{i\ell kt}$ is equal to 1 if the patient i is scheduled before the patient ℓ in the same IR session (k, t), 0 otherwise. The binary variable $\lambda_{i\ell t}$ is equal to 1 if the patient i precedes the patient ℓ on day t independently of the IR room, 0 otherwise. The real variable $\gamma_i \in \mathcal{R}_+$ models the starting time (in minutes) of the procedure of the patient i. We assume that time starts from 0. The binary variable $\beta_{\alpha it}$ is equal to 1 if the anaesthetist α is assigned to the patient i on day t, 0 otherwise.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Constraints (2) model the selection of the patients from the waiting list (constraints (2a)) and their assignment to an IR session devoted to the specialty of the patient (constraints (2c) in which M^1_j can be set to the maximum number of executable procedures for each IR session (k, t) and for each specialty j) without exceeding its time capacity (constraints (2b)).

$$\sum_{k\in K}\sum_{t\in T} x_{ikt} \leq 1, \qquad i \in I, \tag{2a}$$

$$\sum_{i\in I} p_i x_{ikt} \leq s_{kt}, \qquad k \in K, t \in T, \tag{2b}$$

$$\sum_{i\in I_j} x_{ikt} \leq M^1_j \tau^j_{kt}. \qquad j \in J, k \in K, t \in T. \tag{2c}$$

Constraints (3) model the assignment of the anaesthetist α to patient i on session (k, t), if the procedure requires anaesthesia (constraints (3a)) in such a way as not to exceed the maximum time capacity $\bar{A}_{\alpha t}$ of the anaesthetist α on day t. Constraints (3c) avoid overlapping among patients requiring the same anaesthetist on the same day. If two patients i and ℓ have the same anaesthetist α on day t ($\beta_{\alpha it} + \beta_{\alpha\ell t} = 2$) and they are assigned to two different IR room ($x_{ikt} + x_{\ell ht} = 2$), then one of two has to end the procedure before the other ($\gamma_i + p_i \leq \gamma_\ell$). Note that M^2_i can be set to the maximum s_{kt} value for an IR session (k, t) to which the patient i can be assigned. Constraints (3d) impose a precedence among each pair of patients.

$$\sum_{\alpha\in A} \beta_{\alpha it} = a_i \sum_{k} x_{ikt}, \qquad i \in I, t \in T \tag{3a}$$

$$\sum_{i\in I} \beta_{\alpha it} p_i \leq \bar{A}_{\alpha t}, \qquad \alpha \in A, t \in T \tag{3b}$$

$$\gamma_i + p_i \leq \gamma_\ell + M^2_i(5 - \beta_{\alpha it} - \beta_{\alpha\ell t} - x_{ikt} - x_{\ell ht} - \lambda_{i\ell t}), \qquad i, \ell \in I, i \neq \ell, \; k, h \in K, k \neq h, \; t \in T, \alpha \in A, \tag{3c}$$

$$\lambda_{i\ell t} + \lambda_{\ell it} = 1, \qquad i, \ell \in I, i < \ell, t \in T. \tag{3d}$$

Constraints (4) model the sequencing of the patients assigned to the same OR. Constraints (4a) reinforce constraints (2b) by stating that the starting time of the procedure of the patient i plus the duration should not exceed s_{kt}. Constraints (4b) guarantee that two patients i and ℓ do not overlap in the same OR k on day t. Constraints (4c) and (4d) determine the correct sequencing in accordance with the parameter $u_{i\ell}$.

$$\gamma_i + p_i \leq s_{kt}, \qquad i \in I, k \in K, t \in T, \tag{4a}$$

$$\gamma_i + p_i \leq \gamma_\ell + M^2_i(3 - x_{ikt} - x_{\ell kt} - y_{i\ell kt}), \qquad i, \ell \in I, i \neq \ell, k \in K, t \in T, \tag{4b}$$

$$\gamma_i u_{i\ell} \leq \gamma_\ell(1 - u_{\ell i}) + M^2_i(2 - x_{ikt} - x_{\ell kt}), \qquad i, \ell \in I, i \neq \ell, k \in K, t \in T, \tag{4c}$$

$$y_{i\ell kt} + y_{\ell ikt} = 1. \qquad i, \ell \in I, i < \ell, k \in K, t \in T. \tag{4d}$$

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by Cochraneltalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

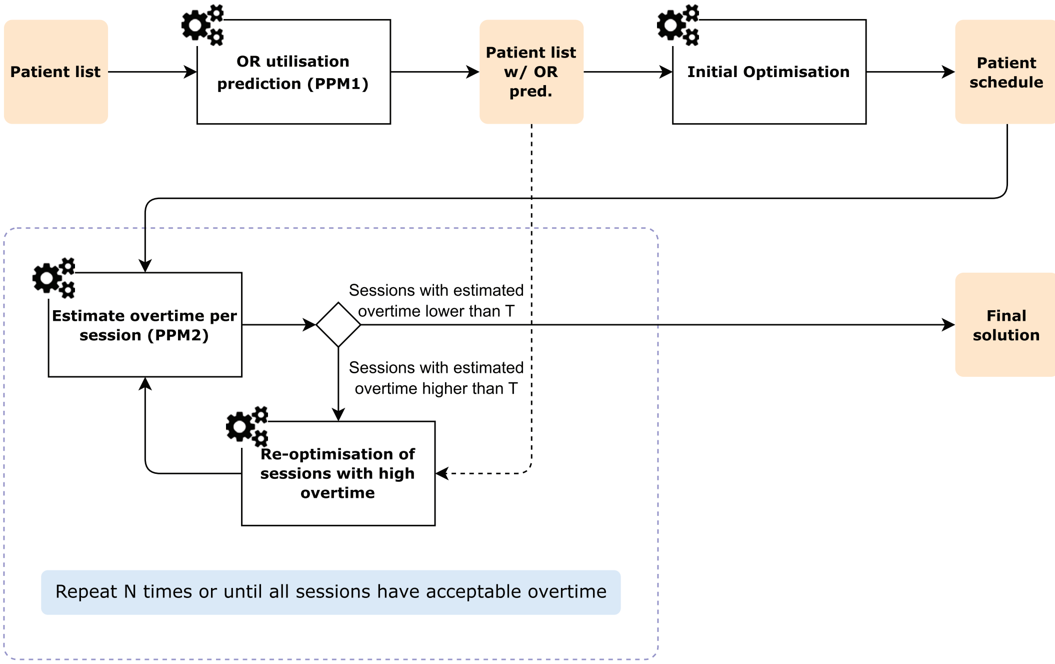


Fig. 1. Optimisation workflow which integrates the PPM models.

Finally, constraints (5) define the domains of the decision variables.

$$x_{ikt},\ y_{i\ell kt},\ \lambda_{i\ell t},\ \beta_{\alpha it} \in \{0, 1\},\ \gamma_i \in \mathcal{R}_+ \quad i, \ell \in I, k \in K, t \in T, \alpha \in A. \tag{5}$$

The objective of the optimisation models is to maximise the utilisation of the IR sessions and, at the secondary level, to select the patients having higher priority r_i, that is maximum waiting time with respect to their maximum time before treatment. We would remark that maximising the waiting time of selected patients is a proxy of the OR utilisation as proved by Aringhieri and Duma (2017) and Aringhieri et al. (2022). Therefore, the objective function of our optimisation model can be simplified as follows:

$$\max z = \sum_{i \in I} r_i \sum_{k \in K} \sum_{t \in T} x_{ikt}. \tag{6}$$

5.3. *Integrating optimisation and predictive process monitoring models*

In this section, we illustrate how the PPM models described in Section 5.1 can be integrated with the optimisation phase described in Section 5.2.

The complete workflow is presented in Fig. 1. The patient waiting list is initially processed by the first predictive model *PPM1*, which adds to every procedure the predicted OR occupancy time. We recall that *PPM1* predicts how much the procedure occupancy time deviates from the estimate given

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Details of training and test datasets

Dataset	Min date	Max date	Weeks	Days	Procedures
Training	Apr 1	Sept 4	23	99 (83%)	324 (81%)
Test	Sept 5	Sept 30	4	20 (17%)	76 (19%)

by the doctors. The waiting list, enriched with this temporal information, is then fed forward to a first optimisation phase that generates an initial schedule. This schedule is subsequently analysed using the second predictive model, *PPM2*, which estimates and quantifies the possible overtime of every *IR session* in the schedule. If all the predicted overtime are lower or equal to a given temporal threshold T, then the schedule is accepted as the final solution. If, instead, some IR sessions have predicted overtime bigger than the threshold T, an additional optimisation phase is performed for these sessions, in which at least one patient is changed or removed in every session, and a new schedule is obtained. This is obtained by adding the constraint

$$\sum_{i \in I} 1 - x_{ikt} \geq 1 \qquad \forall (k, t) \text{ s.t. time estimate} > T,$$

to the model reported in the previous section, fixing all the variables regarding the IR session whose estimate is less than or equal to T, and solving the resulting problem.

The new schedule is again evaluated by the *PPM2* model that predicts the overtime for the adjusted *IR session*. The cycle is iterated until no overtime is bigger than the threshold T or a maximum number of iterations N has been executed.

6. Evaluation

In this section, we evaluate the methodology described in Section 5. To do this, we split our dataset (see Section 4) into two parts: a training set and a test set, so that the latter corresponds to the last four weeks in the dataset. The details of the split are reported in Table 2.

The training set is exclusively used to train the PPM models described in Section 5.1. The test set is used both to assess the PPM model (see the next section), and as a waiting list that is used in the optimisation phase described in Section 5.3. In this section, we give an exhaustive description of the evaluation techniques adopted to test our methodology and the obtained results. Section 6.1 contains the evaluation of the PPM models, while Section 6.2 contains the evaluation of the optimised solutions obtained with the integrated approach, and its comparison with the standard one-shot optimisation which does not use the information from the PPM models.

6.1. Predictive process monitoring

In this section, we describe the evaluation of the two PPM models described in Section 5.1. Even though these two models have been created to support the optimisation phase, as described in Section 5.3, it is still worthwhile to evaluate them independently of the potential improvement of

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3
Hyper-parameter optimisation. The MAE values correspond to the fivefold cross-validation on the training set

Model	MAE	Hyper-parameters			
		n_estimator	max_depth	min_samples_split	min_samples_leaf
PPM1	15.6	50	6	2	2
PPM2 (with OR occupancy estimated by doctors)	22.6	50	6	4	1
PPM2 (with OR occupancy predicted by *PPM1*)	21.4	10	4	8	1

Table 4
Test set evaluation. The MAE is reported together with the mean absolute value of the label for comparison

Model	MAE	Mean absolute label value
PPM1	15.9	38.6
PPM2 (with OR occupancy estimated by doctors)	21.2	337.5
PPM2 (with OR occupancy predicted by *PPM1*)	26.1	249.9

the optimisation performance. The analysis of the combination of PPM and optimisation is done in the next section.

6.1.1. Evaluation setting

The two PPM models are first trained using the training dataset opportunely encoded, as described in Section 5.1. In addition, we perform hyper-parameter optimisation of our predictive models.

Due to the small size of our dataset, we decide to use the same training dataset for both models *PPM1* and *PPM2*. For the same motivation, we opt for not using a dedicated validation set to perform hyper-parameter tuning, instead, we use fivefold cross-validation inside the training set. To build both models we opt for a classical ML regressor, the Random Forest regressor (we employ the scikit-learn implementation of the Random Forest regressor https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html). We selected a *random_state* equal to zero for reproducibility purposes and we performed the hyper-parameter tuning on the following parameters of the model: *n_estimator*, *max_depth*, *min_samples_split* and *min_samples_leaf*. The metric used to validate and test the two models is the MAE.

6.1.2. Evaluation metrics

To evaluate the performance of the regression models, we use the MAE. MAE measures the average absolute difference between the model's predictions and the actual values (ground truth), providing a direct measure of the regressor's accuracy. Specifically, it is calculated as $|\mathrm{PPM}(x) - y|$, where $\mathrm{PPM}(x)$ represents the predicted value from the PPM model for a given input trace x, and y is the corresponding ground-truth label.

6.1.3. Evaluation results

In Table 3 and Table 4 we report the results of the hyper-parameter optimisation and the test set evaluation for the two models, respectively. The model *PPM2* has been evaluated twice, one for

© 2024 The Author(s).

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

every choice of the procedure OR occupancy time estimate – respectively, the estimate by the doctors (first row) and the estimate based on the predictions of *PPM1* (second row). In both cases, the estimates are used both as input to the model and to define the label in Equation (1). From Table 4, we observe that the performance of the two models is not particularly strong. The *PPM1* model, predicting the deviation in the procedure occupancy time with respect to the doctors estimation, has an average MAE of 16 minutes, which corresponds to about one half of the average procedure duration shown in Table 1. The *PPM2* model, which predicts the *total IR time*, obtains a MAE of 21 and 26 minutes in the two variants analysed, respectively, which is about a half of the mean session overtime, as shown in Table 1.

A possible reason for such a result could be the lack of data for training the model. Indeed, we see in Table 2 that we have only 324 traces to train *PPM1* and only 99 traces to train *PPM2*. However, we will see in the next section that, despite the modest performance of the PPM models, the integration of their predictions in the optimisation workflow can indeed increase the quality of the scheduling produced.

6.2. *Optimisation*

In this section, we assess the main contribution of this paper, which is the integration of the PPM models within the optimisation workflow described in Section 5.3. This evaluation aims at answering the following research questions:

- **RQ1**: Does the integration of the PPM models in the optimisation workflow enhance the results of the optimisation?
- **RQ2**: How does each of the two PPM models individually contribute to the outcome of the optimisation?
- **RQ3**: How does the choice of the threshold parameter T (Section 5.3) affect the outcome of the optimisation?

6.2.1. *Evaluation setting*

We apply the optimisation workflow described in Section 5.3 to the waiting list obtained from the test dataset (see Table 2) producing a schedule for 15 days. Such a solution is computed by solving a reduced version of the mathematical model reported in Section 5.2 restricted to only one IR room, namely S_1. We solved it by using a general-purpose solver, such as Cplex.

In order to answer the research questions presented at the beginning of this section, we consider and evaluate the following variants of our optimisation workflow:

- V_one-shot is the standard one-shot optimisation obtained by using the procedure IR room occupancy times estimated by the doctors.
- V_full is the full optimisation workflow of Fig. 1 that uses the procedure IR room occupancy times predicted by the *PPM1* model and the iteration through the *PPM2* model. We consider two values of the threshold T: $T = 15$ (V_full_T15) and $T = 20$ (V_full_T20).
- V_PPM1_one-shot is a one-shot optimisation obtained using the procedure IR room occupancy times predicted by the *PPM1* model.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- V_PPM2 leverages the iterative optimisation of Fig. 1 excluding the *PPM1* module, that is using the procedure IR room occupancy times estimated by the doctors. In this case, also the *PPM2* model is trained without using the information on procedure IR room occupancy time predicted by *PPM1* (see Section 6.1). We consider two values of the overtime threshold T: $T = 5$ (V_PPM2_T5) and $T = 10$ (V_PPM2_T10).

Note that for the variant V_PPM2 we consider smaller values of the overtime threshold T than the one used for the variant V_full. This is because the procedure IR room occupancy times estimated by the doctors are usually overestimated, and therefore the schedule produced is inherently more conservative. In particular, already at the first iteration step of the V_PPM2 workflow, each IR session overtime is already lower than 15 minutes.

For the variants which include the *PPM2* iterative optimisation, we consider a maximum iteration number $N = 100$. In addition, we compare all the schedules obtained against the actual schedule that was used in the historical data during the first 15 days of the test set. We label this schedule as V_historical.

6.2.2. Evaluation metrics

We introduce the two metrics used to assess the schedules produced by the optimisation workflow. First of all, we define some preliminary concepts:

- the *planned IR time* is the time interval in which the IR room is scheduled to operate on;
- the *total IR time* is defined as the time interval from the scheduled opening of the IR room to the end of its restoration after the last procedure. This can be larger than the *planned IR time* due to procedures with delays;
- the *real IR occupancy time* is the total time in which the IR room was actually occupied, either by carrying out a surgical procedure or by a post-operation IR room restoration operation.

The two metrics that we consider are

- *IR usage rate* U: This is the ratio between the *real IR occupancy time* and the *total IR time*. It measures how efficiently the OR is used. A higher value for the IR usage rate is preferred, as it indicates more efficient utilisation of the OR. It is difficult to ascertain an adequate value for the usage rate with precision. This value is significantly influenced by the operative context in relation to the patient case mix and the composition of the patient flow (e.g. elective and non-elective patients) that the operating theatre is required to manage. Nonetheless, values exceeding 75% are in accordance with values deemed adequate in the existing literature (Tyler et al., 2003).
- *overtime* OT: This is the difference between *total IR time* and the *planned IR time*. It reflects resource wastage due to the use of IR rooms beyond scheduled times. Zero overtime is ideal, but overtime below a certain threshold is typically considered acceptable. Also in this case this threshold depends deeply on the operative context. In our case study, overtime of less than 15 minutes is acceptable, as it does not incur extra costs for the hospital.

As an example, consider an IR room, scheduled to operate from 8.00 AM to 1.00 PM. This means that the *planned IR time* is 300 minutes. Two procedures are carried out: the first one goes from 8.00

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 5
Optimised schedule performances compared with the performance of the historical schedule

Optimisation variant	Days	Iterations	Patients	Mean OT (min)	Mean U (%)
V_historical	15	–	67	37	72%
V_one-shot	15	–	60	12	64%
V_PPM1_one-shot	15	–	72	11	81%
V_PPM2_T5	15	37	60	9	64%
V_PPM2_T10	15	3	60	11	64%
V_full_T15	15	62	71	7	81%
V_full_T20	15	21	71	8	81%

AM to 10.30 AM and the second one from 11.00 AM to 1.10 PM, so that the *total IR time* is 310 minutes and the *real IR occupancy time* is 280 minutes. Therefore, $U = 90\%$ and $OT = 10$ minutes.

It is important to stress that these two metrics correspond to different aspects of the schedule performance and therefore a good schedule should improve both of them even if these are two conflicting metrics. Indeed, one could reduce overtime by simply inserting less procedures in an IR session, but this will reduce the IR usage rate. Conversely, a denser schedule will probably increase the IR usage rate at the expense of higher overtime.

We now describe how these two metrics are computed for the schedule produced by our optimisation workflow. The schedule obtained by the optimisation is made of the sequence of the scheduled procedures and their estimated durations that are used for computing the scheduled starting time of each procedure. We consider two possibilities for the procedure estimated duration: the initial estimate made by the doctors and the predicted duration obtained by the *PPM1* model. In order to compute the metrics for the schedule produced by the optimisation workflow, we make the assumption that the procedure actual IR room occupancy time – which takes into account also the time necessary to restore the room after its use – is the one that has been observed historically and that has been extracted from the test dataset. This could be a strong working assumptions since the duration of a procedure may depend on many contingent or external factors. With this assumption, the *real IR occupancy time* is computed as the sum of all the scheduled procedures historical durations.

To compute the *total IR time*, we use the following method: if a procedure ends in advance, the next procedure will still start at the scheduled time, and so this will not impact the *total IR time*. Conversely, if the procedure ends late the next procedure will be delayed, and this will contribute to increasing the *total IR time*. In this way, if a procedure is ahead of its estimated duration it does not reduce, in general, the *overtime* but it does reduce the *IR usage rate*. The *planned IR time* is 300 minutes for all days.

6.2.3. Evaluation results

The performance of the original schedule and the optimised schedules obtained by the optimisation workflow is presented in Table 5. All the schedules have a time horizon of 15 days. For the optimisation variants that include the iterative optimisation leveraging the *PPM2* model, we report how many iterations were necessary to obtain a schedule in which all the daily sessions have a predicted overtime lower than T. For all the variants, we were able to get that all the sessions have an

© 2024 The Author(s).

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

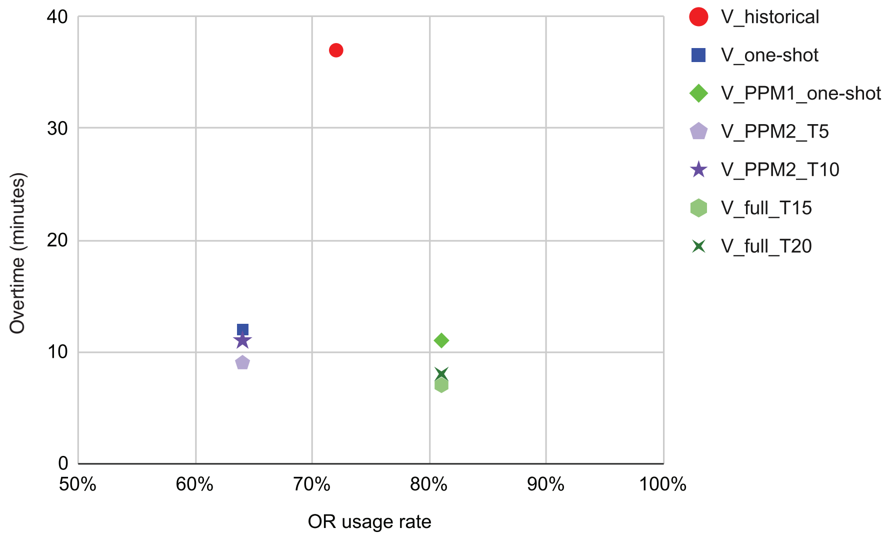


Fig. 2. Mean overtime OT in minutes versus mean usage rate U for the historical and the optimised schedules.

overtime under the threshold so that none of the variants reached the limit of iteration $N = 100$. Moreover, for each variant, we report how many patients the schedule includes in the time horizon, the mean overtime OT and the mean usage rate U.

Note that each optimisation cycle takes at most 60 seconds, which is set as a time limit. Many optimisation cycles are actually much faster than 60 seconds since most of the OR sessions are already fixed so that the optimisation problem is much simpler. For instance, V_full_T15 takes about 283 seconds, with the longer three iterations taking 55 seconds each on average, and the remaining fifty-nine iterations taking less than 2 seconds each.

In the rest of this section, we will discuss the results to address the research questions outlined in Section 6.2.

6.2.4. Answering RQ1

It is evident from Table 5 that all the optimisation variants produce significantly lower overtime than the original schedule (V_historical). However, not all the optimised schedules achieve the same usage rate U of the original schedule. In particular, the variants producing lower usage rates are those that do not employ the *PPM1* model's predictions, as we will explain further while addressing RQ2. Ultimately, we can notice that the schedules produced by the full optimisation workflow, which integrates the two *PPM* models, V_full_T15 and V_full_T20, are the ones with the best performances, obtaining the highest usage rate and producing the lowest overtime. This answers to the first research question RQ1. Namely, the integration of the *PPM* models allows us to obtain an excellent usage rate of over 80%, which is 17 percentage points higher than the ordinary one-shot optimisation, while also reasonably reducing overtime. These results are also displayed in Fig. 2.

Specifically, the integration of the PPM models enables us to achieve an excellent usage rate of over 80.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6.2.5. Answering RQ2

To answer the second research question, we comment on how the different components of our optimisation workflow, presented in Section 5.3, independently affect the performance of the schedule produced. From Fig. 2, it is clear that the effects of *PPM1* and *PPM2* are distinct. Specifically, employing procedure durations predicted by *PPM1* in the optimisation significantly improves the mean IR usage rate. In fact, all optimisation models that incorporate *PPM1*'s predictions (V_PPM1_one-shot, V_full_T15, V_full_T20) achieve a higher usage rate than the original schedule (V_historical). Conversely, optimisation models that do not use *PPM1*'s prediction (V_one-shot, V_PPM2_T5, V_PPM2_T10) obtain a lower usage rate compared to the original schedule. This can be attributed to the fact that these latter models rely on room occupancies estimated by doctors, which are typically overestimated. On the contrary, using the predictions of the *PPM1* model, which are on average closer to the actual duration of the procedures (see Section 6.1) allows the optimisation phase to generate better solutions.

On the other hand, incorporating the room occupancy predictions from *PPM1* does not lead to an improvement in overtime. Conversely, the integration of the iterative optimisation that leverages the predictions of *PPM2* does not affect the IR usage rate but seems to improve appreciably the total overtime, independently of the presence of the *PPM1* module. This behaviour is somehow expected, since the purpose of the *PPM2* module is indeed to re-optimise the IR sessions that, according to the prediction model, generate an important overtime (higher than T).

The overtime reduction might seem modest but is actually significant: for example, the use of the schedule obtained with V_full_T15 saves 70 minutes in 15 days with respect to the one obtained using V_PPM1_one-shot. This allows two IR sessions not to exceed the 15-minute threshold for an acceptable overtime, thus avoiding the related costs. Furthermore, the overtime reduction should be evaluated also considering the increased number of procedures performed, that is 72 for V_PPM1_one-shot and 71 for V_full_T15 and V_full_T20 with respect to the 67 performed on the original schedule (V_historical).

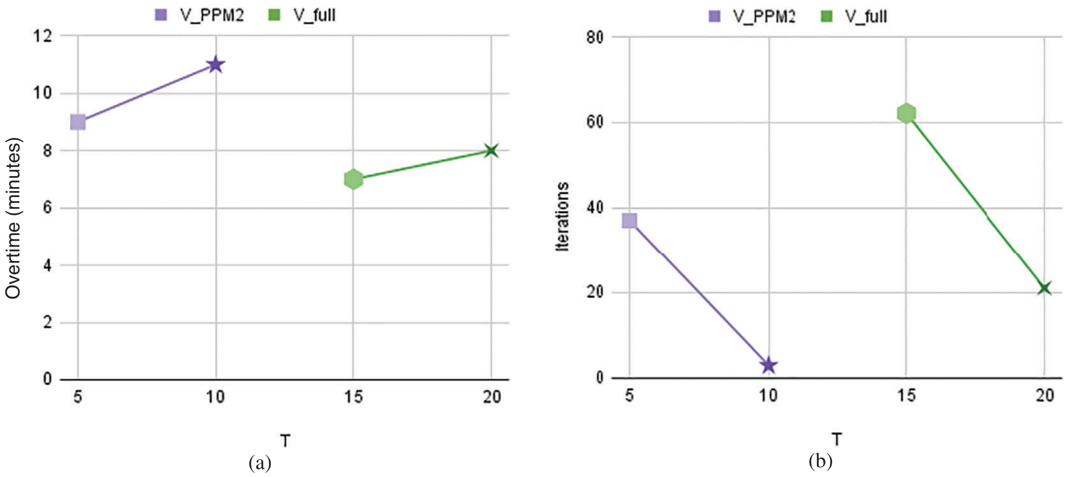
6.2.6. Answering RQ3

Finally, we comment on the role of the overtime threshold T used in the iterative optimisation, thus answering to the third research question. In Fig. 3, we display the dependency of both overtime (Fig. 3a) and the number of iterations (Fig. 3b) on the threshold value T for the two optimisation variants that incorporate the iterative optimisation module based on the *PPM2* predictions (V_PPM2, V_full). We can see that the effect of the parameter T is as expected: reducing the threshold results in schedules with lower overtime, even if the difference in the mean overtime is only a few minutes for both models. The disadvantage of lowering T is that the optimisation workflow needs more iterations to produce the optimal schedule, thus taking a longer time to finish the computation. Specifically, we observe that for V_PPM2 increasing the threshold from 5 to 10 minutes significantly speeds up the convergence of the optimisation workflow, allowing a final acceptable schedule to be reached after only three iterations through the *PPM2* module, with an increase in mean overtime of just two minutes. Similarly, for V_full, increasing the threshold from 15 to 20 minutes reduces the number of iterations by almost a third, with only a 1-minute increase in overtime.

Finally, we comment on the difference in behaviour between V_PPM2 and V_full. It can be observed that the latter requires larger threshold values to achieve a comparable number of iterations

© 2024 The Author(s).

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) (b)

Fig. 3. Dependence of overtime (a) and optimisation iterations (b) on the threshold value T.

relative to the former. This is due to the fact that the integrating *PPM1* predictions in V_full results in tighter schedules, with a higher average number of procedures scheduled per IR session (see Table 5). This likely conditions the *PPM2* model to predict greater overtime, which in turn prolongs the convergence in the iterative cycle.

7. Discussion and limitations

In this section, we discuss the advantages and limitations of the proposed method. Our discussion is organized around two main aspects: the *computational feasibility* of our solution and the potential application of *alternative optimisation approaches* to our problem.

7.0.1. Computational Feasibility

The problem we address falls into the larger family of OR planning problems. Specifically, our problem can be traced back to the surgical case assignment problem (SCAP), which has been shown to be NP-hard in Aringhieri et al. (2015). Nevertheless, we designed our models with computational efficiency in mind, using open-source tools such as Pyomo (https://www.pyomo.org) and Coin-OR (https://www.coin-or.org) to perform the optimization. Given the moderate size of our problem, our methodology runs efficiently on standard PC hardware, which was a key requirement for the research project funding this work.

The ML models employed for the PPM use a Random Forest architecture, which is both quick to train and memory-efficient. Each model can be trained in under two minutes on a standard PC and during each optimisation iteration, PPM inference requires only a fraction of a second, keeping it computationally lightweight within the overall process.

7.0.2. Comparison with alternative methods

In this work, we focused on demonstrating the effectiveness of integrating PPM with optimisation rather than comparing our solution with alternative methods. Given that we are addressing a

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

deterministic version of the problem, a comparison with other methods would add limited value. Using exact optimisation allows us to guarantee an optimal solution, which is feasible within acceptable time limits given our problem size. Although heuristic or non-exact methods can be useful for larger problem instances, they do not guarantee optimality and are generally applied when exact methods become computationally infeasible due to factors like an increase in patients, rooms or anaesthetists. For the current scope and problem instance sizes, our exact method is both feasible and optimal.

Additionally, our solution is agnostic about the choice of ML regressor. However, comparing different ML methods could offer insights into how varying prediction accuracies impact scheduling performance. Due to the limited data available in this project, we chose a Random Forest regressor, which is computationally efficient and performs reliably on small datasets, making it suitable for training and deployment for resource-constrained environments.

8. Conclusion

In this work, we introduced a new optimisation approach to improve the management of an IR operating room computing more robust schedules. The main innovation is the integration inside the optimisation workflow of two PPM models, which help to estimate surgical procedure durations and IR session overtime.

We performed the evaluation of our method using a real-world dataset containing information about four weeks of IR sessions in an IR department. We extracted from the dataset a waiting list of patients and obtained an optimised schedule for fifteen days. We compared this schedule against the original schedule used in reality in terms of mean IR usage rate and mean IR session overtime. We also compared our method with a standard optimisation method that does not leverage PPM models as well as with different variants of our methods that integrate only one of the two PPM models.

We also compared our method with a standard optimisation approach that does not leverage PPM models, as well as with different variants of our method that integrate only one of the two PPM models.

The obtained results show that the proposed approach produces schedules that achieve higher usage rates, lower overtime, and more patients operated on than the original schedule. We also show that the integration of the two PPM models within the optimisation workflow improves the quality of the output schedule with respect to the standard one-shot optimisation. In particular, we show that each of the two PPM models used in the optimisation workflow contributes to improve one of the two dimensions (mean IR usage rate and mean IR session), thus resulting in an improvement on both the dimensions when used together.

As a future direction, we plan to mitigate an inherent limitation of the evaluation carried out in this work. Indeed, in the evaluation, we make the assumption that the durations of the procedures will be exactly the same as the historical data. The duration of a procedure, however, may also very likely depend on external factors (e.g., the day of the execution, the resources available, etc.). We expect that our optimisation approach already takes in part into account these factors. Indeed, the iterative optimisation ensures the robustness of the schedule generated since it leverages on the *PPM2* model that already considers external factors in the prediction of the overtime. To overtake

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by Cochraneltalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

this evaluation limitation and assess the robustness of our optimisation method, one could employ simulation models that reproduce faithfully the time evolution of processes (Camargo et al., 2022; Meneghello et al., 2023). This kind of simulation indeed uses contingent factors (intercase features, resource availability) for the assignment of events' durations. Another direction of investigation is the application of explainable AI (XAI) techniques to the *PPM2* model in the iterative optimisation phase. This integration could speed up this phase, allowing it to converge to the final solution in less iterations. Indeed, the application of XAI techniques to the PPM model (Rizzi et al., 2020) can identify the main cause for an IR session overtime. In this way, we could obtain guidance on the replacements and on the changes to apply to the IR session schedule to reduce the overtime below the T threshold.

From an optimisation perspective, a future direction of investigation is the development of optimisation algorithms to improve the efficiency of the optimisation phase in such a way to deal with larger problems in terms of the number of patients and IR rooms.

A promising future direction for research in the integration of PPM and optimisation models is the replacement of the PPM regressor models with classifiers. Instead of estimating the exact delay, the objective would be to generate an alert that signals the likelihood of a delay with sufficient confidence. This approach could enhance the optimisation process by focusing on identifying potential delays rather than quantifying them. The assumption that a PPM model is available for estimating the delay of individual procedures allows for the development of a robust optimisation model, similar to that proposed by Addis et al. (2016), capable of handling a given number of the procedures with the largest estimated delay.

More generally, it would be valuable to validate the proposed approach to the classical surgery process scheduling problem, as this involves managing multiple ORs and more complex procedures. Such validation could further demonstrate its potential to improve efficiency and resource utilisation in a wider range of medical settings.

Acknowledgements

The authors acknowledge the support of Ilaria Angela Amantea, Paolo Fonio and Marco Grosso for their help in the definition of the problem and for providing the data used in this paper. The research is partially funded by the projects: 'Circular Health for Industry' project funded by 'Compagnia di San Paolo', call 'Intelligenza Artificiale, uomo e società' and 'RIsPOSTE: seRvizi dIgitali Per cOntraSTare l'Epidemia', funded by 'Bando InfraP2' of Regione Piemonte. The research of M. Ronzani is partially funded by the PNRR project FAIR - Future AI Research (PE00000013), under the NRRP MUR program funded by the NextGenerationEU.

Open access publishing facilitated by Universita degli Studi di Torino, as part of the Wiley - CRUI-CARE agreement.

References

Addis, B., Carello, G., Grosso, A., Tànfani, E., 2016. Operating room scheduling and rescheduling: a rolling horizon approach. *Flexible Services and Manufacturing Journal* 28, 1–2, 206–232.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Aringhieri, R., Duma, D., 2017. Patient–centred objectives as an alternative to maximum utilisation: comparing surgical case solutions. In *Optimization and Decision Science: Methodologies and Applications*. Springer, Berlin, pp. 105–112.

Aringhieri, R., Duma, D., Landa, P., Mancini, S., 2022. Combining workload balance and patient priority maximisation in operating room planning through hierarchical multi-objective optimisation. *European Journal of Operational Research* 298, 2, 627–643.

Aringhieri, R., Landa, P., Soriano, P., Tànfani, E., Testi, A., 2015. A two level metaheuristic for the operating room scheduling and assignment problem. *Computer & Operations Research* 54, 21–34.

Beaulieu, I., Gendreau, M., Soriano, P., 2012. Operating rooms scheduling under uncertainty. In Tànfani, E., Testi, A. (eds), *Advanced Decision Making Methods Applied to Health Care*, International Series in Operations Research & Management Science, Vol. 173. Springer Milan, pp. 13–32.

Breuer, D.J., Lahrichi, N., Clark, D.E., Benneyan, J.C., 2020. Robust combined operating room planning and personnel scheduling under uncertainty. *Operations Research for Health Care* 27, 100276.

Camargo, M., Dumas, M., González-Rojas, O., 2022. Learning accurate business process simulation models from event logs via automated process discovery and deep learning. In Franch, X., Poels, G., Gailly, F., Snoeck, M. (eds) *Advanced Information Systems Engineering*. Springer International Publishing, Cham, pp. 55–71.

Cardoen, B., Demeulemeester, E., Beliën, J., 2010a. Operating room planning and scheduling: a literature review. *European Journal of Operational Research* 201, 921–932.

Cardoen, B., Demeulemeester, E., Beliën, J., 2010b. Operating room planning and scheduling problems: a classification scheme. *International Journal of Health Management and Information* 1, 1, 71–83.

Coban, E., Kayış, E., Dexter, F., 2023. The effect of few historical data on the performance of sample average approximation method for operating room scheduling. *International Transactions in Operational Research* 30, 1, 126–150.

Di Cunzolo, M., Guastalla, A., Aringhieri, R., Sulis, E., Amantea, I.A., Ronzani, M., Francescomarino, C.D., Ghidini, C., Fonio, P., Grosso, M., 2022. Combining process mining and optimization: a scheduling application in healthcare. In Cabanillas, C., Garmann-Johnsen, N.F., Koschmider, A. (eds), *Business Process Management Workshops – BPM 2022 International Workshops, Münster, Germany, September 11-16, 2022, Revised Selected Papers*, Springer, Cham, pp. 197–209.

Duma, D., Aringhieri, R., 2015. An online optimization approach for the real time management of operating rooms. *Operations Research for Health Care* 7, 40–51.

Duma, D., Aringhieri, R., 2018. The real time management of operating rooms. In Kahraman, C., Topcu, I. (eds) *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science, Vol. 262. Springer International Publishing, Cham, pp. 55–79.

Duma, D., Aringhieri, R., 2019. The management of non-elective patients: shared vs. dedicated policies. *Omega* 83, 199–212.

Duma, D., Aringhieri, R., 2020. An ad hoc process mining approach to discover patient paths of an emergency department. *Flexible Services and Manufacturing Journal* 32, 1, 6–34.

Duma, D., Aringhieri, R., 2023. Real-time resource allocation in the emergency department: a case study. *Omega* 117, 102844.

Guerriero, F., Guido, R., 2011. Operational research in the management of the operating theatre: a survey. *Health Care Management Science* 14, 89–114.

Guido, R., 2024. Patient admission scheduling problems with uncertain length of stay: optimization models and an efficient matheuristic approach. *International Transactions in Operational Research* 31, 1, 53–87.

Halawa, F., Madathil, S.C., Khasawneh, M.T., 2021. Integrated framework of process mining and simulation-optimization for pod structured clinical layout design. *Expert Systems With Applications* 185, 115696.

Hof, S., Fugender, A., Schoenfelder, J., Jo, B., 2017. Case mix planning in hospitals: a review and future agenda. *Health Care Management Science* 2, 20, 207–220.

Hulsof, P., Kortbeek, N., Boucherie, R., Hans, E., Bakker, P., 2012. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems* 2, 1, 129–175.

Lamiri, M., Xie, X., Dolgui, A., Grimaud, F., 2008. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research* 185, 3, 1026–1037.

Landa, P., Aringhieri, R., Soriano, P., Tànfani, E., Testi, A., 2016. A hybrid optimization algorithm for surgeries scheduling. *Operations Research for Health Care* 8, 103–114.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Landa, P., Sonnessa, M., Tànfani, E., Testi, A., 2018. Multiobjective bed management considering emergency and elective patient flows. *International Transactions in Operational Research* 25, 1, 91–110.

Maggi, F.M., Di Francescomarino, C., Dumas, M., Ghidini, C., 2014. Predictive monitoring of business processes. In *Advanced Information Systems Engineering*, Springer International Publishing, Cham, pp. 457–472.

Meneghello, F., Di Francescomarino, C., Ghidini, C., 2023. Runtime integration of machine learning and simulation for business processes. In *2023 5th International Conference on Process Mining (ICPM)*. IEEE, Piscataway, NJ, pp. 9–16.

Moreira, M.W.L., Rodrigues, J.J.P.C., Korotaev, V., Al-Muhtadi, J., Kumar, N., 2019. A comprehensive review on smart decision support systems for health care. *IEEE Systems Journal* 13, 3, 3536–3545.

Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O.A., Sepúlveda, M., Helm, E., Galvez-Yanjari, V., Rojas, E., Martinez-Millana, A., Aloini, D., Amantea, I.A., Andrews, R., Arias, M., Beerepoot, I., Benevento, E., Burattin, A., Capurro, D., Carmona, J., Comuzzi, M., Dalmas, B., de la Fuente, R., Di Francescomarino, C., Di Ciccio, C., Gatta, R., Ghidini, C., Gonzalez-Lopez, F., Ibanez-Sanchez, G., Klasky, H.B., Prima Kurniati, A., Lu, X., Mannhardt, F., Mans, R., Marcos, M., Medeiros de Carvalho, R., Pegoraro, M., Poon, S.K., Pufahl, L., Reijers, H.A., Remy, S., Rinderle-Ma, S., Sacchi, L., Seoane, F., Song, M., Stefanini, A., Sulis, E., ter Hofstede, A.H., Toussaint, P.J., Traver, V., Valero-Ramon, Z., van de Weerd, I., van der Aalst, W.M., Vanwersch, R., Weske, M., Wynn, M.T., Zerbato, F., 2022. Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics* 127, 103994.

Pesic, M., 2008. Constraint-based workflow management systems: shifting control to users. Ph.D. thesis, Technische Universiteit Eindhoven.

Prodel, M., Augusto, V., Jouaneton, B., Lamarsalle, L., Xie, X., 2018. Optimal process mining for large and complex event logs. *IEEE Transactions on Automation Science and Engineering* 15, 3, 1309–1325.

Prodel, M., Augusto, V., Xie, X., Jouaneton, B., Lamarsalle, L., 2015. Discovery of patient pathways from a national hospital database using process mining and integer linear programming. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, Piscataway, NJ, pp. 1409–1414.

Rizzi, W., Di Francescomarino, C., Maggi, F.M., 2020. Explainability in predictive process monitoring: when understanding helps improving. In Fahland, D., Ghidini, C., Becker, J., Dumas, M. (eds), *Business Process Management Forum*, Springer International Publishing, Cham, pp. 141–158.

Ronzani, M., Ferrod, R., Di Francescomarino, C., Sulis, E., Aringhieri, R., Boella, G., Brunetti, E., Di Caro, L., Dragoni, M., Ghidini, C., Marinello, R., 2022. Unstructured data in predictive process monitoring: lexicographic and semantic mapping to ICD-9-cm codes for the home hospitalization service. In *AIxIA 2021 – Advances in Artificial Intelligence*. Springer International Publishing, Cham, pp. 700–715.

Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., Rademakers, F., 2016. Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling* 19, 5, 493–525.

Schneider, A.J.T., van Essen, J.T., Carlier, M., Hans, E.W., 2020. Scheduling surgery groups considering multiple downstream resources. *European Journal of Operational Research* 282, 2, 741–752.

Siqueira, C.L., Arruda, E.F., Bahiense, L., Bahr, G.L., Motta, G.R., 2018. Long-term integrated surgery room optimization and recovery ward planning, with a case study in the Brazilian National Institute of Traumatology and Orthopedics (INTO). *European Journal of Operational Research* 264, 3, 870–883.

Tyler, D., Pasquariello, C., Chen, C., 2003. Determining optimum operating room utilization. *Anesthesia & Analgesia* 96, 4, 1114–1121.

Valente, R., Testi, A., Tànfani, E., Fato, M., Porro, I., Santo, M., Santori, G., Torre, G., Ansaldo, G., 2009. A model to prioritize access to elective surgery on the base of clinical urgency and waiting time. *BMC, Health Services Research* 9, 1.

van den Broek d'Obrenan, A., Ridder, A., Roubos, D., Stougie, L., 2020. Minimizing bed occupancy variance by scheduling patients under uncertainty. *European Journal of Operational Research* 286, 336–349.

van der Aalst, W., Adriansyah, A., de Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., de Leoni, M., Delias, P., van Dongen, B.F., Dumas, M., Dustdar, S., Fahland, D., Ferreira, D.R., Gaaloul, W., van Geffen, F., Goel, S., Günther, C., Guzzo, A., Harmon, P., ter Hofstede, A., Hoogland, J., Ingvaldsen, J.E., Kato, K., Kuhn, R., Kumar, A., La Rosa, M., Maggi, F., Malerba, D., Mans, R.S., Manuel, A.,

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

McCreesh, M., Mello, P., Mendling, J., Montali, M., Motahari-Nezhad, H.R., zur Muehlen, M., Munoz-Gama, J., Pontieri, L., Ribeiro, J., Rozinat, A., Seguel Pérez, H., Seguel Pérez, R., Sepúlveda, M., Sinur, J., Soffer, P., Song, M., Sperduti, A., Stilo, G., Stoel, C., Swenson, K., Talamo, M., Tan, W., Turner, C., Vanthienen, J., Varvaressos, G., Verbeek, E., Verdonk, M., Vigo, R., Wang, J., Weber, B., Weidlich, M., Weijters, T., Wen, L., Westergaard, M., Wynn, M., 2012. Process mining manifesto. In *Business Process Management Workshops*, Springer Berlin, pp. 169–194.

van der Aalst, W., Carmona, J., 2022. *Process Mining Handbook*. Lecture Notes in Business Information Processing. Springer International Publishing, Cham.

van der Aalst, W.M.P., 2011. *Process Mining – Discovery, Conformance and Enhancement of Business Processes*. Springer, Berlin.

van der Aalst, W.M.P., 2016. *Process Mining – Data Science in Action*, 2nd ed. Springer, Berlin.

van der Aalst, W.M.P., De Masellis, R., Di Francescomarino, C., Ghidini, C., 2017. Learning hybrid process models from events – process discovery without faking confidence. In *Business Process Management – Proceedings of the 15th International Conference, BPM 2017*. Springer, Berlin, pp. 59–76.

Van der Werf, J.M.E., van Dongen, B.F., Hurkens, C.A., Serebrenik, A., 2008. Process discovery using integer linear programming. In *International Conference on Applications and Theory of Petri Nets*,. Springer, Cham, pp. 368–387.

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.

Wang, J., Guo, H., Bakker, M., Tsui, K., 2018. An integrated approach for surgery scheduling under uncertainty. *Computer and Industrial Engineering* 118, 118, 1–8.

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* 3, 37, 757–805.

Zhu, T., Luo, L., Shen, W., Xu, X., Kou, R., 2022. Admission scheduling of inpatients by considering two inter-related resources: beds and operating rooms. *Optimization* 71, 6, 1637–1675.

© 2024 The Author(s).
International Transactions in Operational Research published by John Wiley & Sons Ltd on behalf of International Federation of Operational Research Societies.

14753995, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/itor.13584 by CochraneItalia, Wiley Online Library on [12/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License