ICIS 2022 Proceedings

IS in Healthcare

Dec 12th, 12:00 AM

Leveraging Electronic Health Records to improve Patient Appointment Scheduling: A design-oriented Approach

Janine Rottmann
University of Würzburg, janine.rottmann@uni-wuerzburg.de

Nikolai Stein
Julius-Maximilians-University, nikolai.stein@uni-wuerzburg.de

Leveraging Electronic Health Records to improve Patient Appointment Scheduling: A design-oriented Approach

Completed Research Paper

Janine Rottmann
Julius-Maximilians-University
Wuerzburg, Germany
janine.rottmann@uni-wuerzburg.de

Nikolai Stein
Julius-Maximilians-University
Wuerzburg, Germany
nikolai.stein@uni-wuerzburg.de

Abstract

As demand for healthcare services continues to increase, hospitals are under constant economic pressure to better manage patient appointments. It is common practice in clinical routine to schedule appointments based on average service times, resulting in overtime and waiting times for clinicians and patients. To address this problem, we propose a data-driven decision support system for scheduling patient appointments that accounts for variable service times. We take advantage of the growing amount of patient- and treatment-specific data collected in hospitals. Using a simulation study, we evaluate the decision support system on the practical example of a Gastroenterology facility. Our results demonstrate improved appointment scheduling efficiency compared to the approach currently in use.

Keywords: Appointment Scheduling, Predictive Modeling, Prescriptive Modeling, Healthcare

Introduction

The COVID-19 pandemic has raised awareness in both sociopolitical and scientific contexts about the well-being of healthcare workers under the burden of overtime. Overtime accumulates largely involuntarily and is difficult to reduce (Watanabe and Yamauchi, 2018). A study by Nie et al. (2020) addresses the psychological impact of overtime work and identifies 25.1% of all nurses affected by pandemic-caused overtime work as psychological distress. Overtime has not only been a major problem in the healthcare system since the pandemic outbreak; it has just exacerbated the problem (Holthof and Luedi, 2021). Due to the extensive international interdependencies, it is very likely that a global crisis of this kind will recur in the near future (Eum and Kim, 2022). Not least for this reason, clinical workloads are expected to continue to increase in the coming years (Lorkowski and Jugowicz, 2020). Among the reasons are the continued growth and aging of the population, the expansion of insurance, and the dissatisfaction of medical personnel (Petterson et al., 2012; Winter et al., 2020). Increasing workload causes patients to wait longer for their consultations and sessions to be overrun, negatively affecting both patient satisfaction and staff morale (Zhu et al., 2012). With regard to the increasing demand for healthcare and the increase in expenditure in the healthcare system, hospitals are under constant economic pressure to optimize staff utilization (Gupta and Denton, 2008). Although the shortage of medical personnel remains to be solved in the long term, the more efficient use of medical resources can reduce the workload in the short term. To achieve higher utilization of the clinical workforce, hospitals must better manage patient Appointment Scheduling (AS).

In this paper, we present a data-driven decision support system to improve AS. We test this system in a simulation study based on the workflow in a collaborating Gastroenterology facility. As frequently observed in clinical practice, appointments are scheduled one at a time, with average service times. This approach could be justified in practice if delays caused by excessively long appointments were compensated at the end of the day by appointments ending earlier. However, a recent study by Mandelbaum et al. (2020) refuted this assumption and showed that accounting for service time variability in the AS process leads to a significant reduction in overtime and waiting times. This finding motivates us to further investigate how accounting for service time variability affects AS efficiency.

Previous literature has presented various approaches to address different imponderables in the AS process. Among other complicating factors such as patient and provider preferences, punctuality, cancellations, and no-shows, stochastic service times pose a major challenge for patient AS (Gupta and Denton, 2008). This also applies to the collaborating Gastroenterology facility, which is why we focus on uncertain service times in this article. Previous studies have frequently addressed service time variability in patient AS, and despite the difficulty of this endeavor, progress has been made in addressing this problem. It has been repeatedly demonstrated that individual service times can be estimated based on a priori information regarding patient- and treatment-specific data (Gañan-Cardenas, Jiménez, et al., 2022; Jiao et al., 2020; Podboy and Scheinker, 2020). Furthermore, several studies have found that a more accurate estimate of service times has a positive impact on AS efficiency (Bentayeb et al., 2019; Safdar et al., 2021; Srinivas and Salah, 2021). However, in previous studies, this has only been demonstrated for rule-based AS strategies, where the residual global uncertainty in service time estimates was not accounted for in the model. We address this research gap by developing a decision support system that accounts for global uncertainty in individual service time estimates. To structure our work and emphasize its focus, we address the following research questions:

RQ1 What are the design elements of a data-driven decision support system for patient AS?
RQ2 How does the proposed decision support system compare to AS approaches currently in use?

Addressing these questions, the following section provides an overview of the current state of research in patient AS. After introducing the design-oriented research methodology, we describe the individual design elements of the decision support system. Subsequently, we evaluate each individual module of the decision support system in a discrete event simulation that replicates the processes at the collaborating Gastroenterology facility. We conclude our study with a summary of our findings and an outlook for further research.

Related Work

Decision support systems offer a variety of improvements in clinical routine. To give just a few examples, they may prevent potential errors in medication, provide diagnostic support, aid administrative processes, or improve clinical workflow (Sutton et al., 2020). Since the seminal work of Bailey (1952), there has been a plethora of studies investigating decision support systems for more efficient AS in hospitals. Due to healthcare-specific issues such as soft capacity constraints, scheduling preferences, and the dynamics of healthcare demand, existing models from manufacturing, transportation, or logistic areas cannot be easily transferred to the healthcare sector (Gupta and Denton, 2008). The lack of transferability from generalizable models to individual health care problems has led to a variety of solution approaches presented in previous studies. For an overview of general problem formulations and modeling considerations, we refer the reader to the comprehensive reviews by Cayirli and Veral (2003) and Ala and Chen (2022). Existing approaches can be classified into three different categories according to Ahmadi-Javid et al. (2017): long-term strategic decisions determine the main structure of a decision support system, medium-term tactical decisions refer to the processing of different groups of patients, and short-term operational decisions refer to efficient scheduling and sequencing of individual appointments.

On the strategic level, decision support systems are differentiated by the access policy (same-day or pre-scheduled appointments), the number of servers, the policy of acceptance (urgent or elective patients), and online or offline AS. While offline AS is completed before the start of the session, online AS allows appointments to be adjusted during the session. At the tactical level, decision support systems can be distinguished by the capacity allocation to different groups of patients, constant or variable appointment length, the horizon of the AS window, the number of patients per appointment slot, the number of appointments for each session, the panel size, as well as the priority of different patient groups. Considered from the operational level, decision support systems can be classified as rule-based or optimization-based approaches. The former is a set of easy-to-implement instructions, while the latter seeks to obtain the globally optimal solution for an operational decision. Ahmadi-Javid et al. (2017) identify six operational decisions that have been addressed in previous studies: Server assignment according to patient's preference; Appointment day assignment according to the patient's priority level; Determination of a specific time for the start of treatment; Decision on the acceptance or rejection of patients; Selection of patients from a waiting list; Determining the sequence of appointments. Additionally, they identify various environmental factors considered in the relevant literature. Among others, these include patient and physician tardiness, interruptions, no-shows, and stochastically distributed service times. For a comprehensive taxonomy of complicating factors in patient AS we refer to Gupta and Denton (2008).

We use this framework to position our research for the specific use case of a collaborative clinic. From a strategic perspective, the proposed decision support system is designed to pre-schedule elective patients in an offline multi-server environment. On the tactical level, we take into account variable appointment length for the time horizon of a one-day AS window. Each appointment slot is assigned to one patient only. Panel sizes are distributed equally among physicians, and there is no upper bound on the number of appointments per consultation session. Appointments are considered equally important; therefore, no distinction is made between different groups of patients. From the operational perspective, we pursue an optimization-based approach to determine the optimal start, as well as the sequence of appointments.

According to Ahmadi-Javid et al. (2017), only a few previous studies have addressed scheduling and sequencing of appointments simultaneously. This includes studies by Erdogan et al. (2015), Mancilla and Storer (2012) or Oh et al. (2013), for example. The authors further continue to analyze that little research has been conducted on the simultaneous scheduling and sequencing of appointments taking into account random service times. Among the first Salzarulo et al. (2016) addressed this research gap and took into account individual patient characteristics collected prior to medical consultation. Further studies analyzed the extent to which random service times can be derived from electronic health records. The wealth of electronic health records is widely used in clinical information systems, for example, to support clinical documentation, increase administrative efficiency, and improve patient care quality, safety and coordination (Nguyen et al., 2014). Similarly, electronic health records have been repeatedly proven valuable in predicting service times using machine learning techniques. Of particular note is a study by Podboy and Scheinker (2020) that examines the utility of electronic health records in predicting service times for endoscopic procedures. Studies by Gañan-Cardenas, Jiménez, et al. (2022) and Jiao et al. (2020), among others, suggest that predictions based on state-of-the-art machine learning techniques provide a more accurate approximation of service times compared to clinic estimates, even for sparse data and rare diseases. Against this backdrop, recent studies investigated the impact of individual service-time estimation on AS efficiency. Srinivas and Salah (2021) demonstrate a reduction in patient waiting time and physician idle time. Consistent with these results, Bentayeb et al. (2019) and Safdar et al. (2021) similarly demonstrate higher AS efficiency by considering more accurate service time estimation.

Although the literature is consistent in assuming higher AS efficiency when estimating individual service times, none of the previous studies reports perfect service-time prediction. Given predictive imprecision, mismatches between supply and demand for health care services are likely to persist, although to a lesser extent. However, uncertainty regarding prediction of individual service times is not considered in any of the proposed AS models, which we consider a limitation of previous studies. This limitation is addressed by Mandelbaum et al. (2020). Although they do not individually predict service times from electronic health records, they do account for a global uncertainty regarding stochastically distributed service times in the AS process. In an offline multi-server environment, where service times are uncertain, the approach was found to outperform near-optimal state-of-the-art stochastic and robust optimization techniques.

Thus, we have identified two promising approaches to increase AS efficiency in previous studies, but to the best of our knowledge, they have never been combined before. Building on previous findings, we address this research gap by developing an optimization-based AS model that accounts for global uncertainty in the estimation of individual service times. Following a design-oriented research approach, we present an IT artifact consisting of three modules. First, we predict individual service times employing machine learning techniques while relying on the wealth of data from electronic health records. Second, we propose an allocation mechanism for maximum resource utilization in a multi-server environment. Third, we account for a global uncertainty of service-time estimates by introducing a heuristic modeling approach.

Research Methodology

The elaborations above emphasize the economic pressure to optimize clinical resource utilization and reveal limitations of existing AS approaches. To overcome these limitations, this article proposes an IT artifact that allocates clinical resources more efficiently. Before presenting the design elements of the IT artifact, we briefly outline our methodological research approach.

From a methodological perspective, we chose a design science-oriented research approach, as design science research brings both practical relevance and scientific rigor to information systems research (Baskerville et al., 2018). Design science research involves the development of a wide range of socio-technical artifacts, including decision support systems, modeling tools, governance strategies, methods for information systems evaluation and information systems change interventions (Gregor and Hevner, 2013). Hevner et al. (2004) establish seven guidelines to help researchers gain knowledge and understanding of a design problem and its solution through the creation and application of an IT artifact. As these guidelines are receiving increasing attention in the information systems domain (Goes, 2014; Hevner and Chatterjee, 2010; Peffers et al., 2007), we draw from these to conceptualize the IT artifact in a structured manner:

- **Problem Relevance:** Facing an increasing demand for healthcare services, hospitals are under constant economic pressure to optimize resource allocation. However, this is not an easy endeavor, as uncertain service times complicate AS. For this reason, the development of a decision support system for more efficient resource allocation is essential.

- **Research Rigor:** Our approach consists of three phases: First, we estimate individual service times from electronic health records. Second, we allocate resources according to the predicted service times, striving for maximum resource utilization. Third, we account for the global uncertainty of the predicted service times using a heuristic proposed by Mandelbaum et al. (2020). Finally, we compare the performance of the proposed IT artifact with the clinic's current AS approach.
- **Design as a Search Process:** In developing the individual modules of the IT artifact, we draw on relevant work in the AS literature. This includes the prediction of service times, the development of an allocation mechanism for maximum resource utilization, and the selection of a robust heuristic to account for uncertainty in the prediction of service times.
- **Design as an Artifact:** We design an IT artifact consisting of three modules. Each of its modules is described in detail in the following sections and is implemented in Python.
- **Design Evaluation:** We analyze the performance of the IT artifact at large, as well as its individual design elements, using a discrete event simulation that replicates the processes of a collaborative Gastroenterology facility. Based on the results, we subsequently derive managerial implications for hospitals and discuss the limitations of the proposed IT artifact.
- **Research Contribution:** The primary contribution of this research is the development of a decision support system that accounts for uncertain service times.
- **Research Communication:** Our research invites scholars and healthcare providers to further explore the potential of our data-driven decision support system for AS in the face of increasing demand for healthcare services and prevailing resource scarcity.

In the following section, the IT artifact with its three modules is described in detail. This includes the presentation of inputs and outputs, the assumptions of the underlying parameters, and a mathematical formulation of the allocation mechanism.

Artifact Design

Figure 1 illustrates the information flow of the IT artifact. It receives information from the hospital's Human Resources and Hospital Information Systems, on-site process records, and external parameters. In turn, it outputs an optimized appointment schedule to a simulation module. The resulting appointment schedule comprises a proposed start time and operating room for all patients on a given day. To this end, the IT artifact receives the following input values:

- **Forecasting Module Input:** The forecasting module requires information to estimate the service time for each appointment of the day to be scheduled. For this purpose, it receives resource and appointment information as well as patient characteristics of past and upcoming appointments. It should be noted that only information that is already known prior to an appointment can be used. For a more detailed description of the features used to estimate individual service times, we refer the reader to Table 2.
- **Allocation Module Input:** The collaborative clinic chosen as an example to demonstrate the IT artifact offers both rooms specifically equipped for a particular type of treatment as well as multifunctional rooms. Hence, the allocation module requires information about the type of treatment as well as a detailed representation of which treatment can be performed in which operating room. In order to map the individual equipment of each operating room, a process recording was carried out on site.
- **Heuristic Module Input:** For an optimal determination of appointment start times, the heuristic module receives information about the expected utilization of the clinic on a given day. Moreover, a weighting parameter for the costs of overtime and waiting times is needed.

Based on these inputs, the forecasting module estimates individual service times for each appointment using predictive modeling. Based on these estimates, prescriptive modeling is performed in two steps: Usually several days and weeks before the actual procedure, the allocation module determines the assignment and sequence of patients to operating rooms. In this process, both the estimated service times and the equipment present in each room are taken into account. Few days before the procedure, the heuristic module determines the exact starting time for each appointment, taking into account other external parameters. Operating room assignments typically have a longer lead time, while staffing occurs on a shorter notice. The system then outputs the final appointment schedule, including an assigned operating room and start time for each appointment. To evaluate the system, this information is then transferred to the simulation module, which maps the clinical processes.

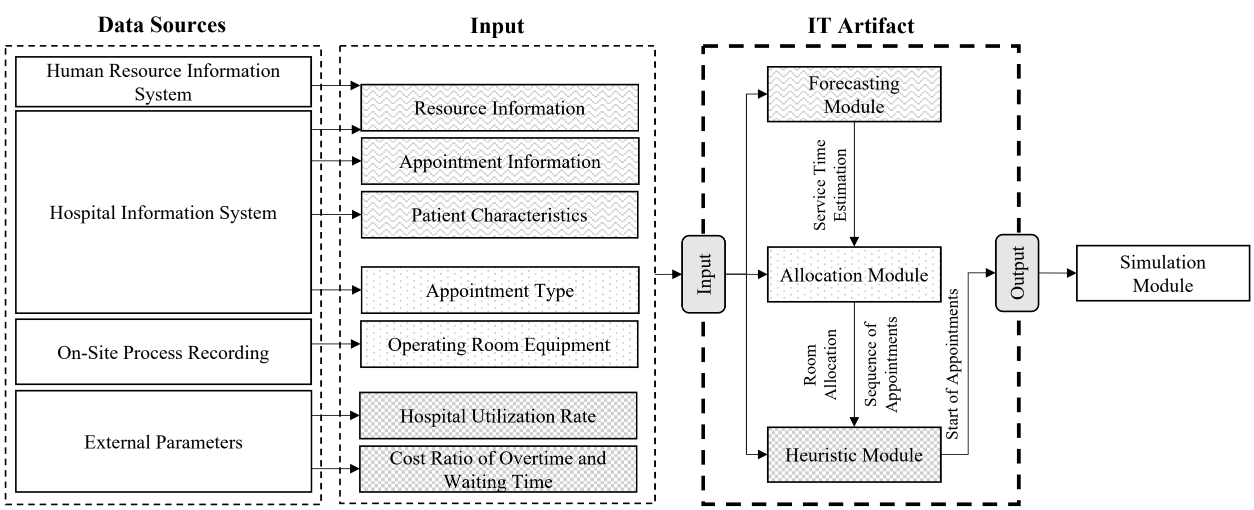

Figure 1. Outline of the IT Artifact information flow.

After a brief description of the data streams and the interaction of the individual modules, the following subsections describe the functionality of the individual modules in detail.

Allocation Module

In this section, we briefly describe the current operational processes in the collaborating Gastroenterology facility. Then we outline the allocation module that automates these processes. We are striving to find a solution that reflects the clinic well while generalizing the decision support system for similar use cases. In the problem setting at hand, patients are treated in five different operating rooms. Two of these operating rooms are multifunctional and the other three are specialized equipped. Appointments are assigned to available resources at least one day in advance. In this process, the scheduler fills a predetermined roster by assigning appointment requests to free time slots in specialized and multifunctional operating rooms in the order of their arrival. Since appointment requests are made asynchronously, allocation is random and not based on the objective of optimal resource allocation.

Scheduling jobs on a shared resource is a well-known problem in the field of operations research and is commonly referred to as *Job-Shop-Scheduling-Problem*. Referring to a comprehensive review by Chaudhry and Khan (2016), the problem at hand can be classified as *Flexible-Job-Shop-Scheduling-Problem* (FJSSP). It has the special requirement that a job can be processed by any resource from a given set of alternative resources. When service times are known in advance, the appointment schedule resulting from the FJSSP is considered optimal. It is further distinguished between total and partial FJSSP. In the former, any job can be executed on any resource, while in the latter, only a subset of resources can be considered for any given job. As some of the operating rooms are multifunctional equipped, the problem can be considered as partial FJSSP. Each job typically consists of a series of tasks that must be performed on different resources and in deterministic processing times in a specified job-dependent sequence. This constraint is relaxed in the problem at hand. For convenience, it is assumed that all jobs associated with an appointment are executed directly one after the other and can therefore be regarded as one process. It is further assumed that pre- and post-endoscopic procedures are routine processes with deterministic service length, while the primary examination is of stochastic service duration.

Variable	Bounds	Definition
e_i^m	$\in [0, T]$	Scheduled end of appointment i on operating room m
s_i^m	$\in [0, T]$	Scheduled start of appointment i on operating room m
T^*	$\in [0, T]$	Makespan
x_i^m	$\in \{0, 1\}$	Indicator whether appointment i is performed in operating room m
$y_{i,j}^m$	$\in \{0, 1\}$	Precedence variable of appointments i and j on operating room m

Table 1. Decision variables of the Flexible-Job-Shop-Scheduling Problem.

The decision variables as well as their respective bounds are defined as summarized in Table 1. The objective is to determine the start of an appointment i in the operating room m within business hours T such that the total length of time that elapses from the start of the first appointment until the end of the last appointment is minimized. This is commonly referred to as the *makespan*, denoted by T^* and given by Equation 1.

$$min\ T^* \tag{1}$$

The makespan is bound to the end of the last appointment, as denoted in Equation 2.

$$T^* \geq e_i^m \quad \forall_{i,m} \tag{2}$$

The estimated end of each appointment is defined as the sum of the individual start time and the estimated service time $\hat{d}_i^m$ obtained by the forecasting module denoted in Equation 3.

$$e_i^m \geq s_i^m + \hat{d}_i^m \quad \forall_{i,m} \tag{3}$$

We introduce a binary variable x_i^m that indicates whether an appointment is performed in an operating room or not. It takes the value 1 if a patient is served in the operating room m from a set of O operating rooms and, conversely, takes the value 0 if not. As each patient must be cared for, the sum of all binary indicators for each patient is equal to one, as denoted by Equation 4.

$$\sum_{m \in O} x_i^m = 1 \quad \forall_i. \tag{4}$$

In each operating room, only one treatment can be performed at a time. We define precedence constraints for every two appointments i and j from a set of N^m appointments scheduled in the same operating room. To this end, we first introduce the binary precedence variable $y_{i,j}^m$. If appointment i is scheduled before appointment j, $y_{i,j}^m$ takes the value 1, conversely 0. For each appointment j scheduled in the operating room m, the sum of all precedence variables $y_{i,j}^m$ of an appointment i must be equal to one, to avoid overlaps. In Equation 5 we denote:

$$\sum_{i \in N^m} y_{i,j}^m = x_j^m \quad \forall_{j,m}. \tag{5}$$

Based on Equation 5, we define in Equation 6 a precedence constraint to avoid overlap between two appointments. To this end, we establish lower bounds on the start time of each appointment. Note that since the objective is to minimize the end of each appointment, start times are scheduled as early as possible. For a sufficiently large number M and a precedence variable equal to one (appointment i before appointment j), the right-hand side of Equation 6 becomes negative. Thus, the lower bound of the start time s_i^m remains equal to 0, as defined in Table 1. In contrast, if the precedence variable $y_{i,j}^m$ is equal to 0, the lower bound for the start time s_i^m is bound to e_j^m. Consequently, appointment i cannot start before appointment j has been completed.

$$s_i^m \geq e_j^m - y_{i,j}^m M \quad \forall_{i,j,m}. \tag{6}$$

To illustrate the allocation module with a minimal example, assume that two appointments A and B are to be scheduled in a single operating room with estimated service times of 20 and 30 minutes, respectively. In accordance with the Equations 1 and 2, the allocation module seeks to minimize the end of the second appointment, which is given by Equation 3 and corresponds to $e_A^1 \geq s_A^1 + 20$ for patient A and $e_B^1 \geq s_B^1 + 30$ for patient B, respectively. Assuming there is only one room available, Equation 4 ensures that $x_A^1 = 1$ and $x_B^1 = 1$. Since both patients must be treated sequentially, Equation 5 states that $y_{A,B}^1 + y_{B,A}^1 = 1$, implying that only one of the two precedence variables can take the value 1. If patient A is treated before patient B $y_{A,B}^1 = 1$ and $y_{B,A}^1 = 0$. In contrast, if patient B is treated before patient A $y_{B,A}^1 = 1$ and $y_{A,B}^1 = 0$. In accordance with Equation 6 we obtain $s_A^1 = 0$ and $s_B^1 \geq 20$ in the first and $s_B^1 = 0$ and $s_A^1 \geq 30$ in the second scenario.

Forecasting Module

Currently, appointments are scheduled with average service times, regardless of the specifics of the procedure or patient. Although routine interventions can be performed in a standardized and predictable manner, complex treatments are difficult to perform and, therefore, more difficult to plan.

The distribution of service times is shown in Figure 2 according to the five different types of treatments practiced in the clinic. The distribution of service times for all treatments (upper left graph) is right-skewed, indicating that a large proportion of treatments are of short duration and only a few treatments are very long. When comparing the distribution of service times across all types of treatments, two characteristic distribution patterns stand out. While gastroscopy and endoscopy treatments are characterized by shorter service times and low standard deviation, endoscopic retrograde cholangiopancreatography (ERCP), colononoscopy, and bronchoscopy procedures exhibit comparatively high service times and standard deviation. Intuitively, it can be assumed that gastroscopy and endoscopy procedures are more routine and thus easier to predict, while all other treatments implicate higher variability and thus present more difficulty in the AS process. With an average service time of 41 min and a standard deviation of 24%, ERCP treatments represent the procedures with the longest service time and the highest variability. At the same time, they account for the smallest proportion (approximately 7%) compared to all other types of treatments. This suggests that ERCP procedures are more complex and less routine.

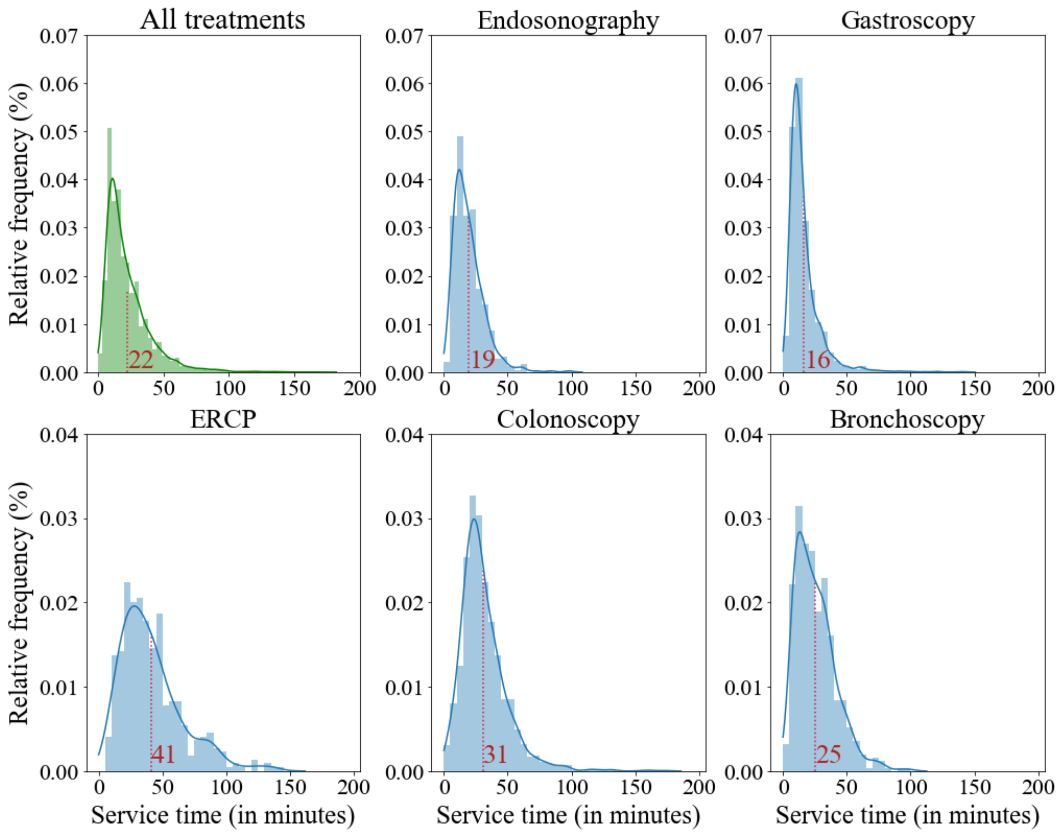

Figure 2. Relative frequency of service time distribution for all treatments (green) and each type of treatment separately (blue) after removal of outliers in 5min intervals. Average service times are highlighted in red.

From the distribution of service times, it can be seen that scheduling appointments with average service times inevitably causes delays in the schedule. In order to better anticipate the length of each appointment, we rely on machine learning as proposed in previous research. Related work has repeatedly demonstrated the superiority of *Extreme Gradient Boosting* (XG-Boost) over other machine learning approaches on tabular data. Shwartz-Ziv and Armon (2022) recently found that XG-Boost outperforms deep learning architectures even on data sets for which previous work suggests using deep learning models. Consistent with these results, XG-Boost has repeatedly proven effective in predicting service times (Bartek et al., 2019; Garside et al., 2021; Zhao et al., 2019). In particular, reference should also be made to a study predicting the duration of endoscopic procedures (Podboy and Scheinker, 2020). With this in mind, we chose XG-Boost as the regression model and predict individual service times based on historical records of $6,347$ interventions recorded over a one-year period.

The facility routinely collects demographic data on its patients, each examination, as well as the attending employees and endoscopes used. Each feature and the corresponding attributes used to predict service length are listed in Table 2. These include patient-related characteristics such as age and gender. The clinic further records the expected length of stay, distinguishing between inpatients, outpatients, and day patients. The latter are considered outpatients staying in the clinic for one night only. In addition, the clinic performs a classification of a patient's physical condition upon arrival on an ordinal scale of one to six, where one corresponds to a healthy patient and six to a deceased patient.

In addition, we are provided with appointment information concerning appointment time, day of week, and month, as well as the type and urgency of each appointment. Urgency is a binary variable that indicates whether it is an emergency or an elective appointment. Furthermore, we use more detailed information on the type of each intervention, which are categorized according to the *German Society for Gastroenterology, Digestive and Metabolic Diseases*.

Lastly, we use resource-specific characteristics about the attending medical team and the endoscope used. Each team consists of one to three physicians as well as one to three nurses. The number of physicians and nurses performing the treatment depends on the roster and availability of staff at a given time.

Predictors	Value	Type
Patient Characteristics		
Age	0 - 97 years	Continous
Gender	Male or Female	Categorical
Length of Stay	Inpatient, Day Patient, Outpatient	Ordinal
Physical Status Classification	1 - 5	Ordinal
Appointment Information		
Appointment Day of Week	Mon, Tue, ..., Fri	Categorical
Appointment Month	Jan, Feb, ..., Dec	Categorical
Appointment Time	0 - 23 h	Continous
Appointment Type	Bronchoscopy, Colonoscopy, Endosonography, ERCP, Gastroscopy	Categorical
Emergency	0 or 1	Binary
Type of Intervention	Classification scheme according to the German Society for Gastroenterology, Digestive and Metabolic Diseases	Categorical
Resource Information		
Endoscope	Designation	Categorical
Nurse	Personnel Number	Categorical
Physician	Personnel Number	Categorical

Table 2. Description of features used for service time estimation.

Heuristic Module

Following the approach of Mandelbaum et al. (2020), the start of each appointment is scheduled to minimize the total cost, taking into account uncertain service-time estimates. Starting from a preliminary schedule obtained by the allocation module, appointments are rescheduled in two stages: First, the clinic's occupancy is approximated throughout the day based on the preliminary schedule. In a second step, appointments are rescheduled to minimize total costs as a function of approximate occupancy. Mandelbaum et al. (2020) suggest to approximate the occupancy of the entire clinic, not each operating room separately, and demonstrate rising AS efficiency as the number of servers increases. Since only two of five operating rooms are multifunctional equipped, we adhere to operating room assignment and estimate the occupancy $Q_i(t)^m$ of each operating room separately. In addition, Mandelbaum et al. (2020) use records of patient punctuality. This is not considered in the present implementation. Occupancy by a single patient is expressed by Equation 7. A patient is expected to arrive on time at $t = s_i^m$ and leave the clinic at time $t = s_i^m + \hat{d}_i^m$, where $\hat{d}_i^m$ represents the estimate of the service time of the appointment i.

$$Q_i^m(t) = \mathbb{1}_{\{s_i^m \leq t \leq s_i^m + \hat{d}_i^m\}} \quad \forall i, m. \tag{7}$$

From Equation 7 we approximate the total occupancy of each operating room in every t ≥ 0, according to Equation 8.

$$Q^m(t) = \sum_{i \in N^m} Q_i^m(t) \qquad \forall_m. \tag{8}$$

In a second step, the total expected costs TC incurred for each operating room are minimized by adjusting the start times for each appointment. Total costs are a function of the cost ratio between waiting time and overtime, denoted as γ, as well as the capacity limit in time t, which is equal to one for each operating room throughout the business day. For convenience, it is assumed that employees are willing to work overtime beyond regular business hours in each operating room T^m. We continue to assume that the cost ratio is known to the healthcare provider. For a function $r : \mathbb{R} \to [0, \infty)$, some cost ratio $\gamma > 0$, the number of patients scheduled in the operating room n^m, and a capacity limit equal to one, the total costs are defined according to Equation 9.

$$TC^m(Q^m) := \int_{-\infty}^{\infty} r\left(\frac{Q^m(t)-1}{n^m}\right) dt + \gamma \int_{T^m}^{\infty} min\{Q^m(t), 1\}\, dt \qquad \forall_m, \tag{9}$$

where the first part of the term represents the waiting time and the second part represents the overtime incurred in each operating room. An increase in γ leads to lower overtime costs and a decrease conversely to higher overtime costs. We assume fixed personnel costs during regular business hours; therefore, idle time is not penalized. To scale our approach to larger clinical problem settings and overcome the associated high computational complexity, we rely on a local search algorithm to solve the problem as suggested by Mandelbaum et al. (2020). To this end, we use *Basin Hopping* because of its ability to find global minima in a high-dimensional feature space.

Performance Evaluation

In this section, we first outline the experimental setup and performance criteria used to evaluate the research question. We then examine the performance of the proposed IT artifact.

Simulation Module

Using the simulation module, we investigate whether the IT artifact serves as a prototype for decision support systems in hospitals. To evaluate the IT artifact in a representative environment, we model the clinical routine of a Gastroenterology facility in a discrete-event simulation. In the process, we generate different scenarios, in each of which we simulate unique sessions from electronic health records.

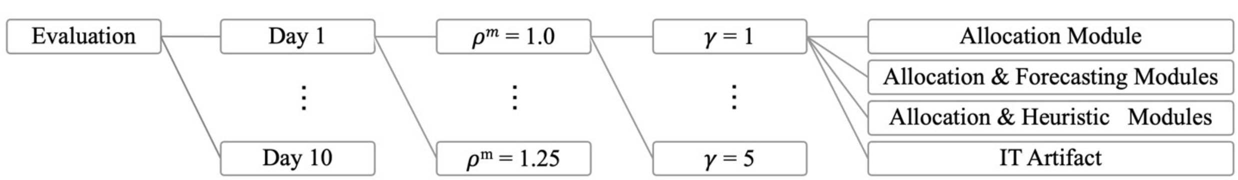

Figure 3. Representation of all evaluation scenarios for a single run. We evaluate each mode of operation for randomly sampled 10 sessions as a function of rising utilization rate ρ^m in each operating room and cost ratio γ.

We investigate the performance of the proposed IT artifact as a function of rising clinical utilization and relative cost increase of overtime as well as different modes of operation. The entire evaluation process is shown in Figure 3. To ensure stable and robust findings, we repeat this evaluation process in 100 runs for 10 sessions each, i.e., for a total time horizon of 1000 days. As a baseline, we schedule appointments deterministically using only the allocation module with historical averages as service times. We then investigate a potential performance improvement by adding the forecasting module for service time estimation. Similarly, we examine the performance of the decision support system incorporating the heuristic module. Finally, we evaluate the IT artifact as a combination of all three modules.

Each mode of operation is evaluated as a function of the average waiting time (WT) for n patients and the average overtime (OT) for o operating rooms, given by Equations 10 and 11, respectively. Following the example of Mandelbaum et al. (2020), the waiting times result from the difference between the actual a_i^m and the scheduled start of each appointment. Any work performed beyond regular business hours is considered overtime.

$$WT = \frac{\sum\limits_{i \in N} \sum\limits_{m \in O} a_i^m - s_i^m}{n}. \tag{10}$$

$$OT = \frac{\sum\limits_{i \in N} \sum\limits_{m \in O} (\hat{d}_i^m - (T^m - a_i^m)^+)^+}{o} \tag{11}$$

Avoiding waiting times and overtime inevitably poses a conflict of interest between clinic staff and patients. Following Mandelbaum et al. (2020), we weight both objectives with the cost ratio γ. It is reasonable to assume that clinics generally perceive overtime as more costly than waiting time, which is why we examine the performance of the IT artifact as a function of the relative increase in overtime costs. Total costs (TC) are obtained from Equations 10 and 11, weighted by the parameter γ, as given in Equation 12.

$$TC = WT + \gamma\, OT \tag{12}$$

In light of the expected future increase in demand for medical health services, we additionally examine AS performance as a function of clinical workload. For a given level of utilization, we determine the business hours for each operating room by Equation 13:

$$T^m = \frac{\sum_{i \in N^m} \hat{d}_i^m}{\rho^m} \qquad \forall_m, \tag{13}$$

where ρ^m specifies the level of utilization in each operating room and $\hat{d}_i^m$ represents the estimated service time for each appointment. Starting from a medium utilization rate ($\rho^m = 1.0$), we increase the level of utilization by up to 125% and evaluate AS performance.

Results

Employing the simulation module, we explore the performance of the IT artifact. First, we investigate the impact of the forecasting module on the allocation module. Figure 4 illustrates the distribution of forecast accuracy measured by *Root-Mean-Squared-Error* (RMSE) and *Mean-Absolute-Error* (MAE). The estimation of service times using the forecasting module has an MAE of 9.01 and an RMSE of 13.48 on average. In comparison, the difference between historical averages and actual service times accounts for an MAE of 10.5 and an RMSE of 14.95. This comparison highlights the value of forecasting versus the estimate of service times based on historical averages. Furthermore, this confirms the results of previous studies and demonstrates that the XG-Boost model can derive service times based on electronic health records.

However, machine learning often turns out to be a ”black box” making it impossible to interpret the prediction results (Goebel et al., 2018). We attempt to overcome this problem and gain an understanding of the decisions upon which service times were predicted. For this purpose, we use a commonly employed explanatory artificial intelligence library, *SHapley Additive exPlanations* (SHAP), developed by Lundberg and

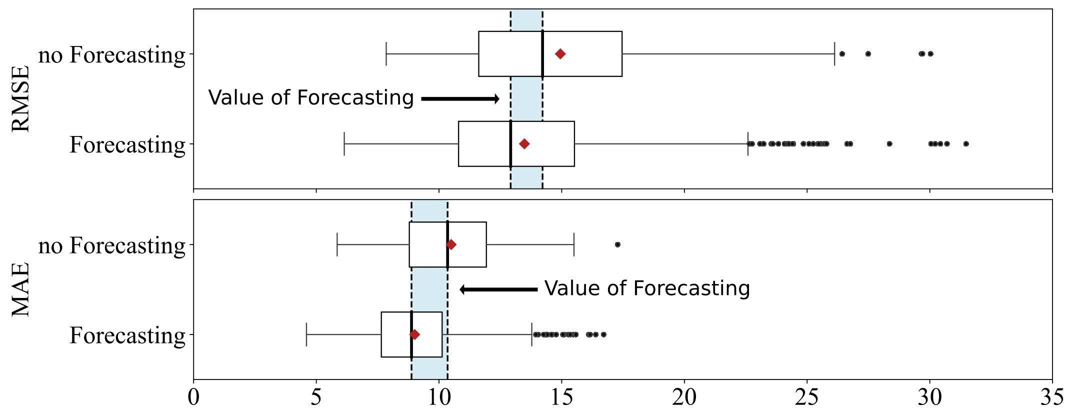

Figure 4. Mean quality of forecasting (red diamond) and value of the forecasting (blue area) measured by the median (black solid line) with the metrics RMSE and MAE in boxplots, each bounded by the upper and lower quartiles.

Lee (2017). Figure 5 illustrates the importance of each feature cluster for prediction, grouped in descending order according to Table 2. The results suggest that the type of appointment has the greatest influence on service time prediction. As previously analyzed, this observation may indicate that gastroscopy and endoscopy procedures are more routine compared to the other types of treatments and therefore are easier to predict. It can also be observed that both the type of treatment and the attending physician are given high weight in the prediction. Surprisingly, emergency treatments have little impact on the prediction. This may be attributable to the small number of training examples in the dataset.

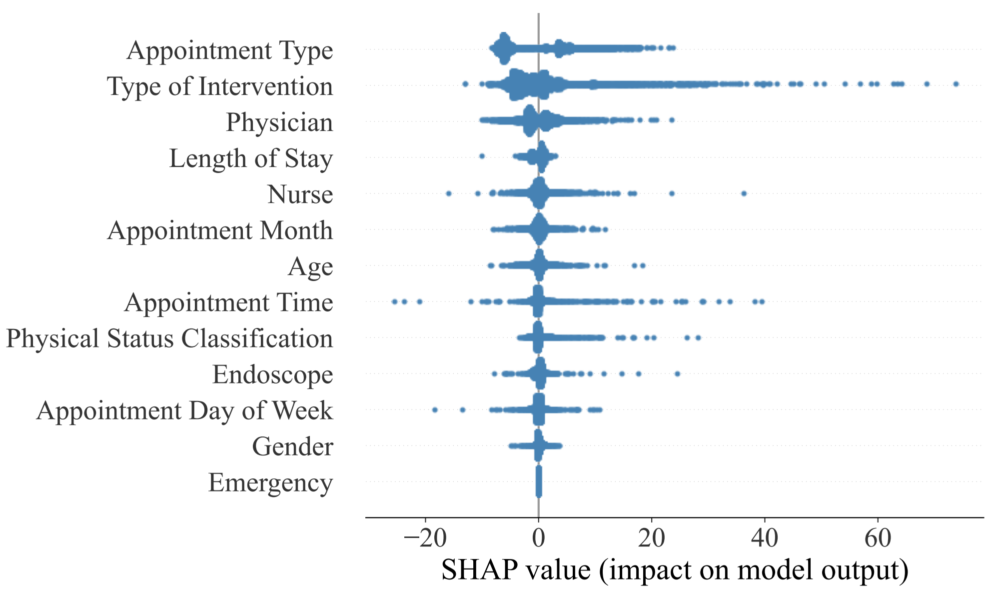

Figure 5. SHAP analysis of the importance of individual feature clusters for service time prediction

We further evaluate the value of the forecasting module on AS efficiency. Figure 6 indicates that adding the forecasting module to the allocation model reduces total costs by 3% to 5% on average. However, the relative reduction in total costs remains constant even as clinical workload and relative overtime costs increase. In addition, we observe a relative reduction in total costs up to 11% when the heuristic module is added to the allocation module. As can be seen in Figure 6, the relative reduction in total cost increases along with the level of utilization of the clinic. It can also be observed that AS efficiency increases as overtime becomes more expensive compared to waiting time. As previously explained, we assume that the hospital's cost ratio is known in advance. Using this information, appointments are rescheduled to overlap previous appointments, thereby reducing server idle time, while at the same time patient waiting times increase.

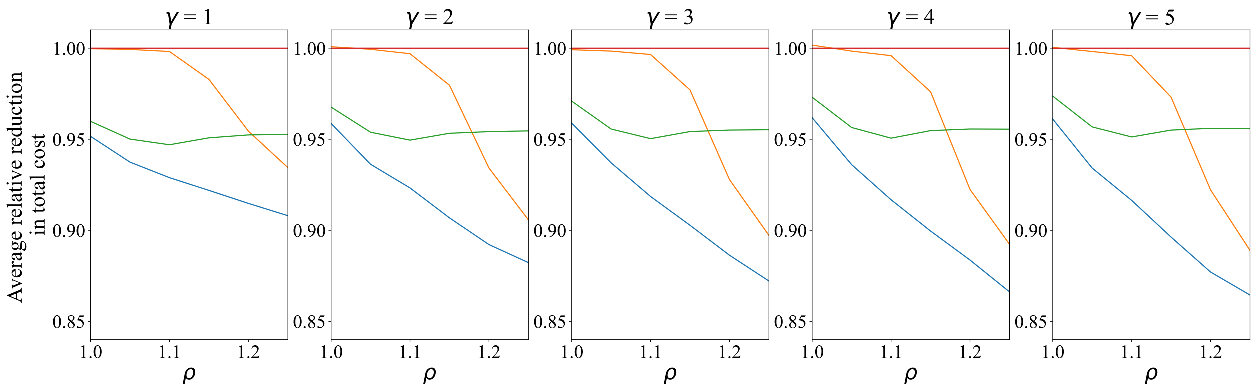

Figure 6. Average relative reduction in total cost of the allocation module (red), the allocation and forecasting modules (green), the allocation and heuristic modules (orange), and the conjunction of all three modules into one IT artifact (blue) as a function of the level of utilization (ρ) and cost ratio (γ).

Having confirmed the potential of the forecasting and heuristic modules separately, we analyze the performance of the IT artifact unifying all three modules. With a balanced cost ratio and low utilization level, the forecasting module exhibits a greater benefit to the allocation module than to the heuristic module. However, the efficiency of the heuristic module increases with the utilization rate and the relative increase in the overtime cost. This trend suggests a greater benefit of the heuristic module compared to the predictive module above a certain utilization rate and cost ratio. In our experiment, we reach this inflection point at a utilization rate of 120% and a cost ratio of 2. However, the greatest improvement over the baseline is achieved by combining the three modules. We observe a continuously increasing cost efficiency with an increasing level of utilization and a relative increase of overtime cost. These results indicate that the proposed IT artifact benefits from both the prediction and heuristics modules.

Conclusion and Outlook

From the simulation study, we gain several managerial insights for clinical practice. The proposed IT artifact addresses key challenges in the development of a data-driven decision support system for hospitals: it enables simultaneous scheduling and sequencing of appointments in a multi-server environment while addressing uncertain service times. By addressing these challenges, the IT artifact enables hospitals to better manage patient appointments and thus better utilize healthcare resources. By studying each module individually, we were able to demonstrate increasing AS efficiency by considering both individual service time estimations as well as a global uncertainty in these estimates. Compared to the clinic's current naïve AS approach, total costs can be reduced between 4% and 14% depending on the level of utilization and cost ratio. Our analyses indicate that taking into account the cost ratio of overtime and waiting times results in additional cost savings. Furthermore, our findings point to increasing cost savings as hospital utilization continues to increase in the future. Against this backdrop, the proposed IT artifact will become increasingly relevant for clinical practice. As mentioned at the outset, this will not solve the acute shortage of healthcare workers in the long term. However, in the short term, the prototype might help hospitals meet the growing demand for health care services if existing resources were used more efficiently. As a result, more patients could potentially be provided access to health care services, and hospital revenues could increase.

These prospects are accompanied by some limitations of the described approach. Currently, the IT artifact considers variable service times, while other imponderables such as provider preferences, punctuality, cancellations, and no-shows remain unaccounted for. Since these confounding factors may have a significant impact on AS efficiency, we encourage further research to incorporate other contingencies in the model design and to evaluate the model in a real-world setting before applying it in practice. Moreover, when assigning patients to operating rooms, the allocation module only considers compatibility in terms of operating room equipment and available servers. In reality, additional resource constraints may have to be taken into account. Examples would be a patients' appointment preferences in terms of day, time or medical team, a patients' rescheduling requests, the expertise of the attending medical team, or the availability of the required endoscopes. Another simplification of the problem is modeling the cost ratio of overtime and waiting time as known, whereas it can at best be estimated in a real-world scenario. Moreover, the heuristic does not account for shared resources. Instead, occupancy for each operating room is estimated separately. Another limitation of the heuristic module entails that it does not account for the estimation error of individual service time predictions and instead only accounts for a global uncertainty. We anticipate that incorporating these considerations would further increase AS efficiency and the transferability of the prototype to a real-world environment.

Despite the potential of information systems, their use in hospitals remains low. Weeger and Gewald (2015) explain this phenomenon, among other things, by the aversion of physicians to computer technology, which can be explained by perceived dependence on information systems, possible restrictions on professional autonomy, or fear of deterioration in quality of care. As a result, physicians often find ways to avoid working with information systems, even though they are usually required to do so (Reiz and Gewald, 2016). These results raise the question of whether high investments in healthcare technology are worthwhile at all. Hah and Bharadwaj (2012) explore this matter and conclude that investments in healthcare technology do indeed improve operational and financial performance. In addition, Venkatesh et al. (2011) find that age has an important influence on whether physicians use information systems and conclude that adoption could increase among the new generation of physicians. These results suggest that more research is needed beyond the mere potential of the proposed IT artifact to evaluate its acceptance by clinicians. For this reason, we encourage future research to test the prototype in a real-world environment.

To exploit the full potential of electronic health records in AS, data protection must be ensured. According to Stark et al. (2018), confidentiality is often difficult to implement in practice because employees need to access patient data across multiple departments, making it difficult to distinguish between legitimate and illegitimate access. To leverage sensitive, fragmented, and nonpublicly accessible data across hospitals, Xu et al. (2021) proposes *Federated Learning*. This involves training a common global model on a central server, while all sensitive data remain in local facilities. We believe that gaining an increasing amount of data-driven insights across multiple facilities will improve the performance of the forecasting module, especially for rare endoscopic procedures. In addition to using hospital-wide data, we encourage future research to identify other sources of information to improve predictive accuracy. In this study, we solely use tabular data from a Hospital Information System. Jiao et al. (2020) find that physician notes have great potential to improve the performance of predictive algorithms. Furthermore, Grisot et al. (2018) suggest having relevant data, such as symptoms, collected by patients themselves at the time of an event via an app and observe increasing accuracy and quality of the data.

Following a design-oriented research approach, this study presents a data-driven decision support system for patient AS. In the course of a simulation study, we demonstrate significant cost reduction compared to the AS approach currently in use. We hope that this study will be a valuable contribution to both researchers and practitioners in managing variable service times in patient AS.

References

Ahmadi-Javid, A., Jalali, Z., and Klassen, K. J. 2017. "Outpatient appointment systems in healthcare: A review of optimization studies," *European Journal of Operational Research* (258:1), pp. 3–34.

Ala, A. and Chen, F. 2022. "Appointment Scheduling Problem in Complexity Systems of the Healthcare Services: A Comprehensive Review," *Journal of Healthcare Engineering* (2022).

Bailey, N. T. 1952. "A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times," *Journal of the Royal Statistical Society: Series B (Methodological)* (14:2), pp. 185–199.

Bartek, M. A., Saxena, R. C., Solomon, S., Fong, C. T., Behara, L. D., Venigandla, R., Velagapudi, K., Lang, J. D., and Nair, B. G. 2019. "Improving operating room efficiency: machine learning approach to predict case-time duration," *Journal of the American College of Surgeons* (229:4), pp. 346–354.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., and Rossi, M. 2018. "Design science research contributions: Finding a balance between artifact and theory," *Journal of the Association for Information Systems* (19:5), p. 3.

Bentayeb, D., Lahrichi, N., and Rousseau, L.-M. 2019. "Patient scheduling based on a service-time prediction model: a data-driven study for a radiotherapy center," *Health care management science* (22:4), pp. 768–782.

Cayirli, T. and Veral, E. 2003. "Outpatient scheduling in health care: a review of literature," *Production and operations management* (12:4), pp. 519–549.

Chaudhry, I. A. and Khan, A. A. 2016. "A research survey: review of flexible job shop scheduling techniques," *International Transactions in Operational Research* (23:3), pp. 551–591.

Erdogan, S. A., Gose, A., and Denton, B. T. 2015. "Online appointment sequencing and scheduling," *IIE Transactions* (47:11), pp. 1267–1286.

Eum, N. J. and Kim, S. H. 2022. "The Role of Information and Communications Technology Policies and Infrastructure in Curbing the Spread of the Novel Coronavirus: Cross-country Comparative Study," *JMIR Public Health and Surveillance* (8:1), e31066.

Gañan-Cardenas, E., Jiménez, J. C., et al. 2022. "Bayesian hierarchical modeling of operating room times for surgeries with few or no historic data," *Journal of Clinical Monitoring and Computing* (36:3), pp. 687–702.

Garside, N., Zaribafzadeh, H., Henao, R., Chung, R., and Buckland, D. 2021. "CPT to RVU conversion improves model performance in the prediction of surgical case length," *Scientific Reports* (11:1), pp. 1–7.

Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., and Holzinger, A. 2018. "Explainable ai: the new 42?," in *International cross-domain conference for machine learning and knowledge extraction,* Springer, pp. 295–303.

Goes, P. B. 2014. "Design science research in top information systems journals," *MIS Quarterly: Management Information Systems* (38:1), pp. iii–viii.

Gregor, S. and Hevner, A. R. 2013. "Positioning and presenting design science research for maximum impact," *MIS quarterly* (37:2), pp. 337–355.

Grisot, M., Lindroth, T., and Islind, A. S. 2018. "Patient Generated Data And Affordances: Conversation Cues In Clinical Decision Making," in *NOKOBIT conference held at Svalbard the 18th-20th of September 2018,* vol. 26, pp. 1–10.

Gupta, D. and Denton, B. 2008. "Appointment scheduling in health care: Challenges and opportunities," *IIE transactions* (40:9), pp. 800–819.

Hah, H. and Bharadwaj, A. 2012. "A multi-level analysis of the impact of health information technology on hospital performance," in *Proceedings of the Thirty Third International Conference on Information Systems, Orlando, FL,* pp. 1–17.

Hevner, A. and Chatterjee, S. 2010. "Design science research in information systems," in *Design research in information systems,* Springer, pp. 9–22.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS quarterly* (28:1), pp. 75–105.

Holthof, N. and Luedi, M. M. 2021. "Considerations for acute care staffing during a pandemic," *Best Practice & Research Clinical Anaesthesiology* (35:3), pp. 389–404.

Jiao, Y., Sharma, A., Ben Abdallah, A., Maddox, T. M., and Kannampallil, T. 2020. "Probabilistic forecasting of surgical case duration using machine learning: model development and validation," *Journal of the American Medical Informatics Association* (27:12), pp. 1885–1893.

Lorkowski, J. and Jugowicz, A. 2020. "Shortage of physicians: a critical review," *Medical Research and Innovation* (1324), pp. 57–62.

Lundberg, S. M. and Lee, S.-I. 2017. "A unified approach to interpreting model predictions," *Advances in neural information processing systems* (30).

Mancilla, C. and Storer, R. 2012. "A sample average approximation approach to stochastic appointment sequencing and scheduling," *IIE Transactions* (44:8), pp. 655–670.

Mandelbaum, A., Momčilović, P., Trichakis, N., Kadish, S., Leib, R., and Bunnell, C. A. 2020. "Data-driven appointment-scheduling under uncertainty: The case of an infusion unit in a cancer center," *Management Science* (66:1), pp. 243–270.

Nguyen, L., Bellucci, E., and Nguyen, L. T. 2014. "Electronic health records implementation: an evaluation of information system impact and contingency factors," *International journal of medical informatics* (83:11), pp. 779–796.

Nie, A., Su, X., Zhang, S., Guan, W., and Li, J. 2020. "Psychological impact of COVID-19 outbreak on frontline nurses: A cross-sectional survey study," *Journal of clinical nursing* (29:21-22), pp. 4217–4226.

Oh, H.-J., Muriel, A., Balasubramanian, H., Atkinson, K., and Ptaszkiewicz, T. 2013. "Guidelines for scheduling in primary care under different patient types and stochastic nurse and provider service times," *IIE Transactions on Healthcare Systems Engineering* (3:4), pp. 263–279.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A design science research methodology for information systems research," *Journal of management information systems* (24:3), pp. 45–77.

Petterson, S. M., Liaw, W. R., Phillips, R. L., Rabin, D. L., Meyers, D. S., and Bazemore, A. W. 2012. "Projecting US primary care physician workforce needs: 2010-2025," *The Annals of Family Medicine* (10:6), pp. 503–509.

Podboy, A. J. and Scheinker, D. 2020. "Machine learning better predicts colonoscopy duration," *Artificial Intelligence in Gastroenterology* (1:1), pp. 30–36.

Reiz, A. and Gewald, H. 2016. "Physicians' Resistance towards Information Systems in Healthcare: the Case of Workarounds." in *PACIS,* p. 12.

Safdar, S., Khan, S. A., and Shaukat, A. 2021. "A Novel Data-Driven Adaptive Technique to Generate a Physician Visiting Schedule for Better Patient Experience," *IEEE Access* (10), pp. 3935–3948.

Salzarulo, P. A., Mahar, S., and Modi, S. 2016. "Beyond patient classification: Using individual patient characteristics in appointment scheduling," *Production and Operations Management* (25:6), pp. 1056–1072.

Shwartz-Ziv, R. and Armon, A. 2022. "Tabular data: Deep learning is not all you need," *Information Fusion* (81), pp. 84–90.

Srinivas, S. and Salah, H. 2021. "Consultation length and no-show prediction for improving appointment scheduling efficiency at a cardiology clinic: a data analytics approach," *International Journal of Medical Informatics* (145), p. 104290.

Stark, B., Gewald, H., Lautenbacher, H., Haase, U., and Ruff, S. 2018. "Misuse of 'Break-the-Glass' Policies in Hospitals: Detecting Unauthorized Access to Sensitive Patient Health Data," *International Journal of Information Security and Privacy (IJISP)* (12:3), pp. 100–122.

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., and Kroeker, K. I. 2020. "An overview of clinical decision support systems: benefits, risks, and strategies for success," *NPJ digital medicine* (3:1), pp. 1–10.

Venkatesh, V., Sykes, T. A., and Zhang, X. 2011. "'Just what the doctor ordered': a revised UTAUT for EMR system adoption and use by doctors," in *2011 44th Hawaii international conference on system sciences,* IEEE, pp. 1–10.

Watanabe, M. and Yamauchi, K. 2018. "The effect of quality of overtime work on nurses' mental health and work engagement," *Journal of Nursing Management* (26:6), pp. 679–688.

Weeger, A. and Gewald, H. 2015. "Acceptance and use of electronic medical records: An exploratory study of hospital physicians' salient beliefs about HIT systems," *Health Systems* (4:1), pp. 64–81.

Winter, V., Schreyögg, J., and Thiel, A. 2020. "Hospital staff shortages: environmental and organizational determinants and implications for patient satisfaction," *Health Policy* (124:4), pp. 380–388.

Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., and Wang, F. 2021. "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research* (5:1), pp. 1–19.

Zhao, B., Waterman, R. S., Urman, R. D., and Gabriel, R. A. 2019. "A machine learning approach to predicting case duration for robot-assisted surgery," *Journal of medical systems* (43:2), p. 32.

Zhu, Z., Heng, B. H., and Teow, K. L. 2012. "Analysis of factors causing long patient waiting time and clinic overtime in outpatient clinics," *Journal of medical systems* (36:2), pp. 707–713.