Received: 28 December 2018 | Revised: 1 June 2020 | Accepted: 8 June 2020 | Published on: 21 July 2020

DOI: 10.1002/nav.21929

RESEARCH ARTICLE

WILEY

Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks

Julian Schiele[1,2] | Thomas Koperna[3] | Jens O. Brunner[1,2]

[1]University Center of Health Sciences at Klinikum Augsburg (UNIKA-T), University of Augsburg, Augsburg, Germany
[2]Chair of Health Care Operations/Health Information Management, Faculty of Business and Economics, University of Augsburg, Augsburg, Germany
[3]Universitätsklinikum Augsburg, Augsburg, Germany

Correspondence
Jens O. Brunner, Chair of Health Care Operations/Health Information Management, Faculty of Business and Economics, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany.
Email: jens.brunner@unikat.uni-augsburg.de

Abstract

In a master surgery scheduling (MSS) problem, a hospital's operating room (OR) capacity is assigned to different medical specialties. This task is critical since the risk of assigning too much or too little OR time to a specialty is associated with overtime or deficit hours of the staff, deferral or delay of surgeries, and unsatisfied—or even endangered—patients. Most MSS approaches in the literature focus only on the OR while neglecting the impact on downstream units or reflect a simplified version of the real-world situation. We present the first prediction model for the integrated OR scheduling problem based on machine learning. Our three-step approach focuses on the intensive care unit (ICU) and reflects elective and urgent patients, inpatients and outpatients, and all possible paths through the hospital. We provide an empirical evaluation of our method with surgery data for Universitätsklinikum Augsburg, a German tertiary care hospital with 1700 beds. We show that our model outperforms a state-of-the-art model by 43% in number of predicted beds. Our model can be used as supporting tool for hospital managers or incorporated in an optimization model. Eventually, we provide guidance to support hospital managers in scheduling surgeries more efficiently.

KEYWORDS

artificial neural network, downstream units, intensive care unit, machine learning, master surgery scheduling, operations research

1 | INTRODUCTION

The health care industry accounts for a large share of expenditures facing ongoing growth in most countries around the globe. In the United States, 3.3 trillion USD or nearly 18% of its gross domestic product (GDP)[1] were spent on health care in 2016 reflecting an annual growth of 4.3% compared to 2015 (Centers for Medicare & Medicaid Services, 2016). A closer look reveals that hospital care is a key driver of health expenditures accounting for 32% in the United States and nearly 40% in the OECD (Centers for Medicare & Medicaid Services, 2016; OECD Publishing, 2017). With rising costs, hospitals are increasingly attracting attention from sponsors in both the governmental and the private sector demanding more cost effectiveness while ensuring the same level of service quality. "Pressures to make operating margins will continue to be at the forefront of most hospital and health system leaders' minds" (Natarajan, Frenzel, & Smaltz, 2017)—particularly, since it seems that nothing "will stop public spending on health care from rising" (Porter, 2013).

Commonly, the response is to cut costs such as payment levels and benefit structures. However, it would be less harmful and more promising to focus on reducing waste, not value-added care. According to Berwick et al., at least 20% of

[1]9% of GDP on average in the OECD countries (OECD Publishing, 2017), 11% in Germany (Eurostat, 2018).

total health care expenditures could be eliminated by addressing overtreatment, failures in coordination and execution of care processes, inefficient pricing, administrative complexity, fraud and abuse (Berwick & Hackbarth, 2012). Among others, particularly health care operations management has emerged as a key discipline to address wasted expenditure founded on a data-driven, mathematical approach (Carter, Hans, & Kolisch, 2012). Contributions to an improved delivery of health care services are manifold, for example, planning of geographic locations for hospitals (Mestre, Oliveira, & Barbosa-Póvoa, 2012), management of scarce resources (Hulshof et al., 2012), analysis of emergency departments (ED) (Saghafian, Austin, & Traub, 2015), and scheduling of staff (Kim & Mehrotra, 2015) and patient appointments (Gupta & Denton, 2008). A further challenge is getting the right balance between efficiency and responsiveness when handling emergency surgeries (Ferrand, Magazine, & Rao, 2014; Sandbaek, Helgheim, Larsen, & Fasting, 2014). One area with significant impact is scheduling of surgeries in the operating room (OR) of a hospital (Gupta, 2007; Li, Gupta, & Potthoff, 2016) or even within a strategic network of multiple hospitals (Roshanaei, Luong, Aleman, & Urbach, 2017).

Next to the OR, the intensive care unit (ICU) is one of a hospital's most expensive resources representing nearly 15% of United States' total hospital expenditures (Halpern & Pastores, 2010). A cost break-down for the United States is depicted in Figure 1. It shows that 3.3 trillion USD were spent on health care including hospital expenditures of 1.1 trillion USD and ICU expenditures of 0.1 trillion USD. Even more importantly, the ICU is an important bottleneck in most hospitals (Litvak, van Rijsbergen, Boucherie, & van Houdenhoven, 2008). If the ICU reaches capacity, other hospital units such as the OR are blocked and inferior patient treatment is fostered, that is, lower probability of ICU admission (McManus et al., 2003), higher discharge rates (Anderson, Price, Golden, Jank, & Wasil, 2011), and increased danger of re-admission (Baker, Pronovost, Morlock, Geocadin, & Holzmueller, 2009). There is even evidence that the mortality decreases with increasing length of stay (LOS) (Bartel, Chan, & Kim, 2017). Consequently, not only OR capacity, but also the closely linked downstream units such as the ICU should be considered in surgery planning.

Following the classification approach proposed in Guerriero and Guido (2011), OR management problems are categorized into three levels according to their decision hierarchy: strategic, tactical, and operational. At the strategic level, surgery time is distributed among different medical specialties, for example, 1 day per week in one OR is allocated to Neurosurgery. Hereafter, we refer to 1 day per week in one OR as one *OR block*. At the tactical level, a master surgery scheduling (MSS) is developed by assigning the given OR blocks of each medical specialty to specific time slots in specific ORs. The suitability of block scheduling is analyzed by van Oostrum, Bredenhoff, and Hans (2010). Usually, a MSS is constructed cyclical, that is, repeating after a fixed cycle.

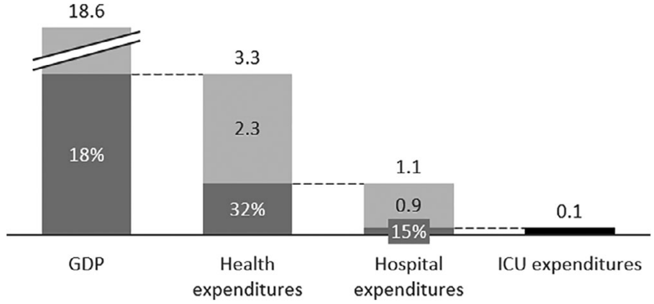

FIGURE 1 Break-down of U.S. health expenditures 2016 in trillion USD. Hospital care accounts for nearly one-third of health costs, thereof about 15% for ICU

An illustrative MSS is presented in Table 8a. At the operational level, specific patients are assigned to given OR blocks.

The purpose of this paper is to predict the impact of surgery decisions on other hospital departments. Our objective is to minimize the deviation between the ICU bed occupancy predicted by our model and the realized ICU bed occupancy. This topic is of high relevance from both a theoretical and a practical standpoint. From a research perspective, it is desirable to achieve prediction results that are as accurate as possible. Our proposed prediction model achieves convincing results and shows higher accuracy than a state-of-the-art approach. Our work is motivated by practical challenges faced in hospitals. Our research partner, the university hospital in Augsburg, is challenged by ICU capacity shortages that cause overtime costs, unsatisfied staff and patients, postponed and canceled surgeries. Based on experience, the hospital management is well aware that allocating OR capacity to different medical specialties has an impact on the resulting occupancy levels of downstream units. Hence, they asked us for a supporting tool in order to address the ICU shortages already on the tactical level during the development and evaluation of a MSS.

We present—to our best of our knowledge—the first prediction model for the integrated OR scheduling problem based on machine learning. We contribute to the research in this field by identifying the key features that are important for a prediction model for the integrated OR scheduling problem and by proposing such a model based on neural networks. In the essential preprocessing step, we retrieve the respective path through the hospital for each patient from the provided electronic hospital records. Furthermore, we formulate the corresponding machine learning problem, derive features and labels, and introduce the memory depth as a new parameter to reflect the impact of previous surgeries. Finally, we configure, train, and deploy a neural network to solve the problem and compare it with alternative machine learning algorithms. In the paper at hand, we consider the ICU as most important supporting unit, but the model can easily be extended to additional units such as ED, general patient ward (hereafter: ward), and intermediate care unit (IMC). We consider elective in- and outpatients as well as urgent patients and reflect all patient paths that occur in the hospital, that is, there is no need to exclude transitions from ICU to ward or patients with multiple surgeries per stay as in state-of-the-art models (see Section 2).

While the application of neural networks is well known for diagnosis in health care and also for forecasting in several industries (see Section 2), we are not aware of previous work applying neural networks to the tactical OR scheduling problem. Machine learning is well suited for this problem since traditional models struggle to reflect the hospital's real-world complexity and its inherent uncertainty. Instead of explicitly modeling the rather complex relationship between inputs, that is, OR blocks per medical specialty, and outputs, that is, number of occupied beds in supporting units, our proposed approach learns automatically from historical data. Our approach is able to reflect a hospitals' real-world complexity in its entirety including more supporting units, patient types, and patient paths than previous work. Neural networks are capable of handling nonlinear relationships. The increasing number of papers that successfully apply neural networks to adjacent fields indicates that the approach is beneficial and worth approaching (Esteva et al., 2017; Goodfellow, Bengio, & Courville, 2016).

Consequently, we provide guidance for hospital managers and show a proof of concept by applying the model to a reference hospital. We preprocess the hospital's real-world data, implement the prediction model, and achieve convincing numerical results. We show that our prediction model outperforms a state-of-the-art model by 43%. Moreover, we demonstrate how one could train the prediction model with ex ante data by bootstrapping from ex post data. The proposed model serves as valuable tool supporting the decision making process of hospital managers in regular discussion rounds to adaptively evaluate a given, feasible MSS with respect to the expected bed occupancy levels in the ICU. Moreover, we incorporate our prediction model in an optimization model and show that the expected ICU bed demand in our partner hospital can be reduced by 8.9% compared to status quo. We also present a comparison with a state-of-the-art optimization model. Our study is based on a large data set covering 7 years of data with nearly 77k patients. Our proposed approach is generalizable to other hospitals since it relies on commonly available electronic hospital records and automatically takes care of all computations from the import of data to the output of the predicted impact on bed occupancy.

The remainder of this paper is organized in five sections. Section 2 provides an overview of previous contributions to OR scheduling and applications of machine learning. Section 3 describes the problem. In Section 4, we present the general prediction model based upon neural networks. In Section 5, we explain the required data input, apply preprocessing and training of the prediction model to a reference hospital, compare our numerical results with state-of-the-art, and discuss applications of our model. In Section 6, we conclude our findings and discuss limitations and managerial insights.

2 | RELATED LITERATURE

In the past six decades, extensive research has been carried out to optimize OR scheduling (Cardoen, Demeulemeester, & Beliën, 2009; Guerriero & Guido, 2011; Samudra et al., 2016). In this section, we first introduce the most relevant contributions to integrated OR scheduling. Then, in order to motivate our approach based upon neural networks, we present previous work addressing neural network applications in the health care sector as well as forecasting models in various industries.

Among other criteria, research on OR scheduling can be differentiated by decision hierarchy (strategic, tactical, operational), patient type (elective vs nonelective, outpatient vs inpatient), performance measure (overutilization vs underutilization), considered supporting units (ED, post-anesthesia care unit [PACU], ward, IMC, ICU), uncertainty (deterministic vs stochastic), research methodology (simulation, mathematical programming, heuristics), and tangibility (theoretic vs real data) (Samudra et al., 2016). Marjamaa, Vakkuri, and Kirvelä (2008) and Marjamaa, Torkki, Hirvensalo, and Kirvelä (2009) provide an overview on OR performance and analyze efficiency gains by parallel processing. OR efficiency without consideration of downstream units has been studied extensively (Batun, Denton, Huschka, & Schaefer, 2011; Dexter & Traub, 2002; Hans, Wullink, Van Houdenhoven, & Kazemier, 2008; Marques, Captivo, & Pato, 2012; Shylo, Prokopyev, & Schaefer, 2013); however, since OR scheduling decisions have also an impact on other departments throughout the entire hospital, an integrated approach that incorporates downstream units seems more suitable to improve their combined performance (Vanberkel, Boucherie, Hans, Hurink, & Litvak, 2010). This interdependency between OR and downstream units, particularly the ICU, is addressed in the following contributions. Adan and Vissers (2002) use integer programming to schedule patients assuming fixed capacities in the OR and downstream units. Their objective is to generate a good admission profile that minimizes the deviation between realized and target resource utilization. Hsu, de Matta, and Lee (2003) focus on bed leveling for outpatients in the PACU. While these two papers model downstream units in a deterministic manner, Beliën and Demeulemeester (2007) present an approach in which both the number of patients and the LOS are assumed to be stochastic. They create MSSs that level the expected bed occupancy in downstream units and are able to calculate the expected bed demand for a given MSS since the probability distribution of arrival times is known. Also, Vanberkel et al. (2011b) propose an analytic model to optimize the MSS and level the expected bed demand. For any given MSS, they are able to calculate the distribution of patients in the wards using binomial distributions and discrete convolutions. A case study where an optimized MSS is implemented successfully in a Dutch cancer center is presented in Vanberkel et al. (2011a).

A question similar to ours is studied by Fügener, Hans, Kolisch, Kortbeek, and Vanberkel (2014) who further extend the previous approach (Vanberkel et al., 2011b) by incorporating multiple downstream units, assessing downstream costs, and maximizing hospital revenues (Fügener, 2015). Based on the probability of an admission to the ICU and the LOS distribution, a three-step approach is presented to derive the occupancy levels for ICU and wards. First, they analyze the pathway of a single patient through the hospital based on historical data to derive the distributions of number of patients resulting from a single OR block (step 1). Second, in order to account for an overlap of patients from previous cycles, the distributions for single OR blocks are convoluted. Third, all blocks of a cyclical MSS are combined to determine the bed demand for the downstream units. In a case study (Fügener et al., 2015), the authors apply the model to a German hospital, predict bed demands for any given MSS, and design two adjusted MSSs resulting in a leveled bed demand and reduced weekend demand, respectively. This approach can be viewed as state-of-the-art since it is integrated, stochastic, based on real data, and considers multiple downstream units. However, their approach differs from our work since their model is constructed in a way such that only a predefined set of patient paths is reflected, that is, OR → (ICU→)ward → discharge, whereas all others are not supported. More precisely, step 1 of their model relies on transition probabilities in a patient flow model that is not able to capture all patient paths (see figure 4 on page 230 in Fügener et al., 2014). This implies for example the loss of information about multiple surgeries, preoperative stays, and transfers from ward to ICU. Moreover, they assume that the probability for a patient to be discharged from the ward after being transferred from the ICU only depends on the time since the transfer from the ICU. For the convolution in their model, they assume that patient paths are independent from each other. In contrast, our approach is based on neural networks which allows us to reflect all patient paths that have occurred in the past. Hence, it seems reasonable that our model is able to reflect the real-world situation even more accurately.

Within the last decade, neural networks have gained momentum across many industries and particularly helped to advance diagnosis applications in health care. Esteva et al. (2017) present a model that outperforms human experts in classifying skin cancer. Other diagnosis applications include the classification of fetal heart rates (Li et al., 2018), prediction of diseases (Chen, Hao, Hwang, Wang, & Wang, 2017), diagnosis of heart diseases under consideration of misclassification costs (Pendharkar & Nanda, 2006), and prediction of colorectal cancer outcome based on tissue samples (Bychkov et al., 2018). Besides for diagnosis, neural networks are also used in the health care sector to support robot-assisted surgeries (Volkov, Hashimoto, Rosman, Meireles, & Rus, 2017), drug discoveries (Wallach, Dzamba, & Heifets, 2015), and a broad variety of additional applications (Hamet & Tremblay, 2017; Natarajan et al., 2017; Shahid, Rappon, & Berta, 2019; Thuemmler & Bai, 2017). Numerous papers can be found using machine learning to address problems on the operational level of a hospital. In particular, the prediction of surgical durations has been studied extensively (Fairley, Scheinker, & Brandeau, 2018; Strum, May, & Vargas, 2000; Tuwatananurak et al., 2019). Unlike them, we consider a problem on the tactical level and do not use individual patient characteristics as input. Across many industries, it has been shown that neural networks are well-suited to tackle forecasting and prediction problems. Some examples include the prediction of rainfall (French, Krajewski, & Cuykendall, 1992), GDP growth (Jahn, 2018), energy consumption in residential buildings (Biswas, Robinson, & Fumo, 2016), useful life of bearings (Guo, Li, Jia, Lei, & Lin, 2017), and railway passenger flow (Toque, Come, Oukhellou, & Trepanier, 2018). Also in hospitals, operational problems have been addressed with machine learning methods. Neural network based models have proven successful for the prediction of hospital admission (Hong, Haimovich, & Taylor, 2018), emergency visits (Hong, Niedzwiecki, Palta, & Tenenbaum, 2018), hospital re-admission (Leung Patrick Cheung & Dahl, 2018), surgery cancellation (Luo, Liu, Hou, & Shi, 2016), surgical durations (Master et al., 2017), clinical deterioration (Churpek et al., 2016), and clinical events (Esteban, Staeck, Baier, Yang, & Tresp, 2016). However, so far we are not aware of any previous work in which neural networks are applied to predict the bed demand for the integrated OR scheduling problem.

Hence, we build upon the existing research by presenting a neural network based approach for the integrated OR scheduling problem that considers OR capacity as well as multiple supporting units (ED, ward, IMC, ICU), is capable of handling uncertainty and complexity, is based on 7 years of real data, and has been designed in close cooperation with key stakeholders of the reference hospital in Augsburg.

3 | PROBLEM DESCRIPTION

In this study, we address the integrated OR scheduling problem. We consider a hospital, in which several medical specialties compete for OR capacity to treat their patients. On the tactical level, hospital managers are asked to develop a MSS by assigning a given number of OR blocks for each medical specialty to specific time slots in specific surgery rooms. As has been emphasized before, this decision affects not only the OR, but also other hospital departments. Hence, an integrated approach to MSS optimization needs to take those effects into account. In order to consider the impact on supporting units, hospital managers require a model that predicts the resulting bed occupancy levels for given surgery decisions.

To understand the impact of surgeries on the supporting units, an individual patient's path needs to be considered. In a rather simple example, a Neurosurgery patient may

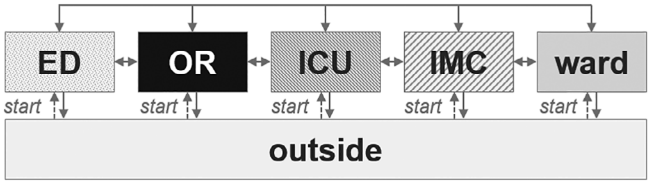

FIGURE 2 Possible patient paths through the hospital. Patients are admitted in any unit of a hospital (OR, ED, ward, IMC, or ICU), transferred among those units in arbitrary order, and finally discharged at any point

be admitted to the ward, then—after a surgery of 3 hours duration—transferred to the ICU for a 3-night stay, then transferred to another ward station for a 5-night stay, and finally discharged from the hospital. Knowing this specific patient path, we are able to draw conclusions from Neurosurgery blocks to the impact on the ICU.

However, in real-life hospitals, not all patients are of the same medical specialty and not all patients of the same medical specialty take the same path. Hence, we deal with a wide variety of different patient paths and different LOS in each unit. This is mainly due to diverse medical conditions, various types of anamneses, different surgical staff and equipment, as well as unforeseen complications during the hospital stay. Figure 2 shows all possible combinations of paths through the units of the hospital and is explained in the following. Self-loops are possible as well, for example, transfers from one ward station to another ward station, but are not depicted in Figure 2 since we directly aggregate stations that belong to the same unit.

To formalize the following steps, we use the notation for sets and indices as described in Table 1. We consider $r \in \mathcal{R}$ surgery rooms that are combined to one common OR capacity which is depicted as a box in Figure 2. In the same way, we refer to the supporting units as $u \in \mathcal{U}$. Figure 2 shows boxes for the supporting units ED, ICU, IMC, and the ward. Each supporting unit is composed of several stations $h_u \in \mathcal{H}_u$. For example, the supporting unit $u = 1$ might be composed of three stations, that is, stations with the numbers $\mathcal{H}_1 = \{1,2,3\}$. In this study, one "hyper-ICU" is considered to pool the capacity of all relevant ICU stations. This assumption is reasonable since the ICU stations are interdisciplinary and, hence, utilization can be balanced. The same holds true for ED, IMC, and the ward, respectively. Even in case of fixed allocation of hospital beds to medical specialties, patients might overflow to beds of related specialties if the designated ones are fully occupied, that is, "off-service placement" (Song, Tucker, Graue, Moravick, & Yang, 2019). For hospitals where this is not acceptable, additional supporting units can be introduced in order to pool the capacity of selected (or even single) ward stations, for example, $\mathcal{H}_{1'} = \{1, 2\}$, $\mathcal{H}_{1''} = \{3\}$. A patient can be admitted to any unit of the hospital (indicated by the dotted arrow called *Start*), transferred between any units (indicated by arrows), and discharged from any unit (indicated by the arrows ending in the unit called outside). The outside unit is used to locate a patient that is currently not within the hospital. Knowing the time of the

TABLE 1 Sets and indices

Description	Index ∈ set
ORs	$r \in \mathcal{R} = \{1, \ldots, R\}$
Supporting units	$u \in \mathcal{U} = \{1, \ldots, U\}$, eg, where $1 \widehat{=}$ICU
Stations belonging to unit u	$h_u \in \mathcal{H}_u = \{1, \ldots, H_u\}$
Patients	$p \in \mathcal{P} = \{1, \ldots, P\}$
Patient paths	$f \in \mathcal{F} = \{1, \ldots, F\}$
Surgeries	$s \in \mathcal{S} = \{1, \ldots, S\}$
Medical specialties	$j \in \mathcal{J} = \{1, \ldots, J\}$, eg, where $1 \widehat{=}$ Cardiothoracic Surgery
Samples/days	$m \in \mathcal{M} = \{1, \ldots, M\}$, where $M = M_{\text{train}} + M_{\text{val}} + M_{\text{test}}$
Features	$n \in \mathcal{N} = \{1, \ldots, N\}$
Memory depth	$d \in \mathcal{D} = \{D^-, \ldots, 0, \ldots, D^+\}$
Days of the week	$e \in \mathcal{E} = \{1, \ldots, 7\}$, where $1 \widehat{=}$ Monday, … , $7 \widehat{=}$ Sunday

admission to the hospital as well as the LOS for each unit, a patient's position within the hospital can be reconstructed for any given date. Accumulating over all patients allows us to draw conclusions for the bed occupancy level in each unit. Furthermore, we refer to patients as $p \in \mathcal{P}$, surgeries as $s \in \mathcal{S}$, patient paths as $f \in \mathcal{F}$, and medical specialties as $j \in \mathcal{J}$.

To predict the impact of surgery decisions on the ICU, we develop and implement a neural network based model which is described in Section 4. In this model, we define one sample $m \in \mathcal{M}$ as 1 day chosen from the set of available days. The description of the features $n \in \mathcal{N}$ is explained in Section 4.3. As not only surgery decisions made for today, but also surgeries conducted on preceding and subsequent days affect today's bed occupancy level, we introduce a new parameter called memory depth $d \in \mathcal{D}$. This parameter regulates how many preceding days D^- and subsequent days D^+ are considered by the model for the prediction of the bed occupancy level on day $m \in \mathcal{M}$.

By means of this study, we want to support hospital managers in making substantiated surgery decisions by providing them with a model that predicts the resulting bed occupancy. Our objective is to minimize the deviation between predicted and actual ICU occupancy levels.

4 | NEURAL NETWORK MODEL FOR PREDICTION OF ICU BED OCCUPANCY

In this section, we formulate the mathematical model to predict the ICU bed occupancy for a given MSS. Even though the proposed prediction model can be applied to any up- or downstream unit, we believe that it is reasonable to focus on one unit for this study. The ICU is most critical since it is most expensive, represents an important bottleneck in the hospital, and bears the risk of blocking ORs.

This section is organized in three parts. First, in Section 4.1, we introduce the concept of neural networks which serves as foundation for our model (readers familiar with machine learning might skip this part). Second, we describe how

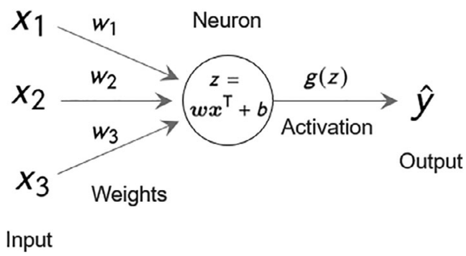

FIGURE 3 A neuron mapping three inputs to one output. Neurons are the fundamental building block of neural networks

to preprocess the hospital records and retrieve the required inputs for the model in Section 4.2. Third, we formulate the neural network problem to predict the ICU bed occupancy (see Section 4.3).

4.1 | Neural networks

The problem at hand corresponds to finding a computable function $\boldsymbol{x} \mapsto f(\boldsymbol{x}) = y$ where we know the input vector $\boldsymbol{x}$, that is, the surgery decisions, and want to draw conclusions for the output value y, that is, the resulting bed occupancy. However, the relationship between inputs and output is not trivial for the underlying hospital setting and hence, difficult to model. Even state-of-the-art literature struggles to represent the real-world complexity in a well-defined model (see Section 2). Instead of modeling the relationship with numerous constraints and restricting the validity to a narrow scope, our model is developed directly from the underlying real-world data.

Supervised learning is a well-suited method to comprehend complex relationships based on historical data. Particularly the field of deep learning[2] has made significant progress in the past decades and—combined with today's computing capacity—offers entirely new possibilities. A comprehensive overview on deep learning is provided in Goodfellow et al. (2016) and LeCun, Bengio, and Hinton (2015).

Neuron. The basic building block of each neural network is a *neuron*. It models the relation between inputs, the so-called *features*, and the output, also called *label*. Figure 3 depicts a neuron with three features and one label for one sample $m = 1$. The input values x_1, x_2, and x_3 are mapped to the output value $\hat{y}$. As we consider only one sample, we omit the sample index $m = 1$.

In a neuron, three mathematical operations are performed. In Equation (1), the dot product between the input vector $\boldsymbol{x}$ and the corresponding weight vector $\boldsymbol{w}$ is computed and a bias b is added. In Equation (2), an activation function $g(z)$ is applied to the term resulting in the prediction $\hat{y}$.

$$z = \boldsymbol{w}\boldsymbol{x}^{\mathrm{T}} + b. \tag{1}$$

$$\hat{y} = g(z). \tag{2}$$

The activation function in (2) introduces nonlinearity in the model and is one of the main reasons for the improved performance of neural networks compared to classical inference approaches such as linear regression. Additionally, they may also be understood in terms of mapping the results of calculations to their natural domains, for example, for probabilities to the interval between zero and one. Common activation functions are sigmoid, ReLu and tanh (see Chapter 6 in Goodfellow et al., 2016).

Network of neurons. Simple neurons are then combined to larger structures. With multiple neurons, an additional layer (called *hidden layer*) can be formed where each neuron is connected to all inputs in the first layer (called *input layer*) and to the neuron in the last layer (called *output layer*). The neuron in the output layer is fed with the outputs of the neurons in the previous layer. Hence, more options for weighting the features are provided and more complex functions can be represented. Kolmogorov's theorem states that any continuous real function on the n-dimensional unit cube is representable by sums and superpositions of continuous real functions of one variable (Kolmogorov, 1957). This implies that any continuous function with N inputs can be represented by an hidden layer comprising exactly $2N + 1$ neurons (Hecht-Nielsen, 1987). However, one often achieves better performance by appending even more layers of neurons. The topology of a neural network is defined by the number of layers and the number of neurons in each layer. As the input layer is defined by the number of features and the output layer is defined by the type of the problem,[3] it remains to choose the number of hidden layers and the number of neurons in each hidden layer. Hereafter, we use the form "1:2:3" to indicate the number of neurons in each hidden layer. For example, the topology "16:4" would describe a neural network with two hidden layers comprising 16 and 4 neurons, respectively. The model is fed by N features $x_1, \ldots, x_N$ depicted in the input layer. In each hidden layer, each neuron receives all values of the previous layer, weights them accordingly, adds a bias term, and applies the activation function. The output of each neuron is then propagated to the next layer. Finally, the prediction $\hat{y}$ in the output layer is computed as the weighted sum of the values from the last hidden layer.

The neural network requires training to learn the relationship between inputs and output. During the training, the network is shown many samples, that is, multiple instances of input values and the corresponding output values. Thus, the features are described by the feature matrix $\boldsymbol{X}$ and the labels by the label vector $\boldsymbol{y}$. In order to evaluate the performance of the prediction, a loss function is used measuring the deviation between the original output vector $\boldsymbol{y}$ and the prediction vector $\hat{\boldsymbol{y}}$. Furthermore, we summarize the weights and biases of a neural network as parameter vector $\boldsymbol{\theta}$. An overview of the notation commonly used in machine learning is found in Appendix A.

[2] A deep neural network is characterized by multiple layers between the input layer and the output layer, also called *hidden layers*.

[3] In regression problems, the output layer comprises exactly one neuron.

TABLE 2 Patient flow table for a single patient $p \in \mathcal{P}$

Index	Timestamp	Type	From_unit	From_station	To_unit	To_station
0	2 January 2014 2:58 PM	Admission	Outside	—	Ward	051
1	2 January 2014 4:31 PM	Start of surgery	—	—	—	—
2	2 January 2014 7:46 PM	End of surgery	—	—	—	—
3	2 January 2014 7:46 PM	Transfer	Ward	051	ICU	031
4	5 January 2014 09:59 AM	Transfer	ICU	031	Ward	053
5	10 January 2014 10:17 AM	Discharge	Ward	053	Outside	—

Abbreviations: This table containing all movements through the hospital is reconstructed for each patient.

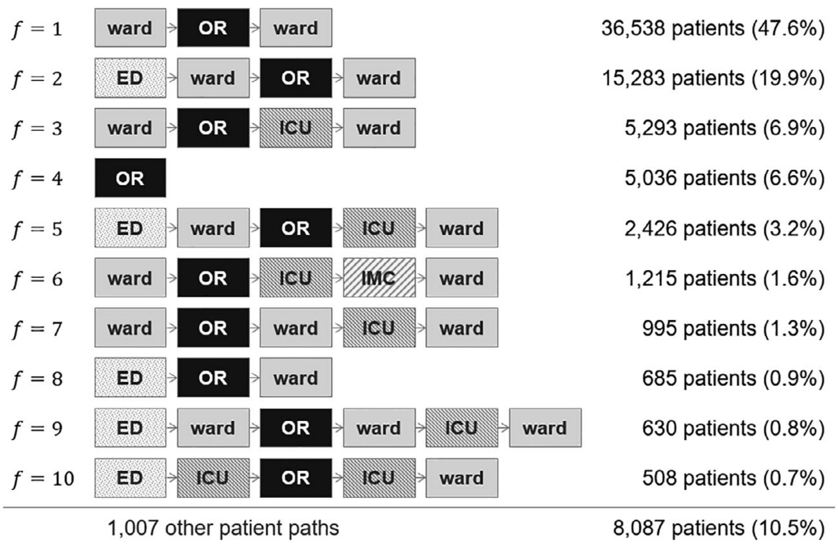

FIGURE 4 Common patient paths and their corresponding patient numbers. Nearly 90% of all patients choose one of the 10 most common patient paths and 7% of the patients are outpatients

Two steps are required, to develop and implement a neural network that predicts the bed occupancy. First, features and labels need to be provided to feed the model. Second, the model needs to determine the best parameter $\boldsymbol{\theta}$ that minimizes the deviation between real and predicted bed occupancy. These two steps are covered in the next two parts.

4.2 | Data preprocessing for extraction of features and labels

In this part, we describe how to retrieve the surgery schedules as well as the resulting bed occupancy levels from historical hospital records. Since hospital records are usually patient-specific, we first need to reconstruct each individual patient's path through the hospital before drawing conclusions for the entire hospital.

Patient flow. For each patient $p \in \mathcal{P}$, a patient flow table is computed containing timestamps, origins, and destinations for relevant transactions that occur during a patient's journey through the hospital. In particular, the admission to the hospital, transfers within the hospital, start and end of a surgery, and the discharge from the hospital are reconstructed. Table 2 shows an illustrative patient flow table for a single patient. Each row corresponds to a specific event. This patient has been admitted to the hospital on 2 January 2014, at 14:58 (which is 2:58 PM). After the surgery, which took place between 2:31 PM and 7:46 PM, the patient has been transferred to the ICU, and afterwards to the ward. Finally, the patient has been released from the hospital on 10 January at 10:17 AM. Depending on medical specialty, medical conditions, surgical staff and equipment, and additional external factors, different patient paths $f \in \mathcal{F}$ through the hospital are encountered. Figure 4 in Section 5.2 depicts the 10 most common paths identified for the reference hospital. Luckily, we do not need to treat all these cases individually as the model will be able to generalize once it has encountered such cases during the training.

4.2.1 | Acquiring the label data: occupancy level

Based on the patient path of an individual patient, we are able to reconstruct its exact location within the hospital for any given date. By superimposing the locations of all patients, we obtain the bed occupancy levels for a hospital unit. The bed occupancy in supporting unit $u \in \mathcal{U}$ on day $m \in \mathcal{M}$ is defined as:

$$\sum_{h_u \in \mathcal{H}_u} \sum_{p \in \mathcal{P}} \delta_{h_u,p,m}, \quad \forall u \in \mathcal{U}, \quad m \in \mathcal{M}, \tag{3}$$

where $\delta_{h_u,p,m} = 1$ denotes that a patient p has been present in station h_u on day m. In this study, we focus on the ICU.

The presented approach can easily be extended to additional supporting units or individual stations.

In close alignment with hospital managers we made two important assumptions. First, patients who are transferred to the ICU within 3 hours after their surgery are regarded as directly transferred, even if the records would suggest an interim ward-stay between surgery and ICU. In most cases, the delayed transfer in the records results from the delayed manual registration rather than from an actual stay in the ward. Thus, it is fair to assume that the patients went to the ICU directly after surgery. Second, we assume that a bed can be occupied by at most one patient per day unless the previous patient is leaving before 11 AM, that is, the so-called *hotel principle*. This assumption is well established within the tactical OR literature (Fügener et al., 2014; Fügener et al., 2015; Shi, Helm, Deglise-Hawkinson, & Pan, 2019; Vanberkel et al., 2011b) since metrics are considered on a daily basis. Furthermore, this assumption is also supported by the data of the reference hospital (in over 98% of the considered bed-days) and consistent from a practical point of view since a bed is reserved for every day of a patient's stay and first needs to be cleaned and prepared before being ready for a new patient.

4.2.2 | Acquiring the feature data: surgery schedule

On a tactical level, one focuses on the surgery decisions as main drivers for the occupancy level in the ICU. Hence, this approach would ideally be based on ex ante surgery schedules. In fact, ex ante data account for the additional uncertainty given that a significant number of scheduled surgeries are canceled in advance (Dexter, Maxbauer, Stout, Archbold, & Epstein, 2014). However, due to the limited data availability, we base this approach on ex post data in which cancellations, no shows, and rescheduling are already reflected. In particular, we focus on the historical surgery records to retrieve the allocation of OR blocks (r, m) for all medical specialties. For each room $r \in \mathcal{R}$ in which a medical specialty has performed surgeries on a given day $m \in \mathcal{M}$, one OR block (r, m) is assigned to the respective specialty. If the same room has been utilized by multiple specialties on the same day, the OR block is allocated to the involved medical specialties according to their accumulated surgery durations. The number of OR blocks assigned to specialty $j \in \mathcal{J}$ on day $m \in \mathcal{M}$ is defined as:

$$\sum_{r\in\mathcal{R}} \frac{\sum_{s\in\mathcal{S}} t_{s,j,r,m}}{\sum_{j\in\mathcal{J}}\sum_{s\in\mathcal{S}} t_{s,j,r,m}}, \quad \forall j \in \mathcal{J}, m \in \mathcal{M}, \tag{4}$$

where $t_{s,j,r,m}$ denotes the duration of surgery s of a patient p of medical specialty j in the surgery room r on day m.

Table 3 shows an illustrative surgery schedule. For example, no OR block was allocated to the medical specialty $j = 1$ on 26 December 2016, and three OR blocks on 27 December. On 28 December, three full OR blocks were assigned to $j = 1$ and an additional room was shared with the medical specialty $j = 3$, where $j = 1$ used 27% of the total surgery time in this room on this day.

4.3 | Neural network problem formulation

After preprocessing, we formulate the supervised regression problem: "predict the number of occupied beds in the ICU on day $m \in \mathcal{M}$ for a given assignment of OR blocks (r, m) to J medical specialties on day m, up to D^- days before, and up to D^+ days after." Labels and features were chosen in close alignment with hospital managers as follows:

Label	$y^{(m)} \in \mathbb{R}_0^+$	Number of occupied beds in the ICU on day m, $\forall m \in \mathcal{M}$
Features	$x_{(e)}^{(m)} \in \{0, 1\}$	One-hot encoding for the **day of the week** e, $\forall e \in \mathcal{E}, m \in \mathcal{M}$
	$x_{(j,d)}^{(m)} \in \mathbb{R}_0^+$	Number of **OR blocks** allocated to specialty j on day $m + d$, $\forall j \in \mathcal{J}, m \in \mathcal{M}, d \in \mathcal{D}$

Hence, we obtain the number N of feature vectors $\boldsymbol{x}_n$, $n \in \mathcal{N} = \{1, \ldots, N\}$, in Equation (5).

$$N = E + J|\mathcal{D}|. \tag{5}$$

The features are labeled by $n \in \mathcal{N}$ in the subscript. The memory depth $d \in \mathcal{D}$ describes the number of previous days D^- and subsequent days D^+ which are taken into account for each sample. For instance, $\mathcal{D} = \{-1, 0\}$ indicates that today's and yesterday's surgeries are reflected in the feature matrix. Further assuming $J = 8$ specialties, we achieve $N = 23$ feature vectors in this example. Additional features are discussed in Section 6.

4.3.1 | Defining features and data set

Full data set. Juxtaposing all feature vectors yields the feature matrix as

$$\boldsymbol{X} = [\boldsymbol{x}_1 \;\; \ldots \;\; \boldsymbol{x}_N] = \left[\begin{array}{c|c} \underbrace{\begin{matrix} x_{e=1}^{(1)} & \cdots & x_{e=7}^{(1)} \\ x_{e=1}^{(2)} & \cdots & x_{e=7}^{(2)} \\ \cdots & \cdots & \cdots \\ x_{e=1}^{(M)} & \cdots & x_{e=7}^{(M)} \end{matrix}}_{\text{Weekday features}} & \underbrace{\begin{matrix} x_{j=1,d=D^-}^{(1)} & \cdots & x_{j=1,d=0}^{(1)} & \cdots & x_{j=1,d=D^+}^{(1)} \\ x_{j=1,d=D^-}^{(2)} & \cdots & x_{j=1,d=0}^{(2)} & \cdots & x_{j=1,d=D^+}^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{j=1,d=D^-}^{(M)} & \cdots & x_{j=1,d=0}^{(M)} & \cdots & x_{j=1,d=D^+}^{(M)} \end{matrix}}_{\text{OR block features for specialty } 1} \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \end{matrix} \underbrace{\begin{matrix} \cdots & x_{j=J,d=D^+}^{(1)} \\ \cdots & x_{j=J,d=D^+}^{(2)} \\ \cdots & \cdots \\ \cdots & x_{j=J,d=D^+}^{(M)} \end{matrix}}_{\text{OR block features for } J} \end{array}\right]. \tag{6}$$

A sample with index m is defined as a set consisting of the respective label for the selected day m as well as the respective values of the feature vectors (ie, one row of $\boldsymbol{X}$) given in (7).

$$(y^{(m)}, x_1^{(m)}, \ldots, x_N^{(m)}). \tag{7}$$

All samples are combined to a data set $(\boldsymbol{y}, \boldsymbol{X})$, also called *full data set* as it covers all selected specialties. Table 3 shows 3 days (27–29 December) of the feature matrix with memory depth $\mathcal{D} = \{-1, 0\}$ for the historical surgery schedule depicted in Table 4. For example, three OR blocks were assigned to the medical specialty $j = 1$ on 27 December 2016. The fact that this day was a Tuesday, is indicated with a one in the second column. If looking at the next sample, that is, 28 January 2016, we notice that the three OR blocks are also found in column $\boldsymbol{x}_{j=1, d=-1}$, which denotes the surgeries conducted on the previous day.

4.3.2 | Increasing the number of samples

Isolated data set. While the previous approach provided a perspective on the joint effect of all medical specialties $j \in \mathcal{J}$, most hospital records allow also a dedicated break-down to patients of each individual medical specialty. By doing so, we generate more samples and enable the neural network to further learn about the underlying structure. The simple example below illustrates that additional information can be extracted by breaking down the full data set covering all medical specialties into *isolated data sets* for the individual medical specialties.

$$\text{Full data set}: \quad (y, x_{j=1}, x_{j=2})_{\text{full}} = (15, 2, 3)$$
$$\iff 2\,\text{blocks for spec.}\ 1 + 3\ \text{blocks for spec.}\ 2 \mapsto 15\ \text{ICU beds.}$$

$$\text{Isolated data set}\ 1: \quad (y, x_{j=1}, x_{j=2})_{j=1} = (5, 2, 0)$$
$$\iff 2\ \text{blocks for specialty}\ 1 \mapsto 5\ \text{ICU beds.}$$

$$\text{Isolated data set}\ 2: \quad (y, x_{j=1}, x_{j=2})_{j=2} = (10, 0, 3)$$
$$\iff 3\ \text{blocks for specialty}\ 2 \mapsto 10\ \text{ICU beds.}$$

In the full data set of this example it remains unclear which share of the joint bed demand of 15 beds is caused by which of the two involved medical specialties. Numerous solutions are possible, for example, six ICU beds for each OR block of specialty 1 and one bed for each OR block of specialty 2. The isolated data sets reveal the function between OR block and resulting bed demand without ambiguity.

Combined data set. To leverage the full information of the hospital records, we derive the full data set $(\boldsymbol{y}, \boldsymbol{X})_{\text{full}}$ as well as the isolated data set $(\boldsymbol{y}, \boldsymbol{X})_j$ for each medical specialty $j \in \mathcal{J}$ and append all sets into large, *combined data set* given in (8).

$$(\boldsymbol{y}, \boldsymbol{X}) = \begin{bmatrix} (\boldsymbol{y}, \boldsymbol{X})_{\text{full}} \\ (\boldsymbol{y}, \boldsymbol{X})_1 \\ (\boldsymbol{y}, \boldsymbol{X})_2 \\ \cdots \\ (\boldsymbol{y}, \boldsymbol{X})_J \end{bmatrix}. \tag{8}$$

The resulting feature matrix $\boldsymbol{X}$ is defined in (9) and the corresponding $\boldsymbol{y}$ is composed of the respective bed occupancy of the full set followed by the bed occupancy of each isolated medical specialty. One could also think of adding all combinations of specialties to the data set which we did not explore in this study.

$$\boldsymbol{X} = \left[\begin{array}{cccccccccc}
x^{(1)}_{e=1} & \cdots & x^{(1)}_{e=7} & x^{(1)}_{j=1,d=D^-} & \cdots & x^{(1)}_{j=1,d=0} & \cdots & x^{(1)}_{j=1,d=D^+} & \cdots & x^{(1)}_{j=J,d=D^+} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
x^{(M)}_{e=1} & \cdots & x^{(M)}_{e=7} & x^{(M)}_{j=1,d=D^-} & \cdots & x^{(M)}_{j=1,d=0} & \cdots & x^{(M)}_{j=1,d=D^+} & \cdots & x^{(M)}_{j=J,d=D^+} \\
x^{(1)}_{e=1} & \cdots & x^{(1)}_{e=7} & x^{(1)}_{j=1,d=D^-} & x^{(1)}_{j=1,d=0} & x^{(1)}_{j=1,d=D^+} & \cdots & \cdots & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
x^{(M)}_{e=1} & \cdots & x^{(M)}_{e=7} & x^{(M)}_{j=1,d=D^-} & \cdots & x^{(M)}_{j=1,d=0} & \cdots & x^{(M)}_{j=1,d=D^+} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
x^{(1)}_{e=1} & \cdots & x^{(1)}_{e=7} & 0 & \cdots & 0 & \cdots & 0 & \cdots & x^{(1)}_{j=J,d=D^+} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
x^{(M)}_{e=1} & \cdots & x^{(M)}_{e=7} & 0 & \cdots & 0 & \cdots & \cdots & \cdots & x^{(M)}_{j=J,d=D^+} \\
\multicolumn{3}{c}{\underbrace{\hphantom{xxxxxxxxxxxx}}_{\text{Weekday features}}} & \multicolumn{5}{c}{\underbrace{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxx}}_{\text{OR block features for specialty 1}}} & \multicolumn{2}{c}{\underbrace{\hphantom{xxxxxxxx}}_{\text{OR block features for } J}}
\end{array}\right]
\begin{array}{l}
\\ \\ \text{Full set} \\ \\ \text{Isolated sets} \\ \text{for } j = 1 \\ \cdots \\ \text{for } j = J \\ \\ \\ \\
\end{array} \tag{9}$$

TABLE 3 Three days of the feature matrix with memory depth $\mathcal{D} = \{-1, 0\}$ corresponding to the surgery schedule depicted in Table 4

Sample	Date	$x_{e=\text{Mon}}$	$x_{e=\text{Tue}}$	$x_{e=\text{Wed}}$	…	$x_{j=1,d=-1}$	$x_{j=1,d=0}$	…	$x_{j=8,d=-1}$	$x_{j=8,d=0}$
2	27 December 2016	0	1	0	…	0.00	3.00	…	0.00	1.00
3	28 December 2016	0	0	1	…	3.00	3.27	…	1.00	0.00
4	29 December 2016	0	0	0	…	3.27	2.00	…	0.00	0.00

Abbreviations: The first seven columns represent the one-hot encoded days of the week $e \in \mathcal{E}$, the remaining columns contain the number of assigned OR blocks for each specialty $j \in \mathcal{J}$ and each day $m \in \mathcal{M}$.

TABLE 4 Historical surgery schedule in number of OR blocks

Sample	Date	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$	$j = 8$
1	26 December 2016	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
2	27 December 2016	3.00	1.00	1.00	1.00	0.00	1.00	2.00	1.00
3	28 December 2016	3.27	0.00	0.73	4.00	1.00	1.00	3.00	0.00
4	29 December 2016	2.00	1.00	1.00	2.00	0.00	1.00	3.00	0.00

Abbreviations: For each OR that is exclusively occupied by one medical specialty, one OR block is assigned to the specialty.

4.3.3 | Training the neural network

The resulting data set is split randomly into three subsets, that is, *training set*, *validation set*, and *test set*. Since our model is based on ex post data as reconstructed from the historical records, the OR capacity for each medical specialty differs between individual days such that variability is reflected when randomly assigning samples to sets. Note that ex ante data would be preferred over ex post data since a significant number of scheduled surgeries is canceled or rescheduled in advance (Dexter et al., 2014). However, due to the limited data availability, this model is based on ex post data. The training set is used to train the neural network, that is, to fit the parameters $\boldsymbol{\theta}$ of the model such that the prediction error of the loss function $J(\boldsymbol{\theta})$ is minimized. The validation set is used to evaluate the performance during training and adjust the hyperparameters accordingly. The test set contains M_{test} "fresh" samples that are excluded from the training process and hence ensure an unbiased performance measure.

After discussions with the hospital management, we decided that robustness of the model has highest priority and consequently selected the mean squared error (MSE) as loss function for this study. This choice implies limitations for the prediction accuracy for high congestion periods as the MSE metric averages both congested and noncongested periods instead of only focusing on congested periods. Using a model $\hat{\boldsymbol{y}}$ with the parameters $\boldsymbol{\theta}$ as prediction for $\boldsymbol{y}$, the loss function $J(\boldsymbol{\theta})$ is given in Equation (10).

$$J(\boldsymbol{\theta}) = \frac{1}{M} \sum_{m=1}^{M} (\hat{\boldsymbol{y}}^{(m)} - \boldsymbol{y}^{(m)})^2. \tag{10}$$

Then, the neural network problem is formulated as in (11).

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}). \tag{11}$$

During the training, gradient descent is used to minimize the loss function by fitting the parameters to the data set.

5 | NUMERICAL RESULTS FOR REFERENCE HOSPITAL

In this section, we apply the proposed model to real data retrieved from a reference hospital in order to show proof of concept and provide guidance for other hospital managers. This section is structured in six parts. First, we provide detailed information about the particular setting in the reference hospital and explain the required input data. Second, we apply preprocessing to reconstruct historical patient paths, ex post surgery schedules, and utilization levels and finally derive features and labels. Third, we train the prediction model, deploy it to test samples, and present numerical results for various constellations. Fourth, we compare our prediction model with the results of alternative algorithms as well as the state-of-the-art model by Fügener et al. (2014). Fifth, we show how the model can be trained with bootstrapped ex ante data. Sixth, we present two valuable options to inform the decision making process using the prediction model, that is, as supporting tool for the hospital management and as objective function in an optimization model.

5.1 | Reference hospital and required input data

Although levels of detail, professionalism, and standardization, as well as types of systems and tools employed for OR scheduling vary significantly between different hospitals, most hospitals keep track of conducted surgeries as well as patient movements. The proposed model requires two sets of input data:

- *Surgery records*. Documentation of details for conducted surgeries. This data set contains patient-related information such as unique patient identifiers, patient types, levels of urgency, and medical specialties, as well as detailed timestamps for each surgery-related step.

- *Supporting units records*. Documentation of patient movements within and between supporting units, that is, ED, wards, IMC, and ICU. This data set contains information such as unique patient identifiers as well as timestamps, station numbers, and details for admission, transfer, and discharge.

Based on our experience, these data sets are commonly available in most hospitals and can be pulled directly from the hospital information system, for example, Agfa Healthcare Orbis[4] in the reference hospital. Universitätsklinikum Augsburg is a 1700-bed, maximum-care university hospital located in Southern Germany. Each year, around 250 000 patients are treated by more than 700 doctors and 2000 nurses in 25 medical specialties and institutes. For this study, we focus on the central OR department[5] in the central location of Universitätsklinikum Augsburg located in Neusäß. We work with the two aforementioned sets of input data covering 7 years from 2010 to 2016. The two data sets are interlinked by an unique patient identifier which is essential to reconstruct the patient paths. Excerpts of both hospital records are shown in Appendix C. The hospital's surgeries are currently orchestrated via MSS A (based on status quo) which is illustrated in Table 8a. The MSS is evaluated and modified once per quarter by the operating room (OR) manager and co-author of this manuscript, Dr Thomas Koperna. It needs to meet the total weekly number of OR blocks for each specialty and cannot exceed the available OR capacity per day. Key evaluation criteria are the impact on downstream departments, particularly the ICU, as well as the availability of staff. During the considered time period, the strategic case mix as well as the average number of treated patients has remained nearly unchanged (see Appendix D) . However, due to the ample period of 7 years, continuous modifications to the MSS, and particularly the fact that we work with ex post data, diversification is still reflected in the data, that is, each medical specialty has seen surgeries on each day of the week, which is essential for the learning process.

In general hospitals, one typically distinguishes between medical and surgical patients as well as between elective and emergency patients. Since the proposed model aims to predict the impact of OR scheduling, only patients with at least one surgery are relevant. This is also supported from a practical perspective since medical and surgical patients utilize dedicated, autonomously acting up- and downstream units. Furthermore, we consider elective patients since they can be rescheduled and, hence, managed by a MSS. Those patients that can be rescheduled within some hours are referred to as urgent patients. While there is less flexibility for urgent patients, they are also considered in our study since there is still some ability to reschedule the respective surgeries, that is, within 6 to 24 hours in the reference hospital. However, while we include urgent patients, it is not reasonable to include emergency patients in the tactical OR scheduling problem since they cannot be scheduled in advance and must be treated immediately in the next available OR. We suggest to exclude them from tactical planning, but rather ensure sufficient buffer capacity in the respective units on top of the predicted demand to accommodate emergency patients. This approach is well established within the tactical OR scheduling literature (Fügener et al., 2014; Vanberkel et al., 2011a). Our patient selection also includes long-term patients that are rather difficult to predict on a tactical level and have a significant impact on the occupancy level as well as on the performance of the prediction model, that is, a maximum LOS of 289 days was observed in the data set. In order to ensure data consistency and a well-defined scope, it is crucial to focus on patients whose patient paths are entirely reflected in the provided hospital records, that is, all ORs and supporting units. For example, we exclude patients of the specialty Dermatology since they are mainly operated on at a different hospital location in the south of Augsburg which is not entirely covered by our data set. Summing up, we select $P = 77k$ patients based on the following criteria:

- *Time period*. The patient stayed in the hospital within the years 2010 and 2016 and had at least one surgery.
- *OR*. The patient was treated in an OR in scope, that is, one of the ORs in the central OR department.
- *Medical specialty*. The patient was associated with one of eight medical specialties in scope, that is, using the ORs in scope as well as the ICU stations in scope. A list of these specialties is provided in Appendix B.
- *Patient type*. The patient was either elective or urgent, either inpatient or outpatient.

The ORs in the central OR department at the reference hospital are mainly utilized by $J = 8$ medical specialties. Hence, patients belonging to one of these medical specialties are considered for this study, that is, Cardiothoracic Surgery (CAS), General, Visceral, and Transplant Surgery (GES), Gynecology (GYN), Oral and Maxillofacial Surgery (MAS), Neurosurgery (NEU), Traumatology, Orthopedics, and Plastic Surgery (TRA), Urology (URO), and Vascular and Intravascular Surgery (VAS). All stations that belong to the same supporting unit are aggregated to a hyper-unit. We consider $I = 4$ supporting units, namely ICU covering $H_{\text{ICU}} = 9$ stations, IMC covering $H_{\text{IMC}} = 1$ station, ED covering $H_{\text{ED}} = 4$ stations, and the ward covering $H_{\text{ward}} = 79$ stations. The extensive data set used in this study outnumbers comparable contributions. Adan and Vissers (2002) base their work on 1 week of data comprising 760 patients, Fügener et al. (2015) look at 6 months of data comprising 2480 patients in three specialties, and Vanberkel et al. (2011b) consider 1 year of data from a 150-bed hospital. The comprehensive data set of Universitätsklinikum Augsburg provides a solid foundation for the evaluation of our approach. As this study has been conducted in close cooperation with key stakeholders

[4]https://global.agfahealthcare.com/main/hospital-it/orbis/.
[5]This department includes all central ORs ("ZOP"), ORs located on the top floor ("Dach-OP"), and ORs located on the first floor ("EG-OP").

of the reference hospital, profound understanding on either side is ensured and potential benefits are evident. Universitätsklinikum Augsburg plans to introduce a new, revised MSS within 2020 where the results of this study form the foundation.

5.2 | Preprocessing results

Starting from plain hospital records containing information about surgery timestamps and movements in supporting units, we apply preprocessing to reconstruct all individual paths that patients take through the hospital. Analyzing Universitätsklinikum Augsburg's historical data from 2010 to 2016, we found the patient paths as depicted in Figure 4. Those patient paths represent the core of our approach since they are essential to reconstruct occupancy levels and ex post surgery schedules, derive features and labels, and finally train the neural network based prediction model.

Nearly half of all patients follow the same patient path ($f = 1$): admission to the hospital, stay in the ward, transfer to the OR for surgery, return to the ward for another stay, and finally discharge from the hospital (depicted in the first row of Figure 4). The fourth path ($f = 4$) refers to outpatients with surgeries. Overall, we found $F = 1{,}017$ distinct patient paths through the hospital while nearly 90% of the patients choose one of the 10 most common ones. Finally, we derive the occupancy levels in up- and downstream units for each given day by accumulating over all patients whose patient paths indicate a stay in that station on that day. Analogously, we retrieve the ex post surgery schedule from the patient paths such that also cancellations, no shows, and rescheduling are reflected. The occupancy levels and ex post surgery schedule are then transformed to a data set composed of labels and features. Note that ex ante data would be preferred, but we use ex post data due to limited data availability.

We use `scikit-learn's train_test_split` (Pedregosa et al., 2011) to split the resulting data set randomly into three subsets, that is, M_{train} samples are assigned to the training set and M_{val} samples to the validation set. Afterwards, each feature is scaled and normalized individually with `scikit-learn's MinMaxScaler` such that it is in the range between zero and one. When fitting the estimator, we only used the training set in order to prevent a spillover of information, that is, the constants computed with the training set are also used to scale the validation and test set. For an overview of the commonly-used notation, we refer to Appendix A.

5.3 | Prediction results

The proposed neural network model to predict bed occupancy levels has been implemented in Python. For data processing, we use `Pandas`, `NumPy`, `SciPy`, `scikit-learn`, and `Tensorflow`. Preprocessing and training were performed on a dedicated simulation node equipped with 56 physical Intel(R) Xeon(R) Platinum 8176 cores with enabled hyperthreading. The computations are structured along three configuration levels. First, the patient group in focus is selected. Second, preprocessing is performed to develop the data sets. Third, model hyperparameters are selected for the training.

1. *Patient selection*. In the base case (see ID 1 in Table 5), we consider all 77k patients. Furthermore, we also present results for subsets of this patient group, that is, elective patients, inpatients, or selected medical specialties.
2. *Data set selection*. In the base case, we consider a memory depth of $\mathcal{D} = \{-20, 10\}$ and one-hot encoded weekdays resulting in $N = 255$ features. However, we also analyze the impact of the memory depth $\mathcal{D}$ and the weekday feature. All samples are shuffled randomly and split into training, validation, and test sets as described in Section 4.3.
3. *Model selection*. We have identified the best hyperparameters for the neural network by running an exhaustive search with `scikit-learn's GridSearchCV`, that is, a "200:50" topology and ReLu activation for the ICU prediction model. Each model has been trained for up to $N_{\text{epochs}} = 100\,000$ epochs using stochastic gradient descent with a constant learning rate $\alpha = 0.00001$ and MSE as loss function. Furthermore, we also present results for different hyperparameters.

After preprocessing (levels 1-2) and training (level 3), the model is deployed to predict the bed occupancy for the samples in the test data set. Figure 5 compares the predicted ICU bed occupancy level with the measured one. The ICU is most critical since it is most expensive, represents an important bottleneck in the hospital, bears the risk of blocking ORs, and finally fosters inferior patient treatment (see Section 1). On the entire test data set, the model achieves a root mean squared error (RMSE) of 3.46. The overall occupancy level is predicted correctly and also major peaks are reflected well, for example, during weekends. Nevertheless, the two curves slightly differ from each other. This seems reasonable since even for two samples with exactly the same realization of all features, the bed occupancy level might be still different due to individual patient characteristics, uncertain medical conditions, and individual behavior and decision making of the medical staff. These characteristics make it difficult for the model to predict exactly the same curve. Albeit the ICU is most crucial, our model is able to predict occupancy levels for other supporting units as well. Balancing the occupancy level, that is, reducing the maximum number of required beds, is a common goal in MSS optimization. In Section 5.6, we formulate an optimization problem for the respective bed occupancy level in the ICU. In an extensive numerical study, different combinations of patient selection, data set selection,

TABLE 5 Numerical results for ICU occupancy prediction based on neural networks

	Patient selection					Data set selection		Model selection		Test
ID	Years	Type	Urgency	Spec.	LOS	$\mathcal{D}$	WD	Label	Topology	RMSE
1	2010-2016	In./out.	El./ur.	8	All	$\{-20, 10\}$	Yes	ICU	200:50	3.46
2	2010-2016	In./out.	**El.**	8	All	$\{-20, 10\}$	Yes	ICU	200:50	2.97
3	2010-2016	In./out.	**El./ur./em.**	8	All	$\{-20, 10\}$	Yes	ICU	200:50	4.27
4	2010-2016	**Inpatient**	El./ur.	8	All	$\{-20, 10\}$	Yes	ICU	200:50	3.46
5	**2016**	In./out.	El./ur.	8	All	$\{-20, 10\}$	Yes	ICU	200:50	3.76
6	2010-2016	In./out.	El./ur.	**4**	All	$\{-20, 10\}$	Yes	ICU	200:50	2.91
7	2010-2016	In./out.	El./ur.	8	**52**	{-20,10}	Yes	ICU	200:50	3.44
8	2010-2016	In./out.	El./ur.	8	**20**	$\{-20, 10\}$	Yes	ICU	200:50	2.78
9	2010-2016	In./out.	El./ur.	8	All	$\boldsymbol{\{-20, 0\}}$	Yes	ICU	200:50	3.49
10	2010-2016	In./out.	El./ur.	8	All	$\boldsymbol{\{-50, 0\}}$	Yes	ICU	200:50	3.55
11	2010-2016	In./out.	El./ur.	8	All	$\boldsymbol{\{-30, 10\}}$	Yes	ICU	200:50	3.51
12	2010-2016	In./out.	El./ur.	8	All	$\{-20, 10\}$	**No**	ICU	200:50	3.59
13	2010-2016	In./out.	El./ur.	8	All	$\{-20, 10\}$	Yes	ICU	**4:4**	3.53
14	2010-2016	In./out.	El./ur.	8	All	$\{-20, 10\}$	Yes	ICU	**200:200:200**	3.47

Abbreviations: el., elective; em, emergency; in., inpatient; out., outpatient; ur., urgent; WD, weekday feature.

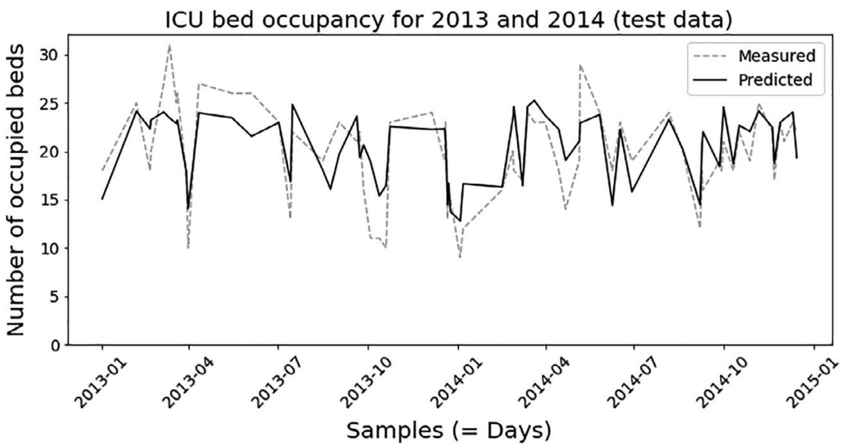

FIGURE 5 Measured and predicted bed occupancy level in the ICU for the test samples within 2013 to 2014

and model selection were computed. Table 5 summarizes our major findings. We have analyzed the impact of different patient selections (ID1 to ID8), various data sets (ID1 and ID9 to ID12), and different model hyperparameters (ID1 and ID13 to ID14). The first row (ID1) in Table 5 represents the base case for this numerical study, that is, the patient selection containing 77k patients. The data set is composed of OR block feature vectors with memory depth $\mathcal{D} = \{-20, 10\}$ and the weekday feature vectors (WD). The model is defined as neural network with topology "200:50", a constant learning rate of $\alpha = 0.00001$, and stochastic gradient descent as optimizer. This base case yields a test error of 3.46 (RMSE). The bold values in each row indicate changes compared to the base case (ID1). Limiting the patient selection to elective patients only (ID2) results in a reduced test error of 2.97 (-14%), while expanding the selection to emergency patients (ID3) yields an increased test error of 4.27 (+23%). As discussed in Section 5.1, it is not recommended to include emergency patients, but we consider elective and urgent patients. Training the network with only 1 year of data results in fewer samples and, hence, an increased test error (ID5). Considering only patients of four specialties, that is, GYN, CAS, NEU, and URO, reduces the test error to 2.91 (ID6). Since long-term patients are rather difficult to predict on a tactical level and have a significant impact on the occupancy level, a better performance for patient selections with a lower maximal LOS is observed (ID7 to ID8). Varying the memory depth in ID9 to ID11 results in slightly increased test errors. Furthermore, the weekday feature in the base case is responsible for a 4%-improvement compared to ID12. Also, varying the topology of the neural network in ID13 to ID14 results in lower performance. We draw four conclusions from the numerical study. First, the selection of the patient group has a major impact on the prediction accuracy. In particular, considering less specialties, only elective patients, or lower maximal LOS results in lower prediction errors. Second, choosing the memory depth wisely improves the prediction results. Third, including the weekday feature achieves

TABLE 6 Numerical results for ICU occupancy prediction based on alternative models

	Patient selection					Data set selection		Model selection		Test
ID	Years	Type	Urg.	Spec.	LOS	D	WD	Label	Model	RMSE
1	2010-2016	In./out.	El./ur.	8	All	{−20, 10}	Yes	ICU	NN (200:50)	3.46
15	2010-2016	In./out.	El./ur.	8	All	{−20, 10}	Yes	ICU	**SVM**	3.72
16	2010-2016	In./out.	El./ur.	8	All	{−20, 10}	Yes	ICU	**KNN**	3.83
17	2010-2016	In./out.	El./ur.	8	All	{−20, 10}	Yes	ICU	**DTR**	5.17
18	2010-2016	In./out.	El./ur.	8	All	{−20, 10}	Yes	ICU	**RFR**	3.73

Abbreviations: el., elective; in., inpatient; out., outpatient; Spec., specialty; ur., urgent; Urg., urgency; WD, weekday feature.

in general better predictions. Fourth, a neural network with topology "200:50" achieved best results.

5.4 | Comparison of predictive power

In order to better contrast the performance of our model to the literature, we compare it with alternative models. First, we keep the preprocessing (levels 1-2) as proposed in this study and only vary the machine learning algorithm (level 3). Second, we compare the entire model with the state-of-the-art model presented in Section 2.

For the first comparison, we implemented various alternative algorithms, trained them on the data sets resulting from the preprocessing process, and evaluated their performance on the test set. The implementation was done using `scikit-learn's SVR`, `KNeighborsRegressor`, `DecisionTreeRegressor`, and `RandomForestRegressor`, respectively. Table 6 summarizes our findings for this numerical comparison. The support vector machine in ID 15 achieves a rather good performance (RMSE of 3.72) being only slightly worse than the proposed neural network (ID 1). Also, the k-nearest neighbors algorithm (ID 16) shows a convincing performance. The decision tree regressor (DTR) in ID 17 achieves a RMSE of 5.17. Using the random forest regressor (RFR) in ID 18 achieves a better RMSE of 3.73. In summary, alternative machine learning algorithms (except for DTR) are also well-suited for the prediction problem at hand, however, do not outperform the proposed neural network (ID 1). Furthermore, even though the presented algorithms could be used as substitute for the neural network in the prediction step, they still rely on the preprocessing steps of our approach.

In the second comparison, we consequently compare our entire model with a state-of-the-art prediction model that does not require our preprocessing. We implemented the stochastic analytical approach proposed by Fügener et al. (2014) using Mathwork's MATLAB[6] and Python. According to the notation by Fügener et al. (2014), $a_j(p)$ reflects the probability that p patients had surgery during one OR block of medical specialty j, b_j represents the probability that a patient of medical specialty j is transferred to the ICU after surgery, and $1 - b_j$ is the probability of a transfer to the ward. $c_j^{WO}(n)$, $c_j^{I}(n)$, and $c_j^{WI}(n)$ represent the probabilities that a patient of medical specialty j stays n days in the ward after surgery, in the ICU after surgery, or in the ward after being released from the ICU, respectively. $d_{j,n}^{I}(n)$ represents the conditional probability that a patient of medical specialty j is transferred from the ICU to the ward on day n, given that he/she was not released before. Likewise, $d_{j,n}^{WO}(n)$, and $d_{j,n}^{WI}(n)$ refer to the probabilities of a discharge from the ward on day n after surgery or after the transfer from the ICU, respectively. The aforementioned parameters were calculated based on the entire data set of Universitätsklinikum Augsburg for the years 2010 to 2016, that is, the base case containing all 77k patients. Due to technical reasons, we even included the test samples into the calculations which results in an advantage compared to our model which has been trained without the test data . Given those parameters, the model is able to predict the resulting ICU bed occupancy for the OR blocks. Figure 6 compares the original bed occupancy (dashed, gray) with the predictions based on our model (solid, black) and the one by Fügener et al. (2014) (dotted, gray). While both models predict the overall level and the trend pretty well, our model is closer to the original data for almost all dates. For the samples depicted in Figure 6, Fügener et al. achieve an RMSE of 5.14 while our model achieves an RMSE of 3.44. On the entire test set, Fügener et al. achieve a RMSE of 6.06. In comparison, our model achieves an prediction error that is 43% lower (RMSE of 3.46).

Given the presented comparisons with alternative algorithms and the state-of-the-art prediction model, we conclude that machine learning is a suitable method to predict bed occupancy levels and that the proposed neural network based approach achieves convincing results. We have shown that our suggested three-step method is suitable for predicting, can accommodate different machine learning algorithms in step 3, and outperforms the current state-of-the-art where a different approach is used for the same task.

5.5 | Training with bootstrapped ex ante data

The numerical results presented in this study are obtained by a prediction model that was trained on ex post data, that is, data on the actual surgeries performed rather than the planned

[6]https://www.mathworks.com/products/matlab.

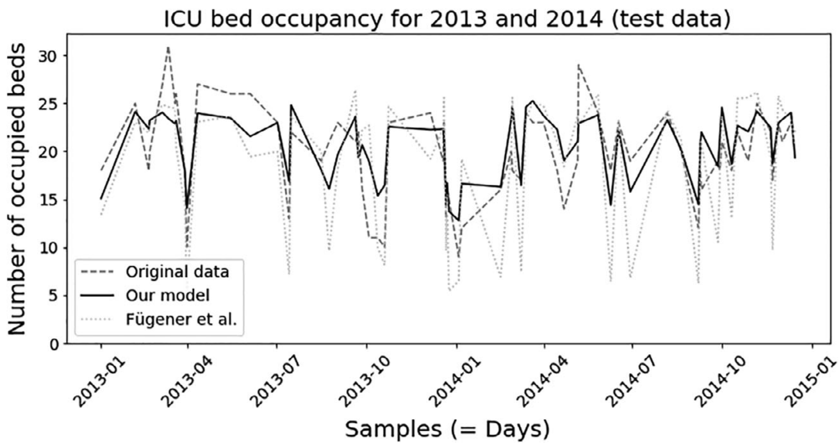

FIGURE 6 Measured ICU bed occupancy compared with the predictions based on A, our model (RMSE: 3.44) and B, the stochastic analytical model (RMSE: 5.14)

surgery schedule. However, a prediction model trained on ex ante surgery schedules would be most useful. Otherwise, it is not clear how the prediction of the ex post behavior will help with evaluating schedules made ex ante. In this section, we address this aspect and show what one would do with ex ante data. First, a data set with bootstrapped ex ante data is generated. Second, the prediction model is trained on the new data set. Third, the prediction results are compared to the aforementioned results obtained with ex post data.

Due to the limited availability of ex ante data in the required granularity for the reference hospital, the data set is generated by bootstrapping ex ante data from the given ex post data set. The same $P = 77$k patients are considered and the same assumptions apply as in the base case described in Section 5.1. We reflect the canceled surgeries in the new data set by amending the ex post data with additional surgeries. To avoid duplicates, each new observation is composed as perturbed combination of existing observations. In fact, we randomly select the characteristics such as medical specialty, OR, date, and duration from the ex post data set. As the procedure durations are drawn from the ex post data set, the model is still learning from the future to some extent. To mitigate this limitation, one might develop a forecasting model for the duration of surgeries in order to bootstrap the ex-ante data. A simple approach could use the historical average duration of all past surgeries or more specifically for the same type of procedure. Following the same procedure as described in Section 4.2, the ex ante surgery schedule is obtained from the bootstrapped data set, that is, the allocation of OR blocks (r, m) for all medical specialties.

Given the bootstrapped ex ante surgery schedule, we create a data set ($\boldsymbol{y}$, $\boldsymbol{X}$) using the same data set selection as in the base case (see Section 4.2), that is, a memory depth of $\mathcal{D} = \{-20, 10\}$ and one-hot encoded weekdays. Finally, the data set is used to train the neural network with a "200:500" topology as in the base case.

Figure 7 compares the original ICU bed occupancy (dashed, gray) with the predictions of our model based on ex post data (solid, black) as well as bootstrapped ex ante data (dotted, black) for a cancellation rate of 10%. While both predictions are quite accurate, the model based on ex post data is slightly closer to the original data for most observations. In fact, the prediction error for the bootstrapped ex ante data (RMSE of 3.60) is 4% higher than for the ex post data (RMSE of 3.46). This seems reasonable since the bootstrapped ex ante data suffer from the additionally introduced uncertainty in form of canceled surgeries. Clearly, the more canceled surgeries reflected in the bootstrapped ex ante data, the higher the prediction error. Table 7 summarizes our findings for a numerical comparison of different cancellation rates. At the reference hospital, we currently observe a cancellation rate of 14% for elective ICU patients, which can reach up to 30% in intense times. This is reflected in our calculations.

We conclude that training the model with ex ante data results in slightly less accurate results than with ex post data. Nevertheless, the performance difference is only minor and the proposed prediction model also yields valuable results if trained with potential ex ante data.

5.6 | Application of prediction model in decision making process

Among other possible applications, the proposed prediction model will be most beneficial for the evaluation of a given MSS. Since the benefits of a new prediction model comes not only from the improvement in prediction quality but also from the improvement in the quality of the decision that the model informs, we present two options to incorporate it into a decision making process in order to inform a better decision. First, in regular discussion rounds of the hospital management, in which the actual MSS is discussed and modified, our prediction model serves as valuable tool to adaptively evaluate a given, feasible MSS with respect to the expected bed occupancy levels in the ICU. Second, the prediction model can be incorporated as the objective function in an optimization model such as a genetic algorithm (GA).

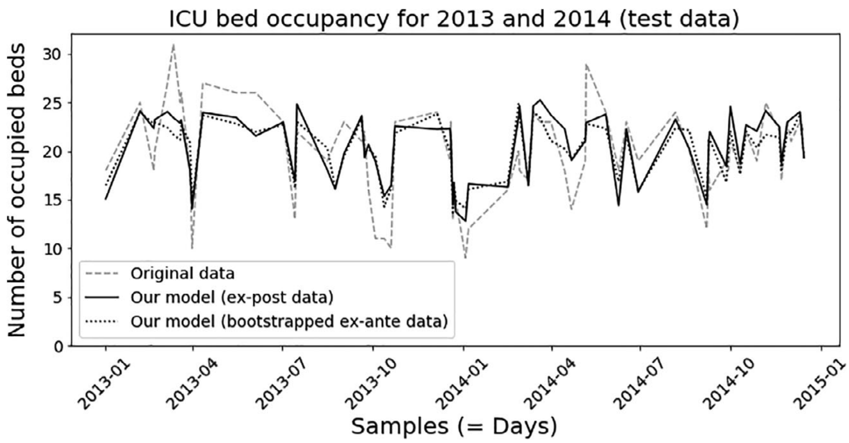

FIGURE 7 Measured ICU bed occupancy compared with the predictions of our model based on A, ex post data and B, bootstrapped ex-ante data for a cancellation rate of 10%

TABLE 7 Numerical results for ICU occupancy prediction based on bootstrapped ex-ante data with different cancellation rates

ID	Patient selection					Ex ante	Data set selection		Model selection		Test
	Years	Type	Urg.	Spec.	LOS	Cancel.	$\mathcal{D}$	WD	Label	Model	RMSE
1	2010-2016	In./out.	El./ur.	8	All	0%	{−20, 10}	Yes	ICU	NN (200:50)	3.46
9	2010-2016	In./out.	El./ur.	8	All	**5%**	{−20, 10}	Yes	ICU	NN (200:50)	3.49
20	2010-2016	In./out.	El./ur.	8	All	**10%**	{−20, 10}	Yes	ICU	NN (200:50)	3.60
21	2010-2016	In./out.	El./ur.	8	All	**30%**	{−20, 10}	Yes	ICU	NN (200:50)	3.64

Abbreviations: Cancel., cancellation rate; el., elective; in., inpatient; out., outpatient; Spec., specialty; ur., urgent; Urg., urgency; WD, weekday feature.

TABLE 8 MSS A and MSS B. Number of OR blocks assigned to each specialty per weekday

$j \in \mathcal{J}$	Mon	Tue	Wed	Thu	Fri	Sat	Sun
(a) MSS A (based on status quo)							
CAS	3	3	3	3	3	0	0
GES	2	3	3	4	4	0	0
GYN	3	3	2	1	2	0	0
MAS	1	0	1	0	0	0	0
NEU	2	2	1	2	2	0	0
TRA	4	4	5	6	4	0	0
URO	1	1	1	1	1	0	0
VAS	2	2	2	1	2	0	0
(b) MSS B (based on discussion with hospital management)							
CAS	2	5	5	3	0	0	0
GES	1	1	4	5	5	0	0
GYN	0	2	2	1	6	0	0
MAS	1	0	0	1	0	0	0
NEU	6	3	0	0	0	0	0
TRA	5	6	5	5	2	0	0
URO	3	1	1	0	0	0	0
VAS	0	0	1	3	5	0	0

MSS A depicted in Table 8a is based on the status quo in the reference hospital. Each row shows the number of OR blocks that are assigned to the respective medical specialty per weekday. For example, GES obtained $d_{\text{GES}} = 16$ weekly OR blocks from strategic planning and uses $\xi_{\text{Mon, GES}} = 2$ rooms each Monday. In total, $R = 18$ surgery rooms are used on each weekday by eight medical specialties. This ensures sufficient capacity for emergency patients and patients of other specialties. Furthermore, the maximum number of assigned OR blocks is $b = 6$.

First, we used the prediction model for the evaluation of a sequence of MSSs suggested by the hospital management

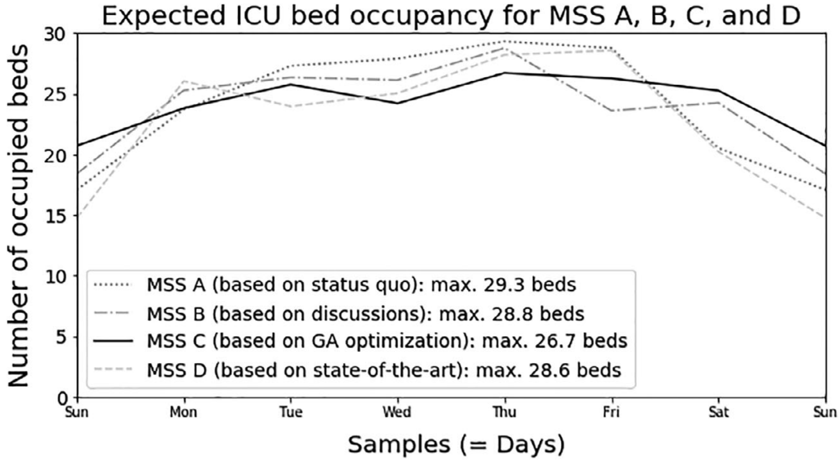

FIGURE 8 Predicted ICU bed occupancy corresponding to MSS A (based on status quo), MSS B (based on discussions with hospital management), MSS C (based on GA optimization), and MSS D (based on state-of-the-art model (Fügener et al., 2014))

with the aim to avoid peaks in the ICU occupancy. In MSS B (see Table 8b), every medical specialty is provided with the same number of OR slots per week as in MSS A. Moreover, the total number of utilized rooms per day remains unchanged, that is, 18 ORs per weekday. The resulting bed occupancy levels for both MSS are depicted in Figure 8.

Second, the prediction model can as well be incorporated as objective function in an optimization model. However, due to the nonlinearity of our model, this integration is not trivial. Being one of the possibilities, we used metaheuristics to incorporate the prediction model. Nonlinearity is introduced in the neural network by the nonlinear activation function $g(z)$ in each neuron.

As motivated in Section 1, one of the most important evaluation criteria for a MSS is the impact on the ICU, particularly the maximum number of occupied beds. Hence, Equation (12) describes a reasonable objective function for a MSS ξ:

$$c(\xi) = \max(\boldsymbol{y}(\xi)), \tag{12}$$

where the vector $\boldsymbol{y}$ denotes the number of occupied beds per day. The max-operator is defined to operate component-wise. The maximum weekly bed occupancy level was chosen since it determines the number of beds that need to be provided in the ICU. Hence, it directly impacts the associated costs for beds and required nursing staff. This is also commonly used in the literature, for example, Beliën and Demeulemeester (2007) and Fügener et al. (2014). The goal of our work is to conduct the same surgical program while reducing the maximum number of required ICU beds per week. In fact, the model does not require any predetermined capacity nor target utilization levels, but rather addresses the shortage of ICU beds which is often caused by a shortage of nursing staff. One of the most critical challenges in German hospitals is to find, recruit, and retain qualified nursing staff. The situation has further intensified in 2019 with the introduction of the German regulation for the threshold of nursing staff.[7] We use a rather simple model to show how the proposed prediction model can be integrated in an optimization framework. A valid MSS needs to meet the total weekly number of OR blocks T_j for each specialty $j \in \mathcal{J}$ obtained from strategic planning and cannot exceed the total available OR capacity R per day. Consequently, we define the problem to minimize the costs $c(\xi)$ resulting from a MSS ξ as in (13):

$$\begin{aligned} \min_{\xi} \quad & c(\xi) \\ \text{s.t.} \quad & \sum_{j \in \mathcal{J}} \xi_{e,j} \leq R, \quad e \in \mathcal{E} \\ & \sum_{e \in \mathcal{E}} \xi_{e,j} \geq T_j, \quad j \in \mathcal{J} \\ & \xi_{e,j} \leq b_{e,j}, \quad e \in \mathcal{E}, j \in \mathcal{J} \\ & \xi_{e,j} \in \{0, \dots, R\}, \quad e \in \mathcal{E}, j \in \mathcal{J} \end{aligned} \tag{13}$$

.To consider availability of staff, rooms, and equipment it might be reasonable to force the MSS to a maximum number of OR blocks $b_{e,j}$ per specialty and weekday. Since $\boldsymbol{y}$ is unknown, we use the approximated bed occupancy $\hat{\boldsymbol{y}}(\xi)$ that is derived by our prediction model.

We obtain a real-valued, multivariate, nonlinear, nondifferentiable objective function that cannot be solved with classic optimization methods such as stochastic gradient descent, quasi-newton methods, or integer programming. Instead, we use a GA that does not require derivatives (Goldberg, 1989). Being an evolutionary algorithm, GA is inspired by different mechanisms present in nature, such as mutation, recombination, and selection. GA optimizes a problem by iteratively creating new candidate solutions based on an existing population and keeping the ones with the best performance. We use the GA provided by MATLAB's Global Optimization Toolbox and ensure feasibility of the solution by incorporating inequality constraints and integers. In future work, one could also develop customized optimization methods for this problem. To find an optimized MSS for the reference hospital, we implemented the aforementioned minimization problem and solved it with GA where the approximated bed

[7] https://www.bundesgesundheitsministerium.de/personaluntergrenzen.

TABLE 9 MSS C and MSS D. Number of OR blocks assigned to each specialty per weekday

$j \in \mathcal{J}$	Mon	Tue	Wed	Thu	Fri	Sat	Sun
(a) MSS C (based on GA optimization)							
CAS	0	6	6	3	0	0	0
GES	0	0	4	6	6	0	0
GYN	0	2	2	1	6	0	0
MAS	1	1	0	0	0	0	0
NEU	6	3	0	0	0	0	0
TRA	6	6	6	5	0	0	0
URO	5	0	0	0	0	0	0
VAS	0	0	0	3	6	0	0
(b) MSS D (based on state-of-the-art model; Fügener et al., 2014)							
CAS	5	2	1	4	3	0	0
GES	3	4	4	2	3	0	0
GYN	1	4	1	2	3	0	0
MAS	0	1	0	1	0	0	0
NEU	3	0	2	1	3	0	0
TRA	5	3	5	5	5	0	0
URO	1	0	3	1	0	0	0
VAS	0	4	2	2	1	0	0

occupancy $\hat{y}$ is computed by deploying our trained prediction model.

Given those parameters and our prediction model, we find that MSS C in Table 9a achieves a particularly low maximum ICU bed occupancy. Note that MSS C provides the same weekly number of OR blocks to every medical specialty and does not exceed the daily number of used ORs compared to MSS A.

Finally, we compare the performance with an optimization algorithm that has been used in the literature. The optimization approach proposed by Fügener et al. (2014) is well suited since it represents state-of-the-art, has the same objective, that is, ward leveling, and similar constraints to ours. We use the prediction model as well as the optimization model by Fügener et al. (2014). We reached out to the corresponding author of this article to ensure that our implementation is in line with their work. The parameters of the prediction model were calibrated to our reference hospital (see Section 5.4). As a straightforward branch-and-bound algorithm based on complete enumeration is only feasible for very small problem instances, we run the optimization using simulated annealing (Aarts, Korst, & Michiels, 2005) with a geometric cooling schedule as proposed by the authors. A swap of two OR blocks is accepted if it decreases the objective function. Otherwise, it is accepted with a probability that decreases over time. Starting with MSS A (based on status quo) as initial solution, the algorithm finally yields MSS D (based on state-of-the-art model) as depicted in Table 9b.

Figure 8 shows the resulting ICU bed occupancy levels for the four MSS A to D evaluated with our prediction model. As all MSSs are designed for a weekly cycle, also the resulting bed occupancy patterns repeat every week. The dotted curve in gray shows the expected bed occupancy level for the currently implemented MSS A. The maximum occupancy level of 29.3 beds is reached on Thursday and the minimum of 17.1 beds on Sunday. The dashdotted curve in gray shows the expected bed occupancy level if we would implement the MSS B. The maximum occupancy level of 28.8 beds is reached on Thursday and the minimum of 18.4 beds on Sunday. The solid curve in black shows the expected bed demand if we would implement MSS C. One observes that the bed occupancy for MSS C shows a better leveling compared to the one of MSS A and MSS B. The peak on Thursday is reduced to 26.7 beds and the utilization of the ICU beds on Monday and the weekend is increased, that is, to 20.7 beds on Sunday. In summary, we can expect a reduction of the maximum ICU bed demand by 8.9% when implementing the new MSS C. The dashed curve in gray shows the expected bed occupancy level if we would implement MSS D. The maximum occupancy level of 28.6 beds is reached on Friday and the minimum of 14.8 beds on Sunday. Hence, our proposed model performs better on the objective of leveling ICU bed occupancy. This is also what we would expect since our model was developed to optimize over the presented metric that is used for the evaluation (our prediction model) and the model by Fügener et al. uses a different metric.

6 | CONCLUSION

In the paper at hand, a neural network based approach for the integrated OR scheduling problem was presented. We have formulated a model to predict the resulting bed occupancy levels in the ICU for a given MSS. The model reflects more supporting units, patient types, and patient paths than any related work. Furthermore, we have applied the model to a 1700-bed maximum-care hospital located in Southern Germany and showed that our model outperforms a state-of-the-art model by 43% in predicting the ICU occupancy level. To conclude this study, we discuss managerial insights, limitations of our model, and options for future research.

6.1 | Managerial insights

We encourage hospital managers to consider the impact of surgery planning on connected departments. In particular, the ICU is one of a hospital's most expensive resources and an important bottleneck. To avoid blocked ORs and inferior patient treatment, also supporting units should be considered for OR scheduling. The proposed model supports hospital managers to predict the consequences of any modifications to the MSS and to develop better ones.

This study is intended to provide guidance for hospital managers. In case of modifications to the hospital data management system, it might be useful to consider the data sources and parameters presented in this study. We have shown, that surgery and ward records serve as valuable resource for further data processing if linked by an unique patient identifier. Given these hospital records, the path of an individual patient can be reconstructed. Accumulating over all patients provides valuable insights into internal processes, occupancy levels, and bottlenecks within the hospital. Hence, we strongly recommend to include historical data into the decision making process. Albeit most hospital managers already do a good job based on their experience, we see four major advantages of formalizing the OR scheduling process. First, additional information might unveil further optimization potential which has not been identified so far. Second, transparency and consistency are important factors to ensure acceptance of the resulting MSS. Third, the valuable expertise is less concentrated on a single person, easier accessible by colleagues, and preserved for successors. Fourth, dedicated steps of the process can be automated more easily freeing up valuable resources and allowing hospital managers to spend more time on the most critical aspects.

Our data-driven model improves OR scheduling and contributes to make hospitals more efficient. In the reference hospital, we expect to reduce the peak ICU bed demand by 8.9%. In hospitals that have a less sophisticated MSS or more distinct specialties, the savings potential might be even higher. On the other side, we might also experience a lower potential for hospitals with a more sophisticated MSS or less distinct specialties. We believe that the presented prediction model serves as valuable resource to support hospital managers in developing a MSS that increases the efficiency in the supporting units. Increased efficiency, reduced peak bed demands, and less surgery cancellations contribute to a safer patient stay.

6.2 | Limitations

The prediction model achieves convincing results for the reference hospital. In the following, we still discuss some potential shortcomings of the proposed model. This study is limited to ex post data since no ex ante data are available for the reference hospital at hand. We added Section 5.5 to demonstrate what one would do with ex ante data. The accuracy of the prediction depends on the amount and variety of samples that are shown to the model during the training process, and hence, the predictive power might be impacted in case of very different MSS patterns. However, we expect only a rather small impact since the training data already reflect high volatility of the MSS during the 7 years. Furthermore, these shortcomings might be mitigated by periodically re-training the model with up-to-date samples. Furthermore, besides the day of the week, there might be additional confounding factors to hospital data that have an impact on the prediction accuracy. For example, there might be operational differences between summer and winter. While changes to the number of OR slots obtained from strategic planning are already directly reflected in the model, the typical impact resulting from one OR block of the same specialty on the same weekday might still be different, which is difficult to capture. However, such confounding factors can be counteracted by introducing additional features. This is also the reason why the introduction of weekday features leads to improved prediction results. For this study, we do not expect unobserved confounding resulting from infrequent visits of surgeons since physicians in Germany—unlike sometimes in the United States (O'Neill & Hartz, 2012)—are directly employed by their respective hospitals. Hence, the German health care system relies on teams of surgeons that are replaceable which also allows to counteract in case of unavailability of OR staff due to sick leaves. The authors' choice of error metric is informed by discussions with the hospital management considering robustness of the model as most important. As a consequence of this choice outliers are not well reflected and the predicted values tend to have a bias towards the mean. This limits the model in making predictions for high congestion periods. One could think of alternative error metrics such as a weighted MSE with more weight on high congestion periods. However, it is difficult to define high congestion periods since actual capacity limits are not considered and an additional parameter for the threshold would be introduced. Research has shown that high utilization levels negatively affect clinical outcomes and patient safety (Kuntz, Mennicken, & Scholtes, 2014). The proposed model, however, might struggle to fully reflect those effects since the real-time bed capacity is not explicitly used as input. As discussed in Section 6.3, we believe that merging operational and tactical planning might be beneficial in many aspects, for example, one could also integrate up-to-date bed occupancy levels in the scheduling process. Finally, we are also aware that the objective function used for the MSS evaluation in the numerical study does not necessarily fit the individual settings of all hospitals. In particular, staff availability might be less flexible in other hospitals, for example, in case of physicians with admitting privileges at multiple hospitals. However, this can be individually adjusted by changing the objective function or the respective parameters in the linear constraints. We are currently working on the introduction of the revised MSS to Universitätsklinikum Augsburg which allows us to evaluate the realized performance of the revised MSS in a future study. During the implementation of a new MSS, also the reasons for cancellations of surgeries might provide valuable information.

6.3 | Future work

The proposed model is tailored to the specific conditions of the integrated OR scheduling problem. However, it is the merit of machine learning that the model can easily be enriched with additional features. Instead of limiting the allocation of OR blocks to medical specialties, one could rather think of smaller patient groups or even individual patients. This would allow to include additional features such as patient type, urgency, diagnostic group, number of previous surgeries, gender, and patient age. Ultimately, we believe that a joint consideration of tactical and operational decision levels would be beneficial. Even today, the allocation of OR blocks to medical specialties could already be amended by additional recommendations, for example, dedicated OR blocks that are mainly intended for the treatment of patients that belong to a specific diagnostic group or show a high probability for being transferred to the ICU.

It is our intention to encourage further research on the intersection between health care operations management and machine learning. In a fully integrated model, also emergency patients should be considered. Neural network based models can be applied to predict their arrival times based on environmental factors such as weather and traffic. Furthermore, surgery duration can be predicted given medical records and information about surgical staff and equipment. Operations research techniques are well suited to determine optimal surgery schedules whereas machine learning is powerful for predictions that can be integrated as constraints into the optimization model. Research on the intersection of those two disciplines will support hospital managers to make OR scheduling even more efficient in the future.

ACKNOWLEDGMENTS

We thank the Institute for Communications Engineering of the Technical University of Munich for using their computational resources while training the neural network model. We thank the AE and the referees for their help and continuing guidance which helped us to make the manuscript much stronger.

ORCID

Julian Schiele https://orcid.org/0000-0002-4118-2889
Jens O. Brunner https://orcid.org/0000-0002-2700-4795

REFERENCES

Aarts, E., Korst, J., & Michiels, W. (2005). *Simulated annealing*. In E. K. Burke & G. Kendall (Eds.), Search methodologies: Introductory tutorials in optimization and decision support techniques (pp. 187–211). New York: Springer.

Adan, I. J. B. F., & Vissers, J. M. H. (2002). Patient mix optimisation in hospital admission planning: A case study. *International Journal of Operations and Production Management*, *22*(4), 445–461.

Anderson, D., Price, C., Golden, B., Jank, W., & Wasil, E. (2011). Examining the discharge practices of surgeons at a large medical center. *Health Care Management Science*, *14*(4), 338–347.

Baker, D. R., Pronovost, P. J., Morlock, L. L., Geocadin, R. G., & Holzmueller, C. G. (2009). Patient flow variability and unplanned readmissions to an intensive care unit. *Critical Care Medicine*, *37*(11), 2882–2887.

Bartel, A. P., Chan, C. W., & Kim, S.-H. (2017). *Should hospitals keep their patients longer? The role of inpatient care in reducing post-discharge mortality*. Working paper.

Batun, S., Denton, B. T., Huschka, T. R., & Schaefer, A. J. (2011). Operating room pooling and parallel surgery processing under uncertainty. *INFORMS Journal on Computing*, *23*(2), 220–237.

Beliën, J., & Demeulemeester, E. (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, *1185–1204*(01), 176.

Berwick, D. M., & Hackbarth, A. D. (2012). Eliminating waste in US health care. *JAMA*, *307*(14), 1513–1516.

Biswas, M. A. R., Robinson, M. D., & Fumo, N. (2016). Prediction of residential building energy consumption: A neural network approach. *Energy*, *117*, 84–92.

Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P. E., Verrill, C., … Lundin, J. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific Reports*, *8*(1), 3395.

Cardoen, B., Demeulemeester, E., & Beliën, J. (2009). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, *201*(3):921–932.

Carter, M., Hans, E. W., & Kolisch, R. (2012). Health care operations management. *OR Spectrum*, *34*, 315–317.

Centers for Medicare & Medicaid Services. (2016). National health expenditures 2016. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata.

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, *5*, 8869–8879.

Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W., & Edelson, D. P. (2016). Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine*, *44*(2), 368–374.

Dexter, F., Maxbauer, T., Stout, C., Archbold, L., & Epstein, R. H. (2014). Relative influence on total cancelled operating room time from patients who are inpatients or outpatients preoperatively. *Anesthesia and Analgesia*, *118*, 1072–1080.

Dexter, F., & Traub, R. D. (2002). How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesthesia and Analgesia*, *94*, 933–942.

Esteban, C., Staeck, O., Baier, S., Yang, Y., & Tresp, V. (2016). *Predicting clinical events by combining static and dynamic information using recurrent neural networks*. In 2016 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 93–101. United States). IEEE Computer Society.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*, 115–118.

Eurostat. (2018). *Healthcare expenditure statistics*. https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_expenditure_statistics.

Fairley, M., Scheinker, D., & Brandeau, M. L. (2018). Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health Care Management Science*, *22*, 756–767.

Ferrand, Y. B., Magazine, M. J., & Rao, U. S. (2014). Managing operating room efficiency and responsiveness for emergency and elective surgeries – A literature survey. *IIE Transactions on Healthcare Systems Engineering*, *4*, 49–64.

French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network. *Journal of Hydrology*, *137*(1), 1–31.

Fügener, A. (2015). An integrated strategic and tactical master surgery scheduling approach with stochastic resource demand. *Journal of Business Logistics*, *36*(4), 374–387, 11.

Fügener, A., Edenharter, G., Kiefer, P., Mayr, U., Schiele, J., Steiner, F., … Blobner, M. (2015). Improving ICU and ward utilization by adapting master surgery schedules: A case study. *Anesthesia and Analgesia Case Reports*, *6*(6), 172–180.

Fügener, A., Hans, E. W., Kolisch, R., Kortbeek, N., & Vanberkel, P. T. (2014). Master surgery scheduling with consideration of multiple downstream units. *European Journal of Operational Research*, *239*(1), 227–236.

Goldberg, D. E. (1989). Genetic algorithms in search, optimization and machine learning (1st ed.). Boston, MA: Addison-Wesley Longman.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. Cambridge, Massachusetts: MIT Press http://www.deeplearningbook.org

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

Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, *240*, 98–109.

Gupta, D. (2007). Surgical suites' operations management. *Production and Operations Management*, *16*, 689–700.

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

Halpern, N. A., & Pastores, S. M. (2010). Critical care medicine in the United States 2000–2005: An analysis of bed numbers, occupancy rates, payer mix, and costs. *Critical Care Medicine*, *38*(1), 65–71.

Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, *69*, S36–S40.

Hans, E. W., Wullink, G., Van Houdenhoven, M., & Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, *185*, 1038–1050.

Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. *IEEE First Annual International Conference on Neural Networks*, *3*, 11–13.

Hong, J. C., Niedzwiecki, D., Palta, M., & Tenenbaum, J. D. (2018). Predicting emergency visits and hospital admissions during radiation and chemoradiation: An internally validated pretreatment machine learning algorithm. *JCO Clinical Cancer Informatics*, *2*, 1–11.

Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLoS One*, *13*(7), 1–13.

Hsu, V. N., de Matta, R., & Lee, C. Y. (2003). Scheduling patients in an ambulatory surgical center. *Naval Research Logistics*, *50*, 218–238.

Hulshof, P., Vanberkel, P., Boucherie, R. J., Hans, E. W., van Houdenhoven, M., & van Ommeren, J. C. (2012). Analytical models to determine room requirements in outpatient clinics. *OR Spectrum*, *34*(2), 391–405.

Jahn, M. (2018). Artificial neural network regression models: Predicting GDP growth. *HWWI Research*.

Kim, K., & Mehrotra, S. (2015). A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Operations Research*, *63*(6), 1431–1451.

Kolmogorov, A. N. (1957). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, *114*, 953–956.

Kuntz, L., Mennicken, R., & Scholtes, S. (2014). Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science*, *61*(4), 707–929.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

Leung Patrick Cheung, B., Dahl, D., & Las Vegas, N. V. (2018). *Deep learning from electronic medical records using attention-based cross-modal convolutional neural networks*. In IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, (pp. 222–225), https://doi.10.1109/BHI.2018.8333409.

Li, F., Gupta, D., & Potthoff, S. J. (2016). Improving operating room schedules. *Health Care Management Science*, *19*(3), 261–278.

Li, J., Huang, L., Shen, Z., Zhang, Y., Fang, M., Li, B., … Wang, H. (2018). Automatic classification of fetal heart rate based on convolutional neural network. *IEEE Internet of Things Journal*, *6*(2), 1394–1401.

Litvak, N., van Rijsbergen, M., Boucherie, R. J., & van Houdenhoven, M. (2008). Managing the overflow of intensive care patients. *European Journal of Operational Research*, *998*(1010), 998–1010.

Luo, L., Liu, H., Hou, X., & Shi, Y. (2016) Machine learning methods for surgery cancellation. *13th International Conference on Service Systems and Service Management (ICSSSM)* (pp. 1–4).

Marjamaa, R. A., Torkki, P. M., Hirvensalo, E. J., & Kirvelä, O. A. (2009). What is the best workflow for an operating room? A simulation study of five scenarios. *Health Care Management Science*, *12*, 142–146.

Marjamaa, R. A., Vakkuri, A., & Kirvelä, O. A. (2008). Operating room management: Why, how and by whom? *Acta Anaesthesiologica Scandinavia*, *52*, 596–600.

Marques, I., Captivo, M. E., & Pato, M. V. (2012). An integer programming approach to elective surgery scheduling. *OR Spectrum*, *34*(2), 407–427.

Master, N., Zhou, Z., Miller, D., Scheinker, D., Bambos, N., & Glynn, P. (2017). Improving predictions of pediatric surgical durations with supervised learning. *International Journal of Data Science and Analytics*, *4*(1), 35–52.

McManus, M. L., Long, M. C., Cooper, A., Mandell, J., Berwick, D. M., Pagano, M., & Litvak, E. (2003). Variability in surgical caseload and access to intensive care services. *Anesthesiology*, *98*(6), 1491–1496.

Mestre, A. M., Oliveira, M. D., & Barbosa-Póvoa, A. (2012). Organizing hospitals into networks: A hierarchical and multiservice model to define location, supply and referrals in planned hospital systems. *OR Spectrum*, *34*, 319–348.

Natarajan, P., Frenzel, J. C., & Smaltz, D. H. (2017). Demystifying big data and machine learning for healthcare. Boca Raton: CRC Press Taylor & Francis Group.

O'Neill, L., & Hartz, A. J. (2012). Lower mortality rates at cardiac specialty hospitals traceable to healthier patients and to doctors' performing more procedures. *Health Affairs (Project Hope)*, *31*(4), 806–815.

OECD Publishing. (2017). *Health at a glance 2017: OECD indicators*. https://doi.org/10.1787/health_glance-2017-en.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Cournapeau, D. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pendharkar, P., & Nanda, S. (2006). A misclassification cost-minimizing evolutionary-neural classification approach. *Naval Research Logistics (NRL)*, *53*(5), 432–447.

Porter, E. (2013). *A world of rising health care costs*. http://economix.blogs.nytimes.com/2013/06/27/a-world-of-rising-health-care-costs/.

Roshanaei, V., Luong, C., Aleman, D. M., & Urbach, D. R. (2017). Collaborative operating room planning and scheduling. *INFORMS Journal on Computing*, *29*(3), 558–580.

Saghafian, S., Austin, G., & Traub, S. J. (2015). Operations research/management contributions to emergency department patient flow optimization: Review and research prospects. *IIE Transactions on Healthcare Systems Engineering*, *5*(2), 101–123.

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*, 493–525.

Sandbaek, B. E., Helgheim, B. I., Larsen, O. I., & Fasting, S. (2014). Impact of changed management policies on operating room efficiency. *BMC Health Services Research*, *14*(1), 224.

Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS One*, *14*(2), e0212356.

Shi, P., Helm, J., Deglise-Hawkinson, J., & Pan, J. (2019). Timing it right: Balancing inpatient congestion versus readmission risk at discharge. Rochester, NY: SSRN Scholarly Paper, Social Science Research Network.

Shylo, O. V., Prokopyev, O. A., & Schaefer, A. J. (2013). Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing*, *25*(4), 682–692.

Song, H., Tucker, A., Graue, R., Moravick, S., & Yang, J. (2019). Capacity pooling in hospitals: The hidden consequences of off-service placement. Rochester, NY: SSRN Scholarly Paper, Social Science Research Network.

Strum, D. P., May, J. H., & Vargas, L. G. (2000). Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *Anesthesiology*, *92*(4), 1160–1167.

Thuemmler, C., & Bai, C. (2017). Health 4.0: How virtualization and big data are revolutionizing healthcare. Switzerland: Springer International Publishing.

Toque, F., Come, E., Oukhellou, L., & Trepanier, M. (2018). *Short-term multi-step ahead forecasting of railway passenger flows during special events with machine learning methods*. In ASPT 2018, Conference on Advanced Systems in Public Transport and Transit Data, Brisbane, Australia, (p. 15). hal-01891446.

Tuwatananurak, J. P., Zadeh, S., Xu, X., Vacanti, J. A., Fulton, W. R., Ehrenfeld, J. M., & Urman, R. D. (2019). Machine learning can improve estimation of surgical case duration: A pilot study. *Journal of Medical Systems*, *43*(3), 44.

van Oostrum, J. M., Bredenhoff, E., & Hans, E. W. (2010). Suitability and managerial implications of a master surgical scheduling approach. *Annals of Operations Research*, *178*(1), 91–104.

Vanberkel, P., Boucherie, R. J., Hans, E. W., Hurink, J. L., van Lent, W. H. M., & van Harten, W. H. (2011a). Accounting for inpatient wards when developing master surgical schedules. *Anesthesia and Analgesia*, *112*(6), 1472–1479.

Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., & Litvak, N. (2010). A survey of health care models that encompass multiple departments. *International Journal of Health Management and Information*, *1*(1), 37–69.

Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., van Lent, W. A. M., & van Harten, W. H. (2011b). An exact approach for relating recovering surgical patient workload to the master surgery schedule. *Journal of the Operational Research Society*, *62*(10), 1851–1860.

Volkov, M., Hashimoto, D. A., Rosman, G., Meireles, O. R., & Rus, D. (2017). *Machine learning and coresets for automated real- time video segmentation of laparoscopic and robot-assisted surgery*. In 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, pp. 754–759, https://doi.org/10.1109/ICRA.2017.7989093.

Wallach, I., Dzamba, M., & Heifets, A. (2015). *AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery*. arXiv:1510.02855.

How to cite this article: Schiele J, Koperna T, Brunner JO. Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. *Naval Research Logistics* 2021;68:65–88. https://doi.org/10.1002/nav.21929

APPENDIX A: MACHINE LEARNING NOTATION

The notation commonly used in machine learning is listed in Table A1.

TABLE A1 Machine learning notation

Description	Notation
Label	$y^{(m)}$
Feature	$x_n^{(m)}$
Data set	$(\boldsymbol{y}, \boldsymbol{X})$
Sample	$(y^{(m)}, x_1^{(m)}, \dots, x_N^{(m)})$
Sample index	$m \in \mathcal{M} = \{1, \dots, M\}$, where $M = M_{\text{train}} + M_{\text{val}} + M_{\text{test}}$
Feature index	$n \in \mathcal{N} = \{1, \dots, N\}$
Size of training data set	M_{train}
Size of validation data set	M_{val}
Size of test data set	M_{test}
Model	$\hat{\boldsymbol{y}}$
Model parameter	$\boldsymbol{\theta}$
Loss function	$J(\boldsymbol{\theta})$
Learning rate	α
Regularization parameter	λ
Number of epochs	N_{epochs}

APPENDIX B: MEDICAL SPECIALTIES

In this study, the following eight medical specialties of Universitätsklinikum Augsburg were considered.

CAS	Cardiothoracic Surgery
GES	General, Visceral, and Transplant Surgery
GYN	Gynecology
MAS	Oral and Maxillofacial Surgery
NEU	Neurosurgery
TRA	Traumatology, Orthopedics, and Plastic Surgery
URO	Urology
VAS	Vascular and Intravascular Surgery

APPENDIX C: HOSPITAL RECORDS

The hospital records provided by Universitätsklinikum Augsburg serve as foundation for this study. The data set comprises more than 600k patients covering the years 2010 to 2016. Table C1 illustrates the surgery records and Tables C2 to C4 show the records for wards including IMC, ICU, and ED stations.

TABLE C1 Excerpt of surgery records

PAT	Type	Date	Urgency	OR	Specialty	Incision	Suture
510033	Inpatient	1 April 2010 12:05 PM	Elective	10	GES	1 April 2010 12:05 AM	1 April 2010 2:20 PM

TABLE C2 Excerpt of ward admission records

Station	PAT	Specialty	Room	Admission	Discharge	Admission category
313	510033	GES	1067	1 April 2010 10:05 AM	4 April 2010 4:15 PM	Inpatient

TABLE C3 Excerpt of ward transfer records

PAT	Admission	Transfer	Origin station	Origin room	Dest. station	Dest. room
510033	1 April 2010 10:05 AM	1 April 2010 8:15 PM	45	4132	50	5135

TABLE C4 Excerpt of ward discharge records

Station	PAT	Specialty	Room	Admission	Discharge	Discharge category
107	510033	GES	10245	1 April 2010 10:05 AM	4 April 2010 4:15 PM	Regular

APPENDIX D: HISTORIC DATA FOR REFERENCE HOSPITAL

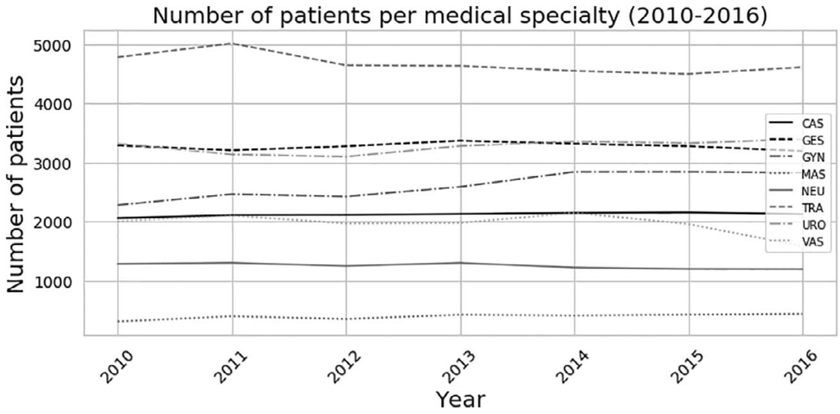

FIGURE D1 Number of patients for selected specialties at reference hospital remained nearly unchanged

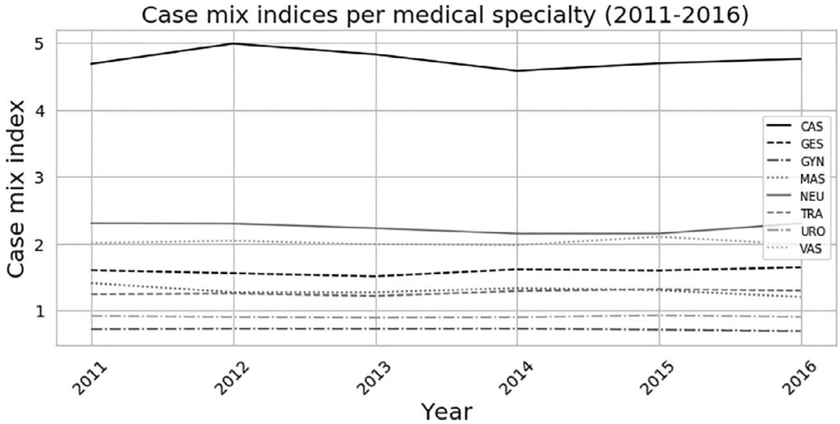

FIGURE D2 Case mix indices for selected specialties at reference hospital remained nearly unchanged