

Predicting the Number of Open Operating Rooms to Enhance Resource Allocation and Patient Quality of Care

Narges Shahraki

shahraki.narges@mayo.edu

Mayo Clinic

Stephen D. Hawley

Mayo Clinic

Ryan M. Schoer

Mayo Clinic

Nageswar R. Madde

Mayo Clinic

Pradeep K. Alam

Mayo Clinic

Travis R. Muller

Mayo Clinic

Aimee L. Tillman

Mayo Clinic

Ellory L. Steinbauer

Mayo Clinic

Matthew N. Vogt

Mayo Clinic

Daryl J. Kor

Mayo Clinic

Research Article

Keywords: Operating rooms, Surgical demand, Resource allocation, Staff planning, Deep learning, Wide-and-deep architecture

Posted Date: July 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10072309/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Predicting the Number of Open Operating Rooms to Enhance Resource Allocation and Patient Quality of Care

Narges Shahraki^{1*}, Stephen D. Hawley¹, Ryan M. Schoer²,
Nageswar R. Madde¹, Pradeep K. Alam³, Travis R. Muller⁴,
Aimee L. Tillman⁵, Ellory L. Steinbauer¹, Matthew N. Vogt⁴,
Daryl J. Kor⁴

^{1*}Center for Digital Health, Mayo Clinic, Rochester, MN, USA.

²Hospital Operations, Mayo Clinic, Rochester, MN, USA.

³Department of Information Technology, Mayo Clinic, Rochester, MN,
USA.

⁴Department of Anesthesiology and Perioperative Medicine, Mayo
Clinic, Rochester, MN, USA.

⁵Administration Project Management Operations, Mayo Clinic,
Jacksonville, FL, USA.

*Corresponding author(s). E-mail(s): shahraki.narges@mayo.edu;

Abstract

Background: The dynamic nature of surgical demand poses significant challenges in planning and scheduling resources to staff the surgical suite. Uncertainty in workload often leads to conservative upstream resource allocation, resulting in staffing misallocation. We propose a novel metric, the daily maximum number of open operating rooms (ORs), as an interpretable resource demand proxy to support staff planning.

Methods: Using historical surgical data from Mayo Clinic in Rochester, Minnesota, from May 5, 2018 through October 8, 2023 (332,904 surgical cases), we developed a deep learning framework based on a wide-and-deep architecture to predict the daily maximum number of open ORs. The model incorporates features capturing historic OR use trends, seasonality effects, and anesthesia division-specific characteristics. Predictions are generated for forecast horizons of up to 90 business days.

Results: Our results demonstrate strong predictive performance, with campus-level root mean square error values ranging from 5 to 7 available ORs across all forecast horizons. Validation across 9 anesthesia divisions confirms the framework’s capability to capture complex, nonlinear dependencies within longitudinal surgical datasets.

Conclusions: This study introduces a novel operational metric and a robust predictive framework for perioperative resource management. By accurately forecasting daily maximum open ORs, health care managers can better align staff capacity with anticipated demand, potentially enhancing staff use and improving the patient care experience by minimizing delays.

Keywords: Operating rooms, Surgical demand, Resource allocation, Staff planning, Deep learning, Wide-and-deep architecture

1 Introduction

Staff scheduling is challenging; it is computationally complex and difficult to achieve due to uncertainty in the workload. There are multiple objectives, constraints, and preferences to be considered and, to make the problem more difficult, scheduling decisions are generally made in multiple stages, with outcomes of the first stage becoming a constraint on the following stages. We investigate the question of how to improve the upstream stages of this process, where staff allocations are made and become constraints to later operational scheduling. Decisions made upstream have a considerable impact on operations, staff well-being, and patient outcomes, and as such tend to be made conservatively. Misallocation is common, resulting in either understaffing or overstaffing.

Improving the performance of upstream resource allocation decisions is a promising avenue to enable managers to: 1) provide sufficient roster capacity; 2) reduce understaffing and the associated costs in overtime and staff burnout; 3) minimize overstaffing; and 4) increase provider satisfaction by granting greater control over their schedules.

A clear factor in the performance of the staff allocation decisions is the uncertainty in demand. In health care, demand is ultimately about patient volume, translating volume to required staff via an easily interpretable metric can aid administrators in their staffing decisions. For example, in anesthesiology, the staff allocation decision can easily be related to the number of operating rooms (ORs) that need coverage. A typical staff ratio could have an anesthesiologist covering 2 ORs but can go as high as 1 per 4 ORs [1]. Minimum staffing requirements can easily be determined knowing the allowed and desired ratio and the number of open ORs. Based on these considerations, we considered a novel metric, daily maximum number of open ORs, and developed a predictive model for use by managers making intermediate-term roster management decisions for scheduled staff in the perioperative setting.

In the following section, we review related work on planning and scheduling, roster management, demand forecasting, and how artificial intelligence/machine learning works in block scheduling and surgery duration prediction. First, we introduce the

challenges related to scheduling of surgical cases, which is the linkage between patients and required staff. Second, we discuss the scheduling literature that addresses assignment of personnel to specific coverage responsibilities (e.g. shifts or specific assigned roles), often called rostering. Third, we review related work on demand forecasting in the surgical context and the opportunity with respect to forecasting coverage requirements. Finally, we discuss related work in artificial intelligence/machine learning for block scheduling and surgery duration prediction.

2 Literature Review

2.1 Planning and Scheduling in the Surgical Setting

Managing and allocating resources in surgical settings, particularly ORs, is challenging due to the highly variable nature of surgical volumes. This complexity is further exacerbated by uncertainties in surgery durations, emergency surgeries, and resource availability. Daily fluctuations and uncertainties in surgical demand and resource availability require scheduling strategies that are both flexible and predictive to ensure optimal resource use [2–4].

Traditional scheduling approaches often result in either overstaffing or understaffing. Overstaffing causes staff idle time and understaffing creates an excessive workload, leading to burnout and affecting morale. Additionally, non-OR days, such as paid time off and research days, are better arranged when the time is granted far enough in advance to organize plans. Optimizing resource allocation in OR environments is crucial for enhancing operational efficiency and staff satisfaction.

For healthcare administrators, investments in schedule optimization aim to meet the most patient needs possible within available healthcare resources. Typically, administrators break the problem into multiple decision stages across time and functional areas of scheduling, reflecting its scale and complexity. OR scheduling focuses on sequencing surgical cases across physical spaces [5, 6]. In contrast, staff scheduling, or rostering, involves allocating personnel to tasks, such as assigning anesthesiologists to cover operating rooms, and is the focus of this paper.

2.2 Roster Management

Scheduling is also called *rostering* when assigning staff to work locations and shifts over a time period of several weeks to a few months. Rostering in health care has long been studied, with one of the first applications to receive significant attention being nursing shifts, which came to be known as the Nurse Rostering Problem [7–11]. More recently, a variant of the problem for anesthesia providers has been studied, labeled the Anesthetist Rostering Problem [12, 13].

In their review of general staff scheduling, Ernst et al. [14] developed a flexible classification scheme defining 6 subproblems composing the rostering problem. These subproblems are 1) demand modeling, 2) days off scheduling, 3) shift scheduling, 4) line of work construction, 5) task assignment, and 6) staff assignment. Our work specifically addresses demand modeling, which remains an underexplored area within the larger staff scheduling problem.

In their recent review of the physician scheduling problem, Erhard et al. [15] highlight the need for a rigorous investigation of realistic physician demand in the literature. They note that in most works reviewed demand uncertainty is neglected, and instead, demand is a deterministic input parameter. Planners and managers need advanced insight of the upcoming demand so they can grant staff requests for paid time off, research time, or other nonclinical activities, while maintaining an adequate pool of staff for schedulers to meet the needs of the practice as the case volume becomes better known.

Many tools exist for solving these medium-term staff allocation problems. However, their effectiveness in practice rests on the accuracy of demand estimates. As a direct input to the scheduling process, high-quality demand forecasting is a critical complementary discipline that enables planners to align staffing to case volume.

2.3 Demand Forecasting, Block Scheduling, and Surgery Duration Prediction

Demand forecasting in health care has gained increasing attention in recent years, particularly within surgical and perioperative settings, as hospitals seek to better plan and allocate clinical and surgical staff. For example, Hassanzadeh et al. [16] focused on predicting daily surgical case volumes. Using historical data and machine learning methods, they demonstrated how hospitals can anticipate daily workload and allocate resources more effectively. Building on this line of work, Singh et al. [17] developed a machine learning pipeline that combined forecasts of surgical volumes with predicted patient length of stay to estimate hospital bed use. This integration highlights the interconnected nature of perioperative demand forecasting, where surgical case volumes not only affect OR scheduling but also downstream inpatient capacity planning and staffing.

Related studies examine block scheduling and tactical OR capacity planning and forecasting, where “blocks” refer to preassigned allocations of OR time to services or surgeons [6, 18–21]. While block forecasts and allocations are valuable for planning, they are not synonymous with the daily number of open ORs, which reflect actual operational capacity and must account for the uncertainty in surgical demand and case duration. This distinction is important: blocks represent planned allocations of OR time, given to surgical services, specialties, or individual surgeons, whereas open ORs represent the expected concurrent rooms to cover patient care demand. Predicting open ORs provides a direct basis for estimating the number of surgical staff needed, shifting the focus from preassigning staff to specific rooms toward dynamically aligning staffing with actual operational needs. Despite extensive literature on surgical volume and block allocation, explicit prediction of daily open ORs is largely absent. Addressing this gap, we introduce daily maximum number of open ORs as an operationally interpretable metric and develop predictive models to support intermediate-term roster management decisions in perioperative care.

Another area of research focuses on predicting surgical case durations, which has direct implications for staffing, as well as for reducing overtime and procedural delays. For example, Jiao et al. [22] developed probabilistic models to forecast surgical case durations, emphasizing the importance of capturing uncertainty in procedure length

rather than relying solely on point estimates. Their findings demonstrated that modeling probability distributions can improve OR scheduling and mitigate downstream delays. Later, Jiao et al. [23] applied a neural network model to predict surgical durations, demonstrating that deep learning approaches can further decrease overtime, reduce surgical expenses, and inform staffing allocation schedulers. Similarly, Abbou et al. [24] evaluated a gradient boosting model to predict surgery durations based on patient demographics, surgery type, and surgeon experience, showing reductions in surgery delays, improved OR use, and support for staffing decisions. Earlier work by Devi et al. [25] proposed a framework for predicting surgery time in ophthalmology using features such as surgeon and anesthetist experience, staff tenure, and anesthesia type.

Overall, these studies primarily aimed to improve surgical scheduling and staffing by more accurately predicting procedure durations and optimizing case sequencing. However, they do not address the need to forecast surgical staffing requirements well in advance. This limitation stems from the fact that detailed information about scheduled surgical cases—necessary for predicting procedure times and sequencing—is typically available only shortly before the surgery date.

2.4 Study Contributions

In this study, we propose a predictive modeling framework to estimate the daily maximum number of open ORs, with explicit consideration of variability in surgical demand. Our work contributes to the literature on perioperative operations management by introducing a novel operational metric, daily maximum number of open ORs, and developing machine learning models to predict this measure.

The proposed metric, daily maximum number of open ORs, quantifies the peak concurrent use of ORs within a given day, thereby capturing the point of highest resource intensity. Unlike traditional measures, such as scheduled block allocations or aggregate surgical volume, this metric directly reflects the operational requirement most relevant for staffing decisions. By predicting this metric, we provide a principled basis for aligning staff capacity with anticipated demand, with implications for both efficiency and workforce sustainability.

To the best of our knowledge, no prior studies have explicitly modeled or predicted the daily maximum number of open ORs to inform staffing decisions. We address this gap by applying deep neural networks to predict the daily maximum number of open ORs up to 90 business days in advance. Deep learning is particularly well-suited to this task, as it accommodates complex, nonlinear relationships in longitudinal surgical datasets. Our models are trained on approximately 5 years of data from Mayo Clinic’s campus in Rochester, Minnesota, and generate predictions at both anesthesia division and aggregate (hospital-wide) levels. Although the empirical application is specific to Mayo Clinic, the underlying conceptual framework and methodologic approach are generalizable across health care systems and surgical contexts.

3 Data

This study uses historic surgical data collected from May 5, 2018 through October 8, 2023, comprising a total of 332,904 surgical anesthesia cases. To reduce the influence of atypical scheduling patterns, data from weekends, holidays, and the COVID-19 period (March 1, 2020 to May 31, 2020) were excluded from the analysis. The dataset is divided into a training set covering 1028 surgical days, a validation set covering 254 surgical days, and a test set covering 257 surgical days. The objective is to predict the daily maximum number of concurrently open ORs within individual anesthesia divisions, up to 90 business days in advance of the scheduled surgery date. To support this prediction task, various features were engineered from the dataset, as outlined in Table 1.

Table 1: List of Features

Feature	Values
Number of listings	Calculated for Windows 1, 2, 3, ..., 90
Number of unique surgeons	Calculated for Windows 1, 2, 3, ..., 90
Avg, min, max, and SD of open ORs	Calculated for Windows 1, 2, 3, ..., 90 over past 5, 10, 20, 30 days
Anesthesia division	Division 1 to 9
Day of the week	Monday / Tuesday / Wednesday / Thursday / Friday
Surgical day code	Blue or Orange
Week number	1 to 53
Day number	1 to 31
Month number	1 to 12
Pre/post holidays	0 or 1
Spring break	0 or 1

Abbreviations: Avg, average; max, maximum; min, minimum; ORs, operating rooms.

The number of surgical listings and the number of unique surgeons are the most important features and are calculated over a 90-day window (Windows 1 to 90), where Window 1 represents the previous business day before surgery and Window 90 represents 90 business days before the surgery day. For example, the number of listings in Window 1 reflects the total surgical listings up to 1 day before the surgery, and the number of unique surgeons in Window 1 represents the count of surgeons scheduled based on those listings. The number of cases listed generally begins to appear 91 days before the surgery date, with most cases being scheduled within 30 days of surgery.

In addition to these core features, we have engineered new features to capture surgical demand trends. These include the average, minimum, maximum, and SD of the daily maximum number of open ORs over the previous 5, 10, 20, and 30 days. By incorporating these features, the model better reflects recent shifts in demand.

We have also incorporated features that account for seasonal and temporal effects. These include the day of the week, surgical day color code, week number, day in

month number, month number, and indicators for pre/post holidays and spring break. The surgical day code used at Mayo Clinic to assign different specialties to perform surgeries on different days is labeled as blue or orange days. The pre/post-holiday feature is a binary variable set to 1 if the surgery occurs the day before or after a holiday and 0 otherwise. Similarly, the spring break feature is set to 1 if the surgery falls within the spring break period and 0 otherwise. These features enhance the model’s ability to capture seasonality, providing a better context for fluctuations in surgical demand.

Anesthesia division is also a key factor in our model. Each anesthesia division faces unique surgical demands and performs different types of surgeries, directly impacting the number of open ORs. At Mayo Clinic in Rochester, the surgical department is divided into 9 main anesthesia divisions (eg, cardiovascular surgery, neurologic surgery, pediatric surgery).

Using these features, we have built a robust dataset based on historic data from May 5, 2018 through October 8, 2023. For each surgery date, the dataset includes the corresponding feature values and the label for that day, representing the daily maximum number of open ORs. This dataset is used to train the AI model to capture the complex relationships between the features and the label. To ensure the model’s robustness and accuracy, we develop a continuous and streamlined data pipeline that maintains data consistency across all stages of the model’s lifecycle—spanning training, validation, testing, and prediction. This pipeline automates the entire data flow, from raw data collection to feature engineering, ensuring that the model always receives high-quality, up-to-date information.

4 Methodology

In this section, we present the architecture of our model, detail the hyperparameters tuned during training, and describe the procedures for training and evaluation in this study.

4.1 Model Architecture: Wide and Deep Network

We formulate a deep neural network with a deep and wide architecture. By combining the wide component with the deep component, we aim to leverage the strengths of both components and build a robust and accurate model. Here we describe the 4 elements of the network: 1) wide, 2) deep, 3) input layer, and 4) output layer.

4.1.1 Wide Component

The wide component is crucial for safely integrating all categorical features into our model. Our dataset contains various categorical features with multiple levels, like day numbers with 31 levels and week numbers with 53 levels, which can result in a sparse input dataset after one-hot encoding. If we only use the deep component of the network, we may lose the original input information by branching into a sparse dataset. Incorporating a wide component helps to safely consider all categorical features in the model. Through the wide component, the inputs are passed to the output layer with

specific weights, allowing the model to remember important details [26]. Given the memorization abilities, we also categorize and include the most important numeric features in the wide component. Additionally, we create feature crossings and incorporate them into this component. A list of all the features incorporated into the wide component includes 1) temporal and seasonal markers such as week, day, and month numbers, day of the week, surgical day code, and indicators for holidays and spring break; 2) categorical identifiers like anesthesia division; 3) interaction terms, specifically crosses of month and day numbers, day and week numbers, and day of the week with surgical day code; and 4) discretized numeric features, including the number of listings and unique surgeons partitioned into 20 buckets. The choice of 20 buckets is arbitrary and can be tuned as a hyperparameter during model optimization.

4.1.2 Deep Component

The deep component allows the network to learn complex interactions and relationships between features. Deep neural networks have been shown to be effective in capturing nonlinear relationships and extracting high-level representations from data [26]. The deep component of our network consists of multiple dense layers with Rectified Linear Unit activation, followed by dropout regularization to prevent overfitting. This regularization technique enhances the model's robustness and generalization capabilities. We input all the numerical features along with selected categorical features to the deep component. The features integrated into the deep component consist of: 1) primary numerical inputs, specifically the number of listings and unique surgeons; 2) trend-based features, such as the rolling average of open ORs over the previous 30 days; and 3) key categorical variables, including anesthesia division, day of the week, surgical day code, pre/postholiday indicators, and week numbers.

4.1.3 Input Layers

All features, categorical and numeric, are fed into the input layer. We've normalized the numerical features and used Keras preprocessing layers to one-hot encode the categorical features through the input layers. Additionally, through the input layer, we categorize the number of listings and number of unique surgeons and create feature crossings.

4.1.4 Output Layer

The output layer generates the final prediction and consists of a single neuron. In our case, we use a regression task with Rectified Linear Unit activation to predict our target.

4.2 Hyperparameters Tuning

In this study, we use Bayesian optimization to tune a comprehensive set of hyperparameters, including learning rate, depth, width, dropout rate, and L2 regularization rate. These hyperparameters play crucial roles in influencing the model's performance. The learning rate governs the adjustment rate of model parameters during training, while depth and width determine the architecture of the deep component. The

dropout rate helps prevent overfitting by randomly dropping out neurons during training. Finally, L2 regularization rate specifies the strength of L2 regularization applied to the model's weights to further mitigate overfitting.

4.3 Training Procedure and Model Evaluation

For model training, we employ the Adam optimizer with a mean squared error (MSE) loss function and root mean square error (RMSE) is used to quantify the model's predictive accuracy. To optimize performance, we train the model over 500 epochs. Upon completing the training process, the model with the lowest RMSE is identified as the best-performing model and saved for future use.

This rigorous training and evaluation procedure ensures that the deployed model possesses strong predictive capabilities and can effectively generalize unseen data. The selection of RMSE as the evaluation metric aligns with the objective of minimizing the difference between predicted and actual values, providing a reliable measure of the model's accuracy.

4.4 Task Modeling and Division Grouping

In this study, our goal is to predict the target metric across anesthesia divisions at Mayo Clinic in Rochester. Specifically, we focus on anesthesia divisions distributed across 3 OR sites: Mayo Clinic Hospital – Rochester, Methodist campus (RMC), Mayo Clinic Gonda Building, and Mayo Clinic Hospital – Rochester, Saint Marys campus (SMC). Collectively, these sites operate more than 100 ORs on a typical day, contributing to over 60,000 anesthesia cases performed annually.

To improve prediction accuracy, we have grouped the anesthesia divisions into 5 categories based on shared operational and use patterns. Within each category, fluctuations in total OR use tend to influence one another. By training a separate model for each group, we aim to enhance predictive performance and better capture the unique underlying dynamics of each group.

4.5 Model Deployment

Departmental leaders and operational managers access model predictions through a Tableau dashboard [27]. Our automated pipeline generates daily predictions for the dashboard. We implemented the inference pipeline in custom Python code [28] using public cloud services and open-source software. The system requirements and technical details of the pipeline are described in the following paragraphs.

The system requirements are notable in 2 respects: 1) a low prediction volume and 2) minimal compute and memory requirements per individual prediction. Given the modest scale and latency requirements, our design objective is an operationally efficient deployment. Thus, we developed the inference pipeline as a batch job run on serverless cloud compute. This approach minimizes idle resources and operational support requirements. Our cloud environment is Google Cloud Platform, but similar services are available from other public cloud vendors.

After training, selected models are saved in Hierarchical Data Format 5 and stored in Google Cloud Storage. The inference pipeline uses Google's Python Client library

[29] to load models from Google Cloud Storage and inference data from Google BigQuery.

The batch job runs in a Docker container [30]. The Docker container is based on Python 3.9-slim and configured for compatibility with local security policies. Our Python package is installed from our private Google Artifact Registry. The container build process has been automated for reproducibility. The version of the inference code that is deployed is pinned to a specific version of the project Python package. This allows for updating system dependencies in the container for security while ensuring we use the exact same version of project code for reproducibility. The project code was packaged for Python using PDM and published in a private Google Artifact Registry [31, 32].

The container was deployed as a Google Cloud Run Job [33] with a scheduled execution trigger, daily at 1:30 PM to allow for all upstream ETLs (extract, transform, load operations) to complete. The predictions are written to a BigQuery table, which Tableau queries.

4.6 Model Validation

The released inference code and models were validated prior to deployment and general availability by comparison between predictions made with the original inference script/notebook and the predictions made by the packaged code. Any discrepancies in features, predictions or CIs between the original data science code and the cloud code were analyzed and resolved before deployment.

The model is deterministic and there were a few problems identified in the prediction after translating to the cloud. However, we also provide Confidence Intervals as part of the daily predictions which are generated by random sampling, and these were more interesting. Fixing the random seed allows for reliable comparison with less computation than needed for statistical convergence of the distribution estimate.

As the need arises, any changes to the inference code package are validated using an automated test suite and comparison to historical predictions. Running in deterministic mode is valuable for this aspect. This validation and deployment process is repeated for any newly developed or trained model to ensure that what we deploy is performing as designed.

5 Results

The results of this study are presented and analyzed using RMSE, which measures the average magnitude of the errors between predicted and actual values. Additionally, visualizations of predicted vs. actual values are provided, allowing for a more intuitive understanding of the model's performance. The frequency and distribution of errors are also examined, offering insights into the types of errors the model is most prone to.

5.1 RMSE Analysis Across Windows

The analysis of RMSE shows significant variations across different models. The model for the total for Rochester campus has the highest variability, ranging from 5 to 7

rooms. The models for SMC, RMC and Gonda show fluctuations ranging from 4 to 5 rooms, 2 to 4 rooms, and 1 to 3 rooms respectively. To provide clarity, the plot in Figure 1 illustrates the RMSE for RMC; as the results show, the trend for RMSE is increasing as we move to further windows. As expected, the best results are observed for smaller window sizes.

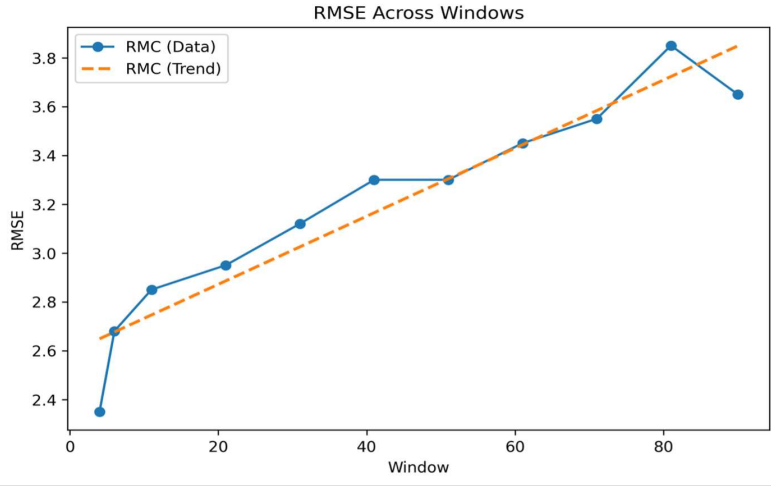


Fig. 1 Root mean square error (RMSE) for Mayo Clinic Hospital – Rochester, Methodist campus (RMC).

5.2 Actual vs Predicted

The daily actual and predicted maximum used ORs are plotted for different windows and anesthesia divisions. The findings show the prediction models can predict daily drops, peaks, and overall trends accurately for most of the windows and divisions. For instance, the plot for RMC for window 4 (Figure 2) illustrates this effectively, where the predicted closely tracks the fluctuations observed in the actual values.

5.3 Error Distribution

For each data point in the test set, the error is calculated as the actual value minus the predicted value. The results show for each division and window, the distribution of errors was approximately normal and centered around zero. For instance, the distribution of errors for RMC and Window 4 is almost symmetrical around zero as shown in Figure 3. The average error was -0.27, with an SD of 2.34, indicating that most errors fell within the range of -2.61 and 2.07.

6 Conclusion

Fluctuating surgical demand poses persistent challenges for perioperative resource management, with direct implications for staffing efficiency, provider well-being, and

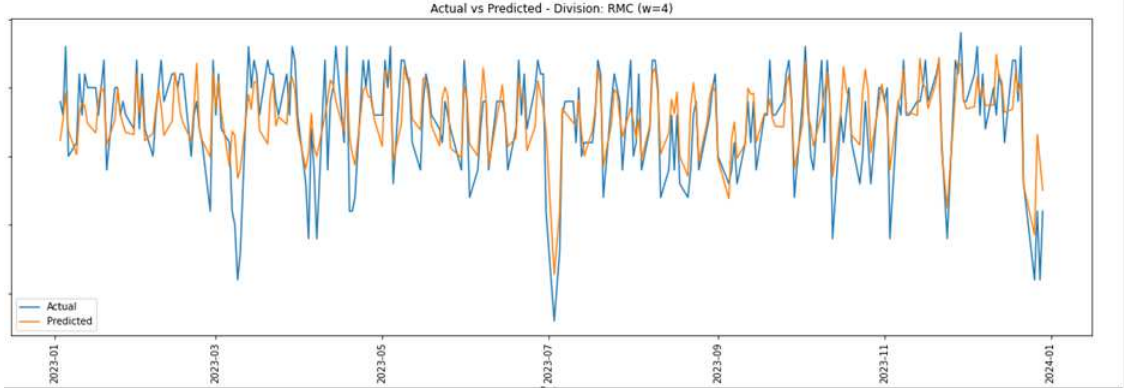


Fig. 2 Actual (blue line) vs. predicted (orange line) daily maximum number of open operating rooms for Mayo Clinic Hospital – Rochester, Methodist campus (RMC) at Window 4.

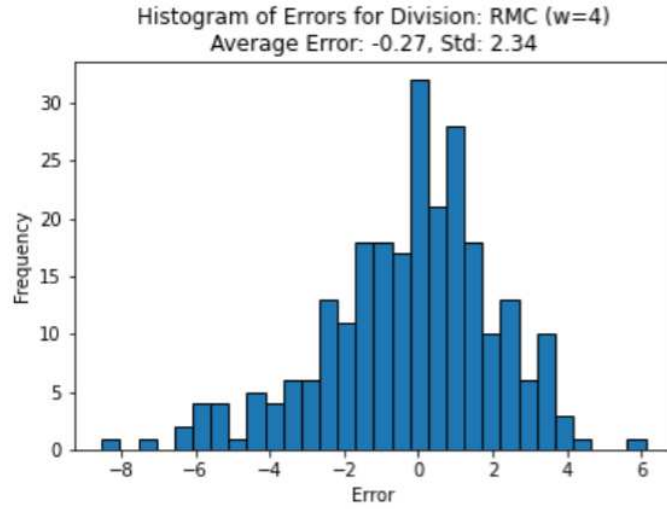


Fig. 3 Frequency and distribution of prediction errors for Mayo Clinic Hospital – Rochester, Methodist campus (RMC) at Window 4.

patient care quality. In this study, we introduce a novel operational demand metric—the daily maximum number of open ORs—and demonstrate that it can be predicted with meaningful accuracy using deep learning models over intermediate- and longer-term forecasting horizons.

Our results show that the proposed models effectively capture daily peaks and drops in OR use across anesthesia divisions, providing actionable estimates of staffing demand well in advance of finalized surgical schedules. Importantly, model performance remains stable as the forecasting horizon extends to several weeks and months, underscoring the robustness of the approach for medium-term roster planning. This

capability is particularly valuable in perioperative environments, where staffing decisions related to paid time off, academic duties, and administrative responsibilities must be made before detailed case-level information becomes available.

These findings extend the existing literature on surgical demand forecasting and perioperative planning in several important ways. Prior studies have largely focused on predicting surgical case volumes [16, 17, 34], block utilization [6, 18–21], or surgical durations [22–25], frequently treating demand as a fixed or deterministic input to downstream scheduling and rostering models [15]. While such approaches have advanced OR scheduling and reduced overtime and delays, they are less well suited to informing upstream staffing allocation decisions that require an interpretable, aggregate measure of concurrent demand. By contrast, the daily maximum number of open ORs directly reflects peak staffing requirements for anesthesia services and other perioperative teams. Accordingly, our work complements and bridges existing research by translating uncertain surgical demand into a staffing-relevant metric that aligns naturally with roster management practices.

From an operational perspective, the ability to anticipate peak OR concurrency offers tangible benefits. More accurate demand estimates can reduce conservative overstaffing while mitigating understaffing that contributes to overtime, burnout, and disruptions in patient care. By improving the quality of demand inputs to rostering systems, the proposed framework supports more balanced staffing decisions, greater provider control over schedules, and improved workforce sustainability.

This study has several important limitations that should be considered. The proposed framework relies on deep neural network models, and although the selected architectures achieved strong and consistent performance across divisions and forecasting windows, further gains may be possible through more extensive architectural exploration and hyperparameter tuning. In practice, however, the modeling space expands rapidly due to the structure of the problem: separate models are trained for each forecasting window and anesthesia division group, resulting in many model configurations. Comprehensive manual tuning across architectures following by hyperparameter tuning is therefore computationally intensive. Consequently, model design choices reflect a balance between predictive performance, scalability, and maintainability rather than exhaustive optimization of individual models.

To address this limitation, future work could focus on developing an automated training and evaluation pipeline capable of systematically exploring multiple model architectures and hyperparameter configurations and identifying best-performing models across divisions and forecasting horizons. In this study, we manually evaluated several model architectures—specifically, determining which features should be fed into the deep and wide components of the neural network—and selected an architecture that performed reasonably well. Automating this process to evaluate combinations of model architectures followed by hyperparameter optimization would enable more efficient model selection at scale, while preserving consistency across the overall modeling framework. Such an approach may further enhance predictive accuracy without sacrificing operational feasibility. Focusing on the refinement of the architecture, Neural Architecture Search methods could be explored and have commercial tooling available [35, 36]. Alternatively, agentic artificial intelligence approaches could be explored, in

which autonomous agents iteratively search for optimal network structures and adapt model architectures over time.

In summary, this study demonstrates that predicting the daily maximum number of open ORs is both feasible and operationally valuable. By introducing a novel demand metric and a scalable deep learning framework, this work advances peri-operative demand modeling and provides a practical foundation for more informed, proactive, and sustainable surgical staffing decisions.

Acknowledgment

The Scientific Publications staff at Mayo Clinic provided copyediting support.

Funding

The authors received no specific funding for this work.

Ethics and Consent to Participate

Not applicable.

References

- [1] Burns, M.L., Saager, L., Cassidy, R.B., Mentz, G., Mashour, G.A., Kheterpal, S.: Association of Anesthesiologist Staffing Ratio With Surgical Patient Morbidity and Mortality. *JAMA Surgery* **157**(9), 807–815 (2022) <https://doi.org/10.1001/jamasurg.2022.2804> . Accessed 2025-07-07
- [2] Guerriero, F., Guido, R.: Operational research in the management of the operating theatre: a survey. *Health Care Management Science* **14**(1), 89–114 (2011) <https://doi.org/10.1007/s10729-010-9143-6> . Number: 1. Accessed 2025-05-02
- [3] Dexter, F., Traub, R.D., Macario, A.: How to release allocated operating room time to increase efficiency: predicting which surgical service will have the most underutilized operating room time. *Anesthesia & Analgesia* **96**(2), 507–512 (2003)
- [4] Dexter, F., Macario, A., Traub, R.D., Hopwood, M., Lubarsky, D.A.: An operating room scheduling strategy to maximize the use of operating room block time: computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time. *Anesthesia & Analgesia* **89**(1), 7–20 (1999)
- [5] Al Amin, M., Baldacci, R., Kayvanfar, V.: A comprehensive review on operating room scheduling and optimization. *Operational Research* **25**(1), 3 (2024) <https://doi.org/10.1007/s12351-024-00884-z> . Number: 1. Accessed 2025-05-01
- [6] Cardoen, B., Demeulemeester, E., Beliën, J.: Operating room planning and scheduling: A literature review. *European Journal of Operational Research*

- 201**(3), 921–932 (2010) <https://doi.org/10.1016/j.ejor.2009.04.011> . Number: 3. Accessed 2025-04-30
- [7] Warner, D.M.: Scheduling Nursing Personnel According to Nursing Preference: A Mathematical Programming Approach. *Operations Research* **24**(5), 842–856 (1976) <https://doi.org/10.1287/opre.24.5.842> . Accessed 2025-07-02
 - [8] Miller, H.E., Pierskalla, W.P., Rath, G.J.: Nurse Scheduling Using Mathematical Programming. *Operations Research* **24**(5), 857–870 (1976). Accessed 2025-07-02
 - [9] Warner, D.M., Prawda, J.: A Mathematical Programming Model for Scheduling Nursing Personnel in a Hospital. *Management Science* **19**(4-part-1), 411–422 (1972) <https://doi.org/10.1287/mnsc.19.4.411> . Accessed 2025-07-02
 - [10] Liebman, J.S., Young, J.P., Bellmore, M.: Allocation of Nursing Personnel in an Extended Care Facility. *Health Services Research* **7**(3), 209–220 (1972). Accessed 2025-07-02
 - [11] Burke, E.K., De Causmaecker, P., Berghe, G.V., Van Landeghem, H.: The State of the Art of Nurse Rostering. *Journal of Scheduling* **7**(6), 441–499 (2004) <https://doi.org/10.1023/B:JOSH.0000046076.75950.0b> . Number: 6. Accessed 2025-05-12
 - [12] Abdullah, N., Ayob, M., Sabar, N.R., Lam, M.C., Yong, L.C.: An On-Call Shift for Anaesthetist Rostering Problem (ARP): A Mathematical Model. In: 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (ICAIAET), pp. 84–89 (2023). <https://doi.org/10.1109/ICAIAET59451.2023.10291773> . <https://ieeexplore.ieee.org/abstract/document/10291773> Accessed 2025-05-12
 - [13] Abdullah, N., Ayob, M., Lam, M.C., Sabar, N.R., Kendall, G., Chian Yong, L.: Advancing Anaesthetist Rostering Quality: A Practical Approach Toward Fairness and Efficiency. *IEEE Access* **13**, 12692–12708 (2025) <https://doi.org/10.1109/ACCESS.2025.3526842> . Accessed 2025-05-08
 - [14] Ernst, A.T., Jiang, H., Krishnamoorthy, M., Sier, D.: Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research* **153**(1), 3–27 (2004) [https://doi.org/10.1016/S0377-2217\(03\)00095-X](https://doi.org/10.1016/S0377-2217(03)00095-X) . Number: 1. Accessed 2025-05-08
 - [15] Erhard, M., Schoenfelder, J., Fügener, A., Brunner, J.O.: State of the art in physician scheduling. *European Journal of Operational Research* **265**(1), 1–18 (2018) <https://doi.org/10.1016/j.ejor.2017.06.037> . Number: 1. Accessed 2025-05-12
 - [16] Hassanzadeh, H., Boyle, J., Khanna, S., Biki, B., Syed, F.: Daily surgery caseload prediction: towards improving operating theatre efficiency. *BMC Medical Informatics and Decision Making* **22**(1), 151 (2022) <https://doi.org/10.1186/>

s12911-022-01893-8 . Accessed 2025-06-11

- [17] Singh, A., Farmer, P.E., Tully, J.L., Waterman, R.S., Gabriel, R.A.: Forecasting Surgical Bed Utilization: Architectural Design of a Machine Learning Pipeline Incorporating Predicted Length of Stay and Surgical Volume. *Journal of Medical Systems* **49**(1), 67 (2025) <https://doi.org/10.1007/s10916-025-02201-3> . Accessed 2025-06-13
- [18] Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., Rademakers, F.E.: Scheduling operating rooms: Achievements, challenges and pitfalls. *Journal of Scheduling* **19**(5), 493–525 (2016) <https://doi.org/10.1007/s10951-016-0489-6> . Accessed 2025-10-03
- [19] Hosseini, N., Taaffe, K.M.: Allocating operating room block time using historical caseload variability. *Health Care Management Science* **18**(4), 419–430 (2015) <https://doi.org/10.1007/s10729-014-9269-z> . Accessed 2025-06-13
- [20] Luo, Y.Y., Wang, B.: A New Method of Block Allocation Used in Two-Stage Operating Rooms Scheduling. *IEEE Access* **7**, 102820–102831 (2019) <https://doi.org/10.1109/ACCESS.2019.2926780> . Accessed 2025-06-13
- [21] M’Hallah, R., Al-Roomi, A.H.: The planning and scheduling of operating rooms: A simulation approach. *Computers & Industrial Engineering* **78**, 235–248 (2014) <https://doi.org/10.1016/j.cie.2014.07.022> . Accessed 2025-06-13
- [22] Jiao, Y., Sharma, A., Ben Abdallah, A., Maddox, T.M., Kannampallil, T.: Probabilistic forecasting of surgical case duration using machine learning: model development and validation. *Journal of the American Medical Informatics Association* **27**(12), 1885–1893 (2020) <https://doi.org/10.1093/jamia/ocaa140> . Accessed 2025-06-13
- [23] Jiao, Y., Xue, B., Lu, C., Avidan, M.S., Kannampallil, T.: Continuous real-time prediction of surgical case duration using a modular artificial neural network. *British Journal of Anaesthesia* **128**(5), 829–837 (2022) <https://doi.org/10.1016/j.bja.2021.12.039> . Accessed 2025-06-11
- [24] Abbou, B., Tal, O., Frenkel, G., Rubin, R., Rappoport, N.: Optimizing Operation Room Utilization—A Prediction Model. *Big Data and Cognitive Computing* **6**(3), 76 (2022) <https://doi.org/10.3390/bdcc6030076> . Number: 3. Accessed 2025-06-11
- [25] Devi, S.P., Rao, K.S., Sangeetha, S.S.: Prediction of Surgery Times and Scheduling of Operation Theaters in Ophthalmology Department. *Journal of Medical Systems* **36**(2), 415–430 (2012) <https://doi.org/10.1007/s10916-010-9486-z> . Accessed 2025-06-11
- [26] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhya, H.,

- Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., Shah, H.: Wide & Deep Learning for Recommender Systems. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. DLRS 2016, pp. 7–10. Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/2988450.2988454> . <https://dl.acm.org/doi/10.1145/2988450.2988454> Accessed 2025-06-13
- [27] Tableau. Salesforce (2024). <https://www.salesforce.com/analytics/tableau/> Accessed 2026-04-02
- [28] Python Software Foundation: Python Programming Language. Python Software Foundation (2020)
- [29] Google: Python Cloud Client Libraries. <https://cloud.google.com/python/docs/reference> Accessed 2025-06-18
- [30] Merkel, D.: Docker: Lightweight Linux Containers for Consistent Development and Deployment | Linux Journal (2014). <https://www.linuxjournal.com/content/docker-lightweight-linux-containers-consistent-development-and-deployment> Accessed 2025-06-18
- [31] Google: Artifact Registry Documentation. Google. <https://cloud.google.com/artifact-registry/docs> Accessed 2025-06-18
- [32] PDM Contributors: PDM Project Documentation. <https://pdm-project.org/en/latest/> Accessed 2025-06-18
- [33] Google: Cloud Run Documentation. Google. <https://docs.cloud.google.com/run/docs> Accessed 2025-06-18
- [34] Tiwari, V., Furman, W.R., Sandberg, W.S.: Predicting Case Volume from the Accumulating Elective Operating Room Schedule Facilitates Staffing Improvements. *Anesthesiology* **121**(1), 171 (2014) <https://doi.org/10.1097/ALN.000000000000287> . Accessed 2025-08-25
- [35] Elsken, T., Metzen, J.H., Hutter, F.: Neural Architecture Search: A Survey. *Journal of Machine Learning Research* **20**(55), 1–21 (2019). Accessed 2026-01-14
- [36] Google: About Vertex AI Neural Architecture Search. <https://docs.cloud.google.com/vertex-ai/docs/training/neural-architecture-search/overview> Accessed 2026-01-14