Use a QR scanner or camera to access more content from SLS

RESEARCH ARTICLE JSLS

Predicting Robotic Hysterectomy Incision Time: Optimizing Surgical Scheduling with Machine Learning

Vaishali Shah, BA, Halley C. Yung, BS, Jie Yang, MPH, MS, Justin Zaslavsky, BA, Gabriela N. Algarroba, MD, Alyssa Pullano, MD, Hannah C. Karpel, MD, MS, Nicole Munoz, Yindalon Aphinyanaphongs, MD, PhD, Mark Saraceni, Paresh Shah, MD, Simon Jones, PhD, Kathy Huang, MD

ABSTRACT

Background and Objectives: Operating rooms (ORs) are critical for hospital revenue and cost management, with utilization efficiency directly affecting financial outcomes. Traditional surgical scheduling often results in suboptimal OR use. We aim to build a machine learning (ML) model to predict incision times for robotic-assisted hysterectomies, enhancing scheduling accuracy and hospital finances.

Methods: A retrospective study was conducted using data from robotic-assisted hysterectomy cases performed between January 2017 and April 2021 across 3 hospitals within a large academic health system. Cases were filtered for surgeries performed by high-volume surgeons and those with an incision time of under 3 hours (n = 2,702). Features influencing incision time were extracted from electronic medical records and used to train 5 ML models (linear ridge regression, random forest, XGBoost, CatBoost, and explainable boosting machine [EBM]). Model performance was evaluated using a dynamic monthly update process and novel metrics such as wait-time blocks and excess-time blocks.

Results: The EBM model was selected for its superior performance compared to the other models. The model reduced the number of excess-time blocks from 1,113 to 905 ($P < .001$, 95% CI [–329 to –89]), translating to approximately 52-hours over the 51-month study period. The model predicted more surgeries within a 15% range of the true incision time compared to traditional methods. Influential features included surgeon experience, number of additional procedures, body mass index (BMI), and uterine size.

Conclusion: The ML model enhanced the prediction of incision times for robotic-assisted hysterectomies, providing a potential solution to reduce OR underutilization and increase surgical throughput and hospital revenue.

Key Words: Efficiency, Gynecologic surgical procedures, Hysterectomy, Machine learning, Operative time, Organizational.

Department of Obstetrics and Gynecology, NYU Langone Health Grossman School of Medicine, New York, New York, USA. (Drs. V. Shah, Munoz, and Huang)

The City University of New York (CUNY) School of Medicine, New York, New York, USA. (Dr. Yung)

Center for Healthcare Innovation & Delivery Science Faculty, Staff & Internal Advisory Board, NYU Langone Health Grossman School of Medicine, New York, New York, USA. (Mr. Yang)

George Washington University School of Medicine and Health Sciences, Washington, DC, USA. (Mr. Zaslavsky)

Department of Obstetrics, Gynecology and Reproductive Sciences, Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvanina, USA. (Dr. Algarroba)

Department of Obstetrics and Gynecology, NYU Grossman School of Medicine New York, New York, USA. (Mr. Pullano)

Department of Obstetrics and Gynecology, NYU Langone Health, New York, New York, USA. (Dr. Karpel)

Department of Population Health, NYU Langone Health Grossman School of Medicine, New York, New York, USA. (Drs. Aphinyanaphongs and Jones)

NYU Langone Hospitals, New York, New York, USA. (Mr. Saraceni)

NewYork-Presbyterian/Weill Cornell Medical Center, New York, New York, USA. (Dr. P. Shah)

Disclosure: none.

Funding: Dr. Yindalon Aphinyanaphongs was partially supported by the National Institutes of Health Grant 3UL1TR001445-05 and the National Science Foundation Award 1928614.

Acknowledgments: Ethical oversight was provided by the New York University (NYU) Langone Institutional Review Board (IRB) (study i21-01419).

Conflict of interests: none.

Address correspondence to: Halley C. Yung, BS, Townsend Harris Hall, 160 Convent Avenue, New York, NY 10031, USA, E-mail: hyung000@citymail.cuny.edu

DOI: 10.4293/JSLS.2024.00040

INTRODUCTION

Surgical departments strive for efficient utilization of operating rooms (ORs), which are the primary contributors to both revenue and cost for most hospitals. ORs can cost between \$21.80 and \$133.12 per minute, depending on case complexity. Cost reduction strategies, such as reducing 7 minutes per case over 250 cases, can yield up to \$100,000 in savings.[1] Recent studies indicate that OR utilization is well below achievable targets at most hospitals.[2] Improving incision time predictions can remedy utilization deficits and allow for more accurate surgical scheduling. In addition to increasing throughput and reducing costs, optimized scheduling practices can facilitate better resource allocation, decreased patient wait-time, and boosts in both patient and staff satisfaction.[1,3,4]

Machine learning (ML), a form of advanced predictive analytics, employs statistical techniques to equip computer systems with the ability to identify patterns from available data variables, otherwise known as features, to make predictions on outcomes of interest.[3,5,6] One study found a ML model significantly increased the number of accurately booked cases from 148 to 219 (34.9% to 51.7%, $P < .001$). Other studies have demonstrated that ML models can also optimize OR resource use as well as wait-time for surgeons, OR staff, and patients.[3,5,7,8]

Few studies on ML models for predicting operative time have focused on robotic-assisted (RA) gynecologic surgery.[3,9] RA surgery may result in higher costs than traditional laparoscopy due to the initial purchase price and ongoing maintenance expenses.[10] A large cohort study showed that the total costs associated with RA hysterectomies were approximately \$2,189 more per case than for its laparoscopic counterpart.[8] Studies external to gynecology show similar trends when comparing procedures performed robotically versus laparoscopically.[10–13] As robotic surgery becomes increasingly more prevalent, boosting OR efficiency becomes a paramount avenue through which surgical services can potentially offset these associated costs. ML models present an opportunity to achieve this optimization and cost reduction, especially when tailored to procedure- and patient-specific features, which have been established to influence incision times.[3,14,15] For instance, in the setting of a hysterectomy, large uteri and adhesions from pelvic inflammatory disease (PID) or cesarean delivery are associated with longer incision times.[16,17] This study aimed to develop and leverage an operation-specific ML model to more accurately predict incision times for RA hysterectomies.

METHODS

Study Population

RA hysterectomy cases performed across 3 hospitals within a large, academic health system between January 2017 and April 2021 were identified (n = 3,058). Cases were filtered to include only those performed by surgeons who completed more than 50 cases within the study period, and with an incision time of less than 3 hours for a total of 2,702 cases included, to be split across training, validation, and test sets. This decision was based on the observation that 90% of hysterectomies in our sample were less than 3 hours long, and cases less than 3 hours are more representative of an uncomplicated RA hysterectomy operation.[1,18] All data for cases were extracted from the available electronic medical record (EMR). This study was reviewed and approved by the Institutional Review Board (IRB). This study received a Waiver of Authorization/Informed Consent from the NYU Langone IRB, as the study presents no more than minimal risk of harm to subjects.

ML Target Prediction

The ML models were trained to predict total incision time for RA hysterectomies. To clarify, the total time for a surgical case consists of preparation time, incision time, and wrap-up time. Incision time, defined as skin closure time minus initial incision time (in minutes), was selected as it is most influenced by patient, operational, and surgeon features compared to preparation time and wrap-up time.

Feature Definition and Extraction

To identify the features for the model, a team of gynecological surgeons and researchers conducted a literature review and compiled a list of characteristics (features) expected to influence total incision time (**Table 1**). Three categories of features were developed: patient-specific features (e.g., body mass index [BMI], age, etc.), surgeon-specific features (e.g., surgeon speed, surgeon, median incision time of most recent surgeries, etc.), and operational features (e.g., number of additional procedures, the presence of trainees in the OR, etc.).

Many of the features of interest were manually extracted from the EMR. Structured data fields existed for some of the characteristics such as age and BMI. Other variables, such as uterine size, presence of trainees during the operation, surgical history, and previous diagnosis history had to be manually searched for within the EMR (i.e. within the imaging report or the provider's history of present illness note). Continuous features, such as age and BMI, were normalized. Categorical features, such as the operating surgeon and patient diagnosis, were encoded as a 1-hot vector.

Table 1.
List of Features Used for Model Building, Variable Type, and Description

Features	Description
Patient Characteristics	
Age (Continuous)	Patient age at the time of surgery
Body mass index (BMI) (continuous)	Patient BMI at the time of surgery
Malignancy (Boolean)	Presence of gynecologic malignancy at the time of surgery determined by whether surgeon was an oncologist
MRI reported pelvic adhesive disease (Boolean)	Presence of pelvic adhesive disease confirmed by MRI
MRI reported uterine size (continuous)	Size of uterus estimated by MRI (calculated using the standard ellipsoid formula: pi/6 × the 3 dimensions)[26]
Prior open abdominal procedure (Boolean)	Patient history of open abdominal surgery
Prior myomectomy (Boolean)	Patient history of a previous myomectomy
Surgeon-Specific Characteristics	
Surgeon speed (continuous)	Median of last 30 surgeries/uterine size
Surgeon (categorical)	The primary surgeon
Diagnosis (categorical)	Indication for hysterectomy
Median of the last 30 incision time (continuous)	Median of the last 30 incision times by surgeon and by specific procedure
Median of the last 10 incision times (continuous)	Median of the last 10 incision times by surgeon and by specific procedure
Median of the last 5 incision times (continuous)	Patient history of a previous myomectomy
Operational Characteristics	
Presence of trainees (Boolean)	Presence of residents or fellows in OR
Presence of physician assistants (Boolean)	Presence of physician assistants in OR
Number of additional procedures (continuous)	Number of additional procedures performed during the hysterectomy case

For some cases, a natural language processing (NLP) program was used to help extract data from the EMR. The NLP program ingests clinical notes and identifies select target phrases consistent with the presence of that feature.[19] The team worked to identify and code these target and skip phrases into an open-source NLP program. For all cases, manual chart extraction was performed to review and validate NLP performance, and the code was updated after each review (i.e., target and skip phrases were updated accordingly to circumvent false positives and further enhance the program's ability to identify data of interest).

Model Descriptions

A total of 5 ML models were evaluated, including linear ridge regression (LR), random forest (RF), XGBoost (XGB), CatBoost (CB), and explainable boosting machine (EBM). The LR model assumes a linear relationship between the features and the predicted outcome. It incorporates L2 regularization to prevent overfitting by constraining the impact of certain features, which allows for fewer contributing features to have near-zero coefficients (less features would have minimal influence). Additionally, no feature is entirely disregarded, as their coefficients never reach zero. In RF models, several "predictor trees" are utilized collectively. Each predictor tree's decisions depend on a randomly selected subset of features (vectors) chosen independently for each tree. This random selection ensures that every predictor tree offers a distinct perspective while adhering to the same guidelines as all the trees in the "forest".[20] XGB model is a "tree-boosting" system that uses an ensemble method, such that each predictor tree is added to the ensemble sequentially, and their predictions are combined to improve the overall performance and predictive power of the model.[21] Similarly, the CB model is a type of gradient boosting algorithm designed to effectively manage categorical features. It implements a novel technique for computing "leaf values" to minimize overfitting, a phenomenon where the model learns too much from the training data, leading to poor performance on new data.[22] Lastly, the EBM model is tree-based, and uses a special method called cyclic gradient boosting. This is designed to detect and understand how different features interact with each other to make predictions.[23] Based on the preliminary performances of the 5 evaluated models, 1 model would be selected as our final model.

Model Development, Validation, and Testing

Our dataset spanned from January 2017 to April 2021. Initially, we observed that creating a single ML model for this entire timeframe led to inconsistent model performance. This inconsistency depended on how we divided the dataset into training, validation, and test sets. For instance, 1 presumed factor contributing to this variability is the occurrence of events that may have impacted the training dataset but not necessarily the validation or test sets. The COVID-19 pandemic serves as a notable example of such an event. To obtain more consistent model performance, we updated our approach by revising the 5 evaluated models monthly using an ensemble learning method. Each month, we trained the models with all preceding months' data and validated it with the current month's data, the latter was split into a validation dataset and a test dataset. The validation set was used to fine-tune and select the best model based on the primary outcome, and the performance of the selected model on the test set was reported. For example, cases from January 2017 were used to train the February 2017 model, and February 2017 cases were subsequently divided into validation and test sets. Similarly, in March 2017, the model utilized data from January and February 2017 for training, while March's cases were divided for validation and then testing. Our dynamic monthly update process only utilizes up to 18 months of prior data, prioritizing the most recent information and trends. Ultimately, this approach yielded 51 dynamic models from February 2017 to April 2021, progressively incorporating the latest data for consistent model performance and accuracy.

Model Evaluation

Model performance was evaluated with a combination of standard and novel metrics. Standard metrics included the following: the percentage of surgeries predicted within a 15% range of the true time in either direction, and the average difference (in minutes) between the true incision time and the model-estimated incision time. Additionally, 2 evaluative metrics were created: wait-time blocks and excess-time blocks. Since ORs in the participating hospitals schedule surgeries in 15-minute increments, we define wait-time blocks as the 15-minute blocks of time that an upcoming case must wait before starting. They are a consequence of underestimating the duration of the preceding case (i.e. the earlier case took longer than anticipated). Conversely, excess-time blocks are the 15-minute time blocks in which an OR is vacant because an earlier case took less time than anticipated, which is a consequence of overestimating the duration of a preceding case. While our study hypothesized a decrease in both wait-time and excess-time blocks, the primary outcome of interest is the reduction in excess-time blocks due to its impact on increasing OR utilization. For all outcome measurements, model performance was compared to baseline (BL), which we defined as the median incision time calculated from the most recent 30 procedures performed by the operating surgeon. This BL measurement reflects a standard practice used across many institutions and incorporated into the software of commercial surgical scheduling systems offered by EMR vendors (e.g., Cerner, Epic, etc.). These systems leverage surgeon-specific historical data to predict incision times.[24]

RESULTS

Demographic Characteristics

During the study period, a total of 3,058 cases of robotic hysterectomy were identified (n = 3,058). As described, cases were filtered based on surgeon caseloads and average incision times, resulting in 2,702 cases included in the final model. The median age of the cohort was 53 years, with 576 cases (21%) being over the age of 65. Of the total cohort, 1,126 cases (42%) have a BMI greater than 30. Trainees were present during 1,995 cases (74% of the operations). The median incision time for the operating surgeon's last 30 procedures was approximately 2 hours (122 minutes). Further details about the cohort can be found in **Table 2**.

Model Selection and Performance

The performances of the 5 tested ML models are compared in **Table 3**. These results were derived from the combination of validation datasets across the 51-month time span during which the monthly dynamic models were developed (n = 1,333). For all the models, the percentage of incision time predictions that fell within a 15% range of the true incision time (in minutes) increased when compared to BL predictions. Similarly, all model predictions had a smaller average time deviation from the true incision time, compared to BL. A decrease in wait-

Table 2.
Demographics and Descriptive for Variables Used in the Models

Features	Descriptives
Median age	53
Age > 65	576 (21.3%)
BMI > 30	1126 (41.7%)
Malignancy	1441 (53.3%)
Presence of trainees (i.e. residents, medical students)	1995 (73.8%)
Presence of physician assistants	2130 (78.8%)
Prior open abdominal procedure	630 (23.3%)
Prior myomectomy	185 (6.8%)
MRI-reported pelvic adhesive disease	204 (7.6%)
MRI-reported uterine size in CC (median)	431.5 (27%)
More than 1 procedure	57%
Number of additional procedures (median)	2
Surgeon (n)	17
Endometrial hyperplasia	748 (27.7%)
Abnormal uterine bleeding	692 (25.6%)
Leiomyoma	415 (15.4%)
Adenomyosis	277 (10.3%)
Pelvic organ prolapse	127 (4.7%)
Median of the last 30 incision time (in minutes)	122
Median of the last 10 incision time (in minutes)	121
Median of the last 5 incision time (in minutes)	118
Day of the Week	
Monday	398 (14.7%)
Tuesday	528 (19.5%)
Wednesday	439 (16.2%)
Thursday	1,047 (38.7%)
Friday	279 (10.3%)
Saturday	8 (0.2%)
Sunday	3 (0.01%)

time blocks (less incision time underestimation) and excess-time blocks (less incision time overestimation) was also observed for all models. We chose the EBM as our final model because it demonstrated the highest reduction (21% decrease) in excess-time blocks compared to the other models evaluated.

Using the combined 51-month test dataset, the EBM model demonstrated a reduction in excess-time blocks, decreasing from 1,113 to 905 (a decrease of 208 blocks, $P < .001$), as outlined in **Table 4**. Considering each time block as 15-minutes, the reduction of 208 excess-time blocks translates to approximately 52 hours.

DISCUSSION

Considering the potential implications for surgical scheduling, our primary outcome measurement focused on reducing excess-time blocks compared to the BL. The rationale behind this is to minimize the underutilization of ORs, allowing for more cases to be performed and subsequently boosting hospital revenue through increased surgical throughput. Based on our findings, the EBM model successfully reduced the number of excess-time blocks compared to BL, highlighting the model's accuracy in predicting incision times. If deployed into practice, our developed ML model can create potential opportunities to schedule additional OR procedures, increasing surgical throughput and revenue.

Excess-time blocks, often the result of overestimation from traditional scheduling practices, represent a critical measure of the operational impact of improved incision time predictions using ML models. On the contrary, operational complications are a major factor contributing to an increase the number of wait-time blocks. Despite accounting for various features influencing incision time with the assistance of ML, complications remain difficult to predict, thus making them less suitable as a primary outcome.

From our test data set of the EBM model, we observed that BL predictions resulted in more wait-time blocks than excess-time blocks (**Table 4**), suggesting an underestimation of the true incision time. This trend was also seen with the model predictions.

Additional Insights for Robotic Hysterectomy Incision Time Prediction

The EBM model can provide explanations for its predictions, revealing how the model itself operates (model level) and how its predictions relate to a patient's case data (patient level). **Figure 1** shows the relative importance of the features we examined in predicting robotic hysterectomy incision time. The most influential features were the median time of a surgeon's most recent operations (of the last 5, 10, and 30 cases), the operating surgeon themselves, the number of the other procedures being performed (procedures other than the robotic hysterectomy), BMI, and uterine size. We noticed that

Table 3.
Performance of the Dynamic Models in the Validation Datasets Combined Across 51 Months (n = 1,333)

Metric	BL	EBM (vs BL)	LR (vs BL)	RF (vs BL)	XGB (vs BL)	CB (vs BL)
Percent of prediction within 15% of true time	31.88	35.78 (+12%)	36.83 (+16%)	34.58 (+8%)	35.56 (+12%)	35.71 (+12%)
Average deviation from true time (in minutes)	39.61	34.11 (−14%)	33.35 (−16%)	35.17 (−11%)	34.41 (−13%)	34.19 (−14%)
Number of wait-time blocks	2437	2174 (−11%)	2098 (−14%)	2237 (−8%)	2190 (−10%)	2176 (−11%)
Number of excess-time blocks	1080	850 (−21%)	867 (−20%)	899 (−17%)	887 (−18%)	867 (−20%)

BL, retrospective baseline; EBM, explainable boosting machine; LR, linear regression; RF, random forest; XGB, XGBoost; CB, CatBoost.

Table 4.
EBM Model Performance in Test Dataset (n = 1,316)

Metric	BL	Model	Model vs BL	95% CI	*P* value[a]
Percent of prediction within 15% of true time	35.94	33.89	−2.1	(−5.8 to 1.7)	.29
Average deviation from true time (in minutes)	24.5	23.1	−1.4	(−3.4 to 1.4)	.33
Number of wait-time blocks	2,310	2,210	−100	(−429 to 119.4)	.07
Number of excess-time blocks	1,113	905	−208	(−329 to −89)	<.001

BL, baseline. *P* values for the difference in percentages were obtained using the χ^2 test. The other rows were via the Wilcoxon test. Confidence intervals were obtained by bootstrapping.

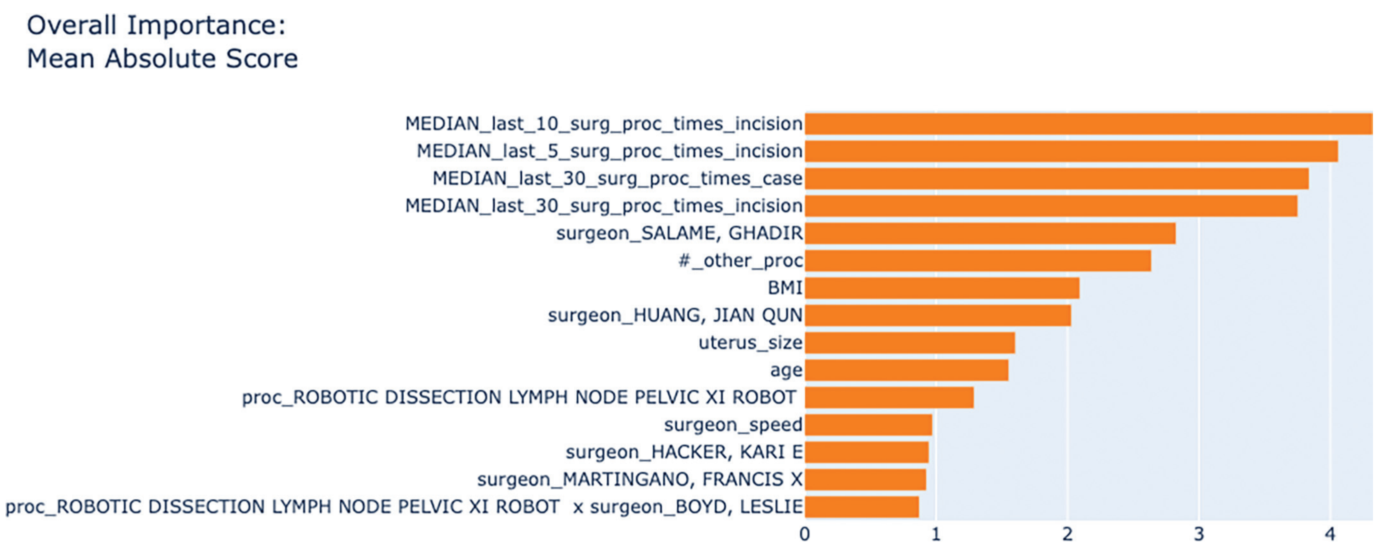

Figure 1. Feature importance in EBM model and mean absolute score. The ordered importance of features incorporated into the final explainable boosting machine (EBM) model are depicted. Surgeon names have been removed from the summary table.

patients requiring multiple procedures tended to lead to longer incision time predictions, as shown in **Figure 2**. Collectively, these results provide insight into the optimization of scheduling. For instance, our model yielded more accurate predictions for cases with multiple procedures compared to standard scheduling practices, which treat each procedure as independent and simply sum their predicted times. Additionally, there was a positive, nonlinear relationship between uterus size and incision time prediction. Although it is well-established that a larger uterus size correlates to a longer incision time, our model analysis revealed more

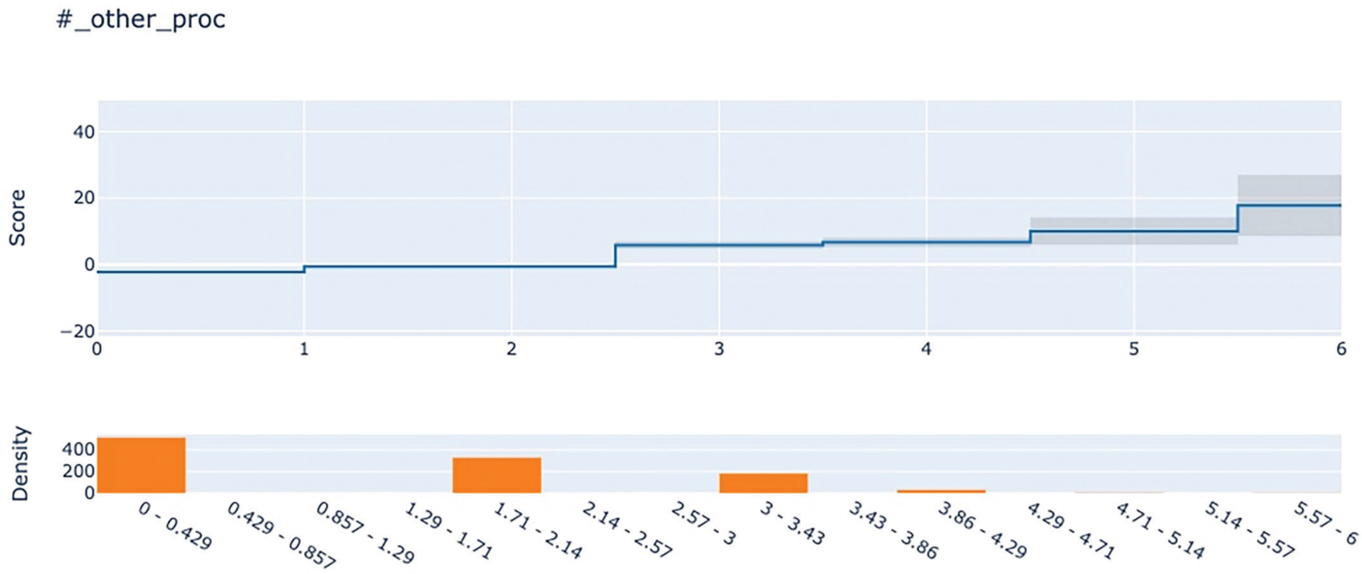

Figure 2. Function of number of other procedures. Demonstrates influence of other procedures in addition to robotic hysterectomy on case time. Score refers to the relative impact on the model prediction; grey blocks indicate error bars. Density refers to the number of cases falling within the specified range.

implications regarding this association. In both literature and billing practice, the threshold for a larger uterus is 250 g, but our data indicates that the most notable change occurs when the uterus size exceeds 343 g.[25,26] Our model predicted that uterus sizes greater than 343 g would result in longer incision times compared to smaller sizes (**Figure 3**). While further investigation is needed to fully interpret this threshold, our results highlight the potential for policy revisions that may better reflect surgical practice.

The benefits of utilizing ML to optimize surgical scheduling extend beyond hospital finances, surgeon scheduling, and patient experiences. More accurate predictions of incision times for robotic hysterectomies can also enhance the efficiency of OR staff allocation, a critical component

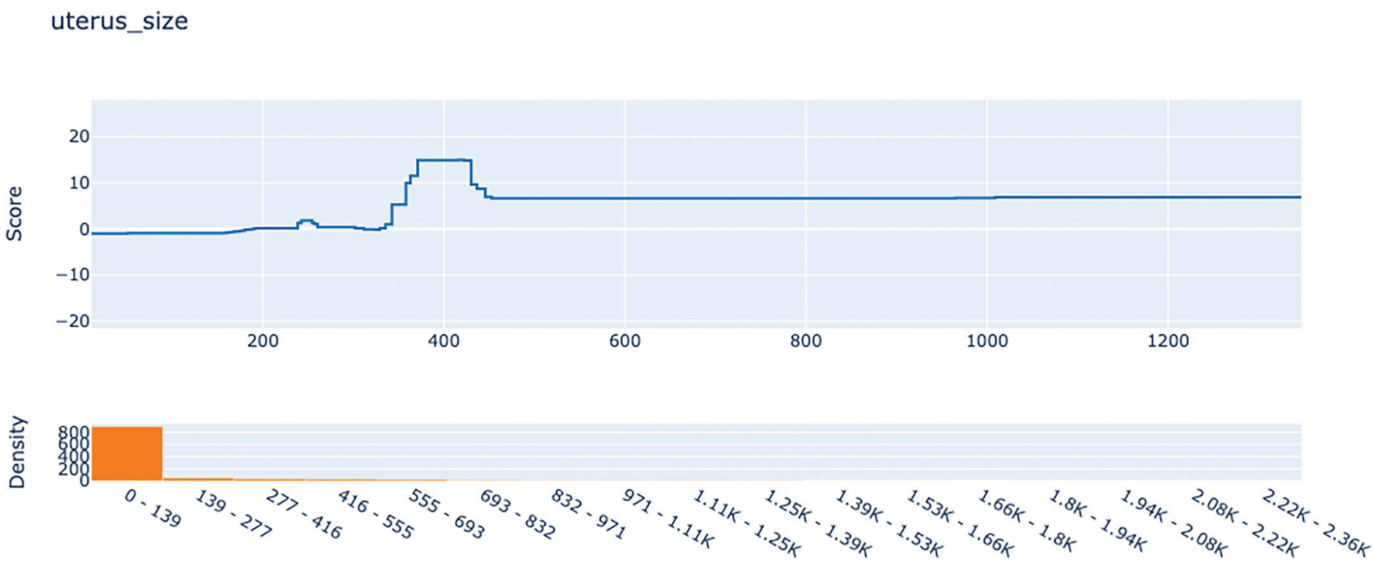

Figure 3. Function of uterine size. Demonstrates influence of uterine size on case time. Score refers to the relative impact on the model prediction; grey blocks indicate error bars. Density refers to the number of cases falling within the specified range.

of successful robotic surgery. Effective scheduling allows for better synchronization between surgeons and other OR staff members, ensuring that the entire surgical team is prepared for subsequent cases without unnecessary delays or idle time.[27]

Additionally, accurate scheduling has a substantial impact on the well-being and job satisfaction of OR nurses. For instance, when surgeries run longer than expected, nurses often face unplanned overtime or extended shifts, leading to fatigue, stress, and lower job satisfaction.[28] Ensuring precise scheduling through ML can alleviate these challenges, contributing to improved staff morale, better work-life balance, and a more efficient workflow for all members of the operating team. Proper planning and allocation of nursing staff are essential not only for seamless operations of robotic surgeries, but also for maintaining a healthy and motivated OR team, which ultimately coincides with more optimal patient care and surgical outcomes.

Limitations

Manual chart review was necessary to pinpoint and extract specific features, such as uterine size, which could not be easily extracted by NLP programs, unlike features with discrete values within the EMR such as age or BMI. Instances where uterine size was not explicitly listed, required manual calculations using the provided dimensions, while in some cases, uterine-related data was entirely absent from the EMR. Given the impact of uterine size on incision time in our model, future efforts should focus on standardizing how this data is listed within EMRs to facilitate a more efficient extraction of variables, whether through NLP programs or manual review. This standardization is especially important for developing similarly structured predictive models that are reliant on EMR data for feature extraction.[29–31]

Moreover, a notable limitation is the exclusion of robotic hysterectomy cases lasting over 3 hours from the dataset utilized for training, validating, and testing our ML model. Numerous factors, such as intraoperative complications, may contribute to prolonged case time, which neither a scheduling team nor our model could reliably anticipate. Further evaluation should involve testing the ML model's capacity to predict cases surpassing the 3-hour threshold. These results may provide further insight into the value of our model's algorithm in forecasting incision times for unforeseen case scenarios, despite initial dataset constraints.

While our model was trained on data from a single cohort of patients and surgeons, utilizing input features specific to our institution, its deployment to external settings is limited by the inherent nature of ML. Nevertheless, our study contributes valuable insights to the literature and highlights the potential of ML models in predicting operating times, thereby advocating for the development of operation-specific models across various institutions.[3,5,8,29] By providing a structured approach and methodology, our study offers guidance for other institutions to adopt and adapt in their own research and integration endeavors.

CONCLUSION

To our understanding, this is the first study to successfully develop a procedure-specific ML model for prediction of RA hysterectomy incision times. Our analysis demonstrated that the ML model substantially outperformed the BL standard by reducing the number of excess-time blocks, potentially translating to the ability to schedule additional procedures, increasing surgical throughput and revenue. We also created 2 novel evaluative metrics, wait-time blocks and excess-time blocks, based on the operational practices of scheduling teams at the institution studied. These metrics facilitate the accurate quantification of model benefits and the associated operational changes in throughput and revenue. The use of these time blocks to evaluate the model's performance is important for the next aspect of this work, the deployment into hospital operations and scheduling. Future applications aim to replace BL practices with ML model predictions to construct surgical schedules, while still analyzing changes based on outcomes of interest. Given the improvement of incision time predictions with our model, we believe implementing a ML program as an OR optimization strategy has promising implications for cost savings, hospital earnings, and improved patient experiences.

References:

1. Shippert RD. A study of time-dependent operating room fees and how to save $100 000 by using time-saving products. *Am J Cosmet Surg.* 2005;22(1):25–34.

2. Doebbeling BN, Burton MM, Wiebke EA, et al. Optimizing perioperative decision making: improved information for clinical workflow planning. *AMIA Annu Symp Proc.* 2012;2012:154–163.

3. Zhao B, Waterman RS, Urman RD, Gabriel RA. A machine learning approach to predicting case duration for robot-assisted surgery. *J Med Syst.* 2019;43(2):32.

4. Macario A. What does one minute of operating room time cost? *J Clin Anesth.* 2010;22(4):233–236.

5. Jiao Y, Sharma A, Ben Abdallah A, Maddox TM, Kannampallil T. Probabilistic forecasting of surgical case duration using machine learning: model development and validation. *J Am Med Inform Assoc.* 2020;27(12):1885–1893.

6. Blum AL, Langley P. Selection of relevant features and examples in machine learning. *Artif Intell.* 1997;97(1–2):245–271.

7. Strömblad CT, Baxter-King RG, Meisami A, et al. Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources: a randomized clinical trial. *JAMA Surg.* 2021;156(4):315–321.

8. Bartek MA, Saxena RC, Solomon S, et al. Improving operating room efficiency: machine learning approach to predict case-time duration. *J Am Coll Surg.* 2019;229(4):346–354.e3.

9. Martinez O, Martinez C, Parra CA, Rugeles S, Suarez DR. Machine learning for surgical time prediction. *Comput Methods Programs Biomed.* 2021;208:106220.

10. Barbash GI, Glied SA. New technology and health care costs–the case of robot-assisted surgery. *N Engl J Med.* 2010;363 (8):701–704.

11. Baek SJ, Kim SH, Cho JS, Shin JW, Kim J. Robotic versus conventional laparoscopic surgery for rectal cancer: a cost analysis from a single institute in Korea. *World J Surg.* 2012;36 (11):2722–2729.

12. Ramsay C, Pickard R, Robertson C, et al. Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol Assess.* 2012;16(41):1–313.

13. Sarlos D, Kots L, Stevanovic N, Schaer G. Robotic hysterectomy versus conventional laparoscopic hysterectomy: outcome and cost analyses of a matched case-control study. *Eur J Obstet Gynecol Reprod Biol.* 2010;150(1):92–96.

14. Traven SA, Reeves RA, Xerogeanes JW, Slone HS. Higher BMI predicts additional surgery at the time of ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc.* 2019;27(8):2552–2557.

15. Consorti F, Milazzo F, Notarangelo M, Scardella L, Antonaci A. Factors influencing the length of the incision and the operating time for total thyroidectomy. *BMC Surg.* 2012;12:15.

16. Hsu WC, Chang WC, Huang SC, Sheu BC, Torng PL, Chang DY. Laparoscopic-assisted vaginal hysterectomy for patients with extensive pelvic adhesions: a strategy to minimise conversion to laparotomy. *Aust N Z J Obstet Gynaecol.* 2007;47(3):230–234.

17. Pogoda KA, Malinowski A, Majchrzak-Baczmanska D, Wosiak A. The analysis of vaginal hysterectomy results depending on the uterine size. *Ginekol Pol.* 2021;92(5):339–343.

18. Roque DR, Robison K, Raker CA, Wharton GG, Frishman GN. The accuracy of surgeons' provided estimates for the duration of hysterectomies: a pilot study. *J Minim Invasive Gynecol.* 2015;22(1):57–65.

19. Swartz J, Koziatek C, Theobald J, Smith S, Iturrate E. Creation of a simple natural language processing tool to support an imaging utilization quality dashboard. *Int J Med Inform.* 2017;101:93–99.

20. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.

21. Chen T, Guestrin C. XGBoost: a Scalable Tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16.* 2016;785–794.

22. Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. *arXiv.* 2018;181011363. https://arxiv.org/abs/1810.11363.

23. Nori H, Jenkins S, Koch P, Caruana R. InterpretML: a unified framework for machine learning interpretability. *arXiv.* 2019190909223. https://arxiv.org/abs/1909.09223.

24. Hosseini N, Sir MY, Jankowski CJ, Pasupathy KS. Surgical duration estimation via data mining and predictive modeling: a case study. *AMIA Annu Symp Proc.* 2015;2015:640–648.

25. Kung FT, Chang SY. The relationship between ultrasonic volume and actual weight of pathologic uterus. *Gynecol Obstet Invest.* 1996;42(1):35–38.

26. Uppal S, Shahin MS, Rathbun JA, Goff BA. Since surgery isn't getting any easier, why is reimbursement going down? An update from the SGO taskforce on coding and reimbursement. *Gynecol Oncol.* 2017;144(2):235–237.

27. Shi Y, Mahdian S, Blanchet J, Glynn P, Shin AY, Scheinker D. Surgical scheduling via optimization and machine learning with long-tailed data: health care management science, in press. *Health Care Manag Sci.* 2023;26(4):692–718.

28. Watanabe M, Yamauchi K. The effect of quality of overtime work on nurses' mental health and work engagement. *J Nurs Manag.* 2018;26(6):679–688.

29. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak.* 2020;20(1):280.

30. Cantor MN, Thorpe L. Integrating data on social determinants of health into electronic health records. *Health Aff (Millwood).* 2018;37(4):585–590.

31. Hemingway H, Asselbergs FW, Danesh J, et al. Innovative Medicines Initiative 2nd programme, Big Data for Better Outcomes, BigData@Heart Consortium of 20 academic and industry partners including ESC. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J.* 2018;39(16):1481–1495.