







Explainable Machine Learning Model to Preoperatively Predict Postoperative Complications in Inpatients With Cancer Undergoing Major Operations

Matthew C. Hernandez, MD¹ ; Chen Chen, MS² ; Andrew Nguyen, MD³; Kevin Choong, MD⁴ ; Cameron Carlin, MS²; Rebecca A. Nelson, PhD⁵ ; Lorenzo A. Rossi, PhD² ; Naini Seth, MBA, BSN, RN⁶; Kathy McNeese, MSN¹; Bertram Yuh, MD¹; Zahra Eftekhari, MS²; and Lily L. Lai, MD¹ 

DOI <https://doi.org/10.1200/JCO.2023.00247>

ABSTRACT

PURPOSE Preoperative prediction of postoperative complications (PCs) in inpatients with cancer is challenging. We developed an explainable machine learning (ML) model to predict PCs in a heterogeneous population of inpatients with cancer undergoing same-hospitalization major operations.

METHODS Consecutive inpatients who underwent same-hospitalization operations from December 2017 to June 2021 at a single institution were retrospectively reviewed. The ML model was developed and tested using electronic health record (EHR) data to predict 30-day PCs for patients with Clavien-Dindo grade 3 or higher (CD 3+) per the CD classification system. Model performance was assessed using area under the receiver operating characteristic curve (AUROC), area under the precision recall curve (AUPRC), and calibration plots. Model explanation was performed using the Shapley additive explanations (SHAP) method at cohort and individual operation levels.

RESULTS A total of 988 operations in 827 inpatients were included. The ML model was trained using 788 operations and tested using a holdout set of 200 operations. The CD 3+ complication rates were 28.6% and 27.5% in the training and holdout test sets, respectively. Training and holdout test sets' model performance in predicting CD 3+ complications yielded an AUROC of 0.77 and 0.73 and an AUPRC of 0.56 and 0.52, respectively. Calibration plots demonstrated good reliability. The SHAP method identified features and the contributions of the features to the risk of PCs.

CONCLUSION We trained and tested an explainable ML model to predict the risk of developing PCs in patients with cancer. Using patient-specific EHR data, the ML model accurately discriminated the risk of developing CD 3+ complications and displayed top features at the individual operation and cohort level.

ACCOMPANYING CONTENT

 Appendix

Accepted March 6, 2024

Published April 22, 2024

JCO Clin Cancer Inform

8:e2300247

© 2024 by American Society of

Clinical Oncology

Creative Commons Attribution
Non-Commercial No Derivatives
4.0 License

INTRODUCTION

In patients undergoing surgery, the risk of postoperative complications (PCs) varies considerably.¹⁻⁶ Patients who develop PC have worse functional outcomes, increased disabilities, and reduced survival.⁷⁻⁹ Moreover, PCs result in longer ICU and hospital days and greater frequency of readmissions—increasing cost for patients and health care systems.¹⁰ Since PCs are burdensome to both patients and health care systems,^{11,12} there is an urgent and unmet need to preoperatively identify patients at higher risk for PC to help prevent and mitigate PC.

Severity of illness, baseline functional performance, and operation type and extent contribute to the risk of developing PC. There is a large volume of existing work identifying factors to improve on risk stratification for postoperative morbidity and mortality.¹³⁻¹⁵ In particular, the American College of Surgeons Surgical Risk Calculator (ACS SRC) estimates preoperative individual- and procedure-specific risk using clinical characteristics.¹⁶ Developed and validated using a heterogeneous patient population, ACS SRC is not specific to patients with cancer and does not incorporate critical features unique to oncology, such as cancer stage and treatments.¹⁷ Moreover, ACS SRC generates a risk score on

CONTEXT

Key Objective

This study discusses the development and testing of an explainable machine learning (ML) model using electronic health record (EHR) data to predict risk of postoperative complications (PCs) in inpatients with cancer who require same-hospitalization operations.

Knowledge Generated

The explainable ML model performs well in the cross-validation and test sets in the prediction of PCs. Features that contributed the most to the PC risk included total hospital days in the past year, vitals, and laboratory values.

Relevance (J.L. Warner)

The model uses patient data from EHRs and demonstrates accuracy in identifying high-risk individuals and highlighting the key factors contributing to this risk at both the individual and group level utilizing explainable ML approaches.*

*Relevance section written by JCO Clinical Cancer Informatics Editor-in-Chief Jeremy L. Warner, MD, MS, FAMIA, FASCO.

the basis of a planned operation using a single CPT code. However, cancer operations often require multiple procedures performed by subspecialty surgeons under one anesthesia setting. Consequently, studies have shown limited performance of the ACS SRC in determining the risk of PC in patients with cancer.^{18,19}

While most current surgical risk calculators are developed using statistical regression analysis, machine learning (ML) algorithms can analyze patient-specific data to generate PC risk prediction models specific to an oncology patient population. The goal of this study is to develop and test an explainable ML model using electronic health record (EHR) data in preoperatively predicting PC specifically in cancer inpatients who require a same-hospitalization operation.

METHODS

Description of Institution

Patients were admitted and treated at an National Cancer Institute–designated Comprehensive Cancer Center. After obtaining Institutional Review Board approval, relevant patient preoperative and outcome data were obtained from the EHR database (Epic Systems Corporation, Verona, WI).

Cohort Selection Criteria

Included in our study were inpatient operations, planned after admission and performed from December 2017 through June 2021 (Fig 1). Inclusions were operations that involve at least one surgeon and completed under general anesthesia. Exclusions were minor operations, outpatient operations, elective operations in which the patient was admitted after the operation (AM Admit operations), and operations requested before patient admission. The final cohort comprised 827 consecutive patients who underwent 988 major operations (Tables 1 and 2).

PC Definitions

PCs were defined according to the Clavien–Dindo (CD) classification system, as described by Dindo et al²⁰ (Table 3; Appendix Table A1). PCs of CD grade 3 or higher (CD 3+) developed within 30 days after the operation were identified by querying the database (Appendix Table A1). Discrepancies were adjudicated by the investigators (K.M., M.C.H., A.N., K.C., and L.L.L.). The outcome was a binary variable of 30-day CD 3+ or no 30-day CD 3+ complication.

Feature Extraction

Variables included in the model were based on clinical relevance and availability within the EHR. Only data obtained before the surgical case request order were included as input variables. As shown in Appendix Table A2, 66 features were included in eight groupings: demographics, diagnosis codes, cancer treatment, previous inpatient admissions and outpatient visits, medication administration records, oxygenation support, vital signs, and laboratory tests. Feature data were collected on the basis of varying and defined time periods of interests (ie, observation windows). Features were aggregated at the operation level within their observation windows to make predictions. The mean and variance of each feature were computed using nonmissing feature values in the training set. All nonmissing feature values in both the training and holdout test sets were standardized using these means and variances to have zero mean and unit variance and were forward filled. Missing feature values were set to zero.

ML Methods

Of 988 operations included in the study, 788 (80%) were assigned to the training set and 200 (20%) were assigned to the holdout test set. The training and holdout test sets

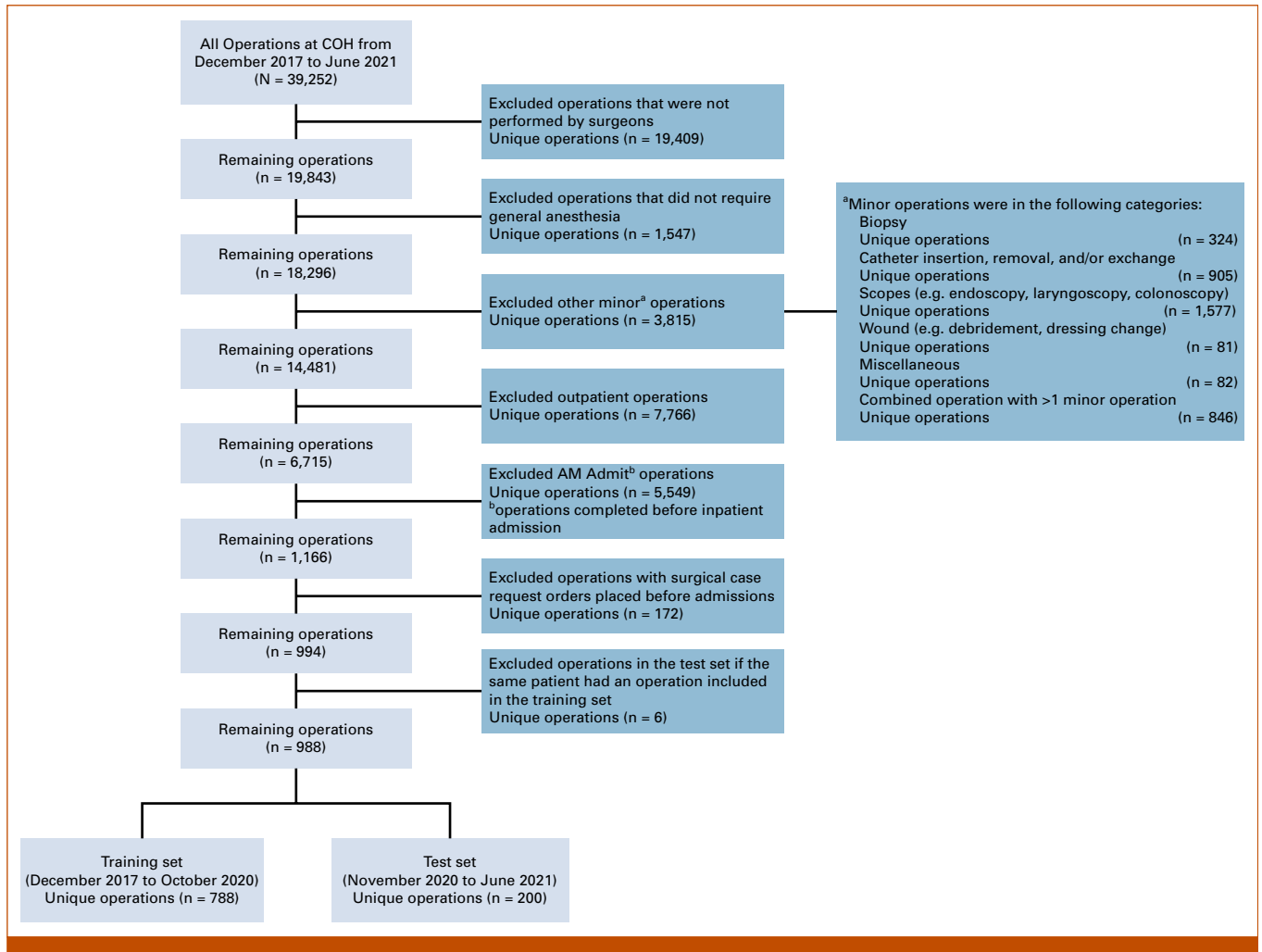


FIG 1. CONSORT diagram of cohort selection criteria. COH, City of Hope National Medical Center.

included operations performed from December 2017 through October 2020 and November 2020 through June 2021, respectively. To mimic real-world scenarios, a temporal split on the retrospective data was used to predict the prospective data. To minimize data leakage, patients who had operations in the training set were not included in the holdout test set (n = 6; Fig 1). A classification model using a gradient boosting decision tree model implemented in the Python XGBoost package was used to predict the likelihood of 30-day CD 3+ PC.²¹ The model produced a score scaled from 0 to 1, with higher scores representing higher risk of PC. To find the optimal hyperparameter set, we performed five-fold cross-validation (CV) with a randomized search on the training set and tuned for the number of trees with other hyperparameters to reduce overfitting (Appendix Table A3). To assess feature contributions, we used Shapley additive explanations (SHAP) for tree-based models.^{22,23} This approach allocated a numerical value (ie, SHAP value) to each feature to measure its impact on model output and to provide a local explanation of an operation. A positive SHAP value

increased the likelihood of a complication, whereas a negative SHAP value reduced the likelihood of a complication. In addition, SHAP values were aggregated across operations and features to obtain a global explanation for the cohort. The ML model development and explanation were implemented using Python 3.8.3, XGBoost 1.5.0, scikit-learn 1.0.2, and SHAP 0.40.0.²⁴

The final model developed from the training set was evaluated on the holdout test set. The model discrimination was assessed using area under the receiver operating characteristic curve (AUROC) and area under the precision recall curve (AUPRC). Model calibration was assessed by plotting mean predicted risk scores against observed complication rates. The calibration plots depict the alignment of the ML model predicted risk with the observed complication rate. In addition, ML model performance metrics such as sensitivity, specificity, and accuracy were calculated at the optimal threshold identified using the Youden Index. Statistical analyses were completed using SAS 9.4 (SAS Institute, Cary, NC).

TABLE 1. Patient Characteristics

| Characteristic | No. (%) or Median (IQR) |
|---------------------------------------|-------------------------|
| Total patients, No. ^a | 827 |
| Age, years, ^b median (IQR) | 59 (48-68) |
| Sex, No. (%) | |
| Female | 438 (53.0) |
| Male | 389 (47.0) |
| Race, No. (%) | |
| White | 594 (71.8) |
| Asian | 123 (14.9) |
| Black or African American | 33 (4.0) |
| Other race | 32 (3.9) |
| Missing | 45 (5.4) |
| Ethnicity, No. (%) | |
| Not Hispanic or Latino | 529 (64.0) |
| Hispanic or Latino | 263 (31.8) |
| Missing | 35 (4.2) |
| Marital status, No. (%) | |
| Married | 508 (61.4) |
| Single | 171 (20.7) |
| Divorced | 68 (8.2) |
| Widowed | 51 (6.2) |
| Other | 24 (2.9) |
| Missing | 5 (0.6) |
| No. of operations, No. (%) | |
| One operation | 719 (86.9) |
| More than one operation | 108 (13.1) |

^aEight hundred twenty-seven patients who underwent 988 operations from December 2017 to June 2021 were included.

^bAge was reported on the basis of the data collected at the first operation.

RESULTS

Cohort Demographics, Operation Characteristics, and Complication Rates

Table 1 outlines the demographic characteristics of 827 patients who underwent 988 operations from December 2017 to June 2021. Most patients (86.9%) underwent one operation; 13.1% (n = 108) of patients underwent more than one operation. The patients, whose median age was 59 years (IQR, 48–68), were predominately women (438, 53.0%), White (594, 71.8%), non-Hispanic (529, 64.0%), and married (508, 61.4%).

As shown in **Table 2**, 451 (45.6%) operations involved more than one procedure and 157 (15.9%) involved more than one surgical subspecialty team. Operations were completed in patients receiving active cancer treatment (373, 37.8%) including chemotherapy (293, 29.7%), immunotherapy (116, 11.7%), and radiation therapy (78, 7.9%). Most operations were performed in patients with newly diagnosed cancer

TABLE 2. Operation Characteristics

| Characteristic | No. (%) |
|-----------------------------------|------------|
| Total operations ^a | 988 |
| Procedure | |
| Single procedure performed | 537 (54.4) |
| More than one procedure performed | 451 (45.6) |
| Surgical subspecialty team | |
| Single surgical team | 831 (84.1) |
| More than one surgical team | 157 (15.9) |
| Cancer treatment ^b | |
| Completed | 615 (62.2) |
| Active | 373 (37.8) |
| Chemotherapy | 293 (29.7) |
| Immunotherapy | 116 (11.7) |
| Radiation | 78 (7.9) |
| Year of cancer diagnosis | |
| <1 year | 523 (52.9) |
| ≥1 year and <5 years | 319 (32.3) |
| ≥5 years and <10 years | 121 (12.2) |
| Missing | 25 (2.5) |
| Malignancy type ^c | |
| Solid malignancy | 738 (74.7) |
| Stage 0 | 4 (0.4) |
| Stage I | 111 (11.2) |
| Stage II | 162 (16.4) |
| Stage III | 144 (14.6) |
| Stage IV | 314 (31.8) |
| Missing | 3 (0.3) |
| Hematologic malignancy | 250 (25.3) |

^aEight hundred twenty-seven patients who underwent 988 operations from December 2017 to June 2021 were included.

^bActive cancer treatment was defined as the receipt of chemotherapy, immunotherapy, and/or radiation therapy within 3 months before the operation.

^cMalignancy type and cancer stage were manually curated. Additional details of malignancy type are presented in Appendix **Table A4**.

(523, 52.9%), solid tumor malignancies (738, 74.7%), and advanced disease (stage III/IV, 458, 46.4%). Preoperative comorbidities are listed in Appendix **Table A4**.

Table 3 lists the type and frequency of the 30-day PC. The overall CD 3+ complication rate was 28.3%. The CD 3+ PC rate was 28.6% in the training set and 27.5% in the holdout test set (**Table 3**). **Table 3** also lists the PC by grade and frequency.

Training and Holdout Test ML Models Have Comparable Performance

Figure 2 depicts the model performance of the training and holdout test sets. Using the training set, the ML model to predict CD 3+ complications achieved an overall CV AUROC of 0.77, ranging from 0.75 to 0.79 with each fold (**Fig 2A**). The

TABLE 3. Thirty-Day Postoperative Complications of CD Grade 3 or Higher

| Complication | Training Set, December 2017-October 2020 | Test Set, November 2020-June 2021 | Total |
|------------------------------------|--|-----------------------------------|------------|
| Total operations, No. | 788 | 200 | 988 |
| Complication, ^a No. (%) | 225 (28.6) | 55 (27.5) | 280 (28.3) |
| CD 3a, No. (%) | 77 (9.8) | 22(11.0) | 99 (10.0) |
| Abscess drain placement | 29 (3.7) | 5 (2.5) | 34 (3.4) |
| Aspiration | 6 (0.8) | 1 (0.5) | 7 (0.7) |
| Bronchoscopy | 9 (1.1) | 3 (1.5) | 12 (1.2) |
| Chest tube placement | 4 (0.5) | 0 (0.0) | 4 (0.4) |
| Embolization | 1 (0.1) | 1 (0.5) | 2 (0.2) |
| Endoscopy | 5 (0.6) | 3 (1.5) | 8 (0.8) |
| Hemorrhage | 12 (1.5) | 5 (2.5) | 17 (1.7) |
| Incision and drainage | 1 (0.1) | 0 (0.0) | 1 (0.1) |
| Thoracentesis | 19 (2.4) | 7 (3.5) | 26 (2.6) |
| Urinary leak | 9 (1.1) | 1 (0.5) | 10 (1.0) |
| CD 3b, No. (%) | 32 (4.1) | 9 (4.5) | 41 (4.1) |
| Bronchoscopy | 3 (0.4) | 0 (0.0) | 3 (0.3) |
| Chest tube placement | 1 (0.1) | 0 (0.0) | 1 (0.1) |
| Endoscopy | 8 (1.0) | 0 (0.0) | 8 (0.8) |
| Hematoma | 2 (0.3) | 6 (3.0) | 8 (0.8) |
| Hemorrhage | 3 (0.4) | 0 (0.0) | 3 (0.3) |
| Incision and drainage | 9 (1.1) | 5 (2.5) | 14 (1.4) |
| Laparoscopy | 5 (0.6) | 1 (0.5) | 6 (0.6) |
| Laparotomy | 5 (0.6) | 2 (1.0) | 7 (0.7) |
| CD 4a, No. (%) | 33 (4.2) | 7 (3.5) | 40 (4.0) |
| Heart failure | 3 (0.4) | 0 (0.0) | 3 (0.3) |
| Liver failure | 1 (0.1) | 2 (1.0) | 3 (0.3) |
| Renal failure | 4 (0.5) | 0 (0.0) | 4 (0.4) |
| Respiratory failure | 25 (3.2) | 5 (2.5) | 30 (3.0) |
| CD 4b, No. (%) | 2 (0.3) | 1 (0.5) | 3 (0.3) |
| CD 5, No. (%) | 81 (10.3) | 16 (8.0) | 97 (9.8) |

Abbreviation: CD, Clavien-Dindo.

^aFor operations with multiple complications, only the complication with the highest CD grade was included.

overall CV AUPRC was 0.56, ranging from 0.52 to 0.63 with each fold (Fig 2C). The ML model achieved an AUROC of 0.73 and an AUPRC of 0.52 in the holdout test set (Figs 2B and 2D). In addition, the model demonstrated good calibration as the mean predicted risk scores were similar to the observed complication rates at different risk strata in both the training and holdout test sets (Figs 2E and 2F). At the optimal threshold determined using the CV set that maximized the Youden Index, the ML model had a sensitivity of 0.78, a specificity of 0.63, and an accuracy of 0.67. Application of the same analysis to maximize the Youden Index in the holdout test set resulted in a sensitivity of 0.67, a specificity of 0.66, and an accuracy of 0.67.

Novel Features With the Highest Contribution to Risk Score

Features that contributed to the model prediction at the cohort level were evaluated using SHAP summary plots (Fig 3).

These plots combine feature contributions and use feature values to visualize the impact directionality has on the ML model. To visualize the relationship between feature values and SHAP values, a SHAP dependence plot was created for the top five features (Appendix Fig A1). The cutoff values for the features negatively affecting and increasing the likelihood of CD 3+ PC were as follows: total hospital length of stays >13 days in the past year, lymphocyte <10.6%, hemoglobin <9.1 g/dL, heart rate >103 beats/min, and/or percentage of abnormal laboratory test results >52% in the past 90 days.

To visualize how different features combine to predict an individual's PC risk, a randomly selected individual risk profile was plotted (Appendix Fig A2). The selected patient was at high risk of PC with a predicted score of 0.74. The elevated risk score resulted from the following individual features: high respiratory rate (28 breaths/min), high heart rate (120 beats/min), high percentage of abnormal laboratory test results in the past 90 days (73%), and low

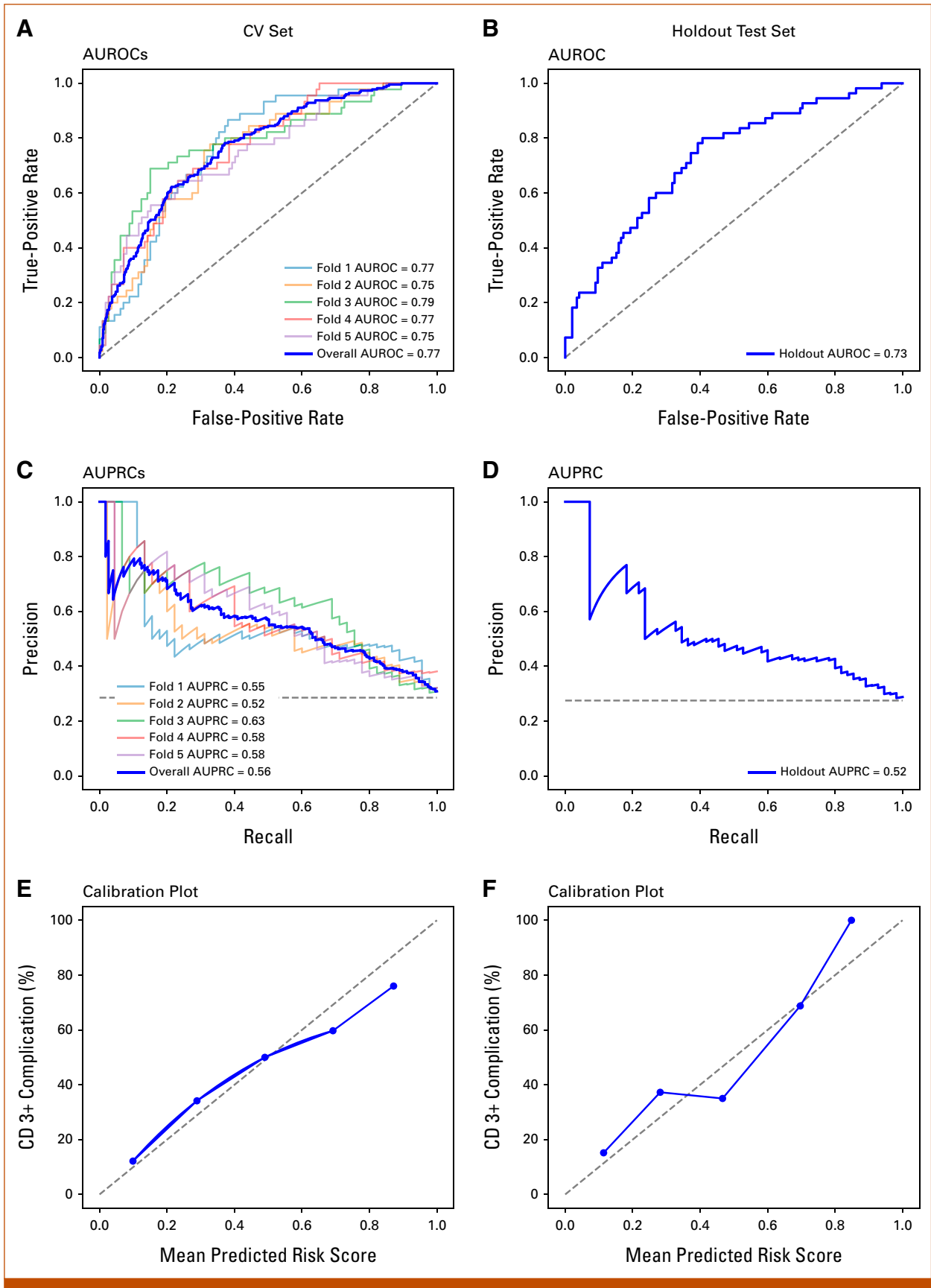


FIG 2. Model performance of the CV set (left column) and the holdout test set (right column). AUROCs for (A) the CV set and (B) the holdout test set. AUPRCs for (C) the CV set and (D) the holdout test set. Calibration plots for (E) the CV set and (F) the holdout test set. The overall AUROC and AUPRC values were computed after combining five folds. The calibration (continued on following page)

FIG 2. (Continued). plot visualizes how correctly the predicted risk by the model estimates the observed CD 3+ complication rate. (E and F) Operations were grouped into five bins with equal widths on the basis of the ascending order of the predicted risk scores; the x-axis of each point represents the mean predicted risk score in each bin; the y-axis of each point represents the CD 3+ complication rate in each bin; the gray dashed 45-degree line represents a perfect calibrated model. AUPRC, area under the precision recall curve; AUROC, area under the receiver operating characteristic curve; CD 3+, Clavien-Dindo grade 3 or higher; CV, cross-validation.

hemoglobin (7.0 g/dL). Together, these features contributed to the individual’s high-risk score that predicted the development of CD 3+ PC.

DISCUSSION

In this study, we developed an explainable ML model using patient-specific, preoperative EHR data to generate a risk prediction model for the development of severe (CD 3+) PC in

inpatients with cancer who underwent same-hospitalization major operations. The trained model performed well when tested in the holdout test set with an AUROC of 0.73 and an AUPRC of 0.52. In addition, there was good calibration between the predicted risk and observed complication rates. Contributing features were identified and explored through the SHAP method. Our findings support continuing development and use of ML models for preoperative risk prediction of PC in patients with cancer.

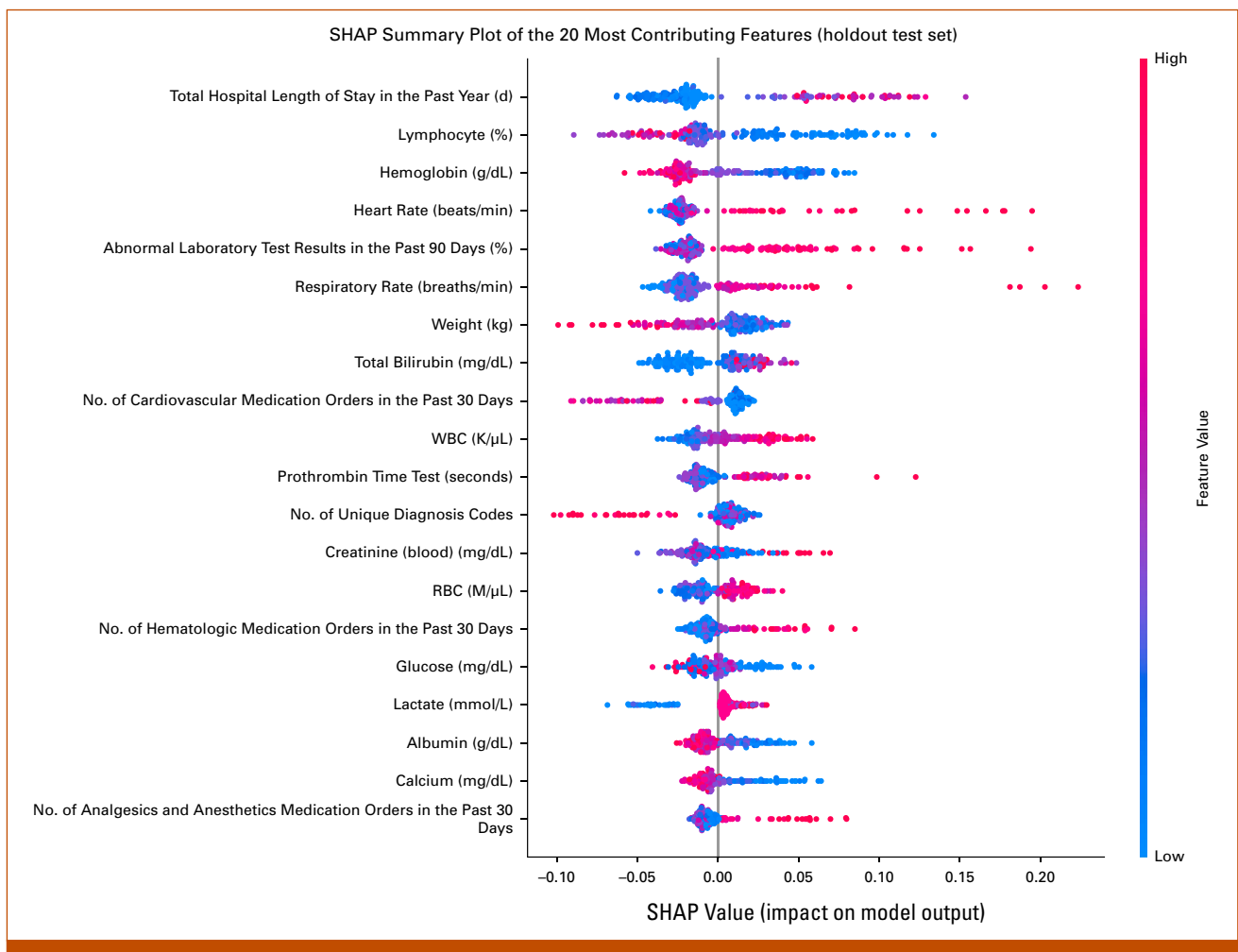


FIG 3. SHAP summary plot of the 20 most contributing features ranked by descending contribution to the prediction of 30-day Clavien-Dindo grade 3 or higher postoperative complications on the basis of the holdout test set. Each dot in the plot represents a SHAP value for a feature and an operation in the holdout test set. The colors represent the feature values with blue for lower values and red for higher values. The negative SHAP value indicates the reduced likelihood of complication, and the positive SHAP value indicates the increased likelihood of complication. The directionality of the feature impact could be evaluated on the basis of the distribution of the red and blue dots. The features were ranked according to the mean absolute SHAP values aggregated across operations in the holdout test set. SHAP, Shapley additive explanations.

Multiple models exist to assess perioperative morbidity and mortality including the ACS SRC.^{16,25} Other prediction tools used to calculate patient risk add features such as vital signs and operative details to improve performance.^{14,26} Overall, these scores perform modestly in patients with cancer. In addition, the prediction tools are not integrated into EHR systems for easy implementation into the clinical workflow.¹⁷ Recently, using administrative data sets, cancer registries, and EHR data, AI techniques to recognize patterns, learn from experience without user input, and detect complex variable interactions to enable the development of algorithms involving nonlinear relationships of features have been harnessed to predict PC.²⁷⁻²⁹ For example, MySurgeryRisk capitalizes on the availability of big data to develop ML model risk scores for eight PCs.^{25,30} However, at this time, the platform is a mobile application which requires manual input of data and is not integrated into inpatient EHRs, limiting the accessibility and utilization in a preoperative inpatient setting.

ML model performance for risk prediction has also been studied in smaller cohorts of patients with cancer limited to a single disease site, operation, or institution. Pera et al³¹ reported the use of ML in the prediction of 90-day mortality after gastrectomy. The ML model performed well with an AUROC of 0.79–0.84. In patients undergoing cytoreductive surgery, the ML model outperformed multiple logistic regression models for postoperative risk prediction with an AUROC of 0.75.³² Similarly, in a small single-center cohort of outpatients with cancer undergoing surgery, the ML model for PC prediction had an AUROC of 0.71.²⁹ To date, none of the risk prediction models have been used in inpatients with cancer agnostic of the site of disease and operation type.^{30,33-35}

Our single-institutional study provides contemporary risk estimation of severe (CD 3+) PC in inpatients with cancer regardless of the type of cancer or operation. Our study includes patients with cancer with hematologic or solid cancer diagnosis and complex operations requiring multiple procedures and different subspecialties. Finally, our ML model, unlike other predictive models,^{29,31,32} incorporates cancer treatments such as days of radiation treatment and number of chemotherapy and immunotherapy medication orders, features which contribute to postoperative outcomes.

AUROC for our training set and holdout test set ML model of 0.77 and 0.73, respectively, are comparable in performance with previously published models. However, we also report AUPRC, a metric which evaluates model interactions between precision (positive predictive value) and recall (true positive rate) at different thresholds and assesses model performance with imbalanced data sets. Our ML model results in an AUPRC of 0.56 in the training set and 0.52 in the holdout test set, enabling discrimination of patients with a

higher risk of complication above the baseline PC rate of 28.3%. In addition, we report calibration plots to compare the ML model predicted risk with the observed PC rate. As such, a patient with a predicted risk score of 0.6 has an observed complication rate of approximately 60%. The results of our calibration plots argue for good alignment between predicted risk and observed rate of PC in our patient population.

To better understand the feature impact from the optimized ML model, we used SHAP values as previously reported in other risk assessment ML models.^{27,35} The SHAP method explains risk prediction by listing the contribution of each feature driving the risk of PC in an individual patient. Using this method, we identified novel features that had not been previously reported. The top contributors to increased PC risk include longer total length of hospital stay in the past year, low lymphocyte level, low hemoglobin level, high heart rate, and high percentage of abnormal laboratory test results in the past 90 days. The features affecting the PC risk can potentially be addressed and optimized preoperatively.^{36,37} In addition, steps of an enhanced recovery pathway after an operation can be adjusted to address features identified by the ML model to improve patient-specific outcomes.

Because our ML model uses readily available EHR data, risk assessment of PC can be seamlessly integrated into the preoperative workflow. Many of the available risk stratification systems, and notably the ACS SRC, are based on static data at a single timepoint in the clinical care timeline.¹⁶ Static timepoints can lead to degradation in risk assessment performance in patients with dynamic clinical status. Without the need for a separate device or web-based application and without the need for manual input of features, our model assigns a risk score automatically using EHR data and makes the risk score immediately available for the clinician to review within the EHR. Ongoing development of this ML model and integration into the EHR environment will enable real-time automated risk assessment as part of the surgeon preoperative workflow.

Finally, and perhaps most importantly, in cancer care, individualized risk can serve as a key component in a multifaceted approach to assist decision making. In patients with advanced cancer, better risk prediction of PC might allow for a more nuanced discussion that balances short- and long-term outcomes against quality-of-life concerns. Preoperative discussions with more accurate and individualized risk assessments may facilitate better education and help set realistic expectations, enabling patient and provider shared decision making.

This study has several limitations. Foremost, the model was validated using a holdout test set that included patients hospitalized during the COVID-19 pandemic. The changes

because of hospital policies surrounding capacity concerns might have resulted in patient demographic changes, which could affect model performance. Ongoing prospective validation will determine performance. In addition, the ML model was limited to inpatients. We chose to begin ML model development for risk prediction in the inpatient cohort because of the availability and quality of the EHR inpatient data. Ongoing work will include outpatients with cancer scheduled for elective operations. Finally, many other factors including hospital- and surgeon-specific variables are not included in our patient-specific ML model. Additional studies incorporating other features to improve performance are warranted. In addition, validation in larger and external

data sets is critical to assess the generalizability of the ML model across cancer populations.

In conclusion, our study developed an explainable ML model that accurately predicted risk of development of CD 3+ PC in patients who were admitted to a tertiary care comprehensive cancer center before undergoing a same-hospitalization major operation. The model incorporates EHR data inputs to generate individual operation-level risk prediction and model explanation that can be readily available to clinicians as part of the presurgical workflow within the EHR system. Further development of the risk model is warranted.

AFFILIATIONS

¹Department of Surgery, University of New Mexico, Albuquerque, NM

²Department of Applied AI and Data Science, City of Hope National Medical Center, Duarte, CA

³Department of Surgery, City of Hope National Medical Center, Duarte, CA

⁴Department of Surgery, Division of Oncology, Primus Health, University of South Carolina Medical School, Greenville, SC

⁵Department of Computational and Quantitative Medicine, Division of Biostatistics, City of Hope National Medical Center, Duarte, CA

⁶Department of Clinical Informatics, City of Hope National Medical Center, Duarte, CA

CORRESPONDING AUTHOR

Lily L. Lai, MD; e-mail: llai@coh.org.

DISCLAIMER

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

EQUAL CONTRIBUTION

M.C.H. and C.C. contributed equally to the completion of this research.

PRIOR PRESENTATION

Presented in part at the 2022 Society of Surgical Oncology Annual Meeting, Dallas, TX, March 9-12, 2022.

SUPPORT

Supported in part by the National Cancer Institute of the National Institutes of Health under award number P30CA033572.

AUTHOR CONTRIBUTIONS

Conception and design: Matthew C. Hernandez, Chen Chen, Andrew Nguyen, Lorenzo A. Rossi, Naini Seth, Kathy McNeese, Zahra Eftekhari, Lily L. Lai

Financial support: Lily L. Lai

Administrative support: Rebecca A. Nelson, Zahra Eftekhari, Lily L. Lai

Provision of study materials or patients: Lily L. Lai

Collection and assembly of data: Matthew C. Hernandez, Chen Chen, Andrew Nguyen, Cameron Carlin, Kathy McNeese, Bertram Yuh, Lily L. Lai

Data analysis and interpretation: Matthew C. Hernandez, Chen Chen, Andrew Nguyen, Kevin Choong, Cameron Carlin, Rebecca A. Nelson, Lorenzo A. Rossi, Kathy McNeese, Bertram Yuh, Lily L. Lai

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Lorenzo A. Rossi

Patents, Royalties, Other Intellectual Property: I'm coinventor in a pending patent application: machine learning enabled prognosis of patient mortality. The patent was submitted as provisional to the USPTO on January 30, 2023, by City of Hope National Cancer Center. To this date, I have not received any royalties

Naini Seth

Employment: City of Hope

Kathy McNeese

Employment: Banner Health

Zahra Eftekhari

Consulting or Advisory Role: Mydayda

No other potential conflicts of interest were reported.

REFERENCES

1. Hyder O, Pulitano C, Firoozmand A, et al: A risk model to predict 90-day mortality among patients undergoing hepatic resection. *J Am Coll Surg* 216:1049-1056, 2013
2. Wegner RE, Verma V, Hasan S, et al: Incidence and risk factors for post-operative mortality, hospitalization, and readmission rates following pancreatic cancer resection. *J Gastrointest Oncol* 10:1080-1093, 2019
3. Bogach J, Cordeiro E, Reel E, et al: Axillary surgery and complication rates after mastectomy and reconstruction for breast cancer: An analysis of the NSQIP database. *Breast Cancer Res Treat* 192:501-508, 2022
4. Tokunaga M, Kurokawa Y, Machida R, et al: Impact of postoperative complications on survival outcomes in patients with gastric cancer: Exploratory analysis of a randomized controlled JCOG1001 trial. *Gastric Cancer* 24:214-223, 2021
5. Snyder RA, Hao S, Irish W, et al: Thirty-day morbidity after simultaneous resection of colorectal cancer and colorectal liver metastasis: American College of Surgeons NSQIP analysis. *J Am Coll Surg* 230:617-627.e9, 2020
6. Sesti J, Almaz B, Bell J, et al: Impact of postoperative complications on long-term survival after esophagectomy in older adults: A SEER-Medicare analysis. *J Surg Oncol* 124:751-766, 2021
7. Tevis SE, Cobian AG, Truong HP, et al: Implications of multiple complications on the postoperative recovery of general surgery patients. *Ann Surg* 263:1213-1218, 2016
8. Khuri SF, Henderson WG, DePalma RG, et al: Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Ann Surg* 242:326-343, 2005
9. McIsaac DI, Bryson GL, Van Walraven C: Association of frailty and 1-year postoperative mortality following major elective noncardiac surgery: A population-based cohort study. *JAMA Surg* 151:538-545, 2016
10. Short MN, Aloia TA, Ho V: The influence of complications on the costs of complex cancer surgery. *Cancer* 120:1035-1041, 2014
11. Stitzenberg KB, Chang YK, Smith AB, et al: Exploring the burden of inpatient readmissions after major cancer surgery. *J Clin Oncol* 33:455-464, 2015
12. Zafar SN, Shah AA, Channa H, et al: Comparison of rates and outcomes of readmission to index vs nonindex hospitals after major cancer surgery. *JAMA Surg* 153:719-727, 2018
13. Menke H, Klein A, John KD, et al: Predictive value of ASA classification for the assessment of the perioperative risk. *Int Surg* 78:266-270, 1993
14. Copeland GP, Jones D, Walters M: POSSUM: A scoring system for surgical audit. *Br J Surg* 78:355-360, 1991
15. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: A severity of disease classification system. *Crit Care Med* 13:818-829, 1985
16. Bilimoria KY, Liu Y, Paruch JL, et al: Surgical risk calculator: A decision aide and informed consent tool for patients and surgeons. *J Am Coll Surg* 217:833-842.e3, 2013
17. El Asmar A, Hafez K, Fauconnier P, et al: The efficacy of the American College of Surgeons Surgical Risk Calculator in the prediction of postoperative complications in oncogeriatric patients after curative surgery for abdominal tumors. *J Surg Oncol* 126:1359-1366, 2022
18. Schwartz PB, Stahl CC, Ethun C, et al: Retroperitoneal sarcoma perioperative risk stratification: A United States Sarcoma Collaborative evaluation of the ACS-NSQIP risk calculator. *J Surg Oncol* 122:795-802, 2020
19. Dave A, Beal EW, Lopez-Aguilar AG, et al: Evaluating the ACS NSQIP risk calculator in primary pancreatic neuroendocrine tumor: Results from the US Neuroendocrine Tumor Study Group. *J Gastrointest Surg* 23:2225-2231, 2019
20. Dindo D, Demartines N, Clavien PA: Classification of surgical complications: A new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 240:205-213, 2004
21. Chen T, Guestrin C: XGBoost: A scalable tree boosting system. *Proceeding of the 2nd ACM International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA, Association for Computing Machinery, August 13-17, 2016, pp 785-794
22. Lundberg SM, Erion G, Chen H, et al: From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2:56-67, 2020
23. Lundberg SM, Nair B, Vavilala MS, et al: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2:749-760, 2018
24. Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825-2830, 2011
25. Bihorac A, Ozrazgat-Baslanti T, Ebadli A, et al: MySurgeryRisk: Development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Ann Surg* 269:652-662, 2019
26. Cutti S, Klersy C, Favalli V, et al: A multidimensional approach of surgical mortality assessment and stratification (Smatt score). *Sci Rep* 10:10964, 2020
27. Jiang F, Jiang Y, Zhi H, et al: Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc Neurol* 2:230-243, 2017
28. Syeda-Mahmood T: Role of big data and machine learning in diagnostic decision support in radiology. *J Am Coll Radiol* 15:569-576, 2018
29. Gonçalves D, Henriques R, Santos LL, et al: On the predictability of postoperative complications for cancer patients: A Portuguese cohort study. *BMC Med Inform Decis Mak* 21:200-213, 2021
30. Ren Y, Loftus TJ, Datta S, et al: Performance of a machine learning algorithm using electronic health record data to predict postoperative complications and report on a mobile platform. *JAMA Netw Open* 5:e2211973, 2022
31. Pera M, Gibert J, Gimeno M, et al: Machine learning risk prediction model of 90-day mortality after gastrectomy for cancer. *Ann Surg* 276:776-783, 2022
32. Deng H, Eftekhari Z, Carlin C, et al: Development and validation of an explainable machine learning model for major complications after cytoreductive surgery. *JAMA Netw Open* 5:e2212930, 2022
33. Merath K, Hyer JM, Mehta R, et al: Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery. *J Gastrointest Surg* 24:1843-1851, 2020
34. Tseng PY, Chen YT, Wang CH, et al: Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care* 24:478, 2020
35. Wen R, Zheng K, Zhang Q, et al: Machine learning-based random forest predicts anastomotic leakage after anterior resection for rectal cancer. *J Gastrointest Oncol* 12:921-932, 2021
36. Kalbfell E, Kata A, Buffington AS, et al: Frequency of preoperative advance care planning for older adults undergoing high-risk surgery: A secondary analysis of a randomized clinical trial. *JAMA Surg* 156:e211521, 2021
37. Klaiber U, Stephan-Paulsen LM, Bruckner T, et al: Impact of preoperative patient education on the prevention of postoperative complications after major visceral surgery: The cluster randomized controlled PEDUCAT trial. *Trials* 19:288, 2018

APPENDIX

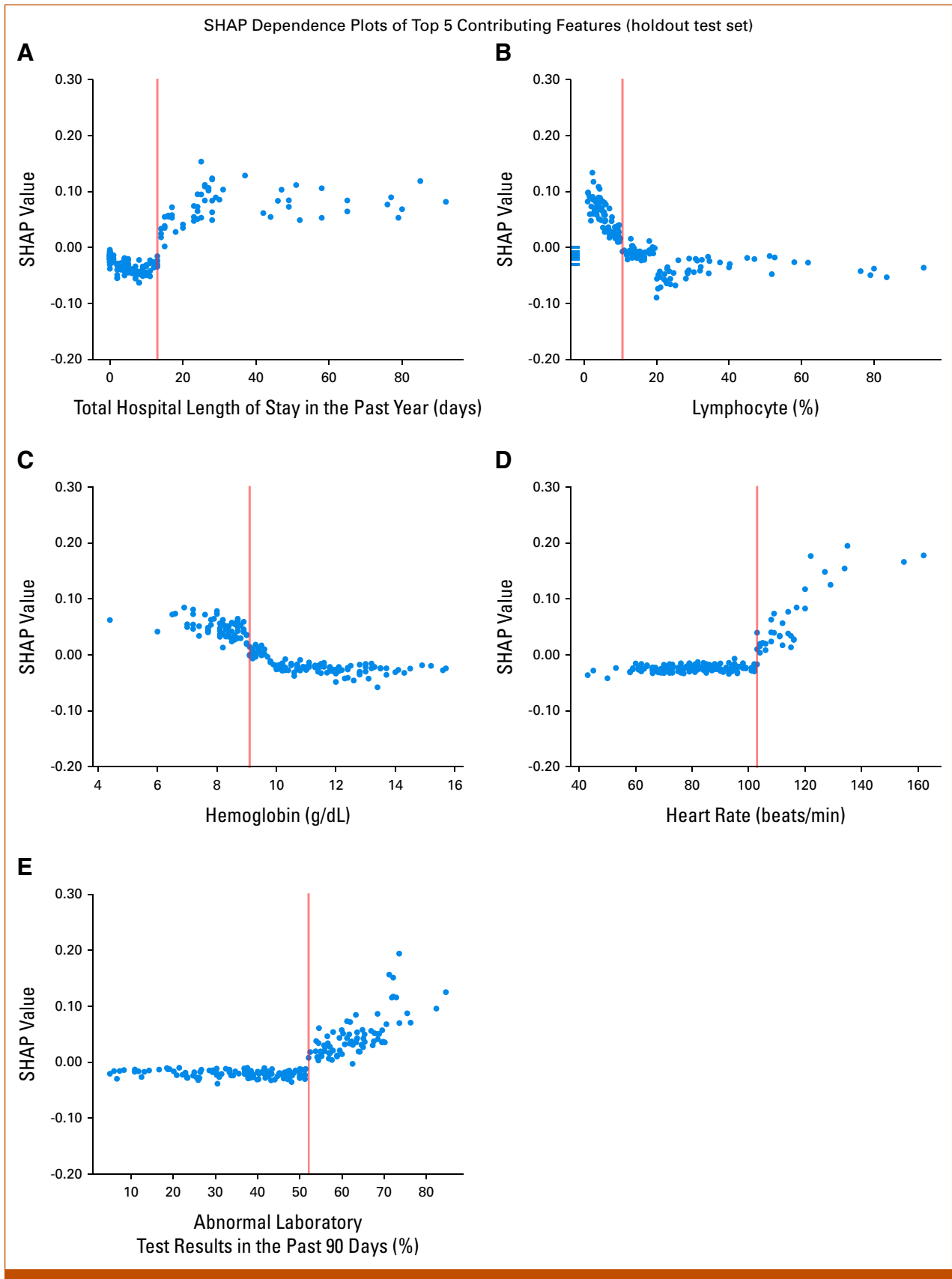


FIG A1. SHAP dependence plots on the basis of the holdout test set to show the effects of the top five contributing features on the prediction of 30-day Clavien-Dindo grade 3 or higher postoperative complications. A SHAP dependence (continued on following page)

FIG A1. (Continued). plot is a scatter plot with each dot representing an operation in the holdout test set. The SHAP dependence plot could be used to visualize the relationship between values of a feature on the x-axis and SHAP values of that feature on the y-axis. The red vertical line in each plot indicates the cutoff value of a feature that separates the positive and negative SHAP values. For (A-E) the top five contributing features, SHAP values were positive and indicated the increased likelihood of complications when total hospital length of stays >13 in the past year, lymphocyte <10.6%, hemoglobin <9.1 g/dL, heart rate >103 beats/min, and/or percentage of abnormal laboratory test results >52% in the past 90 days. SHAP, Shapley additive explanations.

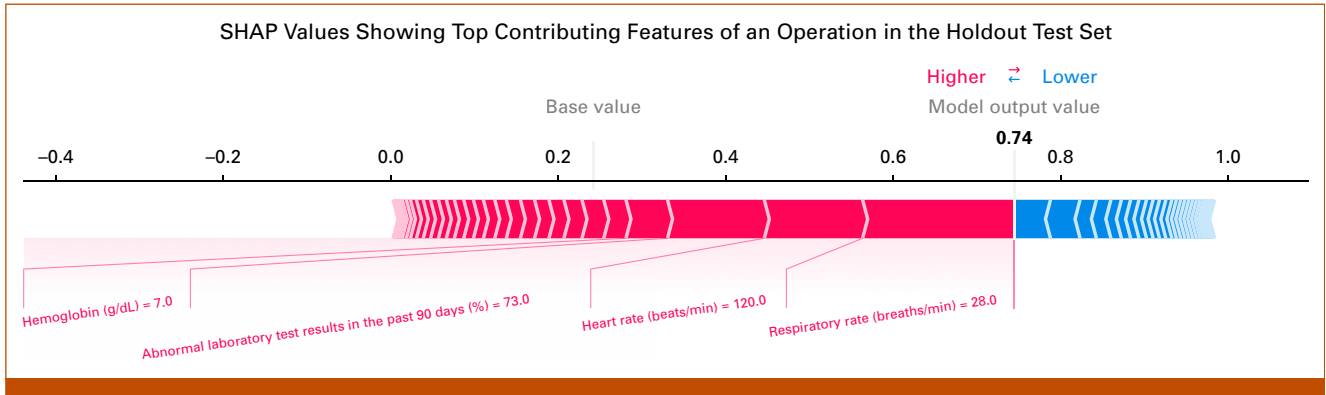


FIG A2. For an operation in the holdout test set, the SHAP values showing top contributing features to the prediction of 30-day Clavien-Dindo grade 3 or higher postoperative complications. The red color represents features that increase the likelihood of complication, whereas the blue color represents features that reduce the likelihood of complication. The size of each arrow bar indicates the magnitude of the feature contribution on the basis of the SHAP value of the feature. The base value is 0.24, which represents the mean of predicted scores in the holdout test set. In this example of a patient scheduled for an operation, the risk score is 0.74 (shown in bold). The top four features contributing to the risk of a complication in this patient are labeled in red. SHAP, Shapley additive explanations.

TABLE A1. Definitions of CD Grade 3 or Higher Complications

| Complication | Description | CD Classification |
|---|---|---|
| Abscess drain placement Aspiration Embolization | Abscess drain placement, aspiration, and embolization should be performed without general anesthesia | 3a |
| Thoracentesis | Thoracentesis should be performed without general anesthesia. It was not a complication if the patient had thoracentesis before the operation | 3a |
| Urinary leak | If the creatinine in body fluid was at least 1.5 times higher than the creatinine in blood and no procedure under general anesthesia was performed to handle the leak | 3a |
| Bronchoscopy Chest tube placement | Bronchoscopy and chest tube placement could be performed with or without general anesthesia. It was not a complication if the patient had bronchoscopy or chest tube placement before the operation | 3a (no general anesthesia) 3b (general anesthesia) |
| Endoscopy Hemorrhage Incision and drainage | Endoscopic procedures, procedures to control hemorrhage, and incision and drainage procedures could be performed with or without general anesthesia | 3a (no general anesthesia) 3b (general anesthesia) |
| Hematoma Laparoscopy Laparotomy | Hematoma, laparoscopy, and laparotomy should be performed under general anesthesia | 3b |
| Heart failure | Troponin I was >1 ng/mL | 4a |
| Liver failure | Total bilirubin was >7 mg/dL, and the INR was >2 | 4a |
| Renal failure | Dialysis was performed on a patient who did not require dialysis preoperatively | 4a |
| Respiratory failure | Intubation procedure was performed in an urgent situation | 4a |
| Multiorgan dysfunction | Two or more CD 4a complications (ie, heart failure, liver failure, renal failure, and respiratory failure) | 4b |
| Death | Death during or after the operation | 5 |

Abbreviations: CD, Clavien-Dindo; INR, international normalized ratio.

TABLE A2. Model Features

| Types of EHR Data | Feature | Observation Window ^a |
|---|---|---|
| Demographics | Continuous: age | NA |
| | Binary: male, White, Hispanic, married | |
| Diagnosis codes ^b | Number of all diagnosis records, number of unique diagnosis codes | All historical data leading to surgical case request time |
| | Number of all diagnosis records for respiratory problems, number of unique diagnosis codes for respiratory problems | |
| Cancer treatment | Number of days on radiation treatment | 1 year leading to surgical case request time |
| | Number of chemotherapy and immunotherapy medication orders | |
| Previous inpatient admissions and outpatient visits | Total hospital length of stay, number of inpatient admissions, number of outpatient visits | 1 year leading to current hospital admission time |
| Medication administration records | Number of all medication orders | 30 days leading to surgical case request time |
| | Number of medication orders for each therapeutic class: anti-infective agents, antineoplastic agents, endocrine and metabolic drugs, cardiovascular agents, respiratory agents, GI agents, genitourinary products, CNS drugs, analgesics and anesthetics, neuromuscular drugs, nutritional products, hematologic agents, topical products, miscellaneous products, unknown class ^c | |
| Oxygenation support | Highest oxygenation support level ^d | 30 days leading to surgical case request time |
| | Number of days on oxygenation support, number of oxygenation support records | 1 year leading to surgical case request time |
| | Most recent values: FiO ₂ , flow rate, PEEP, PIP, tidal volume | Most recent values |
| Vital signs | Most recent values: systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, SpO ₂ , temperature, weight | Most recent values |
| Laboratory tests | Most recent values: albumin, alkaline phosphatase, calcium, creatinine (blood), glucose, hematocrit, hemoglobin, lactate, lymphocytes percent, platelet, prothrombin time, RBC, RDW, segmented neutrophils percent, sodium, total bilirubin, WBC | Most recent values |
| | Number of laboratory test results, number of normal laboratory test results, number of abnormal laboratory test results, percentage of abnormal laboratory test results | 90 days leading to surgical case request time |

Abbreviations: BiPAP, bilevel positive airway pressure; CPAP, continuous positive airway pressure; EHR, electronic health record; FiO₂, the fraction of inspired oxygen; NA, not available; PEEP, positive end-expiratory pressure; PIP, peak inspiratory pressure; RDW, red blood cell distribution width; SpO₂, oxygen saturation by pulse oximetry.

^aFeatures were created within the observation window and aggregated at the operation level.

^bDiagnosis codes related to respiratory problems (eg, acute respiratory failure, pleural effusion, pneumonia, etc) were identified by clinical experts.

^cOrders that did not have therapeutic class assigned in the EHR were grouped to unknown class.

^dThe oxygenation support level was coded as no oxygenation support/room air = 0, face mask = 1, nasal cannula = 2, CPAP/BiPAP = 3, others = 4, ventilator = 5.

TABLE A3. Hyperparameter Tuning Process

| Hyperparameter | Tuning Process | Value in the Final Model |
|--|--|--------------------------|
| Learning rate | Fixed at 0.1 | 0.1 |
| No. of gradient boosted trees | Set the number of boosting iterations at 1,000, maximum tree depth at 4, subsample ratio of columns at 0.8, and tune the number of gradient boosted trees until the average AUROC value computed over five CV folds did not improve in every 50 rounds (early stopping rounds). The last iteration represented the best number of gradient boosted trees that was used to tune the other hyperparameters below for the final model | 70 |
| Maximum tree depth | Randomly picked an integer between 1 and 30 | 4 |
| Minimum sum of instance weight (hessian) needed in a child | Randomly picked an integer between 1 and 10 | 4 |
| Subsample ratio of the training instance | Uniformly picked a number between 0.1 and 1 | 0.86 |
| Subsample ratio of columns | Uniformly picked a number between 0.3 and 1 | 0.65 |
| L1 regularization term on weights | Uniformly picked a number from 10 numbers spaced evenly on a log scale between 0.01 and 100 | 0.22 |
| L2 regularization term on weights | Uniformly picked a number from 10 numbers spaced evenly on a log scale between 0.01 and 100 | 0.01 |

NOTE. All the other hyperparameters not mentioned here were fixed at default values. The randomized search process was evaluated using cross-entropy loss computed over five CV folds.

Abbreviations: AUROC, area under the receiver operating characteristic curve; CV, cross-validation.

TABLE A4. Operation-Specific Details of Malignancy Type and Comorbidities

| Characteristic | No. (%) |
|---------------------------------|-------------------|
| Total operations | 988 (100) |
| Solid malignancy | 738 (74.7) |
| Breast cancer | 138 (14.0) |
| Colorectal cancer | 92 (9.3) |
| Sarcoma | 66 (6.7) |
| Brain cancer | 48 (4.9) |
| Gastric cancer | 47 (4.8) |
| Lung cancer | 41 (4.1) |
| Ovarian cancer | 39 (3.9) |
| Prostate cancer | 34 (3.4) |
| Bladder cancer | 26 (2.6) |
| Melanoma | 20 (2.0) |
| Renal cell carcinoma | 19 (1.9) |
| Cervical cancer | 18 (1.8) |
| Uterine cancer | 18 (1.8) |
| Head and neck cancer | 17 (1.7) |
| Pancreatic cancer | 14 (1.4) |
| Hepatocellular carcinoma | 11 (1.1) |
| Squamous cell carcinoma | 11 (1.1) |
| Appendiceal cancer | 10 (1.0) |
| Cholangiocarcinoma | 8 (0.8) |
| Esophageal cancer | 8 (0.8) |
| Laryngeal cancer | 8 (0.8) |
| Neuroendocrine tumor | 7 (0.7) |
| Small bowel cancer | 7 (0.7) |
| Others | 31 (3.1) |
| Hematologic malignancy | 250 (25.3) |
| AML | 74 (7.5) |
| ALL | 54 (5.5) |
| Myelodysplastic syndromes | 35 (3.5) |
| Multiple myeloma | 31 (3.1) |
| Diffuse large B-cell lymphoma | 27 (2.7) |
| Chronic myeloid leukemia | 9 (0.9) |
| Non-Hodgkin's lymphoma | 8 (0.8) |
| Chronic lymphocytic leukemia | 7 (0.7) |
| Others | 5 (0.5) |
| Comorbidities | |
| Hypertension, uncomplicated | 236 (23.9) |
| Cardiac arrhythmias | 202 (20.4) |
| Fluid and electrolyte disorders | 156 (15.8) |
| Diabetes, uncomplicated | 103 (10.4) |
| Liver disease | 100 (10.1) |
| Coagulopathy | 94 (9.5) |
| Depression | 93 (9.4) |
| Hypothyroidism | 84 (8.5) |
| Other neurologic disorders | 81 (8.2) |
| Valvular disease | 81 (8.2) |
| Diabetes, complicated | 78 (7.9) |

(continued in next column)

TABLE A4. Operation-Specific Details of Malignancy Type and Comorbidities (continued)

| Characteristic | No. (%) |
|---|----------|
| Obesity | 71 (7.2) |
| Chronic pulmonary disease | 68 (6.9) |
| Weight loss | 47 (4.8) |
| Congestive heart failure | 46 (4.7) |
| Renal failure | 36 (3.6) |
| Deficiency anemias | 33 (3.3) |
| Pulmonary circulation disorders | 33 (3.3) |
| Peripheral vascular disorders | 30 (3.0) |
| Rheumatoid arthritis/collagen vascular diseases | 29 (2.9) |
| Hypertension, complicated | 20 (2.0) |
| Paralysis | 20 (2.0) |
| Blood loss anemia | 17 (1.7) |
| Drug abuse | 17 (1.7) |
| Peptic ulcer disease excluding bleeding | 14 (1.4) |
| Alcohol abuse | 8 (0.8) |
| Psychoses | 6 (0.6) |
| AIDS/HIV | 4 (0.4) |