Prediction of intensive care unit admission (>24h) after surgery in elective noncardiac surgical patients using machine learning algorithms

Lan Lan[1,3,4]* 🆔, Fangwei Chen[1,2,4]*, Jiawei Luo[1,4], Mengjiao Li[1,4], Xuechao Hao[5], Yao Hu[1,4], Jin Yin[1,4,6], Tao Zhu[5] and Xiaobo Zhou[7]

Abstract

Background: To develop a highly discriminative machine learning model for the prediction of intensive care unit admission (>24h) using the easily available preoperative information from electronic health records. An accurate prediction model for ICU admission after surgery is of great importance for surgical risk assessment and appropriate utilization of ICU resources.

Method: Data were collected retrospectively from a large hospital, comprising 135,442 adult patients who underwent surgery except for cardiac surgery between 1 January 2014, and 31 July 2018 in China. Multiple existing predictive machine learning algorithms were explored to construct the prediction model, including logistic regression, random forest, adaptive boosting, and gradient boosting machine. Four secondary analyses were conducted to improve the interpretability of the results.

Results: A total of 2702 (2.0%) patients were admitted to the intensive care unit postoperatively. The gradient boosting machine model attained the highest area under the receiver operating characteristic curve of 0.90. The machine learning models predicted intensive care unit admission better than the American Society of Anesthesiologists Physical Status (area under the receiver operating characteristic curve: 0.68). The gradient boosting machine recognized several features as highly significant predictors for postoperatively intensive care unit admission. By applying subgroup analysis and secondary analysis, we found that patients with operations on the digestive, respiratory, and vascular systems had higher probabilities for intensive care unit admission.

Conclusion: Compared with conventional American Society of Anesthesiologists Physical Status and logistic regression model, the gradient boosting machine could improve the performance in the prediction of intensive care unit admission. Machine learning models could be used to improve the discrimination and identify the need for intensive care unit admission after surgery in elective noncardiac surgical patients, which could help manage the surgical risk.

[1]West China Biomedical Big Data Center, West China Hospital/West China School of Medicine, Sichuan University, Chengdu, China
[2]Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, China
[3]IT Center, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[4]Med-X Center for Informatics, Sichuan University, Chengdu, China
[5]Department of Anesthesiology, West China Hospital/ West China School of Medicine, Sichuan University, Chengdu, China
[6]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
[7]School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA

*Lan Lan and Fangwei Chen contributed equally.

Corresponding authors:
Tao Zhu, Department of Anesthesiology, West China Hospital/ West China School of Medicine, Sichuan University, No. 37 Guoxue Xiang, Chengdu 610041, China.
Email: 739501155@qq.com

Lan Lan, West China Biomedical Big Data Center, West China Hospital/ West China School of Medicine, Sichuan University, No. 37 Guoxue Xiang, Chengdu 610041, China.
Email: 957715082@qq.com.

Introduction

About 312.9 million surgical procedures are undertaken worldwide each year.[1] Postoperative deaths are the third greatest contributor to all deaths, which accounts for 7.7% of all deaths globally.[2,3] Surgical mortality has declined over the last decade,[4] but the number of patients in the need for critical care monitoring is still increasing.[5,6] Intensive care unit (ICU) admission following major surgery is considered a standard of care in many healthcare systems. However, critical care resources are limited and expensive.[7] Therefore, the appropriate utilization of ICU beds is of great importance. Identifying those at the highest risk of death or complications is essential. The need for critical care monitoring after the surgery is influenced by numerous interacting factors which are classified into patients' preoperative health, the type and quality of surgery, and anesthesia.[8,9] The ICU admission decisions making by physicians had significantly different ICU admission rates, which are affected by the type and seniority of physicians.[10–12] The resource availability is associated with better survival and increasing resource availability may improve patients' outcomes.[13] Therefore, the development of an accurate predictive model including objective clinical variables in the preoperative assessment is required to guide the allocation of resources such as ICU beds.

Several predictive models have been developed with significant results, but there are some limitations in these studies. The American Society of Anesthesiologists Physical Status (ASA-PS) scale, which relies on physicians' subjective assessment of patients' preoperative health status, has modest inter-rater reliability in clinical practice.[14] Other scores have their own limitations, such as the inclusion of data that are not available during the preoperative discussion, applicability to only specific patients, moderate accuracy, and precision for prediction.[15–18] Otherwise, models like the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) and MySurgeryRisk were developed as universal surgical scores. They provide specific risks for major complications or death after surgery instead of ICU admission.[19–21] Most surgical risk calculators assume that relationships among independent variables are linear which limits the clinical use of prior models.[22] The best approach to assess the risk of a patient relies on prediction models that simultaneously incorporate a large number of variables and provide estimation of events' risks.[23] There is evidence suggesting that machine learning techniques may offer better predictive performance when data input are abundant and variable interactions are complex.[24–26] In this context, machine learning techniques are increasingly being used in various clinical fields[27] and artificial intelligence should be used to augment operative management.[28,29]

In this study, we developed a highly discriminative machine learning model to predict ICU admission using the easily available information from electronic health records (EHRs), and created a personalized prediction model for a given patient by identifying and utilizing data from similar patients. Further identifying which perioperative factors are associated with postoperative ICU admission may help manage the surgical risk.

Methods

Data source and study population

A single-center cohort analysis was performed, consisting of the patients from a previously assembled cohort of 427,283 inpatients who underwent surgery between 1 January 2014 and 31 July 2018 in West China Hospital, Sichuan University. Patients who underwent cardiac surgery, emergency surgery, ambulatory surgery, and minor surgery requiring no anesthesia were excluded. Admissions without surgery records from 2014 to 2015 for system reasons were also excluded. Additionally, to ensure the independence of data, only the first surgical procedure was included in patients who underwent multiple surgeries. Therefore, 135,442 patients were analyzed in this study (Supplementary Figure S1, supporting information). This study was approved by the ethics review board of West China Hospital, Sichuan University, with a waiver of informed consent because of its retrospective study nature.

Outcome and predictors

Based on an extensive review of all variables in the database and the objective evidence, we selected 99 available preoperative variables which were extracted and integrated with structured query language[30,31] (online Supplementary material, supporting information). The primary outcome of interest was ICU stay >24 h because patients who were

discharged from ICU within the first 24 h may have been safely monitored postoperatively in a lower intensity unit.[17] Variables routinely assessed during the preoperative period such as patients' demographics, comorbidities, operative characteristics, and preoperative laboratory tests, which can influence the outcome, were taken into account.

Preoperative comorbidities were recorded using the International Statistical Classification of Disease and Related Health Problems, 10th Revision (ICD-10) codes, and Charlson's comorbidities.[32] Preoperative laboratory tests were the latest taken before the start time of surgery. Surgery details included anesthesia type (general or regional anesthesia), incision type, estimated healing type, surgery type, planned surgery, surgery class, estimated duration of operation, the antimicrobial used before the operation, ASA-PS class, and surgical procedures. The types of surgical procedures were identified by the primary procedure International Classification of Diseases Clinical Modification of 9th Revision Operations and Procedures (ICD-9-CM) codes. We removed minor endoscopic and interventional radiology procedures requiring no anesthesia.[33] As some procedure codes include only a small number of patients, these procedures were classified into 12 basic groups based on the anatomical location of surgery of the ICD-9-CM classification (online Supplementary material, supporting information). Vital signs, such as systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate, and pulse were measured.

Data preprocessing

Outlier and missing values were taken into consideration in data preprocessing. For continuous variables, observations beyond the top and bottom 1% of the actual distribution were considered outliers which were then imputed with the random numbers from 1% to 25% percentiles and 75% to 99% percentiles, respectively. The multivariate imputation method was used to estimate and impute the missing values with the information of other variables in our dataset.

Statistical analysis

The observations were randomly separated into a training set (70%, n = 94,810) to develop the models and a testing set (30%, n = 40,632) to test the performance of each model.

In order to resolve the class imbalance problem, Synthetic Minority Oversampling Technique (SMOTE) was used in the development cohort,which can achieve better classifying performance than just the copy of existing minority cases.[34]

Categorical variables were expressed as frequencies and percentages and chi-square test was used to test for their differences. Continuous variables were expressed as medians and interquartile ranges (IQRs) and Mann–Whitney U-test was used to test for their differences.

To construct the predictive model, we first identified 99 initial available variables based on clinical knowledge. We then carried out feature selection using recursive feature elimination as a wrapper method on top of random forest to find a subset of predictors that can be used to produce a more parsimonious and accurate model. The optimal number of features was 18 with automatic tuning of the number of features selected with 10-fold cross-validation. As a result, the final machine learning model with 18 input variables only was selected in the subsequent analysis.

The following machine learning algorithms were employed: logistic regression (LR), random forest (RF), adaptive boosting (ADA), and gradient boosting machine (GBM) algorithms. These models were chosen due to their widespread use in the machine learning field. Models were developed on the training set with 10-fold cross-validation. And we produced the same sample to train the model with setting seeds.

For each model, we calculated its sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, F1 score, and the area under the receiver operating characteristic curve (AUROC) on the test set to measure the performance of model. The receiver operating characteristic curve (ROC) plots were used to evaluate the performance of classifiers. At the same time, we used the calibration plot and Brier score to evaluate the calibration of the models.

Secondary analysis

We conducted four secondary analyses after building the models to improve the interpretability of the results.

(1) We estimated the relative influence provided by GBM and reported the impact of different features on predicting ICU admission. (2) According to the results of the testing cohort, the model that attained the best performance was chosen as the final model and the individual patient's predicted probabilities of outcome was calculated. On the basis of predicted probabilities, we explored the distribution of probabilities across the hospital length of stay (LOS). (3) Eighteen features were selected from 99 available variables to construct our final model. In order to assess the performance of ASA-PS in predicting ICU admission, the evaluation metrics were compared between the final model and ASA-PS score. This analysis was chosen because ASA-PS score is a well-recognized and traditional risk stratification method. (4) We also proposed a non-negative matrix factorization (NMF) bi-clustering strategy for our dataset that partitions both the rows and the columns simultaneously, which can improve the interpretability of the model. We then explored their subsequent mortality rate using the Kaplan–Meier method.

Sensitivity analysis

To further validate machine learning prediction performance, we reprocessed the variable measured multiple

times before surgery according to time series. Since pre-operative variables such as laboratory tests were not uniformly sampled, we took Hyland et al.'s approach[35] to process time series data. We replaced each of the original time series variables with five new features: mean, median, maximum, minimum, and range. We then carried out feature selection using recursive feature elimination and used the four machine learning algorithms to predict.

Statistical analyses were performed with R software (version 3.6.2; The Comprehensive R Archive Network: http://cran.r-project.org).

Results

Patients' characteristics

A total of 135,442 patients were analyzed in the study, and Supplementary Figure S1 (supporting information) shows the inclusion/exclusion process. Of these, the median (IQR) age was 52 (42, 64) years, 71,162 patients (52.5%) were male and the median (IQR) LOS was 9 (7, 14) days. Overall, 2702 (2.0%) patients were admitted to the ICU postoperatively.

The demographics and important characteristics of patients were summarized in Table 1. Patients requiring ICU admission were older, more likely to be male, longer hospital LOS, more likely to undergo complex surgeries, and had more comorbidities than the patients who were not admitted to the ICU postoperatively. In terms of laboratory tests, patients admitted to the ICU had lower percent of monocyte, calcium, albumin, total protein, cystatin C, high-density lipoprotein (HDL) and cholesterol, as well as higher prothrombin time (PT), mean corpuscular hemoglobin concentration (MCHC), chloride, and activated partial thromboplastin time (aPTT), Serum inorganic phosphorus, and white blood cells (WBC). All predictors were significant between ICU group and non-ICU group, except for fibrinogen (FIB).

Model performance

The distribution of preoperative features and outcome did not differ between development (n = 94,810) and testing (n = 40,632) cohorts.

We extracted 18 impactful features from the 99 initial available variables based on recursive feature elimination with the visualization (Supplementary Figure S2, supporting information) where the blue line represents the optimal number of trees. The model restricted to the 18 most predictive variables had the highest performance (accuracy: 0.98). We listed all initial inputs in online Supplementary material (supporting information) and parsimonious model with 18 input variables in Supplementary Table S1 (supporting information).

After applying the various machine learning algorithms to the testing cohort, we compared their performance as measured by the ROC plots and other evaluation metrics. Figure 1 shows the ROCs of the four candidate models and Table 2 summarizes their evaluation metrics. Among the four models, the GBM model attained the highest AUROC of 0.90, accuracy of 0.96, and F1 score of 0.34. The Brier scores were 0.1282, 0.0574, 0.0836 and 0.0635 for LR, RF, GBM, and ADA, respectively. The calibration plot is shown in Supplementary Figure S9 (supporting information). Combining discrimination and calibration, GBM had the most outstanding prediction performance.

Feature importance

Supplementary Figure S3 (supporting information) shows the top 10 influential predictors and their relative influence provided by GBM. Across these features, the GBM recognized several features as highly significant predictors of postoperatively ICU admission, including surgical procedure, calcium, percent of monocyte, albumin, and PT.

Comparison of risk groups

We grouped patients by surgical procedure and compared their LOS. GBM algorithm was chosen to calculate the risk probabilities for ICU admission to each patient and we then stratified these risk probabilities into two groups over the spectrum of total hospital LOS (Figure 2).

Stratified by individual surgical procedure, the relationship between the predicted probabilities and the observed clinical decision for postoperative ICU admission remained consistent across all of these procedures, with higher rates of ICU admission for patients with higher predicting probabilities. The predicting risk probabilities for postoperatively ICU admission were distinctly different between high and low-risk surgical procedure groups. However, the number of ICU patients in the low-risk group was usually very small.

The percentage of postoperative ICU admissions for each surgical procedure are shown in Table 3 and their hospital LOS are shown in Supplementary Figure S4 (supporting information). Operations on the digestive system, respiratory system, and vascular had the highest percentage of postoperative ICU admissions. The distribution of hospital LOS for the patients with and without ICU admission is shown in Supplementary Figure S5 (supporting information).

Comparison with ASA-PS model

The ASA-PS scale has been widely used in the preoperative assessment of surgical patients. Therefore, we compared the ASA-PS score only with the final GBM model without ASA-PS (Table 4). Compared to ASA-PS score (AUROC: 0.68), the GBM had the higher discrimination (AUROC: 0.90). Using only physicians' subjective assessment, the ASA-PS attained modest inter-rater reliability in clinical practice (Figure 3). The Brier score of 0.4229 for

Table 1. Baseline characteristics and predictors stratified by ICU status.

Variable	No ICU admission	ICU admission	*P*
Age, years	52.00 [42.00, 64.00]	58.00 [48.00, 67.00]	<0.001
Male, sex	69,492 (52.4)	1670 (61.8)	<0.001
Charlson's comorbidity index	2.00 [1.00, 6.00]	6.00 [2.00, 8.00]	<0.001
Laboratory tests			
PT	11.40 [10.90, 12.10]	12.00 [11.00, 13.30]	<0.001
aPTT	28.2 [25.70, 31.00]	29.80 [26.40, 34.50]	<0.001
Serum inorganic phosphorus	1.08 [0.96, 1.22]	1.13 [0.98, 1.30]	<0.001
Total protein	69.50 [65.00, 73.80]	65.10 [57.05, 71.30]	<0.001
Albumin	42.80 [39.40, 45.70]	38.40 [32.70, 43.20]	<0.001
MCHC	327.00 [319.00, 334.00]	330.00 [321.00, 337.00]	<0.001
WBC	5.80 [4.73, 7.18]	7.16 [5.34, 10.25]	<0.001
Percent of monocyte	6.60 [5.40, 7.90]	5.50 [3.90, 7.10]	<0.001
Cystatin C	0.89 [0.78, 1.02]	0.87 [0.75, 1.01]	<0.001
HDL	1.22 [0.98, 1.50]	1.00 [0.69, 1.30]	<0.001
Cholesterol	4.35 [3.70, 5.05]	3.90 [3.01, 4.82]	<0.001
FIB	2.72 [2.27, 3.33]	2.75 [2.23, 3.58]	0.508
Chloride	103.20 [101.20, 105.00]	103.70 [101.10, 106.60]	<0.001
Sodium	141.20 [139.60, 142.90]	141.20 [139.20, 143.00]	<0.001
Calcium	2.26 [2.17, 2.34]	2.15 [2.02, 2.25]	<0.001
ASA-PS class			<0.001
I	7 251 (8.7)	32 (1.3)	
II	61 657 (73.9)	1 337 (53.8)	
III	13 776 (16.5)	1 032 (41.5)	
IV–VI	759 (0.9)	84 (3.4)	
Anesthesia type			<0.001
GA	90 445 (68.1)	2 649 (98.0)	
RA	42 295 (31.9)	53 (2.0)	

(continued)

Table 1. Continued.

Variable		No ICU admission	ICU admission	*P*
Incision type				<0.001
	Type I	53 849 (58.5)	811 (30.6)	
	Type II	36 595 (39.8)	1 742 (65.7)	
	Type III	1 550 (1.7)	97 (3.7)	
Surgical procedure				<0.001
1	Neurologic surgery	15 093 (11.4)	134 (5.0)	
2	Operations on the endocrine system	6 159 (4.6)	45 (1.7)	
3	Eye	6 429 (4.8)	1 (0.0)	
4	Ear nose throat	4 711 (3.5)	8 (0.3)	
5	Operations on the respiratory system	13 658 (10.3)	925 (34.2)	
6	Vascular surgery	10 043 (7.6)	308 (11.4)	
7	Operations on the hemic and lymphatic system	37 450 (28.2)	1 093 (40.5)	
8	Operations on the digestive system	3 538 (2.7)	22 (0.8)	
9	Urological surgery	9 645 (7.3)	19 (0.7)	
10	Operations on the male and female genital organs	2 380 (1.8)	3 (0.1)	
11	Operations on the musculoskeletal system	16 692 (12.9)	132 (4.9)	
12	Operations on the integumentary system	6 941 (5.2)	12 (0.4)	
LOS		9.00 [7.00, 14.00]	19.00 [14.00, 26.00]	<0.001

For continuous variables, data are presented as medians and interquartile ranges (IQRs) and Mann–Whitney U-test was used to test for differences. For categorical variables, data are presented as frequencies and percentages and chi-square test was used to test for association.
ASA-PS, American Society of Anesthesiologists Physical Status; aPTT, activated partial thromboplastin time; FIB, fibrinogen; GA, general anesthesia; HDL, high-density lipoprotein; ICU, intensive care unit; LOS, length of stay; MCHC, mean corpuscular hemoglobin; PT, prothrombin time; concentration; RA, regional anesthesia; WBC, white blood cells.

ASA-PS was much higher than that of 0.0836 for GBM, and Supplementary Figure S9 (supporting information) shows their calibration plot.

Subgroup analysis

We also grouped patients by the type of surgical procedures and compared their averaged clinical cost across four clusters based on the NMF model. As shown in Supplementary Figure S6 (supporting information), after the permutation of the rows and columns of the dataset, the NMF model selected the ID of samples and surgical procedures within the bi-cluster. Cluster 1 (surgical procedures 1, 7, 10, and 12) had medium cost and wide LOS. Cluster 2 (surgical procedure 8) had medium cost and medium LOS. Cluster 3 (surgical procedures 2, 5, and 6) had above-average cost and medium LOS. Cluster 4 (3, 4, 9, and 11) had wide cost and short LOS (Supplementary Figure S7, supporting information). Surgical procedure 1, 2, 3, 4, 7, 9, 10, and 11 had high survival rate, surgical procedures 6, 8, and 12 had medium survival rate, and surgical procedure 5 had low survival rate (Supplementary Figure S8, supporting information).

Sensitivity analysis

Supplementary Table S2 (supporting information) shows that the AUCs of ADA and RF were slightly higher than

that of GBM. However, GBM had the best accuracy and F1 score. Taken together, these two methods using time series data and one time point data as preoperative variables/features have a very similar accuracy, and GBM has the best prediction performance.

Discussion

In this study, we developed prediction machine learning models for the need for ICU admission postoperatively. Using preoperative variables easily available from EHRs, the GBM model was more accurate than other machine learning models and the ASA-PS score. This study shows that machine learning models could be used to improve the discrimination of the prediction model and identify the need for ICU admission after surgery in elective noncardiac surgical patients, which provides an important reference for surgeons to prepare for surgery.
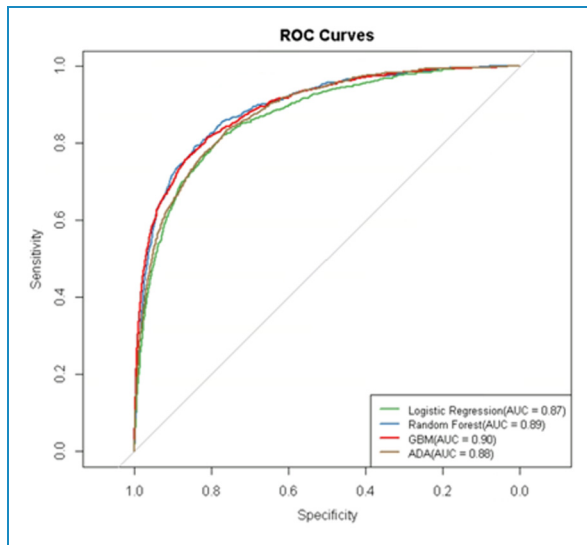
Figure 1. Receiver operating curves (ROCs) of models for the prediction of intensive care unit (ICU) admission.
ADA, adaptive boosting; GBM, gradient boosting machine.

As previously observed, postoperative ICU admission was linked to higher morbidity and mortality. However, most risk scores such as ASA-PS score could not accurately provide the prediction of ICU admission postoperatively. The ability to predict the need for ICU admission after surgery can help develop strategies for a patient's postoperative disposition plan in routine preoperative evaluation, as well as determine protocols directed toward high-risk patients. Previous studies have examined ICU admissions for risk factors[36,37] and in a specific subset of patients.[38] However, there is a lack of studies examining predictions of postoperative ICU admissions. We compared the best model in this study with the performance of ASA-PS score in the same dataset and proved that the machine learning models had better prediction performance than the traditional ASA-PS score in predicting ICU admission.

The application of machine learning to medical and clinical conditions forms a major emerging research trend.[27] First, utilizing machine learning methods could significantly improve the performance of prediction model, which shows that there are a lot of opportunities to improve the performance of clinical prediction model in the field of health.[39] More than 130,000 patients with elective noncardiac surgery were included in this study and they had heterogeneity in demographic characteristics and clinical manifestations. The GBM model, the top performing model in this analysis, still maintained high performance in the prediction of ICU admission. These machine learning approaches may be suited to model the health states in a variety of clinical settings for the relationships of nonlinear and higher dimensional among a large number of variables. A machine learning prediction model has been shown to be more accurate than prior models using LR in prehospital triage in the patients with acute aortic syndrome (AAS).[40] And this study has proved that machine learning methods are more accurate than LR in the prediction of ICU admission. Specifically, our proposed model could be used to identify patients who will be admitted to ICU postoperatively before the operation, and improve the allocation of

Table 2. Model evaluation on testing set for ICU admission.

Model	AUROC	Accuracy	F1 score	Specificity	Sensitivity	PPV	NPV
Logistic regression	0.8680	0.9004	0.2009	0.9917	0.1196	0.6284	0.9059
Random forest	0.8941	0.9429	0.2906	0.9912	0.1932	0.5864	0.9502
GBM	0.8979	0.9597	0.3376	0.9688	0.5148	0.2512	0.9899
ADA	0.8796	0.9227	0.2334	0.9911	0.1455	0.5901	0.9295

ADA, adaptive boosting; AUROC, area under the receiver operating characteristic curve; GBM, gradient boosting machine; ICU, intensive care unit; NPV, negative predictive value; PPV, positive predictive value.
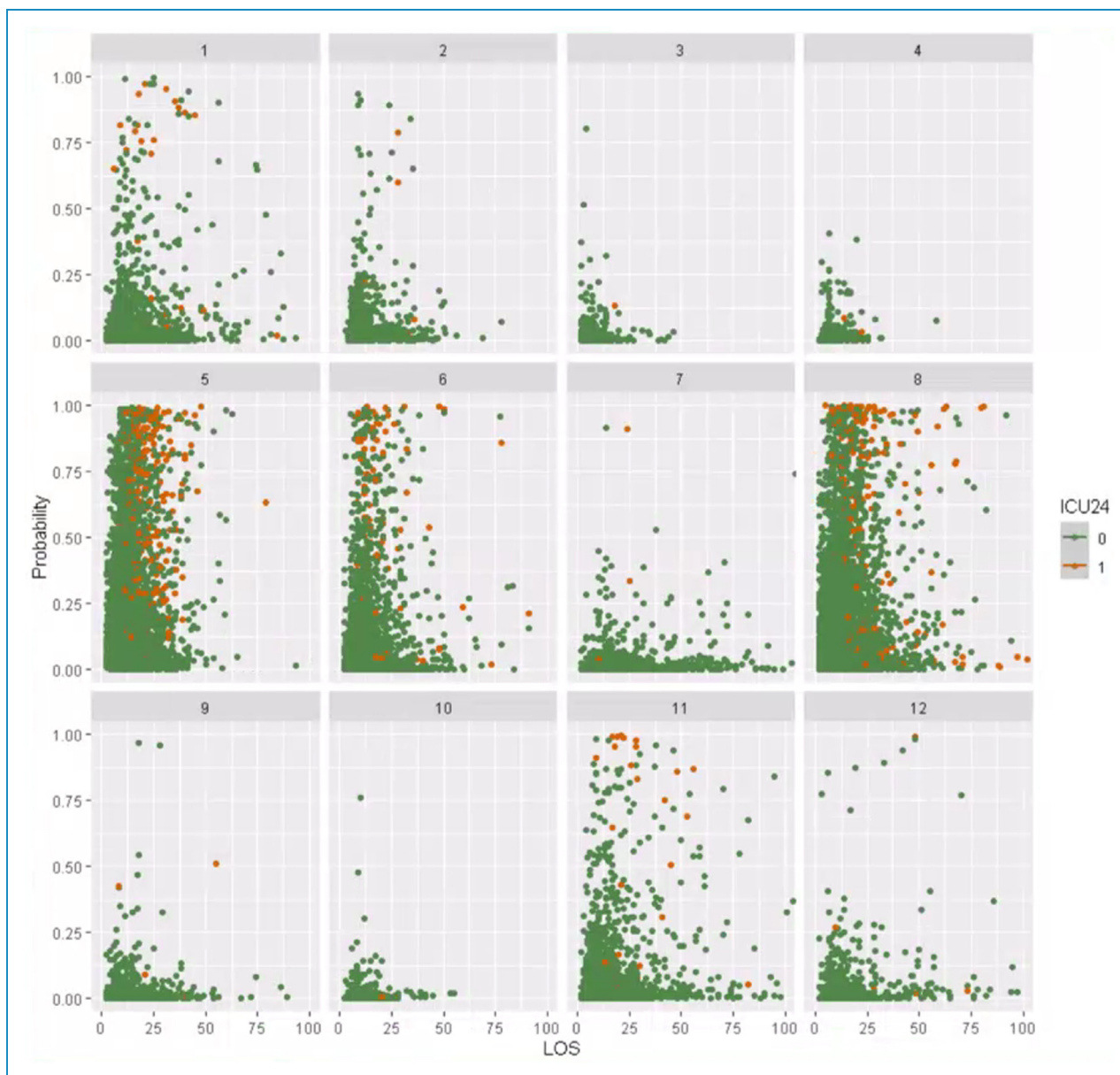
Figure 2. Prediction of intensive care unit (ICU) admission probabilities with hospital length of stay (LOS) among surgical procedures (red represents ICU admission and green represents no ICU admission).

limited ICU resources reasonably. Second, we demonstrate the application of a big data-driven machine learning method in perioperative prediction analysis. Compared with the traditional ASA-PS score, which only focuses on the patients themselves, we consider the variables related to surgery and anesthesia that may be accepted and include the inherent risk of surgery. Our preliminary work included many variables according to the literature and routine available EHRs data; however, the parsimonious model with only 18 variables had nearly the same performance, which indicates that it can be applied more simply and quickly in clinical practice.

Surgical procedure was one of the most important predictors in our study. One strength of this study is the diverse population of patients who underwent surgical procedures. It is common practice to admit patients scheduled for longer and more complex procedures to the ICU by the surgical teams. Of all the surgical procedures analyzed in this study, patients undergoing neurologic, respiratory, cardiovascular, and digestive surgery were the most likely to be admitted to the ICU after surgery. It is unclear if this is due to perceived frailty associated with this population or whether the need for ICU care was actually anticipated by the perioperative personnel.

Table 3. Percentage of cases for each surgical specialty reported to have postoperative ICU admission.

		Training cohort			Testing cohort		
	Surgical procedure	Total	ICU admission n (%)	LOS (median [IQR])	Total	ICU admission n (%)	LOS (median [IQR])
	General	94 810	1 892 (2.0)	9 (7, 14)	40 632	810 (2.0)	9 (7, 14)
1	Neurologic surgery	10 708	97 (0.9)	10 (8, 14)	4 519	37 (0.8)	10 (8, 14)
2	Operations on the endocrine system	4 332	34 (0.8)	9 (7, 11)	1 872	11 (0.6)	9 (7, 11)
3	Eye	4 499	0 (0.0)	6 (4, 7)	1 931	1 (0.1)	6 (4, 7)
4	Ear-nose-throat	3 300	6 (0.2)	7 (6, 7)	1 419	2 (0.1)	7 (6, 7)
5	Operations on the respiratory system	10 221	636 (6.2)	12 (8, 17)	4 362	289 (6.6)	12 (8, 17)
6	Vascular surgery	7 329	214 (2.9)	11 (7, 16)	3 023	94 (3.1)	11 (7, 16)
7	Operations on the hemic and lymphatic system	2 474	15 (0.6)	16 (10, 28)	1 086	7 (0.6)	18 (10, 29)
8	Operations on the digestive system	26 945	770 (2.9)	10 (5, 14)	11 598	323 (2.8)	10 (5, 14)
9	Urological surgery	6 733	12 (0.2)	9 (7, 13)	2 931	7 (0.2)	9 (7, 14)
10	Operations on the male and female genital organs	1 653	2 (0.1)	8 (7, 11)	730	1 (0.1)	8 (7, 11)
11	Operations on the musculoskeletal system	11 763	99 (0.8)	8 (6, 13)	5 061	33 (0.7)	8 (6, 13)
12	Operations on the integumentary system	4 853	7 (0.1)	8 (7, 13)	2 100	5 (0.2)	8 (7, 13)

ICU, intensive care unit; IQR, interquartile range; LOS, length of stay.

Table 4. Prediction performance between machine learning method and traditional score.

Model	AUROC	Accuracy	Specificity	Sensitivity	PPV	NPV
GBM	0.8979	0.9597	0.9688	0.5148	0.2512	0.9899
ASA-PS	0.6829	0.9801	1.0000	0.0000	-	0.9806

ASA-PS, American Society of Anesthesiologists Physical Status; AUROC, area under the receiver operating characteristic curve; GBM, gradient boosting machine without ASA-PS score as input; NPV, negative predictive value; PPV, positive predictive value.

Several limitations are inherent in this study. One is the retrospective design and use of ICD-9 codes to identify patients. In this study, the actual ICU results were used instead of the decision by clinicians,[41,42] and denying ICU admission is common, and age, severity of illness, and diagnosis were important factors in making the decision. Reasons for that include patients too well, patients too sick, lack of beds, and need for more information.[43] The ability to generalize the findings may be limited owing to using the characteristics of patients from a single center. However, the study site is a large tertiary healthcare center with 4300 beds receiving main referrals from southwest of China including Sichuan Province, Chongqing Municipality, Guizhou Province, Yunnan
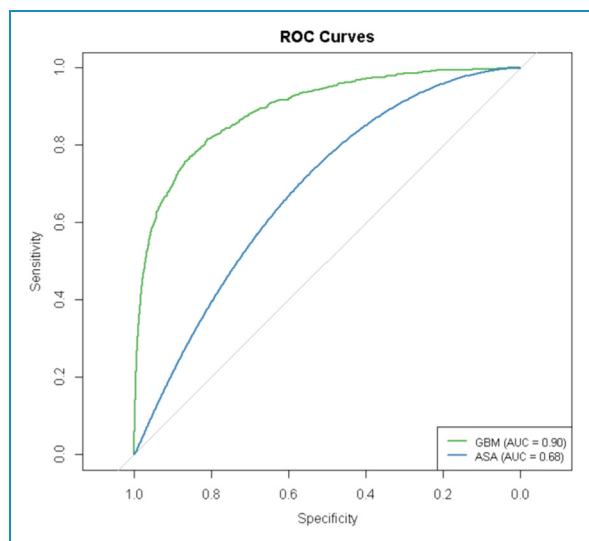
Figure 3. Receiver operating curves (ROCs) for comparing discrimination of gradient boosting machine (GBM) and the American Society of Anesthesiologists Physical Status (ASA-PS) score.

Province, and Tibet Autonomous Region. Therefore, it has a diverse population of patients. And the criteria for ICU admission after surgery might vary based on local needs and resources. We were also unable to define the severity of illness or degree of organ dysfunction for patients that were admitted to the ICU after surgery. Nonetheless, this study serves as the basis of a larger multi-center analysis. Third, obstetrics and gynecology, pediatrics, and stomatology were not included in this study. Lastly, while many approaches based on artificial intelligence to healthcare are criticized for the interpretability, we also visualized the impact of different features to be more explicitly interpretable and intuitive. Further studies are needed to explore the interpretability of machine learning in this area.

Conclusions

In conclusion, the variables in our study are more routinely available and the results based on them can be used more widely. Furthermore, our algorithms included all adult patients with elective noncardiac surgery instead of specific groups, which makes it easier to use. These results provide an opportunity for perioperative optimization interventions in the surgical patients, and further studies are required to assess the impact of empiric admission to ICU after surgery.

References

1. Meara JG, Leather AJM, Hagander L, et al. Global surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *Lancet* 2015; 386: 569–624.
2. Nepogodiev D, Martin J, Biccard B, et al. Global burden of postoperative death. *Lancet* 2019; 393: 401.

3. Biccard BM, Madiba TE, Kluyts H-L, et al. Perioperative patient outcomes in the African surgical outcomes study: a 7-day prospective observational cohort study. *Lancet* 2018; 391: 1589–1598.

4. Fry BT, Smith ME, Thumma JR, et al. Ten-year trends in surgical mortality, complications, and failure to rescue in medicare beneficiaries. *Ann Surg* 2020;271:855–861.

5. International Surgical Outcomes Study. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. *Br J Anaesth* 2016; 117: 601–609.

6. Milbrandt EB, Kersten A, Rahim MT, et al. Growth of intensive care unit resource use and its estimated cost in medicare. *Crit Care Med* 2008; 36: 2504–2510.

7. Ghaffar S, Pearse RM and Gillies MA. ICU admission after surgery: who benefits? *Curr Opin Crit Care* 2017; 23: 424–429.

8. Grocott MP and Pearse RM. Perioperative medicine: the future of anaesthesia? *Br J Anaesth* 2012; 108: 723–726.

9. Short TG, Campbell D, Frampton C, et al. Anaesthetic depth and complications after major surgery: an international, randomised controlled trial. *The Lancet* 2019; 394: 1907–1914.

10. Goel NN, Rodriguez S, Vidal K, et al. Triage decision-making by ICU and emergency medicine physicians: a mixed methods analysis. *Am J Respir Crit Care Med* 2019; 2019: 199.

11. Detsky ME, Harhay MO, Bayard DF, et al. Discriminative accuracy of physician and nurse predictions for survival and functional outcomes 6 months after an ICU admission. *JAMA* 2017; 317: 2187–2195.

12. Does space make waste? The influence of ICU bed capacity on admission decisions. 2013.

13. Motzkus C, Chrysanthopoulou S, Luckmann R, et al. ICU admission source as a predictor of mortality for patients with sepsis. *Chest* 2016; 150: 353A–353A.

14. Sankar A, Johnson SR, Beattie WS, et al. Reliability of the American society of anesthesiologists physical status scale in clinical practice. *Br J Anaesth* 2014; 113: 424–432.

15. Brooks MJ, Sutton R and Sarin S. Comparison of surgical risk score, POSSUM and p-POSSUM in higher-risk surgical patients. *Br J Surg* 2005; 92: 1288–1292.

16. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. Mysurgeryrisk: development and validation of a machine-learning risk algorithm for Major complications and death after surgery. *Ann Surg* 2019; 269: 652–662.

17. Chiew CJ, Liu N, Wong TH, et al. Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission. *Ann Surg* 2020;272:1133–1139.

18. Halpern AB, Othus M, Huebner EM, et al. Developing models to predict intensive care unit admission and mortality for adults with acute myeloid leukemia (AML). *Blood* 2017; 130: 3857.

19. Hyde LZ, Valizadeh N, Al-Mazrou AM, et al. ACS-NSQIP risk calculator predicts cohort but not individual risk of complication following colorectal resection. *Am J Surg* 2019; 218: 131–135.

20. Raymond BL, Wanderer JP, Hawkins AT, et al. Use of the American college of surgeons national surgical quality improvement program surgical risk calculator during preoperative risk discussion: the patient perspective. *Anesth Analg* 2019; 128: 643–650.

21. Lubitz AL, Chan E, Zarif D, et al. American College of surgeons NSQIP risk calculator accuracy for emergent and elective colorectal operations. *J Am Coll Surg* 2017; 225: 601–611.

22. Bertsimas D, Dunn J, Velmahos GC, et al. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive OpTimal trees in emergency surgery risk (POTTER) calculator. *Ann Surg* 2018; 268: 574–583.

23. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models. *Jama* 2017; 318: 1377–1384.

24. Carrano FM, Wang B, Sherman SE, et al. Artificial intelligence outperforms clinical judgment in triage for postoperative ICU care: prospective preliminary results. *J Am Coll Surg* 2019; 229: S141–S142.

25. Bailly S, Meyfroidt G and Timsit J-F. What's new in ICU in 2050: big data and machine learning. *Intensive Care Med* 2018; 44: 1524–1527.

26. Lan L, Guo Q, Zhang Z, et al. Classification of infected necrotizing pancreatitis for surgery within or beyond 4 weeks using machine learning. *Front Bioeng Biotechnol* 2020; 8: 541.

27. Beam AL and Kohane IS. Big data and machine learning in health care. *JAMA* 2018; 319: 1317–1318.

28. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial intelligence and surgical decision-making. *JAMA Surg* 2020;155:148–158.

29. Luo J, Lan L, Peng L, et al. Predicting timing of surgical intervention using recurrent neural network for necrotizing pancreatitis. *IEEE Access* 2020; 8: 207905–207913.

30. Sprung CL, Baras M, Iapichino G, et al. The Eldicus prospective, observational study of triage decision making in European intensive care units: part I–European intensive care admission triage scores. *Crit Care Med* 2012; 40: 125–131.

31. Nates JL, Nunnally M, Kleinpell R, et al. ICU Admission, discharge, and triage guidelines: a framework to enhance clinical operations, development of institutional policies, and further research. *Crit Care Med* 2016; 44: 1553–1602.

32. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. 2005.

33. Bartkowiak B, Snyder AM, Benjamin A, et al. Validating the electronic cardiac arrest risk triage (eCART) score for risk stratification of surgical inpatients in the postoperative setting retrospective cohort study. *Ann Surg* 2019; 269: 1059–1063.

34. SMOTE: Synthetic Minority Over-sampling Technique.

35. Hyland SL, Faltys M, Hüser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat Med* 2020; 26: 364–373.

36. Sobol JB, Gershengorn HB, Wunsch H, et al. The surgical APGAR score is strongly associated with intensive care unit admission after high-risk intraabdominal surgery. *Anesth Analg* 2013; 117: 438–446.

37. Bruceta M, De Souza L, Carr ZJ, et al. Post-operative intensive care unit admission after elective non-cardiac surgery: a single-center analysis of the NSQIP database. *Acta Anaesthesiol Scand* 2020; 64: 319–328.

38. Gray R, Cagliani J, Nauka PC, et al. Evaluation of scoring systems in the early prediction of outcomes in acute pancreatitis. *Am J Gastroenterol* 2017; 112, S13.

39. Pirracchio R, Petersen ML, Carone M, et al. Mortality prediction in intensive care units with the super ICU learner

algorithm (SICULA): a population-based study. *The Lancet Respiratory Medicine* 2015; 3: 42–52.

40. Duceau B, Alsac JM, Bellenfant F, et al. Prehospital triage of acute aortic syndrome using a machine learning algorithm. Br J Surg 2020; 107: 995–1003.

41. Teres D. Civilian triage in the intensive care unit - the ritual of the last bed. *Crit Care Med* 1993; 21: 598–606.

42. Iapichino G, Corbella D, Minelli C, et al. Reasons for refusal of admission to intensive care and impact on mortality. *Intensive Care Med* 2010; 36: 1772–1779.

43. Louriz M, Abidi K, Akkaoui M, et al. Determinants and outcomes associated with decisions to deny or to delay intensive care unit admission in Morocco. *Intensive Care Med* 2012; 38: 830–837.