Development and validation of an explainable machine learning model for predicting myocardial injury after non-cardiac surgery: a retrospective study from two centers in China

Chang Liu, Kai Zhang, Xiaodong Yang, Bingbing Meng, Jingsheng Lou, Yanhong Liu, Jiangbei Cao, Kexuan Liu, Weidong Mi, Hao Li

Submitted to: JMIR Aging
on: November 25, 2023

Table of Contents

Development and validation of an explainable machine learning model for predicting myocardial injury after non-cardiac surgery: a retrospective study from two centers in China

Chang Liu[1]; Kai Zhang[1]; Xiaodong Yang[2]; Bingbing Meng[1]; Jingsheng Lou[1]; Yanhong Liu[1]; Jiangbei Cao[1]; Kexuan Liu[3]; Weidong Mi[1]; Hao Li[1]

[1]Chinese PLA General Hospital Beijing CN
[2]Computer Institute of Chinese Academy of Sciences Beijing CN
[3]Nanfang Hospital, Southern Medical University Guangzhou CN

Corresponding Author:
Hao Li
Chinese PLA General Hospital
Fuxing Road 28th
Beijing
CN

Abstract

Background: Myocardial injury after non-cardiac surgery (MINS) is an easily overlooked complication but closely related to postoperative cardiovascular adverse outcomes, so the improved risk prediction tools are critically needed.

Objective: To develop and validate an explainable machine learning model for predicting MINS in older patients undergoing non-cardiac surgery.

Methods: The retrospective cohort study assessed operations performed on non-cardiac surgical older patients at center 1 in the training set. The least absolute shrinkage and selection operator (LASSO) method and recursive feature elimination (RFE) methods were used to select key features. Prediction performance was measured by the area under the receiver operating characteristic curve (AUC) as the main evaluation metric to select the best algorithms. Validation data from the two datasets were explored to validate the performance of the model and the developed model was compared with the RCRI model. The SHapley Additive exPlanations (SHAP) method was applied to calculate values for each feature, representing the contribution to the predicted risk of complication, and generate personalized explanations.

Results: A total of 12424 patients were included in training set, 4754 in center 1, 2245 in center 2 were included as validating sets. The best-performing model for prediction was CatBoost algorithm, achieving an AUC of 0.805 (95% confidence interval, 0.778–0.831) in the training set, and validating with an AUC of 0.780 in center 1, 0.70 in center 2, with superior performance compared to RCRI (AUC:0.636, P<0.01). The SHAP values indicated the ranking of the level of importance of each variable, and preoperative serum creatinine concentration, red blood cell distribution width, and age accounted for the top three. The results from SHAP method can make predictions towards event with a positive values or non-event with negative value and an explicit explanation of individualized risk prediction.

Conclusions: The Catboost model demonstrated superior capability of predicting individual-level risk of MINS, and the explainable perspective can allow identification of potentially modifiable sources of risk on patient level.

(JMIR Preprints 25/11/2023:54872)
DOI: https://doi.org/10.2196/preprints.54872

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Title page 1

Title: Development and validation of an explainable machine learning model for predicting myocardial injury after non-cardiac surgery: a retrospective study from two centers in China

Running title: Machine learning model for myocardial injury

Chang Liu [1,2,3,*], Kai Zhang [1,2,3,*], Xiaodong Yang [4,*], Bingbing Meng [1,2], Jingsheng Lou [1,2,3], Yanhong Liu [1,2,3], Jiangbei Cao [1,2,3], Kexuan Liu [5], Weidong Mi [1,2,3#], Hao Li [1,2,3#]

[1] Department of Anesthesiology, The First Medical Center, Chinese People's Liberation Army General Hospital (PLA) General Hospital, Beijing, China

[2] Medical School of Chinese PLA General Hospital, Beijing, China

[3] National Clinical Research Center for Geriatric Diseases, Chinese PLA General Hospital, Beijing, China

[4] Institute of Computing Technology Chinese Academy of Science, Beijing, China

[5] Department of Anesthesiology, Nanfang Hospital, Southern Medical University, Guangzhou, China

[*]These authors have contributed equally to this manuscript.

[#]Correspondence to:

Hao Li,

Department of Anesthesiology,

The First Medical Center, Chinese PLA General Hospital,

28th Fuxing Road, Haidian District, Beijing 100853, China.

Telephone:86 15010665099

E-mail: lihao301@126.com

Title: Development and validation of an explainable machine learning model for predicting myocardial injury after non-cardiac surgery: a retrospective study from two centers in China

Background: Myocardial injury after non-cardiac surgery (MINS) is an easily overlooked complication but closely related to postoperative cardiovascular adverse outcomes, so the early diagnosis and prediction are particularly important.

Objectives: To develop and validate an explainable machine learning model for predicting MINS among older patients undergoing non-cardiac surgery.

Methods: The retrospective cohort study included non-cardiac surgical older patients from one northern center and one southern center in China. The datasets from Center 1 were divided into a training set and an internal validation set. And the dataset from Center 2 was used as an external validation set. Before modelling, two methods including the least absolute shrinkage and selection operator (LASSO) and recursive feature elimination (RFE) were used to reduce dimensions of data and select key features from all collected variables. Prediction models were developed based on the extracted features using several machine learning algorithms, including Category Boosting (CatBoost), Random Forest (RF), Logistic Regression (LR), Naïve Bayes, Light Gradient Boosting Machine (LightGBM), eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Decision Tree (DT). Prediction performance was assessed by the area under the receiver operating characteristic curve (AUROC) as the main evaluation metric to select the best algorithms. The model performance was verified by internal and external validation datasets with the best algorithm and compared with the classical Revised Cardiac Risk Index (RCRI). The SHapley Additive exPlanations (SHAP) method was applied to calculate values for each feature, representing the contribution to the predicted risk of complication, and generate personalized explanations.

Results: A total of 19463 eligible patients were included, among those 12464 patients in Center 1 were included as the training set, 4754 patients in Center 1 as the internal validation set, 2245 in Center 2 as the external validation set. The best-performing model for prediction was CatBoost algorithm, achieving the highest AUROC of 0.805 (95% confidence interval, 0.778–0.831) in the training set, and validating with an AUROC of 0.780 in internal validation set, 0.70 in external validation set, also with superior performance compared to RCRI (AUROC:0.636, $P<0.01$). The SHAP values indicated the ranking of the level of importance of each variable, and preoperative serum creatinine concentration, red blood cell distribution width, and age accounted for the top three. The results from SHAP method can make predictions towards event with a positive values or non-event with negative value and an explicit explanation of individualized risk prediction.

Conclusions: The machine learning models can provide a personalized and fairly accurate risk prediction of myocardial injury after non-cardiac surgery, and the explainable perspective can help identification of potentially modifiable sources of risk on patient level.

Keywords: myocardial injury after non-cardiac surgery; older patients; machine learning; personalized prediction

Introduction

Myocardial injury after non-cardiac surgery (MINS), a prominent postoperative cardiovascular complication, occurs in approximately 8% to 22% of patients overall [1]. The Vascular Events in Non-cardiac Surgery Patients Cohort Evaluation (VISION) study showed that MINS was the second most common cause of short-term mortality among eight perioperative adverse events [2-3]. MINS is also reportedly an independent predictor of 1-year or long-term mortality [4]. Nevertheless, 90% MINS events were unrecognized because most patients are not presenting ischemic symptoms and a minority of MINS cases are diagnosed by electrocardiogram (ECG) abnormalities and involve typical chest pain symptoms [5]. Therefore, early prediction and identification of patients at higher risk for MINS is critically important to enhancing the outcomes of this underappreciated complications for older patients.

The most common prediction tool available to make identification of high-risk patients is Revised cardiac risk index (RCRI) [6], a universally used screening tool due to ease of use, but with poor performance in other validation sets. American College of Surgeons National Surgeons Quality Improvement Program (NSQIP) surgical risk calculator [7] and Myocardial Infarction or Cardiac Arrest (MICA) [8] were subsequently developed with higher accuracy than RCRI, designed to predict more severe outcomes including death and myocardial infarction, instead of predicting MINS. Another prediction model was derived from MANAGE cohort [9], using three preoperative risk factors and not considering intraoperative factors. ML has been proven more powerful than conventional logistic regression because it can overcome the limitations of statistical methods and even create personalized risk prediction [10]. Recently, two novel ML models were reported to predict the occurrence of MINS. Oh et.al [11] developed a machine learning model and achieved AUROC of 0.78 using 12 variables. However, the population heterogeneity and lacking external validation may limit its generalization to older patients. Nolde et.al [12] put single-layer and multiple-layer variables to different models and achieved the highest AUROC of 0.77 and accuracy of 0.70. Despite comprehensive included variables, anesthesiologists and surgeons are unable to distinguish modifiable risk factors and make targeted intervention to improve outcomes.

Until now, no validated and accurate risk prediction tools for MINS are currently in use. Therefore, the research purpose was to develop and validate a machine learning model that predicts myocardial injury after non-cardiac surgery risk based on surgery data available on admission and intraoperative period and also make interpretation using Shapley Additive exPlanations (SHAP) method, allowing for targeted interventions to modified risk factors, supporting clinical decision-making.

Methods

Patient cohort

We collected data anonymously from our electronic health record (EHR) system, which was an integrated clinical database containing data on all patients who were admitted to hospitals. The dataset derived from older patients (defined as ≥65 years of age) undergoing non-cardiac surgery from January 2017 to August 2019, and the internal validating dataset from subjects enrolled from July 2020 to July 2021 in Center 1 (Chinese PLA General Hospital, in the northern China). We also included non-cardiac surgical patients in Center 2 (Nanfang Hospital of Southern Medical University, in the southern China) from January 2021 to October 2021 as an external dataset. The uniform exclusion criteria were as follows: excluding patients with American Society of Anesthesiologists (ASA) grade of V, with a short operation interval (scheduled for more than one surgery within a week), who had a short surgery duration (<30 minutes), and with more than 20% of missing data. Patients undergoing either elective or emergency surgery were eligible. The study was approved by the Ethics Committee Board of the First Medical Center of Chinese PLA (No. S2019-311-02), and the requirement for informed consent was waived because this was an observational study with minimal risk to patients. The current study conforms to the principles outlined in the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) Statement.

Data processing

Variables from the following categories were collected: demographics, preoperative comorbid conditions and medications, preoperative laboratory results, vital signs, and intraoperative information. For laboratory testing variables with multiple measurements, we used only the last preoperative measurements within one week before surgery for analysis. Total 118 variables from the electronic database were extracted and listed in eAppendix 1. And additional extraction details are displayed in eAppendix 2 in the Supplement. The least absolute shrinkage and selection operator (LASSO) method, which could solve high dimensionality and multicollinearity between variables was utilized. After the initial screening, recursive feature elimination (RFE) combined with five-fold cross-validation was adopted to re-screen and select the best hyperparameters [13]. After final screening, missing values were imputed using multiple imputation [14].

Outcome

The primary endpoint was the incidence of MINS within the first 30 days after surgery. According to the scientific statement from the American Heart Association [15], MINS was defined as at least one postoperative high-sensitivity troponin T (hs- cTnT) of 20 to <65 ng/L with an absolute change of ≥5 ng/L or a high-sensitivity troponin T concentration ≥65 ng/L; or at least one postoperative measurement of troponin I concentration exceeded the uniform 99th percentile, due to a presumed ischemic etiology irrespective of the presence or absence of clinical symptoms and electrocardiographic changes within the first 30 days after noncardiac surgery.

ML Models

Linear and nonlinear ML Models were applied, including Category Boosting (CatBoost) [16], Random Forest (RF) [17], Logistic Regression (LR) [18], Naïve Bayes [19], Light Gradient Boosting Machine (LightGBM) [20], eXtreme Gradient Boosting (XGBoost) [21], Support Vector Machine (SVM) [22], Decision Tree (DT) [23-24]. Above algorithms were implemented using the Scikit-learn, lightGBM, xgboost and catboost Python packages. Each method is described in detail in eAppendix 3.

Model performance and evaluation

Because of the imbalance between the positive and negative events, the random under-sampling technique was used to avoid overfitting by the rationale of eliminating samples from the majority class to make the majority class equal to the minority class, which is a simple but effective way to treat imbalanced datasets. Eight machine learning models with final indicators were developed to predict outcomes. The area under the receiver operating characteristic curve (AUROC) was used as the evaluation standard of the model performance, and classifiers with larger AUROCs were considered to have better prediction efficiency, and the best-performing ML model was chosen by its AUROC. We also calculated the 95% confidence interval (CI) for each model using the advanced bootstrap method. Similarly, the related sensitivity, specificity, and accuracy were assessed in models conducted. Appropriate figures were produced for these metrics in the best fitting model, including a precision-recall curve and calibration curve, to show the average precision and difference between the predicted risk and actual risk. The AUROCs were also calculated in the validation sets and in the classical revised cardiac risk index (RCRI) model to compare the efficacy.

Model Interpretation

The SHAP method [24] was used to analyze the importance of features in the model because of the limited interpretability in the ML algorithm. SHAP was used as a scoring metric for feature contributions, through determining the difference between the predicted values with and without each feature for all combinations. The greater the influence a particular value of a sample has on the composition of the model, the farther that point deviates from zero on the x-axis. It is calculated by comparing the predicted values for all combinations among all features. Using SHAP values and summary plot, it is thus possible to determine which features have a significant effect on the prediction and whether this contribution is positive or negative. Moreover, SHAP facilitates individual-level risk prediction and stratification, which is straightforward and understandable by doctors.

Statistical analysis

For the baseline data analysis, continuous characteristics were evaluated by the Shapiro-Wilk normality test and analyzed by either the t test for normally distributed variables or the Mann-Whitney U test for skewed data and are reported as means or median. Categorical variables were compared using the chi-square test or Fisher's exact test and are shown as proportions. Because this was a retrospective exploratory study, no attempt was made to estimate the sample size of the study; instead, all eligible patients in the database were included to maximize the statistical power. For all analyses, a two-sided P value of <0.05 was considered statistically significant. All analyses were performed using Python (version 3.6).

Results

Characteristics

In total, we retrospectively enrolled 12,464 patients (median age, 69 years; 54.5% male) who met the inclusion criteria in Center 1 from January 2017 to August 2019 as training dataset. Finally, 884 patients (7.1%) developed postoperative 30-day MINS. 4754 consecutive patients were enrolled in Center 1 as the internal validation dataset with the incidence of MINS of 8.1% (386/4754). 2245 patients in Center 2 were included as the external validation dataset with the incidence of 11.2% (253/2245). The flow charts of patient enrollment in three datasets were shown in **Figure 1**. In the training dataset, patients with postoperative MINS tended to be older; have more chronic conditions such as hypertension, diabetes mellitus, and cerebrovascular diseases; and have more abnormal laboratory test values. The differences in the demographic and other characteristics between patients with and without MINS are summarized in **Table 1**.

Feature selection

Through LASSO, we found that the optimal number of features for model prediction was 27 (**Figure 2A**). The RFE method was used to repeat the model building and feature selecting procedure, finally resulting in 25 features by excluding myocardial infarction history and facility (**Figure 2B**). The features selected by LASSO and RFE are listed in the eAppendix 2.

Model performance and comparison

The training dataset from Center 1 was used to develop the forecast models and MINS was predicted with an AUROC of 0.805 (95% CI, 0.778–0.831) by the best-performing CatBoost method, compared with other seven algorithms. CatBoost revealed a relatively high accuracy (0.730; 95% CI, 0.716–0.745), sensitivity (0.747; 95% CI, 0.694–0.797), and specificity (0.729; 95% CI, 0.714–0.744). The overall AUROC by all algorithms is shown in **Figure 3A**. The average accuracy, sensitivity, and specificity calculated by all the algorithms are summarized in eAppendix 4. The model was well-calibrated with Brier score loss of 0.18, and its calibration plot is depicted in eAppendix 5.

To verify the stability of our model, prediction was validated with an AUROC of 0.794 in the internal validation set of Center 1 and 0.70 in the external validation set of Center 2 respectively, by the method of CatBoost, and their AUROC curve were displayed in the **Figure 3B, 3C**. Bringing the six parameters in the RCRI model into the validation dataset, the poor prediction performance was achieved with an AUROC of 0.636, inferior to our machine learning models ($P<0.01$) (**Figure 3D**).

Model Interpretation

Assisted by the development of explainable ML, the SHAP values for prediction of MINS were calculated. **Figure 4A** shows the 20 most influential factors ranked by the average absolute SHAP value, and **Figure 4B** shows their effect values and interpretations. In the graph, the red dots represent high risk, and the blue dots represent low risk. A higher sCr, higher RDW, older age, increased blood loss, higher blood glucose concentration, higher ASA grade, longer duration of intraoperative hypotension, longer surgery duration, greater infusion of crystalloids or colloids, lower RBC count, lower lymphocyte count, lower albumin, lower sodium, and lower hemoglobin were associated with a higher predicted probability of postoperative MINS. Invasive arterial pressure monitoring, blood transfusion, pre-existing coronary heart disease, and pre-existing hypertension also increased the event risk.

In addition, a visualization method [25] was used to make patient-level prediction interpretations of the model. We provide two examples to illustrate this in **Figure 5**. An 81-year-old patient with ASA grade Ⅲ underwent surgery with a nearly 2.5-hour duration of anesthesia and developed MINS. His preoperative laboratory test values are listed in Figure 5A. The arrows indicate the influence of each feature on prediction; the red arrows suggest an increased risk of the outcome, and the blue arrows suggest a decreased risk. The predicted score of MINS (approximately 3.11) was 30 times higher than the base value predicted by the model (approximately 0.1). Conversely, the second patient with preoperative normal laboratory measurements, with intraoperative blood transfusion, blood loss of 400 ml and intraoperative short hypotension was not suffered from MINS with a predicted score of -0.72, lower than the base value of 0.1.

Discussion

Principal Findings

This cohort study utilized machine learning approaches with multiple demographic and clinical data from EHR to predict the occurrence of postoperative myocardial injury. Catboost algorithm achieved the best predictive performance in the training dataset and validated in both internal and external datasets, with high sensitivity and specificity, also superior to classic RCRI model. The SHAP method also provided information on the contribution of each variable towards event or non-event and thus quantify the association between variables and outcome of a single patient. Our results arose to assist in accurate and timely identification of older patients at high risk of postoperative myocardial injury and provided improved clinical decision support.

The RCRI is considered a conventional predictive model and has been widely used for more than 20 years [6]. Although it has the merit of simplicity with 6 indicators, its use is limited in clinical practice because of its low discriminative ability and lack of specific and sensitive biomarkers for MINS [26]. In our study population, the RCRI model can only achieve an AUROC of 0.636, significantly lower than our model projections. NSQIP surgical risk calculator and MICA were validated to better estimate cardiovascular risk than RCRI, however, NSQIP MICA scores provided only fair discrimination with C-statistic of 0.70 for postoperative myocardial infarction and MINS outcomes in another external validation research [27]. Our study did not compare our models with NSQIP surgical risk calculator and MICA as several key indicators need to be collected prospectively and are not available in our data. Another prediction model by logistic regression was derived from MANAGE cohort, using only three preoperative risk factors, not accounting for intraoperative factors which might be important contributors to adverse outcomes [9]. Therefore, neither of these widely used assessment tools performed by logistic regression statistics has yet been shown to have sufficient predictive strength and applicability.

Recent work has highlighted the strengths of ML algorithms for predicting postoperative complications than classic statistical analysis because it can eliminate nonlinear interactions between clinical variables and resolved the imbalance problem. Oh et.al [11] developed the prediction model using extreme gradient boosting algorithm and achieved an AUROC of 0.78 through 12 variables. There are 5 variables coinciding in Oh's model and ours: operation duration, age, history of chronic kidney disease, history of coronary artery disease, intraoperative red blood cell transfusion. Other inconsistent variables were due to medication differences, uncollected variables and number of events. Furthermore, there were 6811 patients selected from 43019, and the high exclusion rate (84%) and high incidence of MINS (22%) makes the study left in high risk of bias for selection. The potential risk factors in Oh's study may not be generalizable to our dataset all including older patients. Another machine learning model was developed by Nolde et.al [12], through putting single-layer and multiple-layer variables to different models and achieved 0.71 of highest AUROC. But the model with optimal prediction efficacy also included information of postoperative vital parameters and oxygenation within 1-4 days, making it more challenging for anesthesiologists to identify high-risk patients after procedures immediately. Moreover, despite presentation of variable importance ranking, anesthesiologists and surgeons are also unable to distinguish modifiable risk factor and make targeted intervention to improve outcomes.

In our study, we employed several ML approaches based on different principles, and noticed the prediction efficacy of each approach did not greatly differ from each other, suggesting the promising performance of all advanced ML algorithms for the relatively small and low-dimension data. The LR representing the simplest of all classifiers, was chosen to create a reference model against the performance of other machine models. Based on the principle, the CatBoost and RF have relatively good prediction results in our dataset and CatBoost was chosen for the further analysis . And we also noticed that Naïve Bayes algorithm provides the highest accuracy with the disadvantages of the worse classification performance. The reason for this result might be due to different models deal with sample classification in different ways. For the accuracy index, considering only the percentage of correct classification, while the AUROC index reflecting the ability of classification model to discriminate positive and negative samples and considering the influence of the set threshold on prediction results. Though similar or higher accuracy can be achieved, the discrimination and balance of being misjudged were not considered, while AUROC index using as a complementary measure. Combined with the above reasons, we considered that the CatBoost algorithm has a better prediction effect of MINS due to the highest AUROC, and even with much faster speed, and using default parameters.

In addition, our model not only achieves good prediction effect of MINS, but also explored a model-agnostic interpretation technique on how potential variables contribute to adverse outcomes, which was not explored in previous studies. From SHAP values, features with high importance were confirmed and consistent with previously reported related factors, their positive or negative role can also be represented. The top important features contributing to adverse cardiovascular complications, included preoperative renal dysfunction, inflammatory status, glucose metabolism, anemia, electrolyte disturbances. And the intraoperative hemodynamic and other physiological changes are also important contributors to the occurrence of MINS, including more blood loss, prolonged surgery duration, hypotension, greater infusion of fluids and blood transfusion [28-34]. The SHAP plot is a presentation of the predictions for a single sample, in which each eigenvalue is a value that increases or decreases the prediction and its contribution level, providing intuitive explanations for what led to a patient's predicted risk and quantitative prediction at individual levels. For example, in our first sample patient, we recognized that his high preoperative blood glucose concentration played the greatest negative role in the development of complications. Similarly, in the second sample, intraoperative blood transfusion was considered the strongest risk factor for postoperative MINS. Although the complications are unavoidable mainly due to patient themselves and surgical stimuli, others are modifiable through identification of which specific characteristics of patients predispose them to developing an at-risk status and will prompt early targeted prevention or treatments such as administering insulin to patients with a high blood glucose concentration or taking measures to reduce intraoperative blood loss, which may improve the prognosis. The individual risk estimates may provide the modifiable factors through SHAP method, and the information was clinically meaningful and can be used in multiple surgical scenarios.

Limitations

There are limitations to our study. While the model was with high accuracy, it was highly dependent on data from EHR. When one indicator was missing, the true risk of this patient of adverse outcome may not be reflected. Second, the surgical patient data were obtained retrospectively from two hospitals which may have introduced bias because some potential candidates may not be collected in the EHR. Though external validation was conducted in our model, more validation centers are warranted to support the extrapolation and creditability. Third, some variables were excluded before feature selection, especially those laboratory test values with a rate of missing data (>20%), such as NT-proBNP and C-reactive protein, leading to omission and neglect of important indicators. Last, the present study only enrolled elderly Chinese non-cardiac surgical patients from one northern center and one southern center, and whether the results can be extrapolated to other populations remains uncertain.

Conclusions

These findings suggest that the machine learning technique combining the preoperative and intraoperative variables for predicting MINS with a model-agnostic interpretation are potentially efficient management tools for practitioners to guide their postoperative care planning and management.

Acknowledgments

We thank Prof. Lan Sun and Prof. Wei Wei of Hangzhou Lejiu Healthcare Technology for their technical assistance in data extraction. We thank Angela Morben, DVM, ELS, from Liwen Bianji (Edanz) (www.liwenbianji.cn), for editing the English text of a draft of this manuscript.

Authors' Contributions

Chang Liu: Conceptualization, Methodology, Writing-original draft. Kai Zhang: Data curation, Validation. Xiaodong Yang: Methodology, Formal analysis, Validation. Yu Yang: Investigation, Conception and design of study, Data curation. Bingbing Meng: Data curation, Validation. Jingsheng Lou: Supervision. Yanhong Liu: Supervision. Jiangbei Cao: Supervision. Kexuan Liu: Data curation. Weidong Mi: Resources, Project administration, Writing-review & editing, Funding acquisition. Hao Li: Resources, Project administration, Writing-review & editing, Funding acquisition.

Funding

This research was supported by the National Key Research and Development Program of China (2018YFC2001900) and Beijing Nova Program (Z211100002121171). The authors declared no competing financial interests.

Conflicts of Interest

None.

Multimedia Appendix 1

Supplementary table and other material.

References

1. Vascular events In noncardiac Surgery patIents cOhort evaluatioN Writing Group oboTVeInSpceI. Myocardial

injury after noncardiac surgery: a large, international, prospective cohort study establishing diagnostic criteria, characteristics, predictors, and 30-day outcomes. Anesthesiology. 2014;120(3):564-78. PMID: 24534856
2. Gorka J, Polok K, Iwaniec T, Gorka K, Wludarczyk A, Fronczek J, et al. Altered preoperative coagulation and fibrinolysis are associated with myocardial injury after non-cardiac surgery. Br J Anaesth. 2017;118(5):713-9. PMID: 28486646
3. Levy M, Heels-Ansdell D, Hiralal R, Bhandari M, Guyatt G, Yusuf S, et al. Prognostic value of troponin and creatine kinase muscle and brain isoenzyme measurement after noncardiac surgery: a systematic review and meta-analysis. The Journal of the American Society of Anesthesiologists. 2011;114(4):796-806. PMID: 21336095
4. van Waes JA, Grobben RB, Nathoe HM, Kemperman H, de Borst GJ, Peelen LM, et al. One-Year Mortality, Causes of Death, and Cardiac Interventions in Patients with Postoperative Myocardial Injury. Anesth Analg. 2016;123(1):29-37. PMID: 27111647
5. Devereaux PJ, Szczeklik W. Myocardial injury after non-cardiac surgery: diagnosis and management. Eur Heart J. 2020;41(32):3083-91. PMID: 31095334
6. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation. 1999;100(10):1043-9. PMID: 10477528
7. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg. 2013;217(5):833-42.e1-3. PMID: 24055383
8. Pereira-Macedo J, Lopes-Fernandes B, Duarte-Gamas L, Pereira-Neves A, Mourão J, Khairy A, et al. The Gupta Perioperative Risk for Myocardial Infarct or Cardiac Arrest (MICA) Calculator as an Intraoperative Neurologic Deficit Predictor in Carotid Endarterectomy. J Clin Med. 2022;11(21). PMID: 36362595
9. Serrano AB, Gomez-Rojo M, Ureta E, Nuñez M, Fernández Félix B, Velasco E, et al. Preoperative clinical model to predict myocardial injury after non-cardiac surgery: a retrospective analysis from the MANAGE cohort in a Spanish hospital. BMJ Open. 2021;11(8):e045052. PMID: 34348944
10. Dong J, Feng T, Thapa-Chhetry B, Cho BG, Shum T, Inwald DP, et al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care. 2021;25(1):288. PMID: 34376222
11. Oh AR, Park J, Shin SJ, Choi B, Lee JH, Lee SH, et al. Prediction model for myocardial injury after non-cardiac surgery using machine learning. Sci Rep. 2023;13(1):1475. PMID: 36702844
12. Nolde JM, Schlaich MP, Sessler DI, Mian A, Corcoran TB, Chow CK, et al. Machine learning to predict myocardial injury and death after non-cardiac surgery. Anaesthesia. 2023;78(7):853-60. PubMed PMID: 37070957
13. Lei Y, Li YQ, Jiang W, Hong XH, Ge WX, Zhang Y, et al. A Gene-Expression Predictor for Efficacy of Induction Chemotherapy in Locoregionally Advanced Nasopharyngeal Carcinoma. J Natl Cancer Inst. 2021;113(4):471-80. PMID: 33094348
14. Asch DA, Troxel AB, Stewart WF, Sequist TD, Jones JB, Hirsch AG, et al. Effect of Financial Incentives to Physicians, Patients, or Both on Lipid Levels: A Randomized Clinical Trial. Jama. 2015;314(18):1926-35. PMID: 26547464
15. Ruetzler K, Smilowitz NR, Berger JS, Devereaux PJ, Maron BA, Newby LK, et al. Diagnosis and Management of Patients With Myocardial Injury After Noncardiac Surgery: A Scientific Statement From the American Heart Association. Circulation. 2021;144(19):e287-e305. PMID: 34601955
16. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. 2017.
17. Friesner ID, Feng J, Kalnicki S, Garg M, Ohri N, Hong JC. Machine Learning-Based Prediction of Hospitalization During Chemoradiotherapy With Daily Step Counts. JAMA Oncol. 2024. PMID: 38546697
18. Ye G, Zhang C, Zhuang Y, Liu H, Song E, Li K, et al. An advanced nomogram model using deep learning radiomics and clinical data for predicting occult lymph node metastasis in lung adenocarcinoma. Transl Oncol. 2024;44:101922. PMID: 38554572
19. El Morr C, Jammal M, Bou-Hamad I, Hijazi S, Ayna D, Romani M, et al. Predictive Machine Learning Models for Assessing Lebanese University Students' Depression, Anxiety, and Stress During COVID-19. J Prim Care Community Health. 2024;15:21. PMID: 38546161
20. Fu R, Hao X, Yu J, Wang D, Zhang J, Yu Z, et al. Machine learning-based prediction of sertraline concentration in patients with depression through therapeutic drug monitoring. Front Pharmacol. 2024;15:1289673. PMID: 38510645
21. Guo L, Xu X, Niu C, Wang Q, Park J, Zhou L, et al. Machine learning-based prediction and experimental validation of heavy metal adsorption capacity of bentonite. Sci Total Environ. 2024;926:171986. PMID: 38552979
22. Jia M, Lai J, Li K, Chen J, Huang K, Ding C, et al. Optimizing prediction accuracy for early recurrent lumbar disc herniation with a directional mutation-guided SVM model. Comput Biol Med. 2024;173:108297. PMID: 38554662
23. Valsaraj A, Kalmady SV, Sharma V, Frost M, Sun W, Sepehrvand N, et al. Development and validation of echocardiography-based machine-learning models to predict mortality. EBioMedicine. 2023;90:104479. PMID: 36857967

24. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. JAMA Netw Open. 2021;4(3):e212240. PMID: 33783520
25. Athanasiou M, Sfrintzeri K, Zarkogianni K, Thanopoulou AC, Nikita KS. An explainable XGBoost–based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus. 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)2020. p. 859-64.
26. Ruzycki SM, Prystajecky M, Driedger MR, Kachra R. Peri-operative cardiac biomarker screening: a narrative review. Anaesthesia. 2020;75 Suppl 1:e165-e73. PMID: 31903570
27. Fronczek J, Polok K, Devereaux PJ, Górka J, Archbold RA, Biccard B, et al. External validation of the Revised Cardiac Risk Index and National Surgical Quality Improvement Program Myocardial Infarction and Cardiac Arrest calculator in noncardiac vascular surgery. Br J Anaesth. 2019;123(4):421-9. PMID: 31256916
28. Punthakee Z, Iglesias PP, Alonso-Coello P, Gich I, India I, Malaga G, et al. Association of preoperative glucose concentration with myocardial injury and death after non-cardiac surgery (GlucoVISION): a prospective cohort study. Lancet Diabetes Endocrinol. 2018;6(10):790-7. PMID: 30057170
29. Koster A, Zittermann A, Börgermann J, Gummert JF. No Significant Association Between the Transfusion of Small Volumes of Leukocyte-Depleted Red Blood Cells and Mortality Over 7 Years of Follow-up in Patients Undergoing Cardiac Surgery: A Propensity Score Matched Analysis. Anesth Analg. 2018;126(5):1469-75. PMID: 29064873
30. Durmuş G, Belen E, Can MM. Increased neutrophil to lymphocyte ratio predicts myocardial injury in patients undergoing non-cardiac surgery. Heart Lung. 2018;47(3):243-7. PMID: 29500104
31. Smilowitz NR, Redel-Traub G, Hausvater A, Armanious A, Nicholson J, Puelacher C, et al. Myocardial Injury After Noncardiac Surgery: A Systematic Review and Meta-Analysis. Cardiol Rev. 2019;27(6):267-73. PMID: 30985328
32. Qi W, Zhu J, Wu Q, Wang Q, Li X, Yao D, et al. Maize reas1 Mutant Stimulates Ribosome Use Efficiency and Triggers Distinct Transcriptional and Translational Responses. Plant Physiol. 2016;170(2):971-88. PMID: 26645456
33. Salmasi V, Maheshwari K, Yang D, Mascha EJ, Singh A, Sessler DI, et al. Relationship between Intraoperative Hypotension, Defined by Either Reduction from Baseline or Absolute Thresholds, and Acute Kidney and Myocardial Injury after Noncardiac Surgery: A Retrospective Cohort Analysis. Anesthesiology. 2017;126(1):47-65. PMID: 27792044
34. Liu C, Zhang K, Zhang T, Sha X, Xu Y, Gu J, et al. Higher Preoperative Red Blood Cell Distribution Width Increases the Risk of Myocardial Injury After Noncardiac Surgery in Advanced-Age Patients: A Retrospective Cohort Study. Clin Interv Aging. 2023;18:169-79. PMID: 368185

Table 1. Baseline characteristics of patients with or without MINS at Center l in the training set.

Variable	Non-MINS (n=11580)	MINS (n=884)	Total (n=12464)
Age, years (median，IQR)	69(67,73)	72(68,78)	69(67,74)
Hypertension	5225(45.1)	500(56.6)	5725(45.9)
Coronary heart disease	1282(11.1)	206(23.3)	1488(11.9)
Cerebrovascular disease	836(7.2)	140(15.8)	976(7.8)
Myocardial infarction	122(1.1)	35(4)	157(1.3)
Renal insufficiency	117(1)	50(5.7)	167(1.3)
β-blockers	944(8.2)	140(15.8)	1084(8.7)
Diuretics	578(5)	120(13.6)	698(5.6)
Anticoagulants	845(7.3)	164(18.6)	1009(8.1)
HGB (g/L), (median, IQR)	131(120,142)	122(105,136)	131(120,142)
RBC (10^9), (median, IQR)	4.32(3.96,4.65)	4.04(3.54,4.45)	4.3(3.94,4.64)
SCr (umol/L), (median, IQR)	71.2(60.9,82.9)	78.9(64.8,98.175)	71.6(61.1,83.7)
RDW (%), (median, IQR)	12.8(12.3,13.4)	13.3(12.6,14.4)	12.8(12.3,13.4)
Albumin (g/L), (median, IQR)	40.15(37.8,42.7)	38.2(34.7,41.1)	40(37.6,42.6)
Blood glucose	5.08(4.62,5.85)	5.46(4.77,6.63)	5.1(4.63,5.9)
Lymphocyte count	0.3(0.24,0.36)	0.24(0.18,0.32)	0.3(0.24,0.36)
Surgery duration (min), (median, IQR)	144(90,205)	180(120,260)	145(93,210)
ASA grade, n (%)			
I	116(1)	6(0.7)	122(1)
II	9380(81)	485(54.9)	9865(79.1)
III	2034(17.6)	340(38.5)	2374(19)
IV	50(0.4)	53(6)	103(0.8)

Emergency surgery	207(1.8)	76(8.6)	207(1.8)
Colloid input (ml), (median, IQR)	500(0,500)	500(0,1000)	500(0,500)
Crystalloid input (ml), (median, IQR)	1600(1100,2100)	2000(1300,2600)	1600(1100,2100)
Blood loss (ml), (median, IQR)	100(30,200)	150(50,300)	100(50,200)
Blood transfusion, n (%)	1044(9.0)	222(25.1)	1266(10.2)
Duration of intraoperative hypotension (min), (mean, SD)	16.85(37.8)	29.91(57.6)	17.78(42.3)

Abbreviation: HGB: Hemoglobin; RBC: Red Blood Cell; SCr: Serum Creatine; RDW: Red Blood Cell Distribution Count; ASA: American Society of Anesthesiologists; IQR: interquartile range, SD: standard deviation.

Figures legends

Figure 1. The flowchart of participants selection in two centers in China.

Figure 2. Feature selection by LASSO and RFE with 5-fold cross-validation (CV). **(A)**. Through LASSO, the filtered variables were as follows: renal insufficiency, diuretics, cerebrovascular disease, β-blockers, anticoagulants, hypertension, blood transfusion, coronary heart disease, colloid, blood pressure monitoring method, ASA grade, crystalloid, HGB, surgery duration, sodium, age, lymphocyte, anesthesia duration, duration of intraoperative hypotension, RBC, glucose, RDW, blood loss, albumin, SCr, facility, myocardial infarction. **(B)**. RFE with 5-fold CV method filtered features again and removed two (facility, myocardial infarction) and finally remained 25 parameters as above.

Figure 3. Receiver operating characteristics curve of different models. **(A)**. Eight different machine learning prediction models for MINS using the training dataset from Center 1 **(B)**. Model performance in the internal validation dataset from Center 1 **(C)**. Model performance in the external validation dataset from Center 2 **(D)**. Performance of 6 indicators from RCRI in the training dataset.

Figure 4. The model's interpretation. **(A)**. Bar summary of the most important 20 features according to the mean SHAP values. A higher value of a feature has a greater effect on the model's composition, so the further a point deviates from zero on the x-axis. **(B)**. Summary of most impactful features with interpretation. The red dots represent the high-risk value, and the blue dots represent the low-risk value.

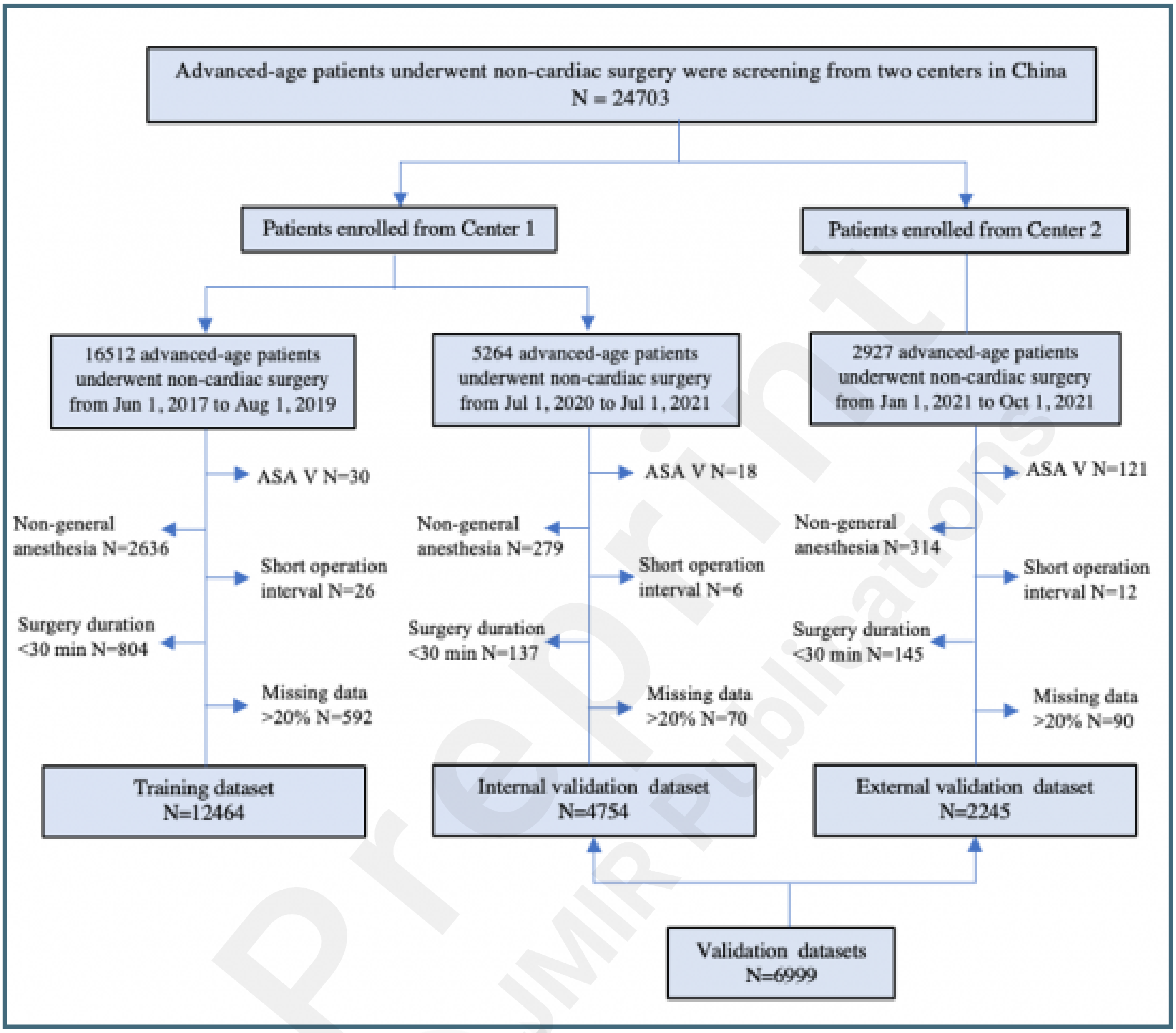
Figure 5. The composition risk of individualized predictions for two patients. A blue arrow indicates that a factor reduced the risk of MINS, whereas a red arrow indicates it increased the risk. (A) An 81-year-old patient with ASA grade Ⅲ underwent surgery with a nearly 2.5-hour duration of anesthesia and developed MINS. (B) A patient with preoperative normal laboratory measurements, with intraoperative blood transfusion, blood loss of 400 ml and intraoperative short hypotension was not suffered from MINS.

Supplementary Files

Untitled.
URL: http://asset.jmir.pub/assets/1bb2e1be8e3ae27868f44512bd935e35.docx

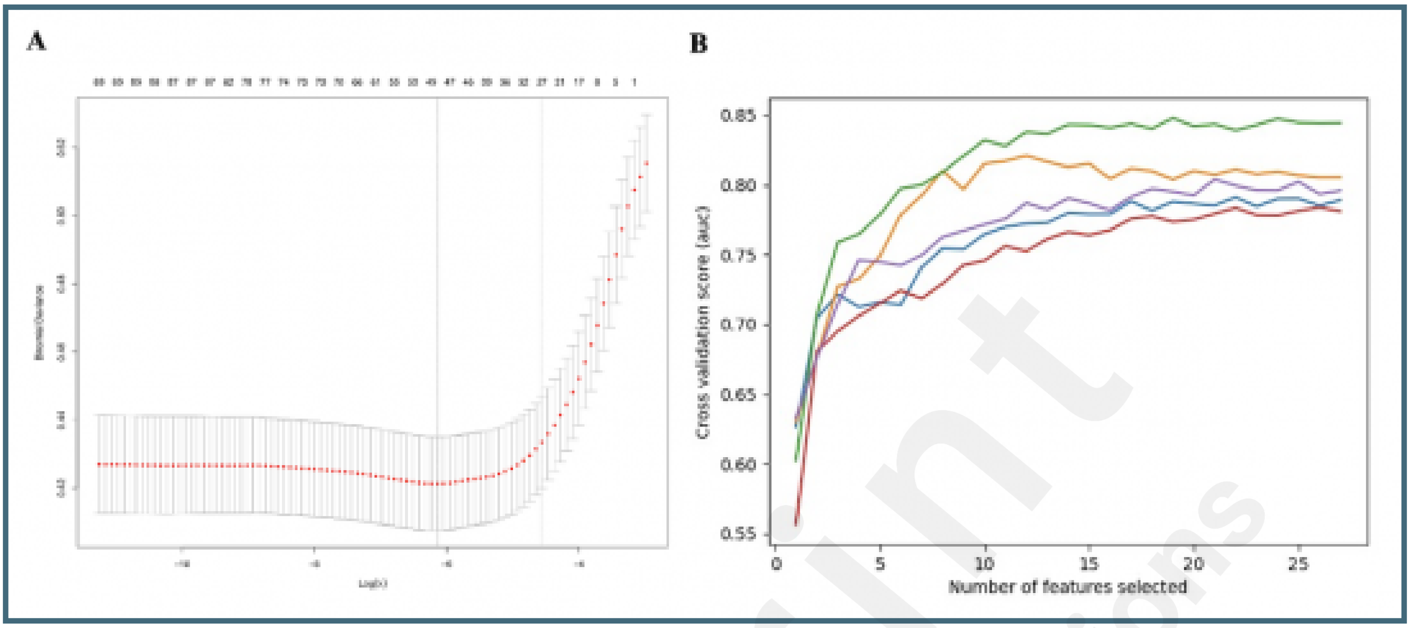
Figures

Untitled.

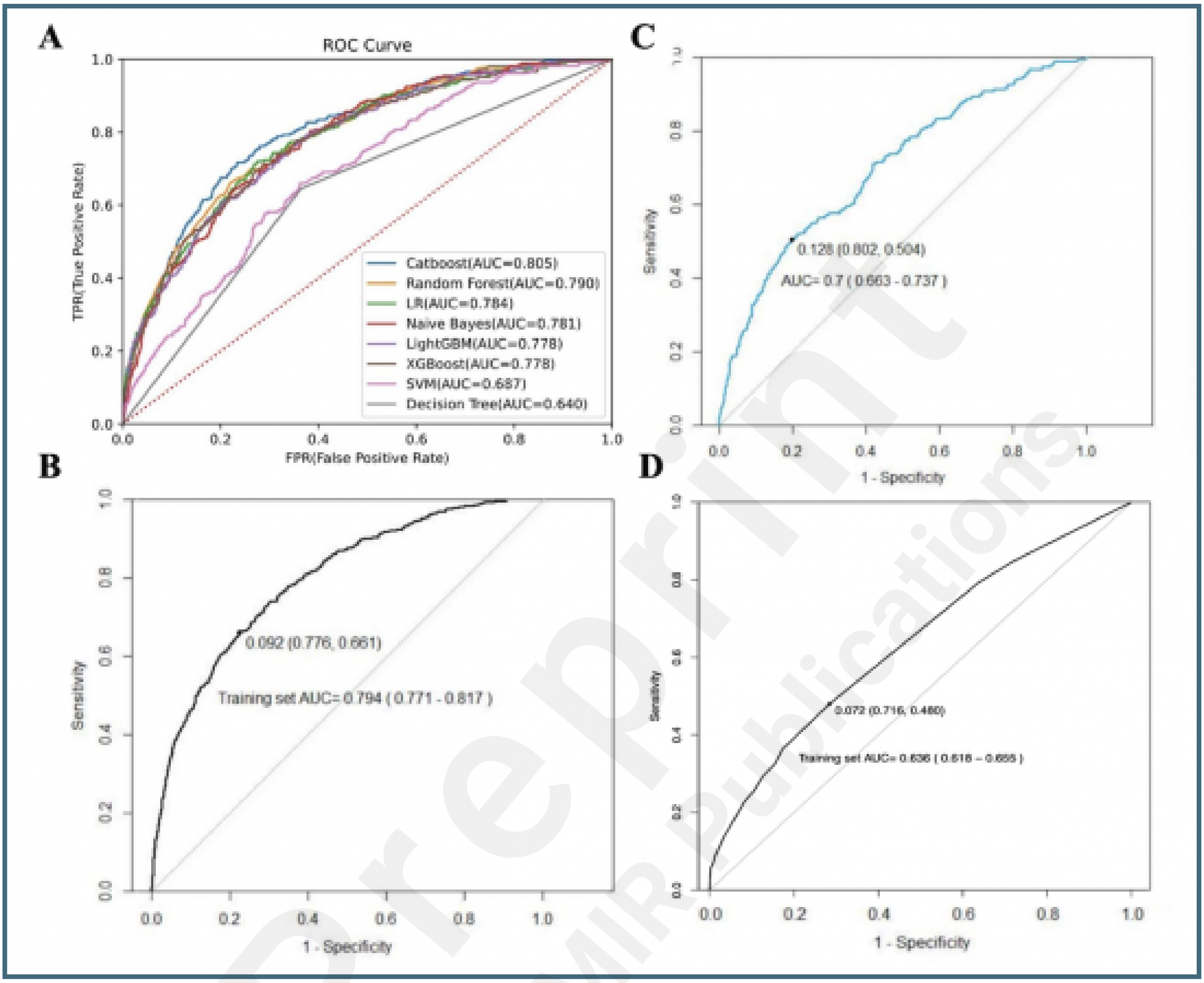


Advanced-age patients underwent non-cardiac surgery were screening from two centers in China N = 24703
Patients enrolled from Center 1
Patients enrolled from Center 2
16512 advanced-age patients underwent non-cardiac surgery from Jun 1, 2017 to Aug 1, 2019
5264 advanced-age patients underwent non-cardiac surgery from Jul 1, 2020 to Jul 1, 2021
2927 advanced-age patients underwent non-cardiac surgery from Jan 1, 2021 to Oct 1, 2021
ASA V N=30
Non-general anesthesia N=2636
Short operation interval N=26
Surgery duration <30 min N=804
Missing data >20% N=592
ASA V N=18
Non-general anesthesia N=279
Short operation interval N=6
Surgery duration <30 min N=137
Missing data >20% N=70
ASA V N=121
Non-general anesthesia N=314
Short operation interval N=12
Surgery duration <30 min N=145
Missing data >20% N=90
Training dataset N=12464
Internal validation dataset N=4754
External validation dataset N=2245
Validation datasets N=6999

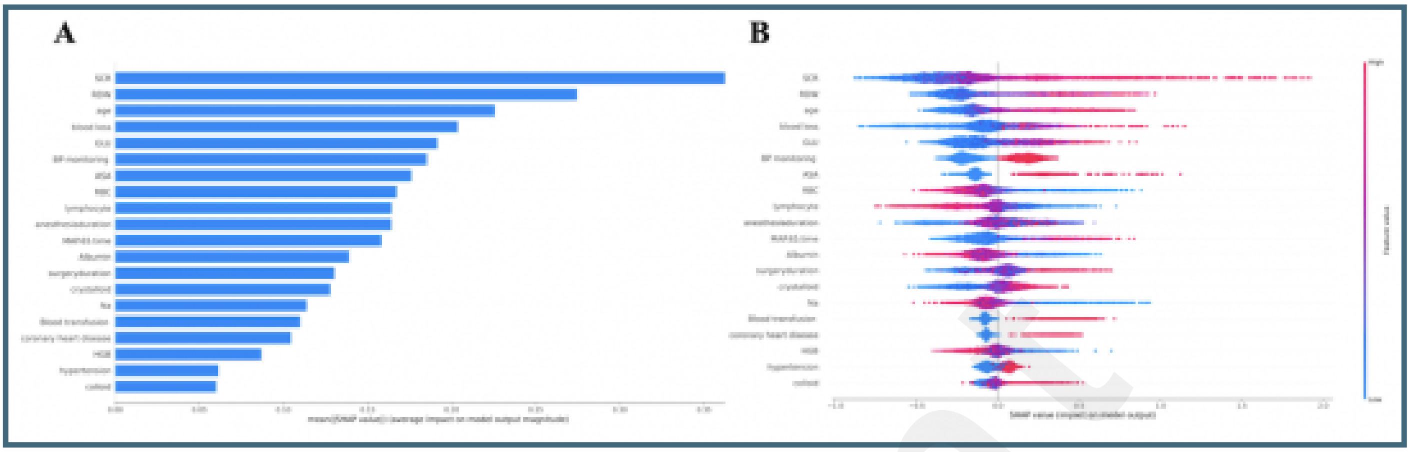
Untitled.



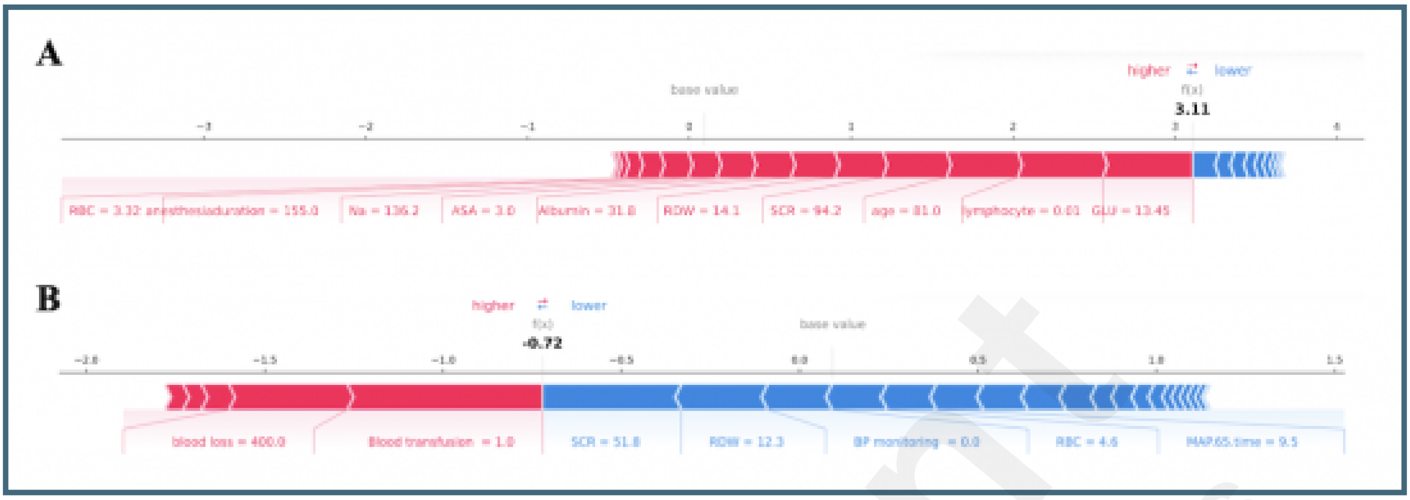
Untitled.


A
ROC Curve
TPR(True Positive Rate)
FPR(False Positive Rate)
Catboost(AUC=0.805)
Random Forest(AUC=0.790)
LR(AUC=0.784)
Naive Bayes(AUC=0.781)
LightGBM(AUC=0.778)
XGBoost(AUC=0.778)
SVM(AUC=0.687)
Decision Tree(AUC=0.640)
B
0.092 (0.776, 0.661)
Training set AUC= 0.794 (0.771 - 0.817)
Sensitivity
1 - Specificity
C
0.128 (0.802, 0.504)
AUC= 0.7 (0.663 - 0.737)
Sensitivity
1 - Specificity
D
0.072 (0.716, 0.480)
Training set AUC= 0.636 (0.618 – 0.655)
Sensitivity
1 - Specificity

Untitled.



Untitled.


A
higher
lower
base value
f(x)
3.11
RBC = 3.32
anesthesiaduration = 155.0
Na = 136.2
ASA = 3.0
Albumin = 31.8
RDW = 14.1
SCR = 94.2
age = 81.0
lymphocyte = 0.01
GLU = 13.45
B
higher
lower
f(x)
-0.72
base value
blood loss = 400.0
Blood transfusion = 1.0
SCR = 51.8
RDW = 12.3
BP monitoring = 0.0
RBC = 4.6
MAP.65.time = 9.5