

---

1           Machine learning-based predictive models for  
2           perioperative major adverse cardiovascular events in  
3           patients with stable coronary artery disease undergoing  
4           non-cardiac surgery

5   Liang Shen<sup>1&</sup>, YunPeng Jin<sup>2&</sup>, AXiang Pan<sup>1</sup>, Kai Wang<sup>2</sup>, RunZe Ye<sup>2</sup>, YangKai  
6   Lin<sup>2</sup>, Safraz Anwar<sup>2</sup>, WeiCong Xia<sup>2</sup>, Min Zhou<sup>1#</sup>, XiaoGang Guo<sup>2#</sup>

7   <sup>1</sup>Department of Information Technology, The First Affiliated Hospital, Zhejiang  
8   University School of Medicine, Hangzhou 310003, China;

9   <sup>2</sup> Department of Cardiovascular Medicine, The First Affiliated Hospital, Zhejiang  
10   University School of Medicine, Hangzhou 310003, China;

11   &. These authors contributed equally to this work and should be considered co-first  
12   authors.

13   #. These authors contributed equally to this work and should be considered  
14   co-corresponding authors.

15   Corresponding authors:

16   Min Zhou, E-mail: minzhou@zju.edu.cn

17   XiaoGang Guo, E-mail: gxg22222@zju.edu.cn

18   **Abstract**

19   **Background:** Machine learning (ML)-based predictive models for perioperative  
20   major adverse cardiovascular events (MACEs) in patients with stable coronary artery  
21   disease (SCAD) undergoing non-cardiac surgery (NCS) have not been reported

22   NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.  
before.

---

23 **Methods:** Clinical data from 9171 consecutive adult patients with SCAD, who  
24 underwent NCS at the First Affiliated Hospital, Zhejiang University School of  
25 Medicine between January 2013 and May 2023, were used to develop and validate the  
26 prediction models. MACEs were defined as all-cause death, resuscitated cardiac arrest,  
27 myocardial infarction, heart failure and stroke perioperatively. Compare various  
28 resampling and feature selection methods to deal with data imbalance. A traditional  
29 logistic regression (the Revised Cardiac Risk index, RCRI) and nine ML models  
30 (logistic regression, support vector machine, Gaussian Naive Bayes, random forest,  
31 GBDT, XGBoost, LightGBM, CatBoost and best stacking ensemble model) were  
32 compared by the area under the receiver operating characteristic curve (AUROC) and  
33 the area under the precision recall curve (AUPRC). The calibration was assessed  
34 using the calibration curve and the patients' net benefit was measured by decision  
35 curve analysis (DCA). Models were tested via 5-fold cross-validation. Feature  
36 importance was interpreted using SHapley Additive explanation (SHAP).

37 **Results:** Among 9171 patients, 514 (5.6%) developed MACEs. The XGBoost  
38 performed best in terms of AUROC (0.898) and AUPRC (0.479), which were better  
39 than the RCRI of AUROC (0.716) and AUPRC (0.185), DeLong test and  
40 Permutation test  $P < 0.001$ , respectively. The calibration curve of XGBoost  
41 performance accurately predicted the risk of MACEs (brier score 0.040), the DCA  
42 results showed that the XGBoost had a high net benefit for predicting MACEs. The  
43 top-ranked stacking ensemble model consisting of CatBoost, GBDT, GNB, and LR  
44 proved to be the best, with an AUROC value of 0.894 (95% CI 0.860-0.928) and an

---

45 AUPRC value of 0.485 (95% CI 0.383-0.587). Using the mean absolute SHAP values,  
46 we identified the top 20 important features.

47 **Conclusion:** The first ML-based perioperative MACEs prediction models for patients  
48 with SCAD were successfully developed and validated. High-risk patients for  
49 MACEs can be effectively identified and targeted interventions can be made to reduce  
50 the incidence of MACEs.

51

## 52 Lay Summary

53 We performed a retrospective machine learning classification study of MACEs in  
54 patients with SCAD undergoing non-cardiac surgery to develop and validate an  
55 optimal prediction model. In this study, we analyzed the data missing mechanism and  
56 identified the best missing data interpolation method, while applying appropriate  
57 resampling techniques and feature selection methods for data imbalance  
58 characteristics, and ultimately identified 24 preoperative features for building a  
59 machine learning predictive model. Eight independent machine learning prediction  
60 models and stacking ensemble models were built, and the models were evaluated  
61 comprehensively using ROC curve, PRC curve, calibration plots and DCA curve.

62 ● We have adopted a series of widely used machine learning algorithms and model  
63 evaluation techniques to build clinical prediction models, and achieved better  
64 performance and clinical practicability than the classical RCRI model, which has  
65 taken the first step to explore the research in this field.

66 ● The prediction results based on the optimal machine learning model are  
67 interpretable, output the importance ranking and impact degree of the top 20

---

68 features of MACEs risk prediction, and are consistent with clinical interpretation,  
69 which is conducive to the application of the model in clinical practice.

70 **Keywords:** prediction model; machine learning; major adverse cardiovascular events;  
71 non-cardiac surgery; imbalance data; feature selection

72

### 73 **Background**

74 Millions of patients undergo non-cardiac surgeries (NCS) worldwide every year  
75 [1], more than 18% of them accompany with stable coronary artery disease (SCAD)  
76 [2]. Perioperative major adverse cardiovascular events (MACEs) occur in 5.7–10.0%  
77 patients with SCAD undergoing NCS [3], which exceeded significantly compared to  
78 only 2.5–3.0% MACEs occurrence rate in the general population [4]. MACEs  
79 represented a significant source of perioperative morbidity and mortality, including  
80 cardiac arrest, myocardial infarction (MI), heart failure (HF) and stroke [5].  
81 Accordingly, it is important to evaluate the risk of MACEs for patients with SCAD  
82 undergoing NCS.

83 Current guidelines highly recommended the use of predictive models to assess the  
84 risk of perioperative MACEs [6, 7]. The most commonly used models are the Revised  
85 Cardiac Risk index (RCRI) [8]. The RCRI is simple and widely validated worldwide.  
86 However, recent large cohort studies have suggested that the RCRI does not have a  
87 strong discriminatory ability [9], especially in patients with known SCAD [10].

88 Machine learning (ML) is an area of artificial intelligence (AI) where algorithms  
89 are employed for identification of patterns in datasets, and have demonstrated superior

---

90 predictive performance on nonlinear data as compared to conventional linear models  
91 such as logistic or cox regression [11]. It learns from the data, as opposed to  
92 regression which stems from theory and assumptions, benefiting from human  
93 intervention and subject knowledge to specify a model [12].As a result, the  
94 application of innovative machine-learning techniques capable of capturing  
95 nonlinearity in clinical practice is imperative.

96 The objective of this study was to derive and validate a ML model based on  
97 easily acquired preoperative clinical data, that can predict perioperative MACEs in  
98 patients with SCAD scheduled for NCS. As far as we know, such a prediction model  
99 has not previously been reported.

## 100 **Methods**

### 101 *Study design and population*

102 We performed a retrospective machine learning classification study (outcomes  
103 were binary categorical) of MACEs in patients with SCAD undergoing NCS to  
104 develop (train) and validate (test) an optimal prediction model. The study design route  
105 flowchart is shown in Figure 1. The machine learning model predicts a future  
106 diagnosis of perioperative MACEs based on features obtained from preoperative usual  
107 clinical care, including demographics, previous diseases, surgical information,  
108 preoperative electrocardiogram (ECG), preoperative echocardiography and  
109 preoperative laboratory tests results such as hemoglobin levels. This study used data  
110 from 9,171 adult patients with SCAD who underwent NCS at the First Affiliated  
111 Hospital, Zhejiang University School of Medicine (FAHZU) between January 2013

---

112 and May 2023.

113 This study was conducted according to Transparent Reporting of Multivariable  
114 Prediction Models for Individual Prognosis or Diagnosis (TRIPOD) and "Guidelines  
115 for Development and Reporting Machine-Learning Predictive Models in Biomedical  
116 Research: A Multidisciplinary View". It complied with the principles of the  
117 Declaration of Helsinki and was approved by the Institutional Ethics Review  
118 Committee of the FAHZU (No. of ethical approval: IIT20230114A). Written  
119 informed consent was waived owing to the nature of the retrospective study design  
120 and the collected data was managed in a de-identified form. This study was executed  
121 and reported in accordance with STrengthening the Reporting of OBservational  
122 studies in Epidemiology (STROBE) guidelines.

### 123 *Inclusion and exclusion criteria*

124 We extracted the study dataset of patients aged 18 years and older who were  
125 hospitalized for surgery with previous SCAD between January 1, 2013 and May 31,  
126 2023 from FAHZU's clinical data warehouse. The types of surgery were elective NCS  
127 based on the American College of Cardiology (ACC)/American Heart Association  
128 (AHA) guidelines of perioperative cardiovascular evaluation [13]. SCAD was  
129 diagnosed if any of the following conditions were met: angiographic demonstration of  
130 coronary stenosis >50%, history of MI (>3 months before enrolment), history of  
131 coronary revascularization (>3 months before enrolment), positive myocardial  
132 perfusion scintigraphy, positive exercise stress test, or typical symptoms of angina  
133 pectoris with simultaneous signs of myocardial ischemia on the ECG [14]. We

---

134 excluded people who underwent cardiac surgery, emergency surgery, day surgery, and  
135 people who underwent multiple surgeries during a single hospital stay. All patients  
136 were evaluated by routine preoperative assessment.

### 137 *Data collection and preprocessing*

138 The electronic medical record system in FAHZU was used in this study. The  
139 International Classification of Diseases, Tenth Edition (ICD-10) has been used to  
140 extract the target population. We identified all discharges over 10 years from the  
141 surgery department with a diagnosis of CAD. Further manual screening of medical  
142 records was performed according to the inclusion/exclusion criteria. We then listed all  
143 available clinical data from the electronic medical record system and performed  
144 feature selection. The study omitted variables with a high rate of missing values (e.g.,  
145 hs-CRP and troponin I). Finally, a total of 64 pre-operative variables were collected,  
146 including patients' demographics (e.g. age, sex and Body Mass Index (BMI)),  
147 pre-existing diseases (e.g. MI, HF, hypertension and diabetes), surgical information  
148 (e.g. surgical type, duration of surgery (DOS) and general anesthesia (GA)),  
149 preoperative ECG (e.g. abnormal Q waves (AQW) and ST-T wave abnormalities  
150 (ST-Ta)), preoperative echocardiography (e.g. left ventricular ejection fraction(LVEF),  
151 regional wall motion abnormality (RWMA), left ventricle diastolic dysfunction  
152 (LVDD) and pulmonary hypertension (PH)), pre-operative laboratory parameters  
153 (e.g. Hemoglobin (Hb), Fasting blood glucose (FBG) and Creatinine (Scr)),  
154 pre-operative drugs (e.g. Nitrates and Insulin), American Society of Anesthesiologists  
155 Physical Status (ASA PS). The putative predictors were chosen on the basis of

---

156 previous studies and the clinical experiences of the investigators.

157 For each feature, we calculated the missing rate in the training dataset, analyzed  
158 the missingness mechanisms, and then selected the appropriate missing value data  
159 imputation method according to the missingness mechanisms. Additionally,  
160 standardization is essential for ensuring that all feature values are on the same scale  
161 and assigned the same weight. All continuous variables (e.g., BMI, laboratory values)  
162 were scaled using StandardScaler or MinMaxScaler in the Scikit-learn package, the  
163 classification of non-binary variables (e.g., surgical type) were one-hot encoded, and  
164 the variables with ordinal characteristics (e.g., ASA PS) were coded with the ordinal  
165 encoder.

#### 166 *Outcomes (Study endpoints and definitions)*

167 The primary outcome was a composite of MACEs (all-cause death, resuscitated  
168 cardiac arrest, MI, HF and stroke) intraoperatively or during hospitalization  
169 postoperatively. Cardiac arrest was defined as the loss of circulation prompting  
170 resuscitation requiring chest compressions, defibrillation, or both [15]. MI was  
171 defined as acute myocardial injury with clinical evidence of acute myocardial  
172 ischemia [16]. Troponin levels were not routinely checked on all enrolled patients.  
173 They were ordered based on routine clinical practice whenever the treating physician  
174 suspected MI based on the clinical status of the patient or ECG findings. HF was  
175 diagnosed mainly by active clinical symptoms or physical examination findings of  
176 dyspnea, orthopnea, peripheral edema, jugular venous distention, rales, third heart  
177 sound, or chest x-ray with pulmonary vascular redistribution or pulmonary edema



---

178 [17]. Stroke was diagnosed by a neurology consultant based on new neurological  
179 findings that were confirmed by imaging studies [18].

### 180 *Class imbalance*

181 The data set of this study included 8657 negative samples (majority class) and  
182 514 positive cases (minority class), with an imbalance ratio (IR) of 16.84:1, indicating  
183 a serious class imbalance. Most standard machine learning algorithms assume or  
184 expect that classification problems have balanced class distributions of equal costs. As  
185 a result, these algorithms are not efficient at handling the complex and imbalanced  
186 data sets that are prevalent in the real world, especially in the medical field. Solving  
187 class imbalance data is mainly realized from two levels of data and algorithm [19]. In  
188 this study, resampling and feature selection are mainly used to process research data,  
189 and ensemble learning models are compared to explore the most appropriate methods  
190 to deal with data imbalance. In order to comprehensively analyze the classification  
191 performance of imbalanced data sets, area under the receiver operating characteristic  
192 curve (AUROC) and area under the precision and recall curve (AUPRC) are  
193 emphasized in model evaluation.

### 194 *Resampling for class imbalance*

195 Resampling is a technique to balance a dataset by reducing the number of  
196 majority classes or increasing the number of minority classes. Among them, Synthetic  
197 minority over-sampling technique (SMOTE) [20] represents the most widely used  
198 method among the resampling methods. Overfitting caused by random oversampling  
199 can be effectively overcome by interpolating new synthetic instances in the line

---

200 between some minority samples and their k-nearest neighbors. The adaptive synthetic  
201 (ADASYN) [21] sampling method is to use a weighted distribution for different  
202 minority class examples according to their level of difficulty in learning, and minority  
203 samples that are more difficult to learn will generate more synthetic data. The  
204 SMOTE+EEN [22] hybrid sampling method uses Edited Nearest Neighbors (ENN)  
205 technology to clean up overlapping samples after the SMOTE algorithm generates a  
206 new synthetic dataset. In this study, SMOTE, ADASYN and SMOTE+ENN sampling  
207 methods were used to resample the training set data. Then, we controlled the sampling  
208 strategy so that the ratio of positive samples to negative samples in the resampling  
209 dataset is 1. Finally, we trained the eXtreme Gradient Boosting (XGBoost) model  
210 with resampling data combined with cross-validation, and compared the model's  
211 performance metrics on the internal validation set. We used correlation functions in  
212 the Python library imbalance-learn to implement resampling.

### 213 *Feature selection for class imbalance*

214 Feature selection is also a feasible technique to deal with imbalanced  
215 classification problems. More representative feature sets are selected to remove  
216 irrelevant and redundant features, thereby improving classification performance and  
217 efficiency. Feature selection is carried out on imbalanced data to optimize the feature  
218 space, find a space that tends to represent concepts of a few classes, and then correct  
219 the classifier's bias towards the majority classes. This study contrasts four widely used  
220 feature selection methods based on different strategies, including the  
221 correlation-based feature selection (CFS) algorithm [23], Boruta algorithm [24],

---

222 BorutaShap algorithm [25], and recursive feature elimination (RFE) [26]. CFS  
223 algorithm is a multivariate filter method that chooses subsets of features that  
224 themselves are uncorrelated but show high correlation with the class, independent of  
225 any learning method, and successfully applied to mortality prediction in three-vessel  
226 disease [27]. Boruta algorithm is to compare the importance of the real predictor  
227 variables with those of random so-called shadow variables using statistical testing and  
228 several runs of XGBoost algorithm. BorutaShap algorithm is an extension of the  
229 Boruta algorithm that leverages the SHapley Additive explanation (SHAP) value as a  
230 measure of feature importance with XGBoost classifier. The RFE algorithm starts  
231 with a base model built on all features. A specific proportion of the least important  
232 features are then removed and a new base model is generated using the remaining  
233 features. These steps are recursively applied until a single feature is left as input. In  
234 this study, XGBoost, Light Gradient Boosting Machine (LightGBM), Random Forest  
235 (RF), support vector machine (SVM) and logistic regression (LR) were selected as the  
236 base models for RFE. Feature selection takes XGBoost model as performance  
237 evaluation. Based on data resampling, AUROC performance is evaluated on  
238 validation set through cross-validation combined with automatic hyperparameter  
239 optimization, and the optimal performance is taken as input feature set of machine  
240 learning model.

#### 241 ***Machine learning model development***

242 We used random stratification to divide the data set into a training dataset (80%)  
243 and a test dataset (20%). Stratification ensured that the proportions of the cases in the

---

244 training datasets and test datasets were equal, which improved the stability of the  
245 model. The training dataset was used for model building, and the test dataset was used  
246 as a hold-out dataset for external validation and did not participate in model  
247 development (including data balancing processing) and hyperparameter selection. We  
248 used randomly stratified 5-fold cross-validation combined with optimal resampling  
249 strategy on the training data set to adjust the hyperparameters in the model and output  
250 the internal validation performance, which can avoid overfitting and assess the  
251 stability of the models. After obtaining the optimal hyperparameters, we used the  
252 model developed in the training dataset for performance evaluation on the hold-out  
253 testing dataset. In order to achieve the best prediction, eight independent models were  
254 built for this study, including LR, SVM , Gaussian Naïve Bayesian (GNB), RF,  
255 gradient boosting decision tree (GBDT), XGBoost, LightGBM, and categorical  
256 boosting (CatBoost). Tree based ensemble models have been applied to other clinical  
257 tasks with excellent performance compared to traditional machine learning algorithms  
258 [28]. Each model provides the same input variables that are optimally selected based  
259 on feature selection, and in order to avoid collinearity between variables affecting the  
260 performance of the prediction model, multicollinearity and correlation analysis are  
261 performed on the optimally selected samples before modeling. We adjusted the  
262 hyperparameters during the model building process based on the Optuna optimization  
263 library of Bayesian optimization [29], where the optimized measure is the AUROC.  
264 Finally, based on 8 independently optimized machine learning prediction models, we  
265 further used the stacking ensemble model, which has been proven to be superior to

---

266 independent machine learning in many fields [30, 31].

### 267 ***Machine learning model evaluation***

268 We developed the models using the training dataset performed 1000 rounds of  
269 bootstrapping on the hold-out testing dataset to report results. We reported numerical  
270 results for accuracy, precision, recall, F1 score, AUROC, and AUPRC. To evaluate the  
271 overall performance, we plotted receiver-operating characteristic (ROC) curves and  
272 precision–recall curves (PRC). The ROC is the ratio of sensitivity to (1-specificity).  
273 According to the AUROC evaluation of model performance, models with a larger  
274 AUROC are considered to have better performance. On the other hand, the PRC  
275 illustrates the trade-off between recall (sensitivity) and precision (positive predictive  
276 value). Models with high performance tend to have a balance of high recall and  
277 precision, yielding large AUPRC values. The statistical comparison of AUROC values  
278 and AURPC values were each computed using Delong Test and Permutation test [32,  
279 33]. A calibration plot was used to evaluate the agreement between the observed and  
280 expected values based on the probability of perioperative MACEs predicted by  
281 various models, and calculated the calibration metrics of the Brier score [34]. The  
282 clinical application value of decision curve analysis (DCA) evaluation model.

### 283 ***Comparison with RCRI***

284 To determine whether the new developed models in our study would improve  
285 upon discrimination of cardiovascular risk prediction, we also developed a baseline  
286 model that mimics the classical clinical scoring system RCRI [8]. The baseline model  
287 was a logistic regression model that included only RCRI. Our newly developed

---

288 machine learning model was compared numerically and statistically with this baseline  
289 model in AUROC and AUPRC performance.

### 290 *Machine learning model explainability*

291 We analyze and visualize the feature importance of the generated predictive  
292 model to comprehend how the model makes predictions and realize an explainable  
293 machine learning model. We used SHAP to analyze and visualize the effect of feature  
294 importance on perioperative MACEs risk based on best-performing predictive models.  
295 The SHAP value represents the effect of features on the prediction in terms of  
296 direction and range by calculating a weighted average and marginal distribution,  
297 which is calculated by comparing the predicted differences in all possible  
298 combinations containing and withholding each feature.

### 299 *Statistical analysis and modeling tools*

300 The normality of the distribution of continuous variables was tested using the  
301 Shapiro–Wilk test. Normally distributed continuous variables were expressed as mean  
302  $\pm$  standard deviation (SD) and compared using the independent samples t-test.  
303 Skewed continuous variables were expressed as median and interquartile range (IQR)  
304 and compared using the Mann–Whitney U-test. Categorical variables are expressed as  
305 frequencies and percentages and using chi-square tests or Fisher's exact probability  
306 tests. The differences were considered to be statistically significant at  $p < 0.05$ .  
307 Machine learning model development and evaluation was performed in python 3.6  
308 using scikit-learn packages.

---

309 **Results**

310 *Participant characteristics*

311 We eliminated 4,486 patients based on exclusion criteria, and ultimately 9,171  
312 patients were included in our study. Among them, 514 (5.6%) patients suffered  
313 perioperative MACEs, as shown in Supplementary Table 1. Table 1 presents baseline  
314 clinical characteristics of the training and testing sets, respectively, and univariate  
315 analyses with and without MACEs. Overall, the baseline clinical characteristics of the  
316 training set and testing set samples appeared to be similar. Patients underwent a wide  
317 range of surgeries as expected in a tertiary referral hospital with a median age of 70  
318 (IQR, 64–77) years. General abdominal, thoracic, and vascular surgeries were most  
319 often performed.

320 *Missing-value characteristics*

---

321 The average proportion of missing values in this study dataset was 7.69%, and the  
322 proportion of the training and validation datasets (7.62%) and the testing dataset  
323 (7.99%) were basically the same (Supplementary Table 2). The missing values of  
324 preoperative laboratory tests, preoperative ECG and preoperative echocardiography  
325 were mainly due to the fact that the patient did not complete the corresponding  
326 examination in FAHZU, while the missing BMI value is mainly due to the patient's  
327 body being unable to measure normally. At the same time, we tested the correlation of  
328 missing values of different variables (Supplementary Figure. 1a), the correlation  
329 between the variables measured as a companion test (e.g., laboratory, ECG and  
330 echocardiogram tests) was high (absolute correlation value  $\geq 0.7$ ), while the  
331 correlation between BMI, laboratory variables, ECG variables, echocardiogram  
332 variables, and DOS variables was not remarkable (absolute correlation value  $\leq 0.2$ ).  
333 Comparison of data missing distribution and completeness of variables  
334 (Supplementary Figure. 1b, 1c), missing values include categorical and continuous  
335 variables, and there is no uniform pattern of missing values for each variable. Based  
336 on the miss at random mechanism of variables, we compared a variety of missing  
337 value imputation algorithm on the XGBoost model, cross-validation on the training  
338 dataset (Supplementary Table 3), and finally selected the k-Nearest Neighbor (KNN)  
339 imputation algorithm with the best performance in this study.

340 ***Resample method***



---

341 The imbalanced training set data is processed by resampling method to overcome  
342 the performance loss caused by data imbalance. Table 2 presents the internal  
343 verification results of LR, RF and XGBoost models in each training set. As shown in  
344 the table, although the three models also achieved high AUROC and AUPRC  
345 performance before balancing the data, the extremely high specificity and extremely  
346 low sensitivity indicated that the classification model without data balance could not  
347 well identify MACEs patients (minority class) due to the inter-class imbalance in the  
348 data. In contrast, after using SMOTE, ANSYN and SMOTE+ENN for data balance,  
349 the sensitivity of the three models has significantly improved, and the indicators of  
350 AUROC and AUPRC have also increased. The results showed that the data balancing  
351 processing can effectively improve the recognition performance of the classification  
352 model for the few class samples. Further contrast the same prediction model based on  
353 different resampling methods to observe the corresponding changes in specificity and  
354 sensitivity. SMOTE+ENN is the best for data balancing and will be applied to model  
355 development and evaluation.

### 356 *Feature selection*

357 The total number of features in the data set after the pre-processing is 75, and the  
358 features are analyzed and selected sequentially based on eight feature selection  
359 methods. Figure 2A corresponds to the optimal feature subset of the eight feature  
360 selection methods and the AUROC performance of the internal validation set after  
361 cross-validation on the training set. The performance of the optimal subset after  
362 feature selection is better than that of the full feature (Supplementary Table 4), but it

---

363 is also related to the selected feature subset, in which the performance of the feature  
364 subset selected by RFE-XGB and RFE-LR methods is the best. Figure 2B shows that  
365 RFE combines five kinds of basis learning models, selects the best subset recursively  
366 based on the feature importance ranking of the learning model, and evaluating the  
367 AUROC performance of the internal validation set after cross-validation on the  
368 training set. When RFE-XGB and RFE- RF are used for feature selection, AUROC  
369 performance is maintained at a relatively high level when the number of features of  
370 the optimal subset is greater than or equal to 3, and the average AUROC is greater  
371 than 8.0.

372 The performance and interpretability of the model are fully balanced, and the  
373 number threshold of features contained in the optimal feature subset is controlled  
374 between 3 and 30, and the feature combination with the best performance is selected  
375 by comparing eight feature selection methods. These features included patients'  
376 demographics [BMI], pre-existing diseases [Ischemic heart disease (IHD), and  
377 Dialysis], surgical information [DOS], preoperative echocardiography [fractional  
378 shortening (FS), left ventricular end systolic dimension (LVDs), LVEF and RWMA),  
379 pre-operative laboratory parameters [Leukocyte, Hb, FBF, Scr, Estimated glomerular  
380 filtration rate (eGFR), Total serum protein (TSP), Albumin (ALB), AST,  
381 Cholinesterase (ChE), Total bilirubin (TB), Total calcium (tCa), Chlorine, APTT,  
382 Fibrinogen (FB) and D-dimer], ASA PS. Subsequently, we carried out  
383 multicollinearity analysis of the selected features. First, correlation coefficients and  
384 corresponding P-values of the features were drawn in the heat map, and it was found

---

385 that TSP and ALB, Scr, eGFR and Dialysis may have collinearity (Supplementary  
386 Figure 2). The variance inflation factor (VIF) was further used for the feature  
387 multicollinearity test, and VIF values less than 5 indicated weak multicollinearity  
388 (Supplementary Table 5). It indicates that the features selected in this study can  
389 effectively avoid the negative effects of feature collinearity on the classification  
390 performance of the model. We ultimately used the above 24 features for model  
391 development and evaluation.

---

392 ***Machine learning model performance***

393 Eight independent candidate models were constructed for perioperative MACEs  
394 prediction using the twenty-four variables mentioned above. Figure 3 A, B presents  
395 the AUROC and the AUPRC of each candidate modeling method in test-set data. All  
396 eight candidate models exhibited superior prediction performances in terms of  
397 AUROC and AUPRC, compared to that of the Baseline-RCRI model, with significant  
398 differences in both AUROC and AUPRC by using the Delong Test and Permutation  
399 test ( $P < 0.001$ ). The XGBoost method delivered the best performance in terms of  
400 AUROC (0.898) and AUPRC (0.479). Table 3 presents the other metrics of mean  
401 values of bootstrapping performance of each model. Further, the DCA curves (Figure  
402 3C) demonstrate that the eight candidate models exhibited a greater net benefit along  
403 with the threshold probability compared with Baseline-RCRI models. The calibration  
404 curve of the eight candidate models is closer to the curve with a slope of  $45^\circ$  than the  
405 Baseline-RCRI model, indicating the better accuracy (Figure 3D), while the Brier  
406 scores were calculated, none exceeding 0.04, and was better than RCRI (0.05).  
407 Machine learning model hyperparameters are listed in Supplementary Table 6.

---

408 The eight prediction models were developed based on the training set data and  
409 5-fold cross-validation. The AUROC and AUPRC per fold on the internal validation  
410 set of the prediction models are shown in Supplementary Table 7. The results show  
411 that the performance is similar in each fold of the verification set, which indicates that  
412 the prediction model has good stability. Further, the internal verification results of  
413 different prediction models were compared with the external verification results.  
414 Figure 4 shows the comparison results of AUROC (A) and AUPRC (B) performance  
415 indicators of the verification set and the test set. The results show that the  
416 comprehensive performance of the validation set and the test set of the prediction  
417 model is similar, and the performance of the test set is slightly better than that of the  
418 verification set, which indicates that various prediction models in this study have  
419 good generalization performance.

#### 420 ***Model explainability***

421 Based on the optimal independent model XGBoost, by using SHAP analysis  
422 (Figure 5 A, B), we determined the top 20 features including IHD, ASA PS, Hb, DOS,  
423 LVDs, D-dimer, ALB, Chlorine, FBG, ChE, Leukocyte, Scr, RWMA, eGFR, BMI,  
424 TSP, APTT, tCa, Dialysis, and LVEF as important features for predicting new onset  
425 MACEs. In Figure 5A, we presented the relationships between their values and the  
426 effect of the model output. Intuitively, IHD, ASA PS, DOS, LVDs, D-dimer, FBG,  
427 Leukocyte, Scr, RWMA, APTT, and Dialysis were positively correlated with the  
428 MACEs, whereas Hb, ALB, Chlorine, ChE, eGFR, BMI, TSP, tCa, and LVEF were  
429 negatively correlated with the MACEs.

---

430 In addition to the overall effect, we applied the SHAP framework to explain  
431 individual cases by providing influential features. Figure 5 shows 2 examples of  
432 random selection - a negative prediction (C) and a positive prediction (D). Features in  
433 blue represent features that contribute to a lower risk while features in red will push  
434 up the risk. These visualizations give users detailed information about how the model  
435 makes predictions and allow them to make appropriate interventions before the new  
436 onset MACEs.

### 437 **Model Stacking**

438 Stacking ensemble models were subsequently developed, and the stacking  
439 ensemble model is a two-layer structure, the first layer is composed of multiple base  
440 models, and the second layer is fixed as a logistic regression model. Based on the 8  
441 independent models of this study, 247 model combinations were listed by exhaustive  
442 method, and then the stacking model and output performance indicators were  
443 established and sorted. Overall, the top-ranked stacked ensemble model consisting of  
444 CatBoost, GBDT, GNB, and LR proved to be the best, with an AUROC value of  
445 0.894 (95% CI 0.860-0.928) and an AUPRC value of 0.485 (95% CI 0.383-0.587).  
446 Compared to the independent optimal prediction model XGBoost, the stacking model  
447 showed slightly higher AUPRC performance and net benefit value (Figure 6), as well  
448 as higher sensitivity (0.788).

### 449 **Discussion**

450 To our knowledge, this is the first study of a systematic framework based on  
451 machine learning techniques to develop multiple models, evaluate performance, and

---

452 select the highest-performing model to predict perioperative MACEs in patients with  
453 SCAD scheduled for NCS. This study shows that compared with the classical  
454 prediction model RCRI [8], the eight independent machine learning prediction models  
455 and the optimal stacking ensemble model have a great improvement in performance  
456 and clinical utility, and have satisfactory generalization, among which XGBoost has  
457 the best performance in the independent machine learning prediction model. IHD,  
458 ASA PS, Hb, DOS, LVDs, D-dimer, ALB, Chlorine, FBG and ChE were the top 10  
459 important characteristics affecting model performance. In addition, the model  
460 combining CatBoost, GBDT, GNB, and LR is considered to be the best model for  
461 stacking ensemble learning, further improving the AUPRC, clinical utility, and  
462 sensitivity of the predictive model. These findings help to identify perioperative  
463 MACEs risk in patients with SCAD scheduled for NCS before surgery and to provide  
464 targeted clinical care through timely intervention.

465 From the data set used in this study, there are some inherent characteristics in the  
466 data set that affect the classification performance of the prediction model, such as data  
467 missing and data imbalance. This paper proposes a series of research strategies to  
468 improve the performance of machine learning classification algorithms based on data  
469 processing. Missing data is a common occurrence in clinical research, and improper  
470 processing will significantly affect the efficacy of the classification model [35]. First,  
471 the identification of missing data mechanism is the basis of selecting missing data  
472 imputation method. There are three typical mechanisms causing missing data: missing  
473 completely at random (MCAR), missing at random (MAR), and missing not at

---

474 random (MNAR). In this study, the missing data mechanism was determined to be  
475 MAR through correlation test and integrity comparison of missing data. On this basis,  
476 different missing data imputation methods were compared, and the most effective one  
477 was selected according to the performance output of the internal validation set, while  
478 directly deleting missing data would lead to estimation bias [36]. Second, in terms of  
479 data balance, this study reconstructs the data set from the data itself. Primarily,  
480 resampling method is used to optimize the sample space, and then feature selection  
481 method is combined to optimize the feature space. After resampling the unbalanced  
482 data, the performance of all models has been significantly improved, especially the  
483 sensitivity of RF model increased from 0 to 0.874 after the SMOTE+ENN method.  
484 SMOTE+ENN method, after SMOTE algorithm generates new synthetic data set,  
485 uses ENN clearing technology to reduce the problem that SMOTE often introduces  
486 more noise and overfitting to some extent [37]. On the basis of selecting suitable  
487 resampling methods, the feature selection method based on AUC evaluation criteria  
488 was adopted in this study. Some studies have concluded through experiments that  
489 feature selection is more important than classification method selection in order to  
490 overcome overfitting problems and achieve better classification performance [38].  
491 The feature selection method is used to delete category-irrelevant features, reduce the  
492 dimension of data, find a space that tends to represent the concepts of minority classes,  
493 correct the classifier's bias to the majority classes, and solve the unbalanced  
494 classification problem with poor classification performance of minority classes [39].  
495 Although univariate analysis of clinical features was performed in this study,



---

496 threshold filtering features were not directly set according to p-value in the process of  
497 feature selection, mainly considering that univariate analysis might ignore the  
498 interaction between features [40]. The 8 feature selection methods in this study  
499 mainly select features based on the relevance and importance of features, and the  
500 results show that the performance of the model after feature selection is indeed  
501 improved. The internal verification performance comparison of different feature  
502 selection methods finally determines 24 effective features. The distribution of feature  
503 categories was fairly balanced, including 1 feature of patients' demographics, 2  
504 features of pre-existing diseases, 1 feature of surgical information, 4 features of  
505 preoperative echocardiography, 15 features of pre-operative laboratory parameters  
506 and ASA PS. The feature subset after feature selection is continued to be  
507 multicollinear detected and processed by correlation coefficient test and VIF method,  
508 so as to reduce the complexity of classification model construction and improve  
509 model stability and generalization ability.

510       Considering the clinical applicability of the prediction model, the prediction  
511 model in this study uses only routine clinical and laboratory data and selects only a  
512 small number of features, which is conducive to the data being automatically  
513 collected through the program and the model being applied to other institutions to  
514 obtain stable performance. In addition, this study only used preoperative data, not  
515 intraoperative data, so it has the ability to predict prognosis before surgery. The use of  
516 intraoperative data may improve prediction accuracy but may lead to an exaggeration  
517 of model performance and delay in implementing interventions to patients.

---

518 In our study, XGBoost provided the best predictive performance among the  
519 independent models built. Compared with the classical model RCRI, which uses  
520 logistic regression with 6 equal weight variables, the main advantages of XGBoost  
521 model are the ability to capture the nonlinear relationships between the model features  
522 and the outcome, as well as having higher order interactions between features.  
523 Evaluates the performance of the machine learning predict model in external test set,  
524 performs 1000 rounds of bootstrapping sampling method to report the result  
525 confidence interval to evaluate the stability of the model parameters. The ROC curve  
526 and its corresponding AUC are a function of the sensitivity and specificity of the  
527 predictive model and are used to quantify the overall ability of the test to correctly  
528 identify normal and abnormal ones. The prediction models developed in this study all  
529 had AUROC values greater than 0.88 on the test set, and the AUROC value of the  
530 XGBoost model was close to 0.9, which means that on average, the test will correctly  
531 predict abnormal outcomes 90% of the time, and the model has excellent prediction  
532 ability. Compared with the ROC curve, the PRC curve assesses the true proportion of  
533 the positive prediction and provides more information on the prediction assessment of  
534 the imbalanced dataset. The AUPRC value is low compared to the higher AUROC  
535 value, and the AUPRC value of the prediction model we developed is between 0.39  
536 and 0.48, but it is still much higher than the classical model RCRI (AUPRC=0.185).  
537 The reason for the low AUPRC value is that the low incidence of perioperative  
538 MACEs leads to the imbalance of research data categories, although we have  
539 performed techniques such as resampling and feature selection before model

---

540 construction to try to reduce the impact of data imbalance on model performance, it is  
541 difficult to significantly increase the AUPRC value. In addition to the evaluation of  
542 the predictive performance of the model, we also reported the calibration of the model,  
543 which was reported in the form of a calibration plot. The calibration plot shows that  
544 the predictive model developed in this study is well calibrated, although it appears to  
545 have a tendency to slightly underestimate the risk of MACEs. ROC curve, PRC curve,  
546 sensitivity, specificity, and calibration for assessing predictive models are reported,  
547 but do not provide answers as to whether models are effective in clinical practice.  
548 Decision analysis attempts to address the question of clinical utility assessment by  
549 combining the clinical outcomes of the model [41]. The DCA curve shows that the  
550 developed prediction model has good clinical practicability and has obvious  
551 advantages over the classical RCRI model. Finally, in order to further improve the  
552 performance of the prediction model, we used a stacking ensemble algorithm based  
553 on the optimization of independent machine learning prediction models [31].  
554 Compared with the XGBoost model, the optimal stacking model combining CatBoost,  
555 GBDT, GNB and LR improved the AUPRC, clinical practicability and sensitivity of  
556 the prediction model.

557       The interpretability of machine learning predictions requires attention, so that  
558 doctors can understand them, trust them and gain useful insights for the clinical  
559 practice. However, the “black box” nature of the ML algorithm and the difficulty for  
560 clinicians to understand and trust the interpretation of the data are still the most  
561 difficult hurdle to overcome [42]. XGBoost was excellent at predicting post-operative

---

562 mortality, with performance comparable to deep learning [11]. Compared to deep  
563 learning, XGBoost has the advantage of using the SHAP to interpret the model output,  
564 demonstrating the possibility of solving the "black box" problem. In this study, taking  
565 XGBoost as an example, we calculated the SHAP values of important features and  
566 used SHAP graphs to intuitively show the impact of features on the prediction model.  
567 IHD, ASA PS, and Hb were the top three important features of the XGBoost. This is  
568 consistent with clinical practice because IHD, ASA PS, and Hb have been used as  
569 important predictive indicators in previous clinical prediction models [43]. It is worth  
570 mentioning that, chlorine was not considered as a predictive indicator in univariate  
571 analysis. However, chlorine was the eighth important feature in XGBoost. In fact,  
572 chlorine has been proved to be related to heart failure [44]. Most of the predictive  
573 indicators in XGBoost can be improved, which means the patients can benefit from  
574 appropriate preoperative intervention.

---

575 This study had several strengths. To begin with, there is a higher incidence of  
576 perioperative MACEs in the patient with SCAD scheduled for NCS compared to the  
577 general population, but few studies have been conducted. We have adopted a series of  
578 widely used machine learning algorithms and model evaluation techniques to build  
579 clinical prediction models, and achieved better performance and clinical practicability  
580 than the classical RCRI model, which has taken the first step to explore the research  
581 in this field. In addition, the prediction results based on the optimal machine learning  
582 model are interpretable, output the importance ranking and impact degree of the top  
583 20 features of MACEs risk prediction, and are consistent with clinical interpretation,  
584 which is conducive to the application of the model in clinical practice. Moreover, we  
585 use Bayesian algorithm to automatically adjust the model hyperparameters, so that the  
586 selection of appropriate missing data imputation method, resampling technology and  
587 feature selection method can be combined with automatic hyperparameter tuning, and  
588 it is also confirmed that the appropriate resampling technology combined with feature  
589 selection can greatly improve the impact of data imbalance on model performance.  
590 Finally, we put forward the stacking ensemble model, and use the exhaustion method  
591 to form 247 stacking models, evaluate the performance of each model in turn, and  
592 select the optimal stacking model. Compared with the optimal independent model  
593 XGBoost, the optimal stacking ensemble model showed slightly higher AUPRC  
594 performance and clinical utility, with higher sensitivity.

---

595 The study has several limitations. First, in terms of feature collection, currently  
596 features are mainly from single text data of electronic medical records, and  
597 high-quality features can be extracted based on image (such as electrocardiogram)  
598 recognition technology. Second, we developed models based on data sets from a  
599 single medical center. Exploring the predictability of this model in other medical  
600 centers could add even more value. However, it should be noted that the data set in  
601 this study was extracted from 10 years of data from a large medical center with  
602 multiple hospitals, and the model evaluation used a testing set that was completely  
603 independent of model development as external validation. Third, as a retrospective  
604 study, the effect of predicting perioperative MACEs risk on prognosis in patients with  
605 SCAD remains unknown.

606 In future studies, we will further develop and validate current machine learning  
607 models based on data from other large, multicentre populations that can predict  
608 different types of MACEs (e.g., all-cause death, resuscitated cardiac arrest, MI, HF  
609 and stroke) and provide interventions accordingly. Another direction is to integrate the  
610 model into the clinician's workflow by designing an interactive interface, integrating  
611 with electronic medical record systems, and further exploring the model's impact on  
612 clinician behavior and patient outcomes.

## 613 **Conclusion**

614 In this study, we analyzed the data missing mechanism and identified the best  
615 missing data interpolation method, while applying appropriate resampling techniques  
616 and feature selection methods for data imbalance characteristics, and ultimately

---

617 identified 24 preoperative features for building a machine learning predictive model.  
618 Eight independent machine learning prediction models and stacking ensemble models  
619 were built, and the models were evaluated comprehensively using ROC curve, PRC  
620 curve, calibration plots and DCA curve. The results show that the machine learning  
621 prediction model developed in this study has better prediction performance and  
622 generalization than the classical RCRI model, and has the potential to be applied in  
623 clinical practice. With further validation and refinement, machine learning predictive  
624 models can help more effectively assess perioperative MACEs risk and target  
625 interventions to at-risk populations, as well as provide better clinical access and ease  
626 of use.

627

628 **List of abbreviations**

629 **NCS:** non-cardiac surgery

630 **SCAD:** stable coronary artery disease

631 **MACEs:** major adverse cardiovascular events

632 **MI:** myocardial infarction

633 **HF:** heart failure

634 **RCRI:** Revised Cardiac Risk index

635 **ML:** Machine learning

636 **AI:** artificial intelligence

- 
- 637 **ECG**: electrocardiogram
- 638 **FAHZU**: First Affiliated Hospital, Zhejiang University School of Medicine
- 639 **TRIPOD**: Transparent Reporting of Multivariable Prediction Models for Individual  
640 Prognosis or Diagnosis
- 641 **STROBE**: STrengthening the Reporting of OBservational studies in Epidemiology
- 642 **ACC**: American College of Cardiology
- 643 **AHA**: American Heart Association
- 644 **ICD-10**: International Classification of Diseases, Tenth Edition
- 645 **BMI**: Body Mass Index
- 646 **DOS**: duration of surgery
- 647 **GA**: general anesthesia
- 648 **AQW**: abnormal Q waves
- 649 **ST-Ta**: ST-T wave abnormalities
- 650 **LVEF**: left ventricular ejection fraction
- 651 **RWMA**: regional wall motion abnormality
- 652 **LVDD**: left ventricle diastolic dysfunction
- 653 **PH**: pulmonary hypertension
- 654 **Hb**: Hemoglobin
- 655 **FBG**: Fasting blood glucose



- 656 **Scr:** Creatinine
  
- 657 **ASA PS:** American Society of Anesthesiologists Physical Status
  
- 658 **IR:** imbalance ratio
  
- 659 **AUROC:** area under the receiver operating characteristic curve
  
- 660 **AUPRC:** area under the precision and recall curve
  
- 661 **SMOTE:** Synthetic minority over-sampling technique
  
- 662 **ADASYN:** adaptive synthetic
  
- 663 **ENN:** Edited Nearest Neighbors
  
- 664 **XGBoost:** eXtreme Gradient Boosting
  
- 665 **CFS:** correlation-based feature selection
  
- 666 **RFE:** recursive feature elimination
  
- 667 **SHAP:** SHapley Additive exPlanation
  
- 668 **LightGBM:** Light Gradient Boosting Machine
  
- 669 **RF:** Random Forest
  
- 670 **LR:** logistic regression
  
- 671 **SVM:** support vector machine
  
- 672 **GNB:** Gaussian Naïve Bayesian
  
- 673 **GBDT:** gradient boosting decision tree
  
- 674 **CatBoost:** categorical boosting

- 675 **ROC:** receiver-operating characteristic
- 676 **PRC:** curves and precision–recall curves
- 677 **DCA:** decision curve analysis
- 678 **SD:** standard deviation
- 679 **IQR:** interquartile range
- 680 **KNN:** k-Nearest Neighbor
- 681 **IHD:** Ischemic heart disease
- 682 **FS:** fractional shortening
- 683 **LVDs:** left ventricular end systolic dimension
- 684 **eGFR:** Estimated glomerular filtration rate
- 685 **TSP:** Total serum protein
- 686 **ALB:** Albumin
- 687 **ChE:** Cholinesterase
- 688 **TB:** Total bilirubin
- 689 **tCa:** Total calcium
- 690 **FB:** Fibrinogen
- 691 **VIF:** variance inflation factor
- 692 **MCAR:** missing completely at random
- 693 **MAR:** missing at random

---

694 **MNAR:** missing not at random

695 **Declarations**

696 *Availability of data and materials*

697 The datasets generated during and/or analyzed during the current study are not  
698 publicly available but are available from the corresponding author at reasonable  
699 request.

700 *Ethics approval and consent to participate*

701 This study was approved by the Institutional Ethics Review Committee of the the  
702 First Affiliated Hospital, Zhejiang University School of Medicine (No. of ethical  
703 approval: IIT20230114A). Written informed consent was waived owing to the nature  
704 of the retrospective study design and the collected data was managed in a  
705 de-identified form.

706 *Consent for publication*

707 Not applicable.

708 *Competing interests*

709 The authors declare that they have no known competing financial interests or personal  
710 relationships that could have appeared to influence the work reported in this paper.

---

711 ***Funding***

712 This work was supported by grants from the National Natural Science Foundation of  
713 China (82170331), Joint Funds from the National Natural Science Foundation of  
714 China (U21A20337), and grants from the Key Research and Development Plan of  
715 Zhejiang Province (2020C03017)

716 ***Authors' contributions***

717 LS and YPJ contributed to the conception, design, analysed data and coding of the  
718 work, and drafted the manuscript. AXP, KW, RZY, YKL, SA and WCX contributed to  
719 the acquisition, analysis, and interpretation of data for the work. MZ and XGG were  
720 responsible for conceptualization and formal analysis, critically revised the  
721 manuscript. All authors read and approved the final manuscript.

722 ***Acknowledgments***

723 This work was supported by grants from the National Natural Science Foundation of  
724 China, Joint Funds from the National Natural Science Foundation of China, and  
725 grants from the Key Research and Development Plan of Zhejiang Province

726 ***Supplementary Material***

727 data supplement.docx

728 **References**

- 729 [1] Weiser TG, Haynes AB, Molina G, et al. Size and distribution of the global volume of surgery  
730 in 2012[J]. Bull World Health Organ,2016,94(3):201-209F. doi:10.2471/BLT.15.159293
- 731 [2] Smilowitz NR, Gupta N, Guo Y, Beckman JA, Bangalore S, Berger JS. Trends in  
732 cardiovascular risk factor and disease prevalence in patients undergoing non-cardiac  
733 surgery[J]. Heart,2018,104(14):1180-1186. doi:10.1136/heartjnl-2017-312391
- 734 [3] Handke J, Scholz AS, Dehne S, Krisam J, Gillmann HJ, Janssen H, Arens C, Espeter F, Uhle F,  
735 Motsch J, Weigand MA, Larmann J. Presepsin for pre-operative prediction of major adverse  
736 cardiovascular events in coronary heart disease patients undergoing noncardiac surgery: Post hoc

- 737 analysis of the Leukocytes and Cardiovascular Peri-operative Events-2 (LeukoCAPE-2) Study. *Eur*  
738 *J Anaesthesiol.* 2020 Oct;37(10):908-919. doi: 10.1097/EJA.0000000000001243. PMID:  
739 32516228.
- 740 [4] Smilowitz NR, Gupta N, Ramakrishna H, Guo Y, Berger JS, Bangalore S. Perioperative Major  
741 Adverse Cardiovascular and Cerebrovascular Events Associated With Noncardiac  
742 Surgery[J]. *JAMA Cardiol*,2017,2(2):181-187. doi:10.1001/jamacardio.2016.4792
- 743 [5] Devereaux PJ, Chan MT, Alonso-Coello P, Walsh M, Berwanger O, Villar JC, et al.  
744 Association between postoperative troponin levels and 30-day mortality among patients undergoing  
745 noncardiac surgery. *JAMA.* (2012) 307:2295-304. doi: 10.1001/jama.2012.5502
- 746 [6] Fleisher LA, Fleischmann KE, Auerbach AD, Barnason SA, Beckman JA, Bozkurt B, et al.  
747 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients  
748 undergoing noncardiac surgery: executive summary: a report of the American College of  
749 Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* (2014)  
750 130:2215-45. doi: 10.1161/CIR.000000000000105
- 751 [7] Halvorsen S, Mehilli J, Cassese S, Hall TS, Abdelhamid M, Barbato E, De Hert S, de Laval I,  
752 Geisler T, Hinterbuchner L, Ibanez B, Lenarczyk R, Mansmann UR, McGreavy P, Mueller C,  
753 Muneretto C, Niessner A, Potpara TS, Ristić A, Sade LE, Schirmer H, Schüpke S, Sillesen H,  
754 Skulstad H, Torracca L, Tutarel O, Van Der Meer P, Wojakowski W, Zacharowski K; ESC  
755 Scientific Document Group. 2022 ESC Guidelines on cardiovascular assessment and management  
756 of patients undergoing non-cardiac surgery. *Eur Heart J.* 2022 Oct 14;43(39):3826-3924. doi:  
757 10.1093/eurheartj/ehac270. PMID: 36017553.
- 758 [8] Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a  
759 simple index for prediction of cardiac risk of major noncardiac  
760 surgery[J]. *Circulation*,1999,100(10):1043-1049. doi:10.1161/01.cir.100.10.1043
- 761 [9] Davis C, Tait G, Carroll J, Wijesundera DN,Beattie WS. The revised cardiac risk index in the  
762 new millennium: a single center prospective cohort re-evaluation of the original variables in 9,519  
763 consecutive elective surgical patients. *Can J Anaesth* 2013;60:855-63.
- 764 [10] Che L, Xu L, Huang Y, Yu C. Clinical utility of the revised cardiac risk index in older Chinese  
765 patients with known coronary artery disease[J]. *Clin Interv Aging*,2017,13:35-41. Published 2017  
766 Dec 22. doi:10.2147/CIA.S144832
- 767 [11] Lee SW, Lee HC, Suh J, et al. Multi-center validation of machine learning model for  
768 preoperative prediction of postoperative mortality[J]. *NPJ Digit Med*,2022,5(1):91.  
769 doi:10.1038/s41746-022-00625-6
- 770 [12] Watson X, D'Souza J, Cooper D, Markham R. Artificial intelligence in cardiology:  
771 fundamentals and applications. *Intern Med J.* 2022 Jun;52(6):912-920. doi: 10.1111/imj.15562.  
772 Epub 2022 May 31. PMID: 34613658.
- 773 [13] Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA guideline on  
774 perioperative cardiovascular evaluation and management of patients undergoing noncardiac  
775 surgery: executive summary: a report of the American College of Cardiology/American Heart  
776 Association Task Force on Practice Guidelines. *Circulation.* 2014;130:2215-45. doi:  
777 10.1161/CIR.000000000000105.
- 778 [14] Fihn SD, Gardin JM, Abrams J, et al. 2012  
779 ACCF/AHA/ACP/AATS/PCNA/SCAI/STSGuideline for the diagnosis and management of  
780 patients with stable ischemic heart disease: executive summary: a report of the American College

- 
- 781 of Cardiology Foundation/American Heart Association task force on practice guidelines, and the  
782 American College of Physicians, American Association for Thoracic Surgery, Preventive  
783 Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions,  
784 and Society of Thoracic Surgeons. *Circulation*. 2012;126:3097-137. doi:  
785 10.1161/CIR.0b013e3182776f83.
- 786 [15] Andersen LW, Holmberg MJ, Berg KM, et al. In-Hospital Cardiac Arrest: A Review. *JAMA*.  
787 2019;321:1200-10. doi: 10.1001/jama.2019.1696.
- 788 [16] Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. *J*  
789 *Am Coll Cardiol*. 2012;60:1581-98. doi: 10.1016/j.jacc.2012.08.001.
- 790 [17] Yancy CW, Jessup M, Bozkurt B., et al. 2013 ACCF/AHA guideline for the management of  
791 heart failure: a report of the American College of Cardiology Foundation/American Heart  
792 Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2013;62:e147-239. doi:  
793 10.1016/j.jacc.2013.05.019.
- 794 [18] Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st  
795 century: a statement for healthcare professionals from the American Heart Association/American  
796 Stroke Association. *Stroke*. 2013;44:2064-89. doi: 10.1161/STR.0b013e318296aeca.
- 797 [19] Zhang J , Cui X , Li J ,et al.Imbalanced classification of mental workload using a  
798 cost-sensitive majority weighted minority oversampling strategy[J].*Cognition Technology and*  
799 *Work*, 2017.DOI:10.1007/s10111-017-0447-x.
- 800 [20] Chawla N V , Bowyer K W , Hall L O ,et al.SMOTE: synthetic minority over-sampling  
801 technique[J].*AI Access Foundation*, 2002(1).DOI:10.1613/JAIR.953.
- 802 [21] He H , Bai Y , Garcia E A ,et al.ADASYN: Adaptive synthetic sampling approach for  
803 imbalanced learning[J].*IEEE*, 2008.DOI:10.1109/IJCNN.2008.4633969.
- 804 [22] Batista G E A P A , Prati R C , Monard M C .A study of the behavior of several methods for  
805 balancing machine learning training data[J].*Acm Sigkdd Explorations Newsletter*, 2004,  
806 6(1):20-29.DOI:10.1145/1007730.1007735.
- 807 [23] Hall M A .Correlation-based Feature Selection for Machine Learning[J].Phd Thesis Waikato  
808 Univer Sity, 2000.
- 809 [24] Kursu M B , Rudnicki W R .Feature Selection with Boruta Package[J].*Journal of Statistical*  
810 *Software*, 2010, 36(11):1-13.DOI:10.18637/jss.v036.i11.
- 811 [25] Caballero, W.N., Gaw, N., Jenkins, P.R., & Johnstone, C. (2023). Toward automated  
812 instructor pilots in legacy Air Force systems: Physiology-based flight difficulty classification via  
813 machine learning. *Expert Syst. Appl.*, 231, 120711.
- 814 [26] Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data  
815 using random forest. *BMC Bioinformatics*. 2006;7:3. Published 2006 Jan 6.  
816 doi:10.1186/1471-2105-7-3.
- 817 [27] Feng X, Zhang C, Huang X, et al. Machine learning improves mortality prediction in  
818 three-vessel disease. *Atherosclerosis*. 2023;367:1-7. doi:10.1016/j.atherosclerosis.2023.01.003.
- 819 [28] Gould, M.K., Huang, B.Z., Tammemagi, M.C., Kinar, Y., & Shiff, R. (2021). Machine  
820 Learning for Early Lung Cancer Identification Using Routine Clinical and Laboratory Data.  
821 *American journal of respiratory and critical care medicine*.
- 822 [29] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation  
823 Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International*  
824 *Conference on Knowledge Discovery & Data Mining*.

- 825 [30] Kardani, N., Zhou, A., Nazem, M., & Shen, S. (2020). Improved prediction of slope stability  
826 using a hybrid stacking ensemble method based on finite element analysis and field data. *Journal*  
827 *of rock mechanics and geotechnical engineering*.
- 828 [31] Wang, J., Chen, H., Wang, H., Liu, W., Peng, D., Zhao, Q., & Xiao, M. (2023). A Risk  
829 Prediction Model for Physical Restraints Among Older Chinese Adults in Long-term Care  
830 Facilities: Machine Learning Study. *Journal of Medical Internet Research*, 25.
- 831 [32] DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under  
832 two or more correlated receiver operating characteristic curves: a nonparametric approach.  
833 *Biometrics*, 44 3, 837-45.
- 834 [33] Chung, E.Y., & Romano, J.P. (2013). EXACT AND ASYMPTOTICALLY ROBUST  
835 PERMUTATION TESTS. *Annals of Statistics*, 41, 484-507.
- 836 [34] Rufibach, K. (2010). Use of Brier score to assess binary predictions. *Journal of clinical*  
837 *epidemiology*, 63 8, 938-9; author reply 939.
- 838 [35] Farhangfar, A., Kurgan, L., & Dy, J.G. (2008). Impact of imputation of missing values on  
839 classification error for discrete data. *Pattern Recognit.*, 41, 3692-3705.
- 840 [36] Collins, L.M., Schafer, J.L., & Kam, C. (2001). A comparison of inclusive and restrictive  
841 strategies in modern missing data procedures. *Psychological methods*, 6 4, 330-51.
- 842 [37] Batista, G.E., Prati, R.C., & Monard, M.C. (2004). A study of the behavior of several  
843 methods for balancing machine learning training data. *SIGKDD Explor.*, 6, 20-29.
- 844 [38] Putten, P.V., & Someren, M.V. (2004). A Bias-Variance Analysis of a Real World Learning  
845 Problem: The CoIL Challenge 2000. *Machine Learning*, 57, 177-195.
- 846 [39] Liu, H., Zhou, M., & Liu, Q. (2019). An embedded feature selection method for imbalanced  
847 data classification. *IEEE/CAA Journal of Automatica Sinica*, 6, 703-715.
- 848 [40] Guyon, I.M., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *J.*  
849 *Mach. Learn. Res.*, 3, 1157-1182.
- 850 [41] Vickers, A.J., Van calster, B., & Steyerberg, E.W. (2016). Net benefit approaches to the  
851 evaluation of prediction models, molecular markers, and diagnostic tests. *The BMJ*, 352.
- 852 [42] Cuocolo, R., Perillo, T., De Rosa, E., Uggia, L., & Petretta, M. (2019). Current applications of  
853 big data and machine learning in cardiology. *Journal of Geriatric Cardiology : JGC*, 16, 601 - 607.
- 854 [43] Halvorsen, Sigrun et al. "2022 ESC Guidelines on cardiovascular assessment and  
855 management of patients undergoing non-cardiac surgery." *European heart journal* (2022): n. pag.
- 856 [44] Zandijk, A.J., van Norel, M.R., Julius, F.E., Sepehrvand, N., Pannu, N.I., McAlister, F.A.,  
857 Voors, A.A., & Ezekowitz, J.A. (2021). Chloride in Heart Failure: The Neglected Electrolyte.  
858 *JACC. Heart failure*, 9 12, 904-915.

## 859 **Figure Legends**

860 **Figure 1.** Flowchart of study design route.

861 **Figure 2.** Performance assessment of feature selection methods. (A) The optimal feature subset of  
862 the eight feature selection methods and the AUROC performance of the internal validation set  
863 after cross-validation on the training set. (B) RFE combines five kinds of basis learning models,

---

864 selects the best subset recursively based on the feature importance ranking of the learning model,  
865 and evaluating the AUROC performance of the internal validation set after cross-validation on the  
866 training set.

867 **Figure 3.** Performance assessment of the models. (A) Receiver operating characteristic curve  
868 (ROC) of MACEs prediction models in testing set. (B) Precision-Recall curve (PRC) of MACEs  
869 prediction models in testing set. (C) Decision curve analysis (DCA) for the nine MACEs  
870 prediction models in the testing set. (D) Calibration plots of MACEs prediction models in the  
871 testing set.

872 **Figure 4.** Compare the performance of validation sets and testing sets on different models. (A)  
873 Compare AUROC values of validation set and testing set. (B) Compare AUPRC values of  
874 validation set and testing set.

875 **Figure 5.** SHAP interprets the XGBoost Predictive model. (A) The SHAP analysis was performed  
876 on the XGBoost. Each row of the graph represents a variable and the horizontal coordinate is the  
877 SHAP value, which represents the distribution of the effect of the variable on the risk of MACEs,  
878 with positive values indicating a risk of MACEs and negative values indicating no risk of MACEs.  
879 A point represents a patient, while red represents a higher value and blue represents a lower value.  
880 (B) The average of the absolute values of the SHAP values for each variable in the XGBoost is  
881 taken as the significance of that variable. (C) Examples of negative predictions for MACEs. (D)  
882 Examples of positive predictions for MACEs.

883 **Figure 6.** The ROC, PR and DCA curve performance of stacking model. ROC: receiver operating  
884 characteristic; PR: precision-recall; DCA: decision curve analysis; AUROC: area under the  
885 receiver operating characteristic curve; AUPRC: area under precision-recall curve.



886 **Tables**

887 **Table 1.** Baseline clinical characteristics of the study population and their association with  
888 perioperative outcomes.

Variables	Total (n=9171)	Training and validation set (n=7336)			Testing set (n=1835)		
		Non-MACEs (n=6925)	MACEs (n=411)	P-value	Non-MACEs (n=1732)	MACEs (n=103)	P-value
<b>Demographic</b>							
Age (years)	70 (63, 76)	70 (64, 76)	72 (63, 79)	0.003	70 (63, 76)	69 (62, 77)	0.845
Male	6133 (66.9)	4597 (66.4)	304 (74.0)	0.002	1155 (66.7)	77 (74.8)	0.090
Body mass index (kg/m <sup>2</sup> )	23.6 (21.5, 25.7)	23.7 (21.6, 25.8)	22.0 (19.9, 24.4)	<0.001	23.6 (21.5, 25.6)	22.6 (19.8, 24.4)	0.003
<b>Comorbidities</b>							
Hypertension	5737 (62.6)	4323 (62.4)	266 (64.7)	0.350	1076 (62.1)	72 (69.9)	0.113
Diabetes mellitus	2534 (27.6)	1880 (27.1)	141 (34.3)	0.002	472 (27.3)	41 (39.8)	0.006
Stroke	863 (9.4)	637 (9.2)	55 (13.4)	0.005	158 (9.1)	12 (11.7)	0.390
Dialysis	196 (2.1)	119 (1.7)	51 (12.4)	<0.001	18 (1.0)	8 (7.8)	<0.001
COPD	214 (2.3)	171 (2.5)	8 (1.9)	0.504	34 (2.0)	1 (1.0)	0.719
<b>Cardiac history</b>							
Ischemic heart disease	3454 (37.7)	2498 (36.1)	252 (61.3)	<0.001	641 (37.0)	63 (61.2)	<0.001
Myocardial infarction	2147 (23.4)	1567 (22.6)	116 (28.2)	0.009	434 (25.1)	30 (29.1)	0.356
Heart failure	491 (5.4)	331 (4.8)	73 (17.8)	<0.001	68 (3.9)	19 (18.4)	<0.001
Atrial fibrillation	416 (4.5)	286 (4.1)	48 (11.7)	<0.001	74 (4.3)	8 (7.8)	0.132
Valvular heart disease	168 (1.8)	122 (1.8)	18 (4.4)	<0.001	22 (1.3)	6 (5.8)	0.004
PTCA	1823 (19.9)	1373 (19.8)	78 (19.0)	0.675	349 (20.2)	23 (22.3)	0.593
CABG	146 (1.6)	112 (1.6)	9 (2.2)	0.376	22 (1.3)	3 (2.9)	0.162
<b>Preoperative blood tests</b>							
Leukocyte(×10 <sup>9</sup> /L)	6.2(5.1, 7.7)	6.2 (5.1, 7.6)	7.4(5.6, 10.3)	<0.001	6.1 (5.0, 7.5)	7.3(5.4, 11.3)	<0.001
Hemoglobin (g/L)	131 (117, 144)	132 (118, 144)	104 (84, 123)	<0.001	132 (118, 144)	100 (82, 119)	<0.001
Platelet (×10 <sup>9</sup> /L)	195 (157, 239)	195 (158, 239)	179 (134,237)	<0.001	198 (158, 239)	182 (133,237)	0.023
Fasting blood glucose (mmol/L)	5.47 (4.83, 6.68)	5.45 (4.82, 6.58)	6.80 (5.20, 9.23)	<0.001	5.43 (4.81, 6.53)	7.05 (5.23, 9.69)	<0.001
Serum creatinine (μmol/L)	77 (65, 93)	77 (65, 92)	94 (70, 176)	<0.001	76 (65, 92)	85 (67, 139)	<0.001
eGFR (mL/min/1.73m <sup>2</sup> )	81 (65, 91)	82 (66, 91)	62 (28, 84)	<0.001	81 (66, 91)	71 (36, 90)	<0.001
Total serum protein (g/L)	69.0 (64.0, 73.2)	69.3 (64.5, 73.4)	63.1 (55.5, 68.4)	<0.001	69.2 (64.1, 73.4)	59.9 (53.8, 68.6)	<0.001
Albumin (g/L)	42.7 (38.6, 45.8)	42.9 (39.1, 45.9)	37.2 (32.4, 41.3)	<0.001	42.9 (39.2, 45.9)	35.5 (31.3, 39.8)	<0.001
Globulin (g/L)	26.1 (23.5, 28.4)	26.2 (23.6, 28.5)	25.3 (21.7, 28.9)	<0.001	26.0 (23.5, 29.0)	25.0 (20.8, 29.2)	0.017

	29.0)	29.0)	28.9)			29.6)	
ALT (U/L)	18 (13, 26)	18 (13, 26)	16 (10, 27)	0.001	18 (13, 27)	17 (11, 31)	0.738
AST (U/L)	20 (16, 26)	20 (16, 25)	20 (15, 31)	0.225	20 (16, 26)	22 (15, 40)	0.078
GGT (U/L)	25 (17, 42)	25 (17, 42)	27 (16, 52)	0.157	26 (17, 43)	29 (19, 68)	0.025
ALP (U/L)	77 (63, 94)	76 (63, 94)	80 (62, 108)	0.043	77 (64, 93)	80 (62, 113)	0.170
Cholinesterase (U/L)	7435 (6151, 8663)	7500 (6280, 8708)	5546 (3976, 7184)	<0.001	7553 (6289, 8751)	5360 (3294, 7124)	<0.001
Total bilirubin (µmol/L)	10.3 (7.6, 14.3)	10.3 (7.6, 14.3)	9.0 (6.2, 14.5)	0.004	10.4 (7.9, 14.1)	8.6 (6.8, 16.7)	0.393
Direct bilirubin (µmol/L)	4.0 (3.0, 5.4)	4.0 (3.0, 5.4)	4.0 (3.0, 6.1)	0.174	4.0 (3.0, 5.5)	4.3 (3.0, 8.2)	0.074
Indirect bilirubin (µmol/L)	6.0 (4.1, 9.0)	6.1 (4.2, 9.0)	5.0 (3.0, 7.4)	<0.001	6.1 (4.4, 9.0)	4.6 (3.0, 8.4)	<0.001
Potassium (mmol/L)	4.13 (3.85, 4.42)	4.14 (3.85, 4.42)	4.10 (3.70, 4.46)	0.081	4.13 (3.85, 4.44)	3.99 (3.70, 4.40)	0.013
Sodium (mmol/L)	142.0 (140.0, 143.0)	142.0 (140.0, 143.0)	140.0 (138.0, 143.0)	<0.001	142.0 (140.0, 143.0)	140.0 (137.0, 142.0)	<0.001
Chlorine (mmol/L)	104.0 (102.0, 106.0)	104.0 (102.0, 106.0)	104.0 (101.0, 107.0)	0.765	104.0 (102.0, 106.0)	104.0 (101.0, 107.0)	0.728
Total calcium(mmol/L)	2.24 (2.14, 2.33)	2.25 (2.15, 2.34)	2.14 (2.00, 2.25)	<0.001	2.25 (2.16, 2.34)	2.10 (2.00, 2.24)	<0.001
Inorganic phosphorus (mmol/L)	1.10 (0.97, 1.22)	1.10 (0.98, 1.22)	1.11 (0.95, 1.33)	0.027	1.09 (0.97, 1.21)	1.06 (0.90, 1.27)	0.629
Uric acid (µmol/L)	316 (256, 381)	315 (257, 380)	300 (230, 389)	0.022	320 (260, 384)	294 (218, 378)	0.026
Triglyceride (mmol/L)	1.22 (0.91, 1.69)	1.22 (0.91, 1.69)	1.16 (0.87, 1.56)	0.019	1.21 (0.91, 1.70)	1.22 (0.90, 1.68)	0.980
Total cholesterol (mmol/L)	3.64 (3.05, 4.37)	3.64 (3.06, 4.36)	3.41 (2.80, 4.15)	<0.001	3.70 (3.09, 4.46)	3.45 (2.90, 4.33)	0.062
LDL-C (mmol/L)	1.83 (1.40, 2.44)	1.83 (1.41, 2.43)	1.73 (1.23, 2.25)	<0.001	1.87 (1.42, 2.51)	1.74 (1.41, 2.55)	0.159
PT (s)	11.0 (10.9, 12.1)	11.4 (10.9, 12.1)	12.0 (11.3, 13.3)	<0.001	11.4 (10.9, 12.0)	12.5 (11.7, 13.5)	<0.001
APTT (s)	27.2 (25.2, 29.4)	27.1 (25.2, 29.3)	28.6 (26.2, 34.2)	<0.001	27.1 (25.1, 29.1)	29.9 (26.5, 35.5)	<0.001
Fibrinogen (g/L)	3.00 (2.53, 3.65)	2.99 (2.53, 3.64)	3.32 (2.49, 4.34)	<0.001	3.00 (2.56, 3.59)	2.93 (2.28, 4.18)	0.717
D-dimer(µg/L FEU)	460 (234, 1055)	434 (223, 962)	1366 (552, 3788)	<0.001	439 (230, 1000)	2340 (704, 5116)	<0.001
Preoperative ECG							
AQW	501 (6.3)	357 (5.9)	32 (10.3)	0.001	104 (6.9)	8 (12.1)	0.134

ST-Ta	3901 (49.0)	2913 (47.9)	204 (65.8)	<0.001	739 (48.8)	45 (68.2)	0.002
Preoperative echocardiography							
AO (mm)	30.0 (27.0, 33.0)	30.0 (27.0, 33.0)	30.5 (28.0, 33.0)	0.108	30.0 (27.0, 32.0)	30.0 (27.5, 32.0)	0.817
IVSd (mm)	9.0 (9.0, 10.0)	9.0 (9.0, 10.0)	10.0 (9.0, 11.0)	<0.001	9.0 (9.0, 10.0)	10.0 (9.0, 11.0)	0.124
LVDd (mm)	48.0 (44.0, 51.0)	47.0 (44.0, 51.0)	48.0 (45.0, 52.0)	0.001	48.0 (44.0, 51.0)	50.0 (45.0, 53.0)	0.035
FS (%)	37.0 (33.0, 40.0)	37.0 (34.0, 40.0)	35.0 (31.8, 39.0)	<0.001	37.0 (34.0, 40.0)	34.0 (31.0, 39.0)	<0.001
LA (mm)	34.0 (30.0, 38.0)	34.0 (30.0, 38.0)	35.0 (31.0, 41.0)	<0.001	34.0 (30.3, 37.0)	35.0 (31.0, 40.0)	0.044
LVPWd (mm)	9.0 (9.0, 10.0)	9.0 (9.0, 10.0)	10.0 (9.0, 11.0)	<0.001	9.0 (9.0, 10.0)	10.0 (8.5, 11.0)	0.334
LVDs (mm)	30.0 (27.0, 33.0)	30.0 (27.0, 32.0)	31.0 (28.0, 35.0)	<0.001	30.0 (27.0, 33.0)	31.0 (27.0, 35.0)	0.055
LVEF (%)	66.0 (62.0, 71.0)	66.0 (62.0, 71.0)	63.0 (58.0, 69.0)	<0.001	66.0 (62.0, 70.0)	62.0 (58.3, 68.8)	<0.001
RWMA	532 (7.5)	348 (6.5)	68 (19.7)	<0.001	90 (6.9)	26 (28.9)	<0.001
PH	326 (4.6)	221 (4.1)	47 (13.6)	<0.001	49 (3.7)	9 (10.0)	<0.001
LVDD	6210 (87.1)	4741 (88.1)	253 (73.1)	<0.001	1155 (88.1)	61 (67.8)	<0.001
ASA class				<0.001			<0.001
II	3887 (42.4)	3072 (44.4)	61 (14.8)		742 (42.8)	12 (11.7)	
III	5210 (56.8)	3834 (55.4)	312 (75.9)		985 (56.9)	79 (76.7)	
IV	74 (0.8)	19 (0.3)	38 (9.2)		5 (0.3)	12 (11.7)	
General anesthesia	6702 (73.1)	5016 (72.4)	338 (82.2)	<0.001	1273 (73.5)	75 (72.8)	0.879
Types of surgery							
General	2823 (30.8)	2082 (30.1)	154 (37.5)	0.002	546 (31.5)	41 (39.8)	0.080
Abdominal	2156 (23.5)	1573 (22.7)	131 (31.9)	<0.001	414 (23.9)	38 (36.9)	0.003
Nonabdominal	667 (7.3)	509 (7.4)	23 (5.6)	0.183	132 (7.6)	3 (2.9)	0.075
Thoracic	1060 (11.6)	805 (11.6)	29 (7.1)	0.005	219 (12.6)	7 (6.8)	0.079
Orthopedic	854 (9.3)	643 (9.3)	46 (11.2)	0.198	154 (8.9)	11 (10.7)	0.538
ENT	236 (2.6)	192 (2.8)	2 (0.5)	0.005	39 (2.3)	3 (2.9)	0.509
Neurological	444 (4.8)	321 (4.6)	31 (7.5)	0.007	80 (4.6)	12 (11.7)	0.001
Gynecologic	171 (1.9)	139 (2.0)	4 (1.0)	0.141	27 (1.6)	1 (1.0)	1.000
Urologic	1550 (16.9)	1185 (17.1)	54 (13.1)	0.037	307 (17.7)	4 (3.9)	<0.001
Ophthalmology	767 (8.4)	603 (8.7)	1 (0.2)	<0.001	162 (9.4)	1 (1.0)	0.001
Vascular	1113 (12.1)	835 (12.1)	87 (21.2)	<0.001	169 (9.8)	22 (21.4)	<0.001
Aortic	113 (1.2)	76 (1.1)	16 (3.9)	<0.001	17 (1.0)	4 (3.9)	0.027
Non-aortic	1000 (10.9)	759 (11.0)	71 (17.3)	<0.001	152 (8.8)	18 (17.5)	0.003
Dental	153 (1.7)	120 (1.7)	3 (0.7)	0.124	29 (1.7)	1 (1.0)	1.000
DOS (min)	87 (50, 147)	84 (49, 141)	140 (80, 211)	<0.001	87 (50, 144)	139(79, 234)	<0.001

---

890 Abbreviations: COPD, chronic obstructive pulmonary disease; PTCA, percutaneous transluminal coronary angioplasty; CABG, coronary  
891 artery bypass graft; eGFR, estimated glomerular filtration rate; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT,  
892 gamma-glutamyl transferase; ALP, alkaline phosphatase; LDL-C, low density lipoprotein cholesterol; PT, prothrombin time; APTT,  
893 activated partial thromboplastin time; AQW, abnormal Q waves; ST-Ta, ST-T wave abnormalities; AO, aorta diameter; IVSd,  
894 interventricular septum thickness at end diastole; LVDD, left ventricular end diastolic dimension; FS, fractional shortening; LA, left atrial  
895 anteroposterior dimension; LVPWd, left ventricular posterior wall thickness at end diastole; LVDS, left ventricular end systolic  
896 dimension; LVEF, left ventricular ejection fraction; RWMA, regional wall motion abnormality; PH, pulmonary hypertension; LVDD,  
897 left ventricle diastolic dysfunction; ASA, American Society of Anesthesiologists; ENT, ear, nose, and throat; DOS, duration of surgery.

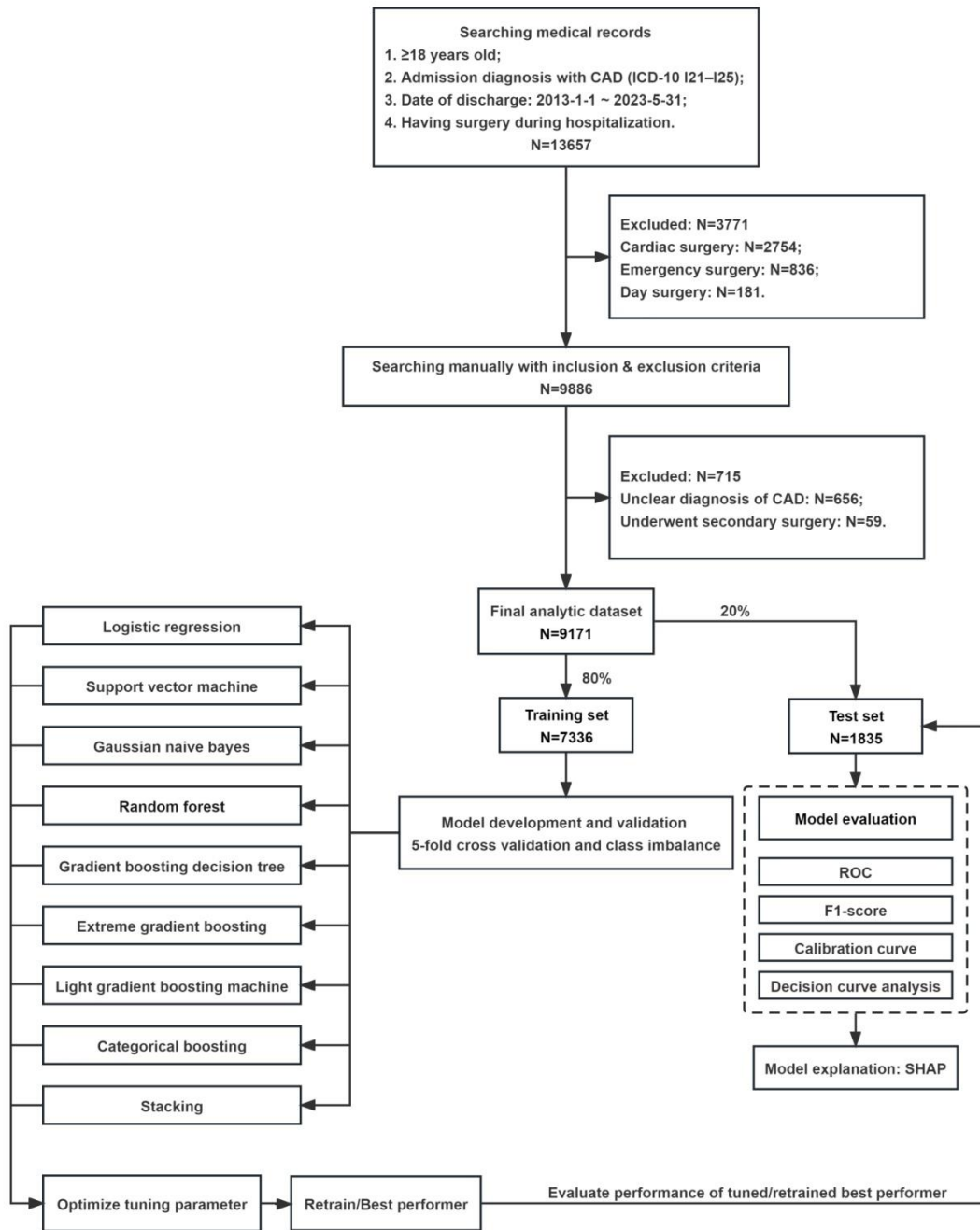
898 **Table 2.** The internal verification results of models trained were obtained by using different  
899 balanced class methods combined with cross-validation using all features.

Models	Methods	AUROC(95%CI)	AUPRC(95%CI)	Accuracy	Specificity	Sensitivity	Youden Index
LR	-	0.879(+/-0.024)	0.451(+/-0.126)	0.879	0.991	0.270	0.261
RF	-	0.864(+/-0.036)	0.414(+/-0.095)	0.944	1.000	0.000	0.000
XGBoost	-	0.862(+/-0.047)	0.403(+/-0.055)	0.946	1.000	0.034	0.034
LR	SMOTE	0.867(+/-0.036)	0.412(+/-0.121)	0.824	0.828	0.757	0.585
RF	SMOTE	0.870(+/-0.047)	0.386(+/-0.105)	0.943	0.979	0.343	0.322
XGBoost	SMOTE	0.875(+/-0.041)	0.427(+/-0.137)	0.942	0.976	0.372	0.348
LR	ANSYN	0.864(+/-0.041)	0.408(+/-0.134)	0.806	0.809	0.764	0.573
RF	ANSYN	0.871(+/-0.044)	0.388(+/-0.104)	0.941	0.973	0.399	0.372
XGBoost	ANSYN	0.877(+/-0.035)	0.415(+/-0.137)	0.947	0.982	0.348	0.330
LR	SMOTE+ENN	0.868(+/-0.036)	0.396(+/-0.105)	0.726	0.719	0.842	0.561
RF	SMOTE+ENN	0.863(+/-0.053)	0.360(+/-0.096)	0.864	0.691	0.874	0.565
XGBoost	SMOTE+ENN	0.872(+/-0.037)	0.397(+/-0.105)	0.894	0.907	0.672	0.579

900 **Table 3.** The predictive performance in the test set of the 9 models. LR: logistic regression; SVM:  
901 support vector machine; GNB: Gaussian Naive Bayes; RF: random forest; GBDT: gradient  
902 boosting decision tree; XGBoost: extreme gradient boosting; LightGBM: light gradient boosting  
903 machine; CatBoost: categorical boosting; AUROC: area under the receiver operating characteristic  
904 curve; AUPRC: area under the precision recall curve.

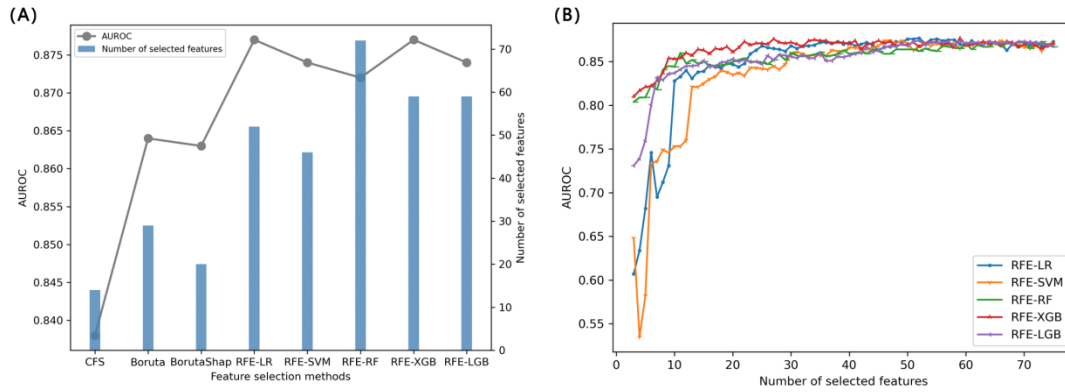
Models	AUROC (95%CI)	AUPRC (95%CI)	Accuracy	Specificity	Precision	Recall (Sensitivity)	F1 score	Youden Index
Baseline-RCRI	0.716(+/-0.045)	0.185(+/-0.078)	0.679	0.680	0.109	0.662	0.187	0.342
LR	0.896(+/-0.033)	0.438(+/-0.106)	0.732	0.722	0.160	0.895	0.271	0.617
SVM	0.892(+/-0.034)	0.431(+/-0.103)	0.780	0.777	0.181	0.837	0.298	0.613
GNB	0.880(+/-0.034)	0.392(+/-0.081)	0.879	0.887	0.279	0.740	0.404	0.627
RF	0.892(+/-0.035)	0.454(+/-0.099)	0.777	0.773	0.181	0.845	0.297	0.618
GBDT	0.895(+/-0.032)	0.460(+/-0.103)	0.901	0.916	0.314	0.652	0.423	0.568
XGBoost	0.898(+/-0.034)	0.479(+/-0.101)	0.874	0.881	0.271	0.748	0.397	0.629
LightGBM	0.897(+/-0.033)	0.445(+/-0.102)	0.845	0.850	0.230	0.758	0.353	0.608
CatBoost	0.890(+/-0.034)	0.448(+/-0.102)	0.804	0.803	0.199	0.826	0.320	0.630

905 **Figures**



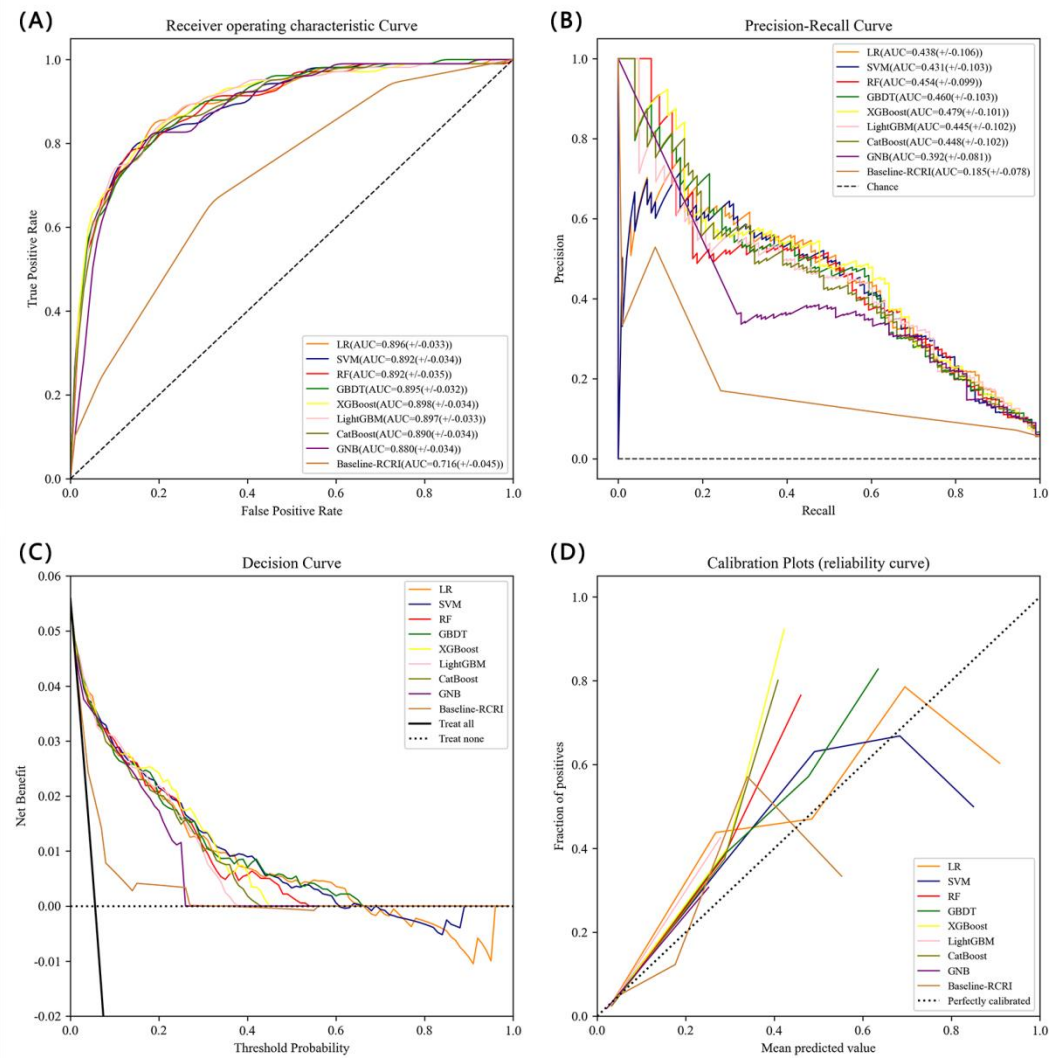
906

907 **Figure 1.**



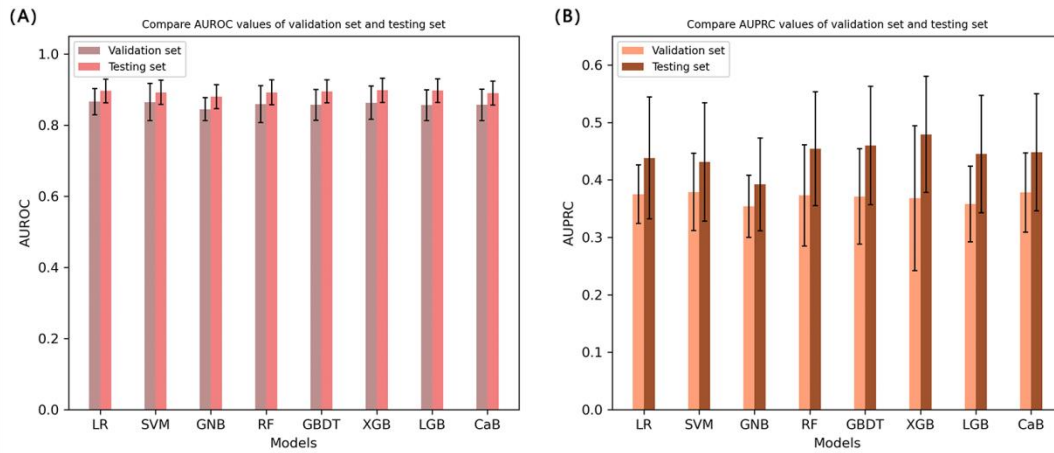
908

909 **Figure 2.**



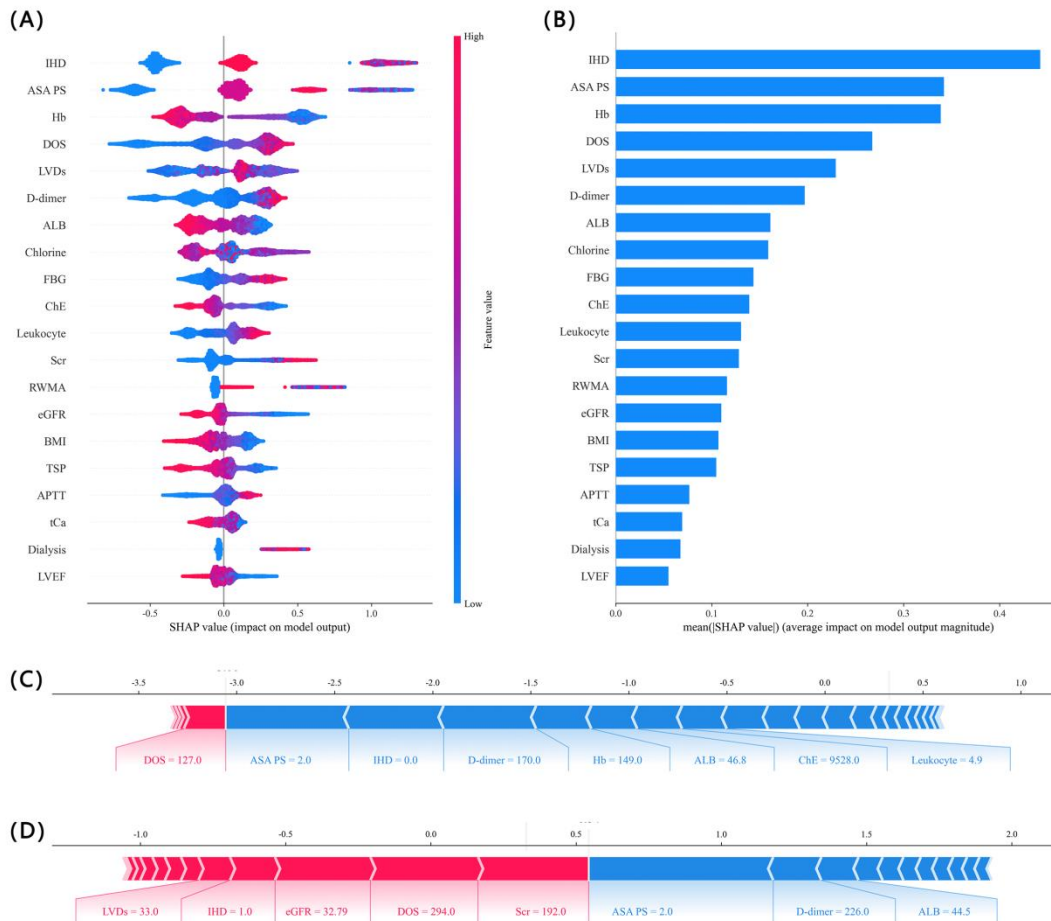
910

911 **Figure 3.**



912

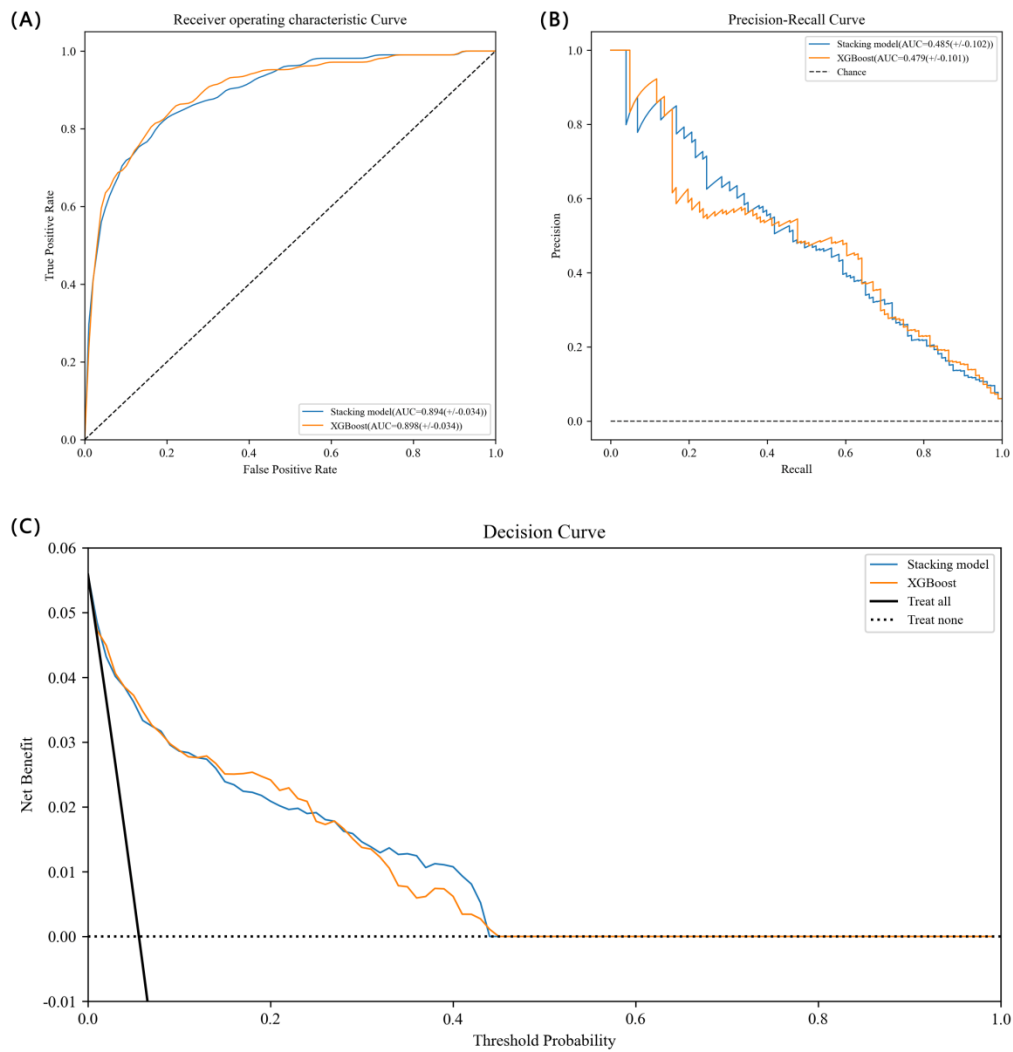
913 **Figure 4.**



914

915 **Figure 5.**

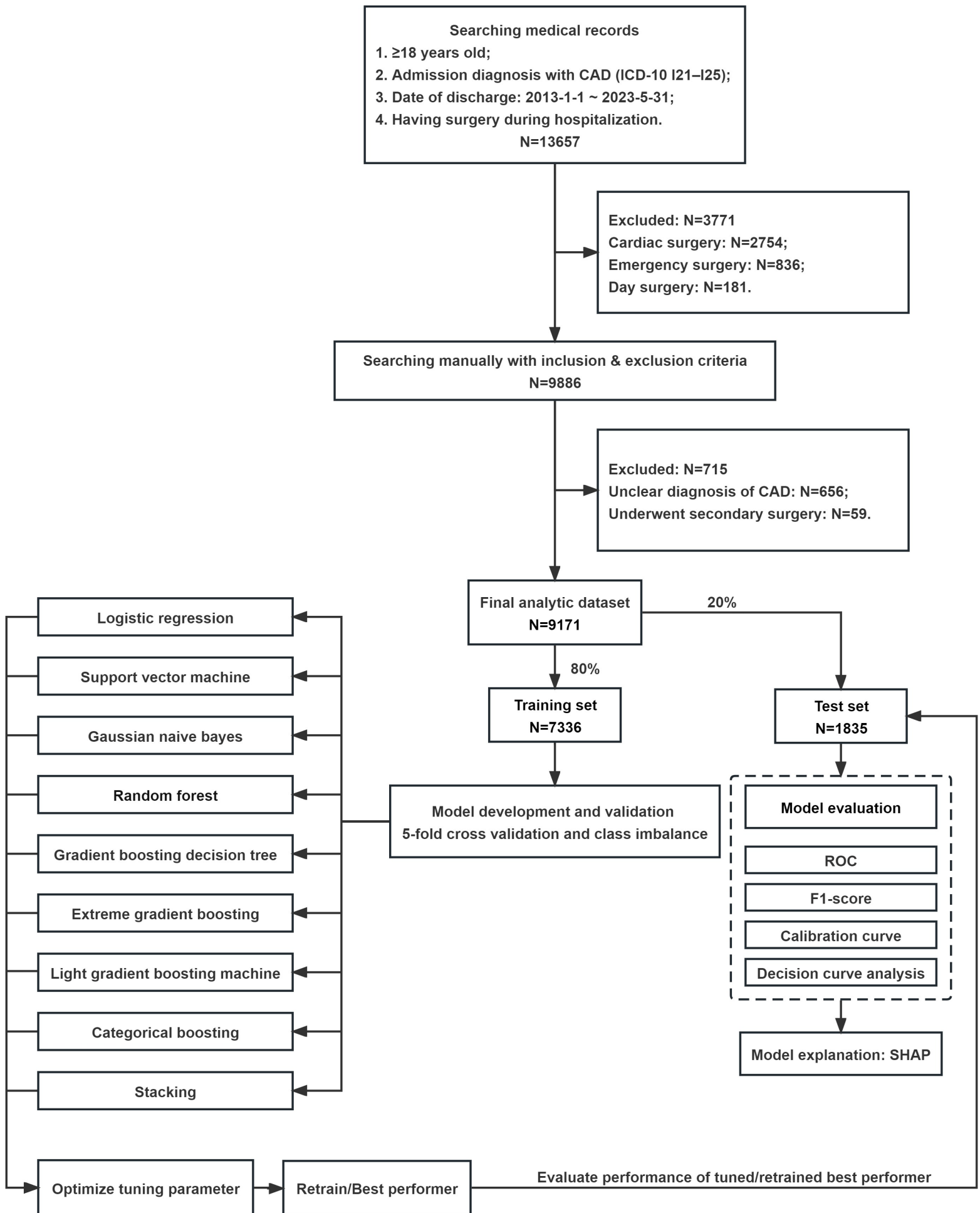


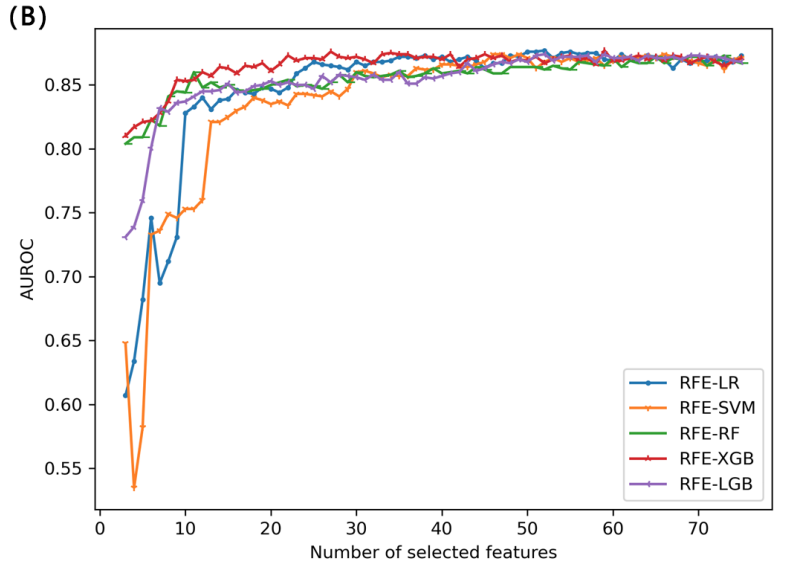
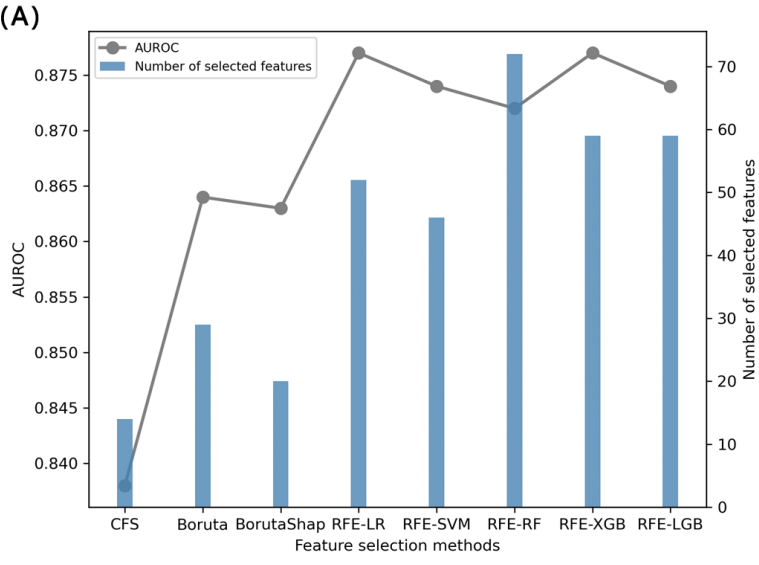


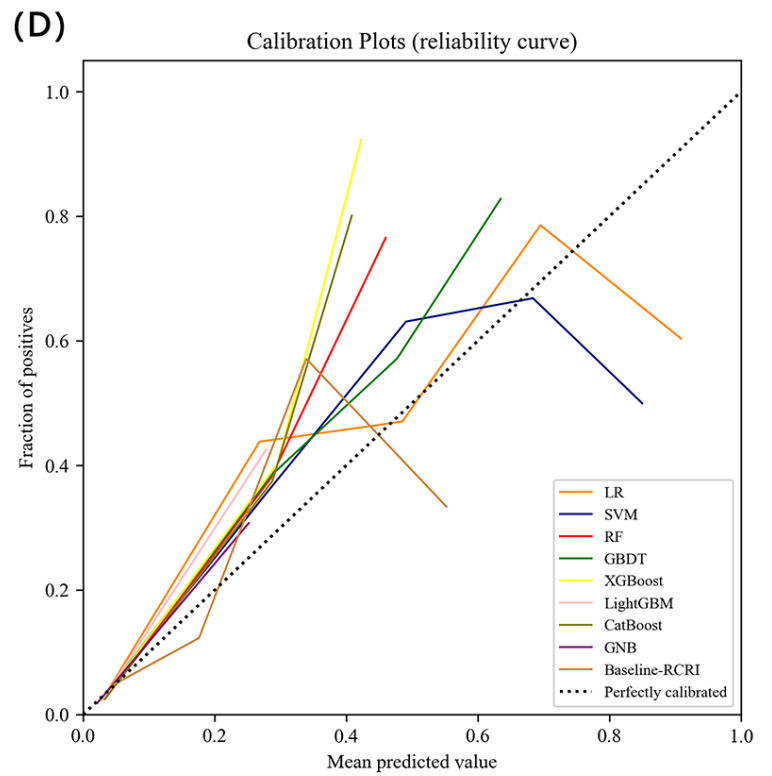
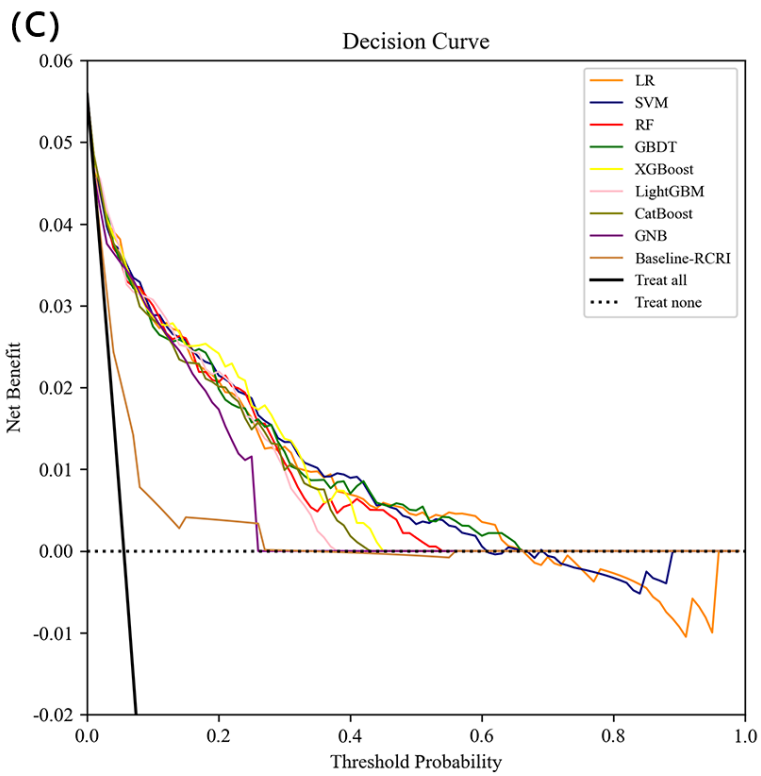
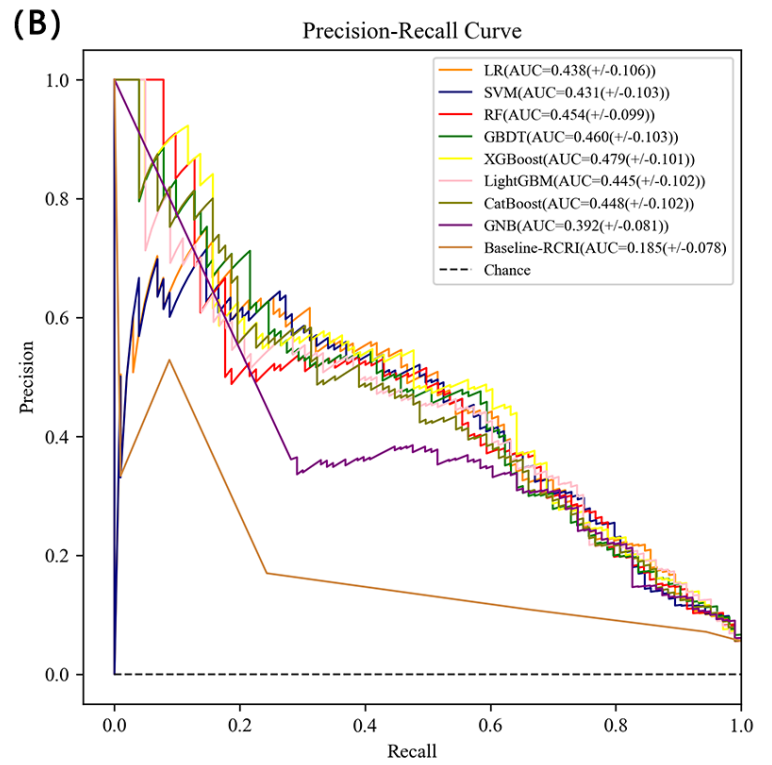
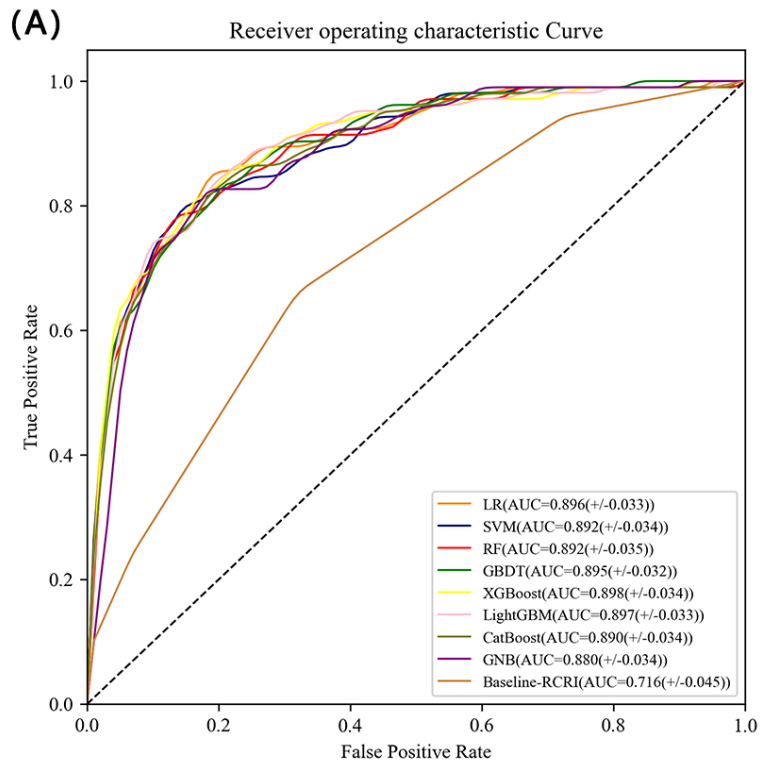
916

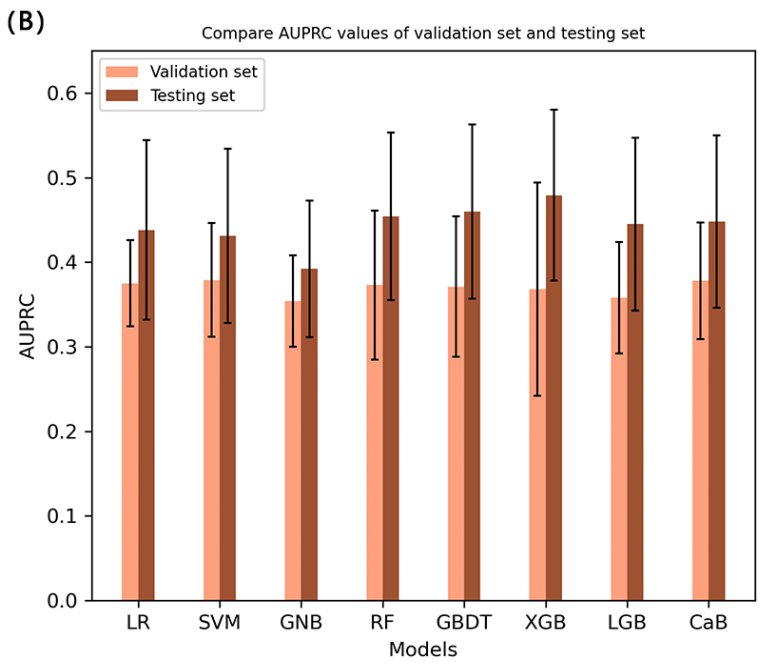
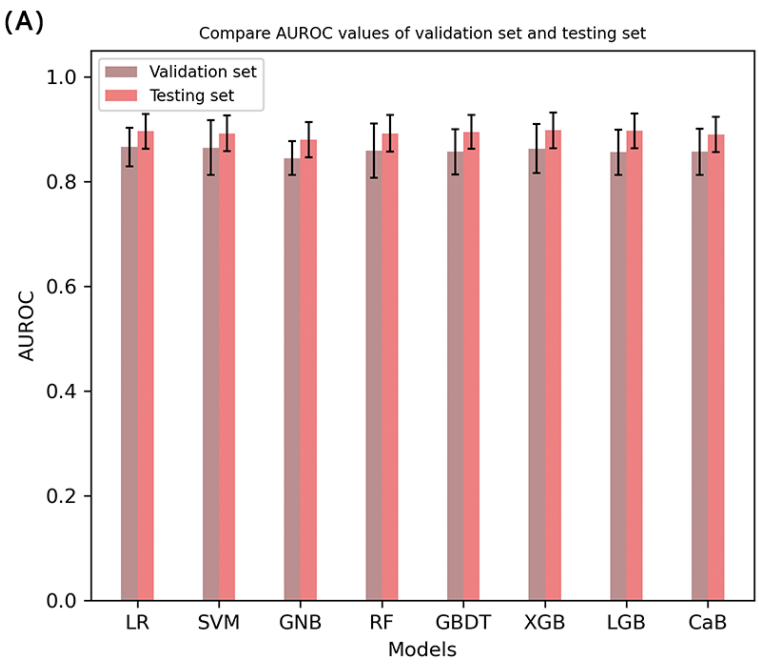
917 **Figure 6.**

918

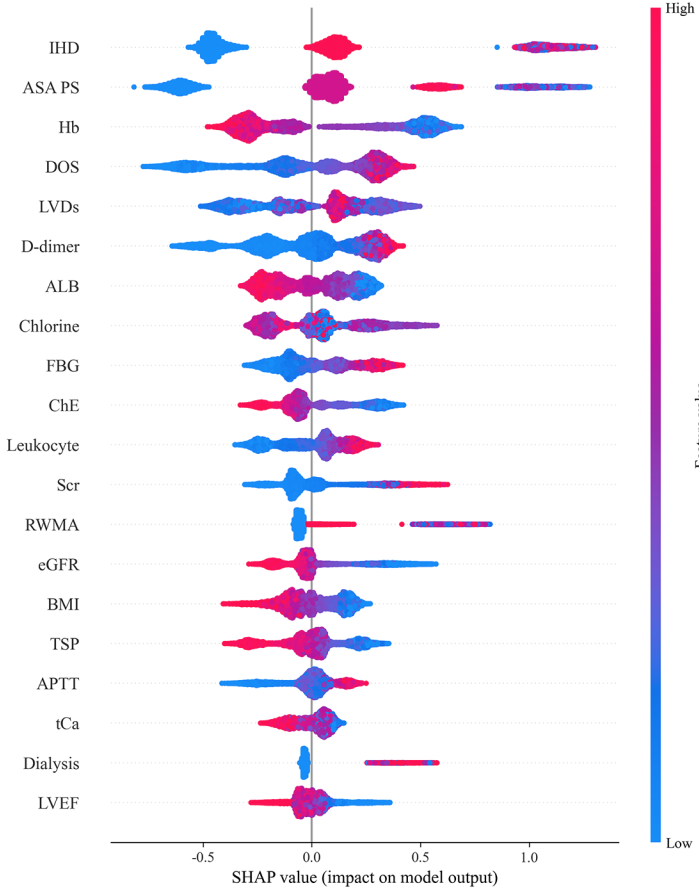




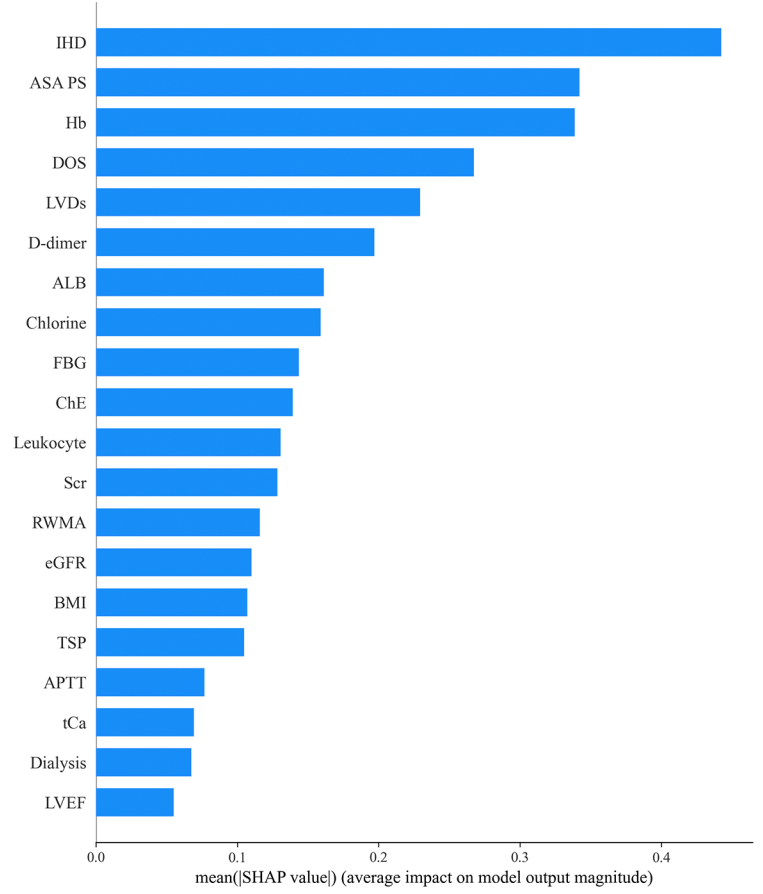




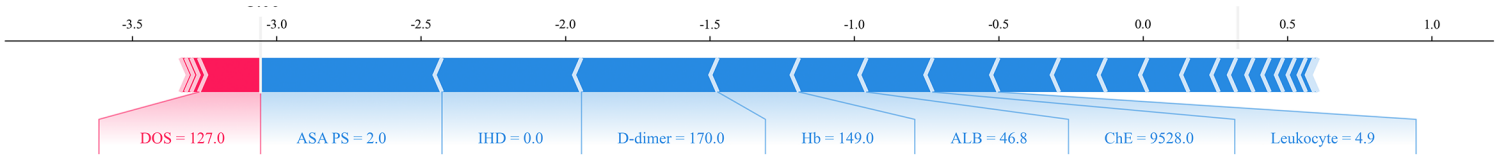
(A)



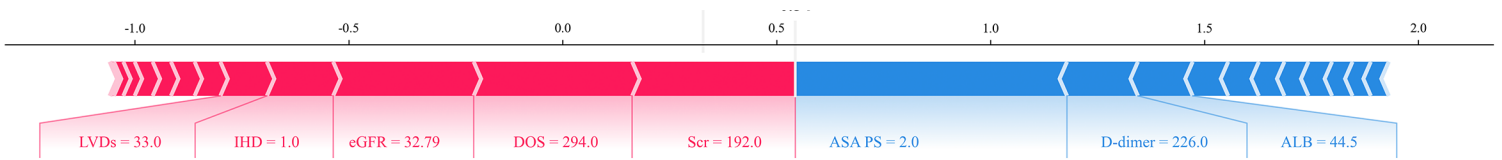
(B)

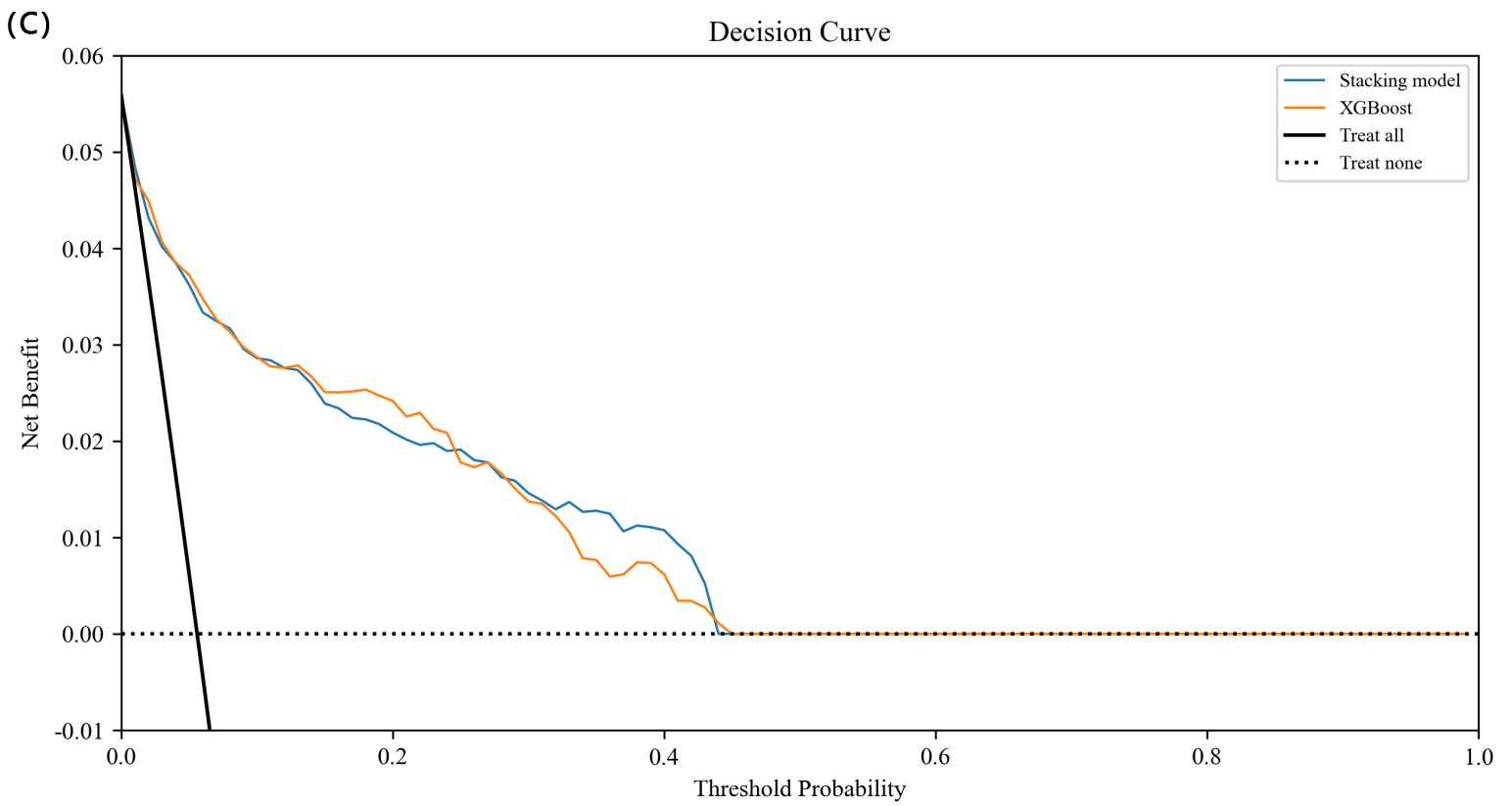
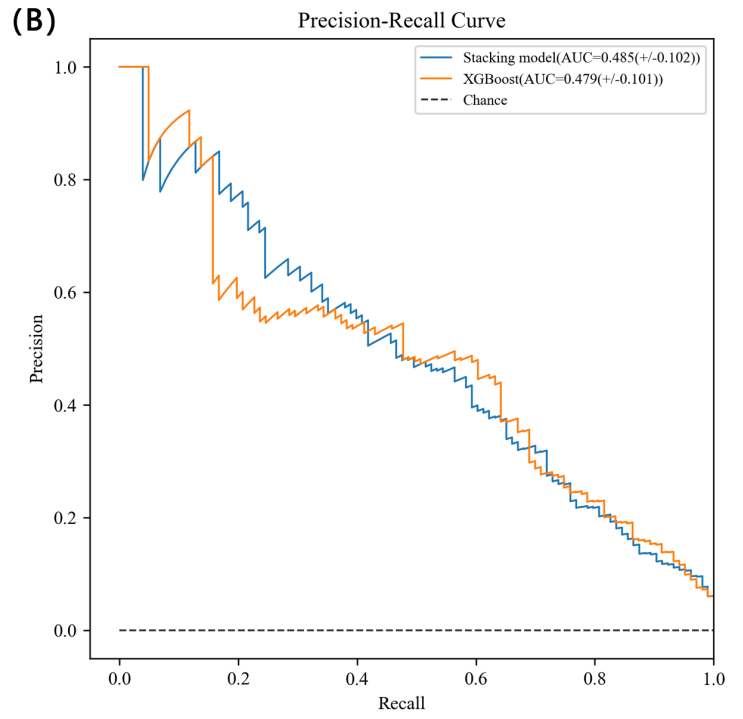
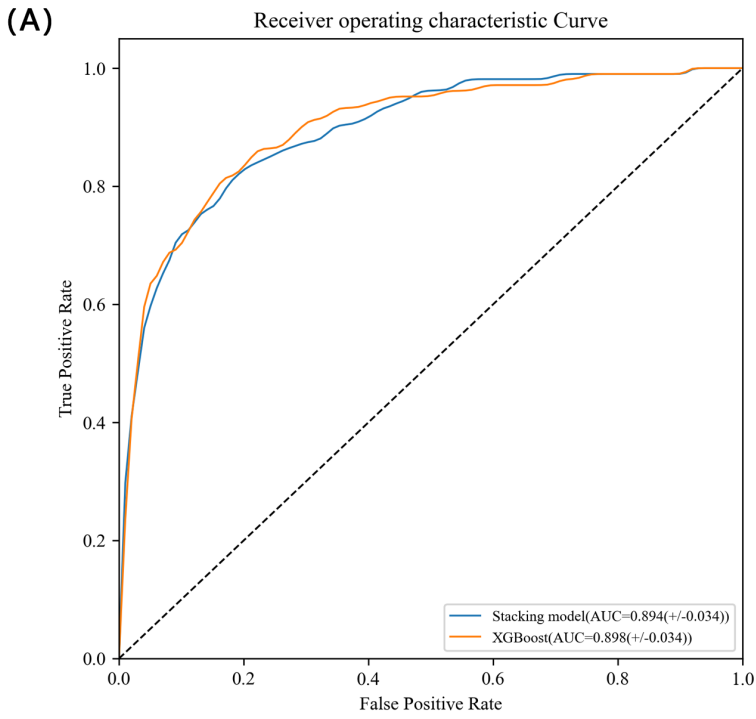


(C)



(D)





**Table 1.** Baseline clinical characteristics of the study population and their association with perioperative outcomes.

Variables	Total (n=9171)	Training and validation set (n=7336)			Testing set (n=1835)		
		Non-MACEs (n=6925)	MACEs (n=411)	P-value	Non-MACEs (n=1732)	MACEs (n=103)	P-value
<b>Demographic</b>							
Age (years)	70 (63, 76)	70 (64, 76)	72 (63, 79)	0.003	70 (63, 76)	69 (62, 77)	0.845
Male	6133 (66.9)	4597 (66.4)	304 (74.0)	0.002	1155 (66.7)	77 (74.8)	0.090
Body mass index (kg/m <sup>2</sup> )	23.6 (21.5, 25.7)	23.7 (21.6, 25.8)	22.0 (19.9, 24.4)	<0.001	23.6 (21.5, 25.6)	22.6 (19.8, 24.4)	0.003
<b>Comorbidities</b>							
Hypertension	5737 (62.6)	4323 (62.4)	266 (64.7)	0.350	1076 (62.1)	72 (69.9)	0.113
Diabetes mellitus	2534 (27.6)	1880 (27.1)	141 (34.3)	0.002	472 (27.3)	41 (39.8)	0.006
Stroke	863 (9.4)	637 (9.2)	55 (13.4)	0.005	158 (9.1)	12 (11.7)	0.390
Dialysis	196 (2.1)	119 (1.7)	51 (12.4)	<0.001	18 (1.0)	8 (7.8)	<0.001
COPD	214 (2.3)	171 (2.5)	8 (1.9)	0.504	34 (2.0)	1 (1.0)	0.719
<b>Cardiac history</b>							
Ischemic heart disease	3454 (37.7)	2498 (36.1)	252 (61.3)	<0.001	641 (37.0)	63 (61.2)	<0.001
Myocardial infarction	2147 (23.4)	1567 (22.6)	116 (28.2)	0.009	434 (25.1)	30 (29.1)	0.356
Heart failure	491 (5.4)	331 (4.8)	73 (17.8)	<0.001	68 (3.9)	19 (18.4)	<0.001
Atrial fibrillation	416 (4.5)	286 (4.1)	48 (11.7)	<0.001	74 (4.3)	8 (7.8)	0.132
Valvular heart disease	168 (1.8)	122 (1.8)	18 (4.4)	<0.001	22 (1.3)	6 (5.8)	0.004
PTCA	1823 (19.9)	1373 (19.8)	78 (19.0)	0.675	349 (20.2)	23 (22.3)	0.593
CABG	146 (1.6)	112 (1.6)	9 (2.2)	0.376	22 (1.3)	3 (2.9)	0.162
<b>Preoperative blood tests</b>							
Leukocyte(×10 <sup>9</sup> /L)	6.2(5.1, 7.7)	6.2 (5.1, 7.6)	7.4(5.6, 10.3)	<0.001	6.1 (5.0, 7.5)	7.3(5.4, 11.3)	<0.001
Hemoglobin (g/L)	131 (117, 144)	132 (118, 144)	104 (84, 123)	<0.001	132 (118, 144)	100 (82, 119)	<0.001
Platelet (×10 <sup>9</sup> /L)	195 (157, 239)	195 (158, 239)	179 (134,237)	<0.001	198 (158, 239)	182 (133,237)	0.023
Fasting blood glucose (mmol/L)	5.47 (4.83, 6.68)	5.45 (4.82, 6.58)	6.80 (5.20, 9.23)	<0.001	5.43 (4.81, 6.53)	7.05 (5.23, 9.69)	<0.001
Serum creatinine (μmol/L)	77 (65, 93)	77 (65, 92)	94 (70, 176)	<0.001	76 (65, 92)	85 (67, 139)	<0.001
eGFR (mL/min/1.73m <sup>2</sup> )	81 (65, 91)	82 (66, 91)	62 (28, 84)	<0.001	81 (66, 91)	71 (36, 90)	<0.001
Total serum protein (g/L)	69.0 (64.0, 73.2)	69.3 (64.5, 73.4)	63.1 (55.5, 68.4)	<0.001	69.2 (64.1, 73.4)	59.9 (53.8, 68.6)	<0.001
Albumin (g/L)	42.7 (38.6, 45.8)	42.9 (39.1, 45.9)	37.2 (32.4, 41.3)	<0.001	42.9 (39.2, 45.9)	35.5 (31.3, 39.8)	<0.001
Globulin (g/L)	26.1 (23.5, 29.0)	26.2 (23.6, 29.0)	25.3 (21.7, 28.9)	<0.001	26.0 (23.5, 29.0)	25.0 (20.8, 29.6)	0.017
ALT (U/L)	18 (13, 26)	18 (13, 26)	16 (10, 27)	0.001	18 (13, 27)	17 (11, 31)	0.738



AST (U/L)	20 (16, 26)	20 (16, 25)	20 (15, 31)	0.225	20 (16, 26)	22 (15, 40)	0.078
GGT (U/L)	25 (17, 42)	25 (17, 42)	27 (16, 52)	0.157	26 (17, 43)	29 (19, 68)	0.025
ALP (U/L)	77 (63, 94)	76 (63, 94)	80 (62, 108)	0.043	77 (64, 93)	80 (62, 113)	0.170
Cholinesterase (U/L)	7435 (6151, 8663)	7500 (6280, 8708)	5546 (3976, 7184)	<0.001	7553 (6289, 8751)	5360 (3294, 7124)	<0.001
Total bilirubin (µmol/L)	10.3 (7.6, 14.3)	10.3 (7.6, 14.3)	9.0 (6.2, 14.5)	0.004	10.4 (7.9, 14.1)	8.6 (6.8, 16.7)	0.393
Direct bilirubin (µmol/L)	4.0 (3.0, 5.4)	4.0 (3.0, 5.4)	4.0 (3.0, 6.1)	0.174	4.0 (3.0, 5.5)	4.3 (3.0, 8.2)	0.074
Indirect bilirubin (µmol/L)	6.0 (4.1, 9.0)	6.1 (4.2, 9.0)	5.0 (3.0, 7.4)	<0.001	6.1 (4.4, 9.0)	4.6 (3.0, 8.4)	<0.001
Potassium (mmol/L)	4.13 (3.85, 4.42)	4.14 (3.85, 4.42)	4.10 (3.70, 4.46)	0.081	4.13 (3.85, 4.44)	3.99 (3.70, 4.40)	0.013
Sodium (mmol/L)	142.0 (140.0, 143.0)	142.0 (140.0, 143.0)	140.0 (138.0, 143.0)	<0.001	142.0 (140.0, 143.0)	140.0 (137.0, 142.0)	<0.001
Chlorine (mmol/L)	104.0 (102.0, 106.0)	104.0 (102.0, 106.0)	104.0 (101.0, 107.0)	0.765	104.0 (102.0, 106.0)	104.0 (101.0, 107.0)	0.728
Total calcium(mmol/L)	2.24 (2.14, 2.33)	2.25 (2.15, 2.34)	2.14 (2.00, 2.25)	<0.001	2.25 (2.16, 2.34)	2.10 (2.00, 2.24)	<0.001
Inorganic phosphorus (mmol/L)	1.10 (0.97, 1.22)	1.10 (0.98, 1.22)	1.11 (0.95, 1.33)	0.027	1.09 (0.97, 1.21)	1.06 (0.90, 1.27)	0.629
Uric acid (µmol/L)	316 (256, 381)	315 (257, 380)	300 (230, 389)	0.022	320 (260, 384)	294 (218, 378)	0.026
Triglyceride (mmol/L)	1.22 (0.91, 1.69)	1.22 (0.91, 1.69)	1.16 (0.87, 1.56)	0.019	1.21 (0.91, 1.70)	1.22 (0.90, 1.68)	0.980
Total cholesterol (mmol/L)	3.64 (3.05, 4.37)	3.64 (3.06, 4.36)	3.41 (2.80, 4.15)	<0.001	3.70 (3.09, 4.46)	3.45 (2.90, 4.33)	0.062
LDL-C (mmol/L)	1.83 (1.40, 2.44)	1.83 (1.41, 2.43)	1.73 (1.23, 2.25)	<0.001	1.87 (1.42, 2.51)	1.74 (1.41, 2.55)	0.159
PT (s)	11.0 (10.9, 12.1)	11.4 (10.9, 12.1)	12.0 (11.3, 13.3)	<0.001	11.4 (10.9, 12.0)	12.5 (11.7, 13.5)	<0.001
APTT (s)	27.2 (25.2, 29.4)	27.1 (25.2, 29.3)	28.6 (26.2, 34.2)	<0.001	27.1 (25.1, 29.1)	29.9 (26.5, 35.5)	<0.001
Fibrinogen (g/L)	3.00 (2.53, 3.65)	2.99 (2.53, 3.64)	3.32 (2.49, 4.34)	<0.001	3.00 (2.56, 3.59)	2.93 (2.28, 4.18)	0.717
D-dimer(µg/L FEU)	460 (234, 1055)	434 (223, 962)	1366 (552, 3788)	<0.001	439 (230, 1000)	2340 (704, 5116)	<0.001
Preoperative ECG							
AQW	501 (6.3)	357 (5.9)	32 (10.3)	0.001	104 (6.9)	8 (12.1)	0.134
ST-Ta	3901 (49.0)	2913 (47.9)	204 (65.8)	<0.001	739 (48.8)	45 (68.2)	0.002
Preoperative echocardiography							

AO (mm)	30.0 (27.0, 33.0)	30.0 (27.0, 33.0)	30.5 (28.0, 33.0)	0.108	30.0 (27.0, 32.0)	30.0 (27.5, 32.0)	0.817
IVSd (mm)	9.0 (9.0, 10.0)	9.0 (9.0, 10.0)	10.0 (9.0, 11.0)	<0.001	9.0 (9.0, 10.0)	10.0 (9.0, 11.0)	0.124
LVDd (mm)	48.0 (44.0, 51.0)	47.0 (44.0, 51.0)	48.0 (45.0, 52.0)	0.001	48.0 (44.0, 51.0)	50.0 (45.0, 53.0)	0.035
FS (%)	37.0 (33.0, 40.0)	37.0 (34.0, 40.0)	35.0 (31.8, 39.0)	<0.001	37.0 (34.0, 40.0)	34.0 (31.0, 39.0)	<0.001
LA (mm)	34.0 (30.0, 38.0)	34.0 (30.0, 38.0)	35.0 (31.0, 41.0)	<0.001	34.0 (30.3, 37.0)	35.0 (31.0, 40.0)	0.044
LVPWd (mm)	9.0 (9.0, 10.0)	9.0 (9.0, 10.0)	10.0 (9.0, 11.0)	<0.001	9.0 (9.0, 10.0)	10.0 (8.5, 11.0)	0.334
LVDs (mm)	30.0 (27.0, 33.0)	30.0 (27.0, 32.0)	31.0 (28.0, 35.0)	<0.001	30.0 (27.0, 33.0)	31.0 (27.0, 35.0)	0.055
LVEF (%)	66.0 (62.0, 71.0)	66.0 (62.0, 71.0)	63.0 (58.0, 69.0)	<0.001	66.0 (62.0, 70.0)	62.0 (58.3, 68.8)	<0.001
RWMA	532 (7.5)	348 (6.5)	68 (19.7)	<0.001	90 (6.9)	26 (28.9)	<0.001
PH	326 (4.6)	221 (4.1)	47 (13.6)	<0.001	49 (3.7)	9 (10.0)	<0.001
LVDD	6210 (87.1)	4741 (88.1)	253 (73.1)	<0.001	1155 (88.1)	61 (67.8)	<0.001
ASA class				<0.001			<0.001
II	3887 (42.4)	3072 (44.4)	61 (14.8)		742 (42.8)	12 (11.7)	
III	5210 (56.8)	3834 (55.4)	312 (75.9)		985 (56.9)	79 (76.7)	
IV	74 (0.8)	19 (0.3)	38 (9.2)		5 (0.3)	12 (11.7)	
General anesthesia	6702 (73.1)	5016 (72.4)	338 (82.2)	<0.001	1273 (73.5)	75 (72.8)	0.879
Types of surgery							
General	2823 (30.8)	2082 (30.1)	154 (37.5)	0.002	546 (31.5)	41 (39.8)	0.080
Abdominal	2156 (23.5)	1573 (22.7)	131 (31.9)	<0.001	414 (23.9)	38 (36.9)	0.003
Nonabdominal	667 (7.3)	509 (7.4)	23 (5.6)	0.183	132 (7.6)	3 (2.9)	0.075
Thoracic	1060 (11.6)	805 (11.6)	29 (7.1)	0.005	219 (12.6)	7 (6.8)	0.079
Orthopedic	854 (9.3)	643 (9.3)	46 (11.2)	0.198	154 (8.9)	11 (10.7)	0.538
ENT	236 (2.6)	192 (2.8)	2 (0.5)	0.005	39 (2.3)	3 (2.9)	0.509
Neurological	444 (4.8)	321 (4.6)	31 (7.5)	0.007	80 (4.6)	12 (11.7)	0.001
Gynecologic	171 (1.9)	139 (2.0)	4 (1.0)	0.141	27 (1.6)	1 (1.0)	1.000
Urologic	1550 (16.9)	1185 (17.1)	54 (13.1)	0.037	307 (17.7)	4 (3.9)	<0.001
Ophthalmology	767 (8.4)	603 (8.7)	1 (0.2)	<0.001	162 (9.4)	1 (1.0)	0.001
Vascular	1113 (12.1)	835 (12.1)	87 (21.2)	<0.001	169 (9.8)	22 (21.4)	<0.001
Aortic	113 (1.2)	76 (1.1)	16 (3.9)	<0.001	17 (1.0)	4 (3.9)	0.027
Non-aortic	1000 (10.9)	759 (11.0)	71 (17.3)	<0.001	152 (8.8)	18 (17.5)	0.003
Dental	153 (1.7)	120 (1.7)	3 (0.7)	0.124	29 (1.7)	1 (1.0)	1.000
DOS (min)	87 (50, 147)	84 (49, 141)	140 (80, 211)	<0.001	87 (50, 144)	139(79, 234)	<0.001

Notes: Results presented as median (IQR), or n (%).

Abbreviations: COPD, chronic obstructive pulmonary disease; PTCA, percutaneous transluminal coronary angioplasty; CABG, coronary artery bypass graft; eGFR, estimated glomerular filtration rate; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT,

gamma-glutamyl transferase; ALP, alkaline phosphatase; LDL-C, low density lipoprotein cholesterol; PT, prothrombin time; APTT, activated partial thromboplastin time; AQW, abnormal Q waves; ST-Ta, ST-T wave abnormalities; AO, aorta diameter; IVSd, interventricular septum thickness at end diastole; LVDD, left ventricular end diastolic dimension; FS, fractional shortening; LA, left atrial anteroposterior dimension; LVPWd, left ventricular posterior wall thickness at end diastole; LVDs, left ventricular end systolic dimension; LVEF, left ventricular ejection fraction; RWMA, regional wall motion abnormality; PH, pulmonary hypertension; LVDD, left ventricle diastolic dysfunction; ASA, American Society of Anesthesiologists; ENT, ear, nose, and throat; DOS, duration of surgery.

**Table 2.** The internal verification results of models trained were obtained by using different balanced class methods combined with cross-validation using all features.

Models	Methods	AUROC(95%CI)	AUPRC(95%CI)	Accuracy	Specificity	Sensitivity	Youden Index
LR	-	0.879(+/-0.024)	0.451(+/-0.126)	0.879	0.991	0.270	0.261
RF	-	0.864(+/-0.036)	0.414(+/-0.095)	0.944	1.000	0.000	0.000
XGBoost	-	0.862(+/-0.047)	0.403(+/-0.055)	0.946	1.000	0.034	0.034
LR	SMOTE	0.867(+/-0.036)	0.412(+/-0.121)	0.824	0.828	0.757	0.585
RF	SMOTE	0.870(+/-0.047)	0.386(+/-0.105)	0.943	0.979	0.343	0.322
XGBoost	SMOTE	0.875(+/-0.041)	0.427(+/-0.137)	0.942	0.976	0.372	0.348
LR	ANSYN	0.864(+/-0.041)	0.408(+/-0.134)	0.806	0.809	0.764	0.573
RF	ANSYN	0.871(+/-0.044)	0.388(+/-0.104)	0.941	0.973	0.399	0.372
XGBoost	ANSYN	0.877(+/-0.035)	0.415(+/-0.137)	0.947	0.982	0.348	0.330
LR	SMOTE+ENN	0.868(+/-0.036)	0.396(+/-0.105)	0.726	0.719	0.842	0.561
RF	SMOTE+ENN	0.863(+/-0.053)	0.360(+/-0.096)	0.864	0.691	0.874	0.565
XGBoost	SMOTE+ENN	0.872(+/-0.037)	0.397(+/-0.105)	0.894	0.907	0.672	0.579

**Table 3.** The predictive performance in the test set of the 9 models. LR: logistic regression; SVM: support vector machine; GNB: Gaussian Naive Bayes; RF: random forest; GBDT: gradient boosting decision tree; XGBoost: extreme gradient boosting; LightGBM: light gradient boosting machine; CatBoost: categorical boosting; AUROC: area under the receiver operating characteristic curve; AUPRC: area under the precision recall curve.

Models	AUROC (95%CI)	AUPRC (95%CI)	Accuracy	Specificity	Precision	Recall (Sensitivity)	F1 score	Youden Index
Baseline-RCRI	0.716(+/-0.045)	0.185(+/-0.078)	0.679	0.680	0.109	0.662	0.187	0.342
LR	0.896(+/-0.033)	0.438(+/-0.106)	0.732	0.722	0.160	0.895	0.271	0.617
SVM	0.892(+/-0.034)	0.431(+/-0.103)	0.780	0.777	0.181	0.837	0.298	0.613
GNB	0.880(+/-0.034)	0.392(+/-0.081)	0.879	0.887	0.279	0.740	0.404	0.627
RF	0.892(+/-0.035)	0.454(+/-0.099)	0.777	0.773	0.181	0.845	0.297	0.618
GBDT	0.895(+/-0.032)	0.460(+/-0.103)	0.901	0.916	0.314	0.652	0.423	0.568
XGBoost	0.898(+/-0.034)	0.479(+/-0.101)	0.874	0.881	0.271	0.748	0.397	0.629
LightGBM	0.897(+/-0.033)	0.445(+/-0.102)	0.845	0.850	0.230	0.758	0.353	0.608
CatBoost	0.890(+/-0.034)	0.448(+/-0.102)	0.804	0.803	0.199	0.826	0.320	0.630