BMC Anesthesiology

RESEARCH

Open Access

Development and external validation of a machine learning model for predicting severe immediate postoperative complications in adults undergoing cardiac surgery

Hyeon Cheun[1†], Jang Ho Ahn[3†], Jin-Woo Park[5†], Seo Hee Ko[6,7†], Jae-Kwang Shim[6,7], Hyeonhoon Lee[4,8,9], Youn Joung Cho[1,2], Karam Nam[1,2], Jae-Woo Ju[1,2], Jaeyeon Chung[1], In Jung Kim[1], Hyung-Chul Lee[1,2,4*] and Yunseok Jeon[1,2*]

Abstract

Background Patients undergoing cardiac surgery are at high risk of severe immediate postoperative complications. Predicting immediate postoperative adverse events remains challenging owing to the complex and nonlinear interplay of numerous risk factors. This study aimed to develop and externally validate a machine-learning (ML) model to predict critical outcomes during the immediate postoperative period after cardiac surgery.

Methods Adult patients who underwent cardiac surgery at Seoul National University Hospital (SNUH) between October 2004 and October 2021 were included for model development and internal validation. Thirty-seven preoperative and intraoperative variables were used as model inputs. The primary outcome was a composite of reoperation for bleeding, death, cardiac arrest, and initiation of mechanical circulatory support with extracorporeal membrane oxygenation (ECMO) or intra-aortic balloon pump (IABP) within 24 h after surgery. An extreme gradient boosting (XGBoost) model was developed as the primary model and compared with a least absolute shrinkage and selection operator (LASSO) regularized logistic regression model. An equal-weighted ensemble of the two models was additionally constructed. Model performance was assessed using the area under the receiver operating characteristic curve (AUROC) with 95% confidence intervals (CIs), and pairwise comparisons were performed using DeLong's test. For external validation, we analyzed data from Seoul National University Bundang Hospital (SNUBH) and Severance Hospital.

Results A total of 7,946 patients from SNUH were used for model development, with external validation performed in 2,270 patients from SNUBH and 1,966 patients from Severance Hospital. The incidence of the composite outcome

[†]Hyeon Cheun, Jang Ho Ahn, Jin-Woo Park and Seo Hee Ko contributed equally to this work.

*Correspondence:
Hyung-Chul Lee
vital@snu.ac.kr
Yunseok Jeon
jeonyunseok@gmail.com

Full list of author information is available at the end of the article

was 5.93%, 1.37%, and 2.49% at SNUH, SNUBH, and Severance Hospital, respectively. The XGBoost model achieved an AUROC of 0.804 (95% CI, 0.707–0.891) in internal validation, and 0.735 (95% CI, 0.636–0.829) and 0.712 (95% CI, 0.630–0.795) in external validations at SNUBH and Severance Hospital, respectively. The XGBoost model demonstrated the highest discrimination in internal validation among the XGBoost, LASSO and ensemble models; however, no statistically significant differences were observed among the three models in external validation.

Conclusions The proposed ML model demonstrated acceptable discrimination in internal validation and maintained baseline discriminative capacity across two independent external cohorts in predicting severe immediate postoperative complications after cardiac surgery. However, precision-oriented metrics declined substantially in external settings, largely attributable to lower event rates, underscoring the need for institution-specific recalibration prior to clinical implementation.

Keywords Cardiac surgery, Machine learning, Postoperative complications, Predictive model, External validation

Introduction

Patients undergoing cardiac surgery are at high risk of major postoperative complications, which are associated with increased risk of multi-organ dysfunction and poor clinical outcomes [1, 2]. A previous study conducted in the United States between 2012 and 2021 reported a 28.5% incidence of major adverse events, including in-hospital mortality and cardiac complications [3].

Among these complications, severe postoperative events — such as cardiac arrest, reoperation for bleeding, initiation of mechanical circulatory support, and death — represent particularly critical events requiring urgent intervention. Cardiac arrest occurs in approximately 0.7% of patients within the first 24 h after cardiac surgery, and reoperation for bleeding in 5–9% within the same period [4, 5]. In a large cohort study, 15% of all in-hospital mortality after cardiac surgery has been reported to occur on postoperative day 0, and postcardiotomy cardiogenic shock requiring mechanical circulatory support complicates 2–6% of cardiac surgical procedures [6, 7].

Effective prevention and timely management of these complications during the immediate postoperative period are essential. The first 24 h after cardiac surgery represent a critical transition phase in which patients either recover sufficiently to be transferred from the intensive care unit to a general ward or require continued intensive monitoring and rapid intervention due to an elevated risk of hemodynamic deterioration [8]. However, conventional methods for predicting such events often rely on clinical intuition, which is limited by subjectivity, variability, and a lack of reproducibility.

Machine learning (ML) algorithms have recently emerged as powerful tools for predictive analytics in perioperative medicine, including risk prediction after cardiac surgery [9]. Several studies have demonstrated that ML-based models can perform comparably or superiorly to traditional risk stratification systems, such as the EuroSCORE [10] and the Society of Thoracic Surgeons (STS) score [11–16]. Nevertheless, most previous ML studies have primarily focused on long-term outcomes, and few have specifically addressed the prediction of severe adverse events occurring in the immediate postoperative period after cardiac surgery. Moreover, owing to their complexity and large number of parameters, ML models are prone to overfitting. Therefore, external validation is essential to assess their generalizability and clinical applicability [17].

Accordingly, this study aimed to develop and externally validate an ML-based predictive model for severe immediate postoperative complications within the first 24 h following cardiac surgery. Accurate and timely identification of patients at high risk for severe immediate postoperative complications may facilitate optimized resource allocation and improve patient outcomes.

Methods

Study design and population

This retrospective study was conducted at three tertiary hospitals; Seoul National University Hospital (SNUH), Seoul National University Bundang Hospital (SNUBH), and Severance Hospital. The study protocol was approved by the institutional review boards of all participating institutions (SNUH, IRB no. 2210-145-1373, approved November 8, 2022; SNUBH, IRB no. B-2309-852-405, approved August 25, 2023; and Severance Hospital, IRB no. 4-2023-0881, approved August 31, 2023). All research procedures were conducted in accordance with the Declaration of Helsinki and relevant institutional guidelines and regulations. The requirement for informed consent was waived by each institutional review board due to the retrospective nature of the study.

Adult patients (≥ 18 years) who underwent cardiac surgery at SNUH between October 2004 and October 2021 were eligible for inclusion in the study. Data from SNUBH were collected for surgeries performed between June 2007 and December 2021 and data from Severance Hospital were collected for surgeries performed between January 2019 and June 2022. We included only major cardiac surgeries lasting more than 4 h, and for patients who underwent multiple cardiac surgeries during the study

period, only the first operation was analyzed. Patients who experienced intraoperative death, or had incomplete vital sign recordings were excluded.

Data collection

A total of 37 variables were used as the input features for the model development. These included demographics data (age, sex, and body mass index [kg/m^2]); preoperative laboratory values such as serum potassium (mEq/L), serum chloride (mEq/L), serum calcium (mg/dL), hemoglobin (g/dL), white blood cell count (x 10^3/μL), platelet (x 10^3/μL), lactate (mmol/L), blood urea nitrogen (mg/dL), blood glucose (mg/dL), serum creatinine (mg/dL), and c-reactive protein (mg/L); and the left ventricular ejection fraction (%) obtained from transthoracic echocardiography (most recent measurement within three months before surgery).

The intraoperative variables included total crystalloid infusion volume (mL), number of red blood cell units transfused (units), estimated blood loss (mL), and cardiopulmonary bypass (CPB) time (minutes). For off-pump procedures, the CPB time was assigned a value of '0'. Mean intraoperative laboratory measurements – including arterial pH, bicarbonate (mmol/L), serum sodium (mmol/L), serum potassium (mmol/L), blood glucose (mg/dL), lactate (mmol/L), base excess in extracellular fluid, partial pressure of oxygen and carbon dioxide (mmHg) – were also incorporated.

The hemodynamic variables included arterial systolic and diastolic blood pressures (mmHg), heart rate (beats/min), oxygen saturation (%), body temperature (℃), and central venous pressure (mmHg). Hemodynamic data were recorded at 1-minute intervals, whereas body temperature was recorded at 5-minute intervals and measured using a nasal temperature probe. Arterial blood pressure was primarily measured via the radial artery; however, when femoral arterial pressure monitoring was performed, femoral measurements were used instead of radial values. Non-invasive blood pressure measurements were not included in the analysis.

Preoperative continuous renal replacement therapy (CRRT) was also recorded. Information regarding the application of extracorporeal membrane oxygenation (ECMO) or intra-aortic balloon pump (IABP) was collected for both the preoperative and intraoperative periods. Implausible hemodynamic data points – such as blood pressure < 20mmHg or > 200mmHg, oxygen saturation < 15% or > 100%, body temperature < 15℃ or > 40℃ - were treated as missing values prior to statistical feature extraction.

Study outcomes

Severe immediate postoperative complications were defined as a composite endpoint comprising immediate postoperative complications that occurred within 24 h after surgery. These included any of the following: (1) reoperation for bleeding control, (2) death, (3) cardiac arrest, or (4) initiation of mechanical circulatory support with ECMO or IABP. Cardiac arrest was defined as the occurrence of asystole or initiation of cardiac massage.

Statistical analysis and model development

Data from each hospital were summarized as mean ± standard deviation (SD), median with interquartile range (IQR), or number (proportion), as appropriate. The normality of continuous variables was assessed using the Shapiro-Wilk test in conjunction with visual inspection of histograms and quantile-quantile (Q-Q) plots. Continuous variables that approximated a normal distribution were compared across hospitals using analysis of variance (ANOVA) and are presented as mean ± SD, whereas variables with non-normal distributions were analyzed using the Kruskal-Wallis test and are reported as median (IQR). Categorical variables were compared using chi-square tests. All statistical analyses were performed using R version 4.3.1 (R foundation for Statistical Computing, Vienna, Austria).

Model development and machine learning analyses were performed using Python 3.9 (Python Software Foundation, Wilmington, DE, USA). Extreme gradient boosting (XGBoost) was implemented as the base machine learning algorithm. XGBoost is an ML-based model that implements a series of decision trees, with each tree correcting the residuals of the previous trees [18]. Missing values were imputed using the iterative imputer from the Scikit-learn library [19].

The SNUH dataset was randomly divided into a development cohort (95.0%) and an internal test cohort (5.0%). Considering the low event rate, we decided on the 95:5 split of the dataset to prioritize the model's representational capacity and generalizability. The development dataset was further partitioned into five patient-stratified folds and underwent five-fold cross-validation. Hyperparameter optimization was conducted using the Optuna framework, with the mean area under the receiver operating characteristic curve (AUROC) across all validation folds [20]. The test dataset was independently imputed and used for internal validation, while datasets from SNUBH and Severance Hospital served as external validation cohorts.

Model performance evaluation

The model performance was evaluated using complementary metrics, including the AUROC, area under the precision-recall curve (AUPRC), F1-score, accuracy and 95% confidence intervals (CIs). AUROC reflects overall discrimination, whereas AUPRC is more informative in imbalanced datasets by emphasizing performance on

positive cases [21]. The F1-score balances precision and recall, and accuracy summarizes the proportion of correctly classified predictions [22]. The use of these metrics together provides a comprehensive assessment of the model performance across different aspects of clinical relevance [22]. The optimal discrimination threshold was determined using Youden's index derived from the SNUH development dataset [23]. Calibration was assessed through spline recalibration and quantified using the Brier score and the integrated calibration index [24–26]. In addition to numerical metrics, we also presented calibration plots to visually assess the agreement between the predicted and observed risks, confusion matrices to summarize the threshold-based classification performance, and test probability histograms to illustrate the distribution of predicted risks across outcome groups.

SHAP-based feature selection and post-hoc analysis

Shapley additive explanations (SHAP) were employed for feature importance analysis and to enhance interpretability of the XGBoost model [27]. Among the 37 candidate variables, features were iteratively selected in descending order of their mean absolute SHAP values. Specifically, this SHAP-based forward selection was performed on the entire development dataset, not nested within a cross-validation loop, to rank features by importance. Forward selection was then carried out while monitoring performance on the internal test cohort. SHAP dependency plot was generated to visualize the relationship between each predictor and the model output, illustrating both the magnitude and direction of each feature's contribution [27].

Decision curve analysis

To evaluate the clinical utility of our predictive model, we performed decision curve analysis (DCA) [28]. It calculates the net benefit of a predictive model across a range of threshold probabilities, comparing it against default strategies of treating all or treating no patients. We generated decision curves for the internal test set and both external validation cohorts. For the external validation cohorts, predicted probabilities were recalibrated using Platt scaling prior to decision curve analysis. Specifically, a logistic regression model was fitted with the log-odds of the original predicted probabilities as the sole predictor, re-estimating both the intercept and slope to align the predicted probabilities with the observed event rates in each external cohort.

Model comparison and ensemble construction

Model performance was compared to a least absolute shrinkage and selection operator (LASSO) regularized logistic regression model developed with the same 37 variables using the same training cohort. Given that XGBoost captures complex nonlinear interactions among variables [18], while LASSO provides a parsimonious linear summary that may generalize more robustly across heterogeneous populations [29], we hypothesized that combining their predicted probabilities may yield more consistent performance across institutions. We additionally constructed an equal-weighted ensemble model by averaging the predicted probabilities of the XGBoost and LASSO models. Equal weighting (w = 0.5) was selected to avoid overfitting to the internal test set; a post-hoc sensitivity analysis across weights ranging from 0 to 1 confirmed that no single weight simultaneously optimized discrimination across all three cohorts, supporting equal weighting as the most generalizable strategy. DeLong's test was performed to compare the AUROC among the three models — XGBoost, LASSO, and the ensemble — across the internal and external validation cohorts.

Results

A total of 8,027 patients were initially screened at SNUH, and 81 patients were excluded based on the predefined criteria. Finally, 7,946 patients were included in the analysis. For external validation, 2,270 patients from SNUBH and 1,966 patients from Severance Hospital were included in the analysis (Fig. 1).

The baseline characteristics and intraoperative profiles of the patients from the three hospitals are summarized in Table 1. The preoperative use of ECMO or IABP was more frequent at SNUH than at the other centers (SNUH: 319 [4.0%], SNUBH: 57 [2.5%] and Severance: 8 [0.4%], $P < 0.001$). A similar pattern was observed for intraoperative ECMO or IABP use (SNUH: 848 [10.7%], SNUBH: 93 [4.1%], and Severance: 13 [0.7%], $P < 0.001$). CPB time was also significantly longer at SNUH (SNUH: mean 187 ± SD 76.9 min, SNUBH: mean 126 ± SD 49.3 min, Severance: mean 105 ± SD 48.7 min, $P < 0.001$). In contrast, patients at SNUBH received larger volumes of intraoperative crystalloid and units of red blood cell (RBC) transfusions compared with the other institutions (SNUH: median 1,350 [IQR 800-2,400] mL and median 2 [IQR 1–3] units, SNUBH: median 2,695 [IQR 2,039 – 3,670] mL and median 4.5 [2.5–7.5] units, Severance: median 1,100 [IQR 900-1,500] mL and median 0 [IQR 0–0] units, $P < 0.001$). Certain intraoperative hemodynamic variables were slightly higher at Severance Hospital than at the other two centers, including systolic blood pressure (SNUH: mean 98.8 ± SD 13.4 mmHg, SNUBH: mean 97.9 ± SD 9.24 mmHg, Severance: mean 101 ± SD 10.6 mmHg, $P < 0.001$), diastolic blood pressure (SNUH: mean 55.7 ± SD 6.94 mmHg, SNUBH: mean 57.5 ± SD 7.43 mmHg, Severance: mean 76.0 ± SD 9.28 mmHg, $P < 0.001$), body temperature (SNUH: median 34.1 [IQR 32.9–35.5] ℃, SNUBH: median 34.9 [IQR 34.2–35.4] ℃, Severance: median 35.5 [IQR 35.1–35.8] ℃, $P < 0.001$),

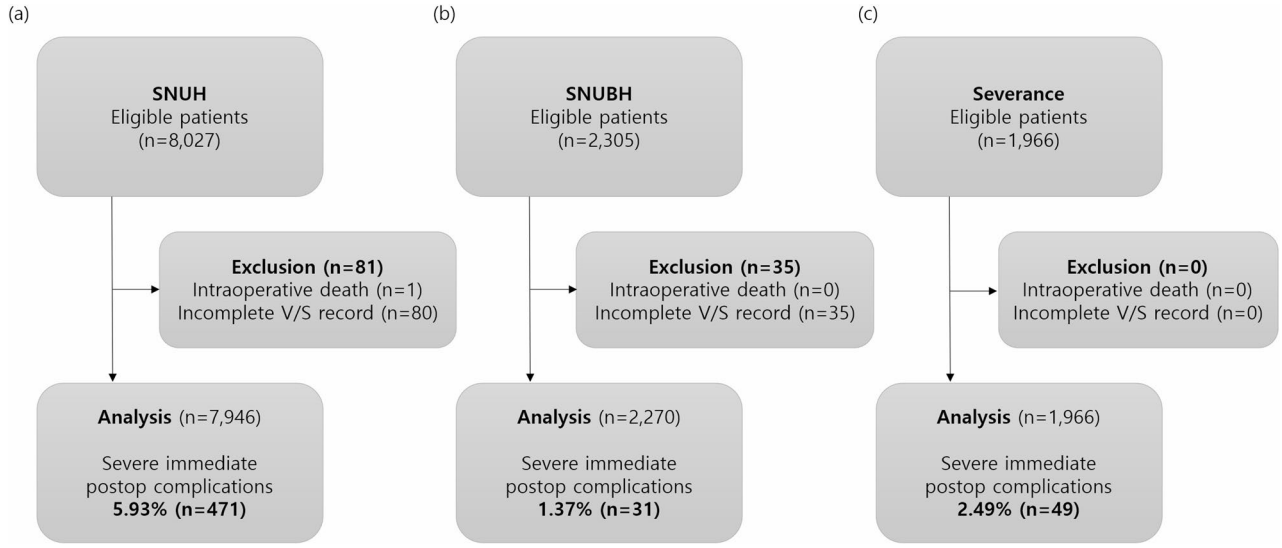

Fig. 1 Flow diagrams of (**a**) Seoul National University Hospital, (**b**) Seoul National University Bundang Hospital, and (**c**) Severance Hospital. SNUH, Seoul National University Hospital; SNUBH, Seoul National University Bundang Hospital; V/S, vital sign; postop, postoperative

and central venous pressure (SNUH: mean 7.75 ± SD 2.48 mmHg, SNUBH: mean 9.22 ± SD 3.10 mmHg, Severance: mean 10.4 ± SD 1.79 mmHg, $P < 0.001$).

The incidence of severe immediate postoperative complications was 5.93% (471/7,946) at SNUH, 1.37% (31/2,270) at SNUBH, and 2.49% (49/1,966) at Severance Hospital. The incidence of each component outcome (reoperation for bleeding, death, cardiac arrest, ECMO or IABP use) is presented in Table 2.

In the internal validation, the model achieved an AUROC of 0.804 (95% CI, 0.707–0.891), an AUPRC of 0.360 (95% CI, 0.194–0.573), an F1-score of 0.222, and an overall accuracy of 0.683. For external validation, performance metrics were as follows: AUROC 0.735 (95% CI, 0.638–0.822), AUPRC 0.076 (95% CI, 0.030–0.188), F1-score 0.036, and accuracy 0.440 for SNUBH and AUROC 0.712 (95% CI, 0.630–0.795), AUPRC 0.204 (95% CI, 0.110–0.341), F1-score 0.104, and accuracy 0.781 for Severance Hospital. The ROC curves and calibration plots are shown in Fig. 2. Confusion matrices and test probability histograms are shown in Supplementary Figure S1.

SHAP analysis was used to identify the 16 most influential features, which were subsequently incorporated into the final model (Fig. 3). CPB time was the strongest predictor and was positively associated with severe immediate postoperative complications. Preoperative ejection fraction and mean intraoperative systolic blood pressure (SBP) were negatively correlated with the outcome. Higher intraoperative lactate levels, greater estimated blood loss, older age, and larger RBC transfusion volumes were associated with an increased risk of severe immediate postoperative complications.

We conducted DCA to assess the clinical utility of the model in the internal validation cohort (SNUH) and two external validation cohorts (SNUBH and Severance), as shown in Supplementary Figure S2. In the internal validation cohort, the XGBoost model provided positive net benefit over both the 'treat-all' and 'treat-none' strategies across a wide range of threshold probabilities. In the external validation cohorts, the model demonstrated positive net benefit over the 'treat-none' strategy; however, the range of clinically beneficial threshold probabilities was substantially narrower, and the magnitude of net benefit was considerably lower than in internal validation, reflecting the markedly lower event rates in these cohorts.

The LASSO model selected 26 of the 37 candidate variables. Compared to the LASSO model, the XGBoost model demonstrated superior discrimination in the internal validation cohort (AUROC 0.804 vs. 0.748, $p = 0.016$). However, in external validation, the performance advantage of XGBoost over LASSO was not statistically significant (SNUBH: AUROC 0.735 vs. 0.771, $p = 0.170$; Severance: AUROC 0.712 vs. 0.730, $p = 0.473$). The equal-weighted ensemble model achieved an AUROC of 0.782 in internal validation, 0.752 in SNUBH, and 0.730 in Severance Hospital. The ensemble model demonstrated statistically significant improvement over LASSO in internal validation ($p = 0.025$), with no statistically significant difference in AUROC compared to either XGBoost or LASSO in external validation cohorts. Overall comparisons of ROC curves and AUROC values across the three models are presented in Supplementary Figure S3 and Supplementary Table S1.

Table 1 Baseline characteristics and operative profiles

	SNUH (*n* = 7,946)	**SNUBH (*n* = 2,270)**	**Severance (*n* = 1,966)**	***P* value**
Demographic data				
Age, years	61.8 ± 13.4	64.5 ± 12.7	64.5 ± 12.0	< 0.001
Male sex	4752 (59.8)	1430 (62.8)	1203 (61.2)	0.020
BMI, kg/m^2	23.4 ± 3.58	24.4 ± 3.47	24.2 ± 3.53	< 0.001
Surgery type				< 0.001
Isolated CABG	2666 (33.6)	1008 (44.4)	418 (21.3)	
Isolated non-CABG	1941 (24.4)	1082 (47.7)	687 (34.9)	
Two procedures with CABG	338 (4.3)	4 (0.2)	72 (3.7)	
Two procedures without CABG	1484 (18.7)	4 (0.2)	417 (21.2)	
≥ 3 procedures	752 (9.5)	1 (0.0)	299 (15.2)	
Others	767 (9.7)	171 (7.5)	71 (3.6)	
Medical history				
Preop ECMO/IABP	319 (4.0)	57 (2.5)	8 (0.4)	< 0.001
Preop CRRT	323 (4.1)	31 (1.4)	17 (0.9)	< 0.001
Baseline laboratory findings				
CRP, mg/L	0.18 (0.06–0.61)	0.60 (0.15–2.24)	0.20 (0.09–0.55)	< 0.001
Lactate, mmol/L	1.1 (0.8–1.7)	1.3 (0.9–2.1)	1.1 (0.8–1.4)	< 0.001
BUN, mg/dL	19.1 ± 10.4	20.2 ± 11.8	20.5 ± 10.8	< 0.001
Serum Cr, mg/dL	0.97 (0.80–1.17)	0.88 (0.72–1.08)	0.87 (0.73–1.05)	< 0.001
Serum Glu, mg/dL	130 ± 55.0	121 ± 36.2	120 ± 44.4	< 0.001
Hemoglobin, g/dL	12.7 ± 1.86	12.6 ± 2.16	12.4 ± 2.09	< 0.001
WBC, x10^3/μL	6.87 ± 2.86	7.25 ± 3.40	6.64 ± 2.91	< 0.001
Platelet, x10^3/μL	209 ± 71.3	214 ± 73.6	207 ± 73.1	0.016
Potassium, mEq/L	4.15 ± 0.436	4.16 ± 0.462	4.24 ± 0.433	< 0.001
Chloride, mEq/L	104 ± 4.04	104 ± 4.63	104 ± 3.87	0.075
Calcium, mg/dL	8.95 ± 0.533	8.86 ± 0.569	8.89 ± 0.533	< 0.001
Preop LVEF, %	55.3 ± 12.3	55.8 ± 12.9	60.0 ± 13.6	< 0.001
Surgery and anesthesia profiles				
Surgery on CPB	5191 (65.3)	1525 (67.1)	1526 (77.6)	< 0.001
†CPB time, min	187 ± 76.9	126 ± 49.3	105 ± 48.7	< 0.001
Crystalloid, mL	1350 (800–2400)	2695(2039–3670)	1100(900–1500)	< 0.001
EBL, mL	900 (500–1500)	700 (500–1000)	550 (380–750)	< 0.001
RBC, units	2.0 (1.0–3.0)	4.5 (2.5–7.5)	0.0 (0.0–0.0)	< 0.001
Intraoperative ECMO/IABP	848 (10.7)	93 (4.1)	13 (0.7)	< 0.001
Intraoperative vital signs (mean)				
SBP, mmHg	98.8 ± 13.4	97.9 ± 9.24	101 ± 10.6	< 0.001
DBP, mmHg	55.7 ± 6.94	57.5 ± 7.43	60.5 ± 5.72	< 0.001
Heart rate, /min	73.8 ± 11.4	76.0 ± 12.4	76.0 ± 9.28	< 0.001
SpO_2, %	99.6 (98.7–99.9)	99.6 (99.1–99.9)	99.7 (99.2–99.9)	< 0.001
Body temperature, ℃	34.1 (32.9–35.5)	34.9 (34.2–35.4)	35.5 (35.1–35.8)	< 0.001
CVP, mmHg	7.75 ± 2.48	9.22 ± 3.10	10.4 ± 1.79	< 0.001
Intraoperative laboratory findings (mean)				
pO_2, mmHg	251 ± 70.3	223 ± 71.8	224 ± 33.3	< 0.001
pCO_2, mmHg	40.7 ± 4.24	35.6 ± 3.74	35.8 ± 3.29	< 0.001
pH	7.40 ± 0.0437	7.39 ± 0.0411	7.40 ± 0.0380	< 0.001
Bicarbonate, mEq/L	24.9 ± 2.16	21.8 ± 1.80	22.3 ± 1.65	< 0.001
BEecf	0.00 ± 2.54	-2.07 ± 1.95	-2.65 ± 1.98	< 0.001
Lactate, mmol/L	1.80 (0.97–2.28)	1.92 (1-2.3)	1.51 (1-1.77)	< 0.001
Sodium, mEq/L	135 ± 5.13	138 ± 2.64	134 ± 3.09	< 0.001

Table 1 (continued)

	SNUH (*n* = 7,946)	SNUBH (*n* = 2,270)	Severance (*n* = 1,966)	*P* value
Potassium, mEq/L	4.15 ± 0.476	4.01 ± 0.414	4.18 ± 0.524	< 0.001
Glucose, mg/dL	130 ± 32.4	142 ± 41.0	143 ± 37.8	< 0.001

Data are expressed as mean ± standard deviation, median (interquartile range), or number (proportion). P-values are derived from ANOVA test or Kruskal-Wallis test for continuous variables, and Chi-square test for discrete variables

SNUH Seoul National University Hospital, *SNUBH* Seoul National University Bundang Hospital, *BMI* Body mass index, *CABG* Coronary artery bypass grafting, *Preop* Preoperative, *ECMO* Extracorporeal membrane oxygenation, *IABP* Intra-aortic balloon pump, *CRRT* Continuous renal replacement therapy, *CRP* C-reactive protein, *BUN* Blood urea nitrogen, *Cr* Creatinine, *Glu* Glucose, *WBC* White blood cell, *LVEF* Left ventricle ejection fraction, *CPB* Cardiopulmonary bypass, *EBL* Estimated blood loss, *RBC* Red blood cell, *SBP* Systolic blood pressure, *DBP* Diastolic blood pressure, *MBP* Mean blood pressure, SpO_2 Oxygen saturation, *CVP* Central venous pressure, pO_2 Partial pressure of arterial oxygen, pCO_2 Partial pressure of arterial carbon dioxide, *BEecf* Base excess in extracellular fluid

†CPB time was summarized exclusively for surgeries in which cardiopulmonary bypass was performed

Table 2 Incidences of composite outcomes

	SNUH (*n* = 7,946)	SNUBH (*n* = 2,270)	Severance (*n* = 1,966)
Re-operation	271 (3.41%)	8 (0.35%)	40 (2.03%)
Death	24 (0.30%)	5 (0.22%)	2 (0.10%)
Cardiac arrest	131 (1.65%)	13 (0.57%)	2 (0.10%)
ECMO or IABP	202 (2.54%)	18 (0.79%)	12 (0.61%)
Composite outcome	471 (5.93%)	31 (1.37%)	49 (2.49%)

Data are expressed as number (proportion). *SNUH* Seoul National University Hospital, *SNUBH* Seoul National University Bundang Hospital, *ECMO* Extracorporeal membrane oxygenation, *IABP* Intra-aortic balloon pump

Discussion

In this study, an XGBoost-based model achieved an AUROC of 0.804 in the internal validation for predicting severe immediate postoperative complications in adults who underwent cardiac surgery, demonstrating a good discriminative performance.

Existing predictive models have primarily focused on general ICU populations or late deterioration (> 24 h) after cardiac surgery [30–32]. To the best of our knowledge, this is the first study to successfully predict severe immediate (≤ 24 h) clinically significant complications in cardiac surgery – a domain that has been underrepresented in prior literature [12–16]. This study achieved comparable performance in a large, multi-institutional cohort.

Instead of relying on blood pressure thresholds or vasoactive medication doses, we defined primary outcome using a composite of objective and clinically significant adverse events occurring within 24 h after cardiac surgery, requiring substantial medical intervention – namely, reoperation for bleeding control, death, cardiac arrest, and initiation of mechanical circulatory support with ECMO or IABP. Although these events represent severe terminal outcomes rather than early warning signs, predicting the risk of these critical events within the first 24 h enables clinicians to proactively identify high-risk states and optimize resource allocation.

In this study, perioperative factors – including demographic characteristics, preoperative laboratory findings, intraoperative parameters, and hemodynamic variables – were analyzed to identify predictors of immediate postoperative complications. Among the 37 variables examined, several features emerged as significant predictors in the SHAP dependency analysis. Longer CPB time, lower preoperative ejection fraction, elevated intraoperative lactate levels, greater estimated blood loss with higher RBC transfusion volume, and advanced age are well-established risk factors for postoperative major complications and mortality following cardiac surgery [33–37]. Intraoperative hypotension, defined as a mean arterial pressure (MAP) < 65mmHg, has been independently associated with adverse outcomes such as stroke, acute kidney injury, and mortality after cardiac surgery [38]. Most prior studies have defined hypotension in cardiac surgery with mean arterial pressure during CPB period [38–41]. However, the relationship between SBP and postoperative outcomes remains controversial. While low intraoperative SBP has been associated with postoperative acute kidney injury after coronary artery bypass grafting [42], elevated SBP (> 140mmHg) or marked SBP variability has been associated with higher 30-day mortality in cardiac surgical patients [43, 44]. In the present study, the mean intraoperative SBP was negatively correlated with severe immediate postoperative complications, suggesting that lower intraoperative SBP may contribute to immediate postoperative deterioration.

Notably, the top predictors identified by our SHAP analysis are established risk factors in cardiac surgery. The fact that our model independently identified these clinically validated variables without prior assumptions serves as evidence of its clinical validity. To assess whether a simplified approach using these established predictors could achieve comparable performance, we constructed a post-hoc logistic regression model incorporating only the top three SHAP-ranked predictors (CPB time, preoperative ejection fraction, and mean value of intraoperative lactate). The simplified model achieved AUROCs of 0.738, 0.696, and 0.714 in internal validation, SNUBH, and Severance Hospital, respectively. The XGBoost model demonstrated statistically superior discrimination over the simplified model in internal validation ($p = 0.034$), while no significant difference

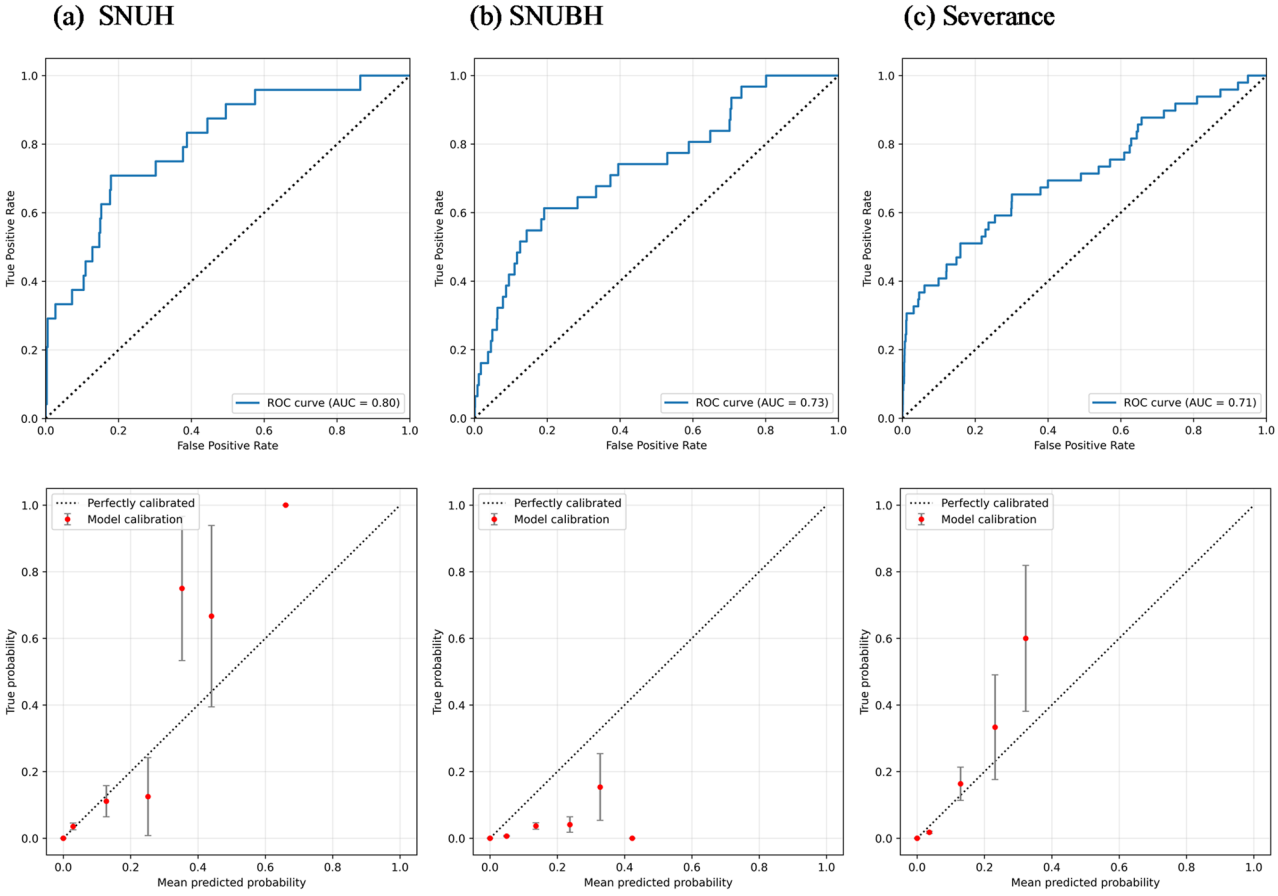

Fig. 2 ROC curves and calibration plots of the XGBoost model across three validation cohorts. ROC curves (top row) and calibration plots (bottom row) for (**a**) Seoul National University Hospital (SNUH), (**b**) Seoul National University Bundang Hospital (SNUBH), and (**c**) Severance Hospital. XGBoost, Extreme gradient boosting; AUC, Area under the receiver operating characteristic curve

was observed in external validation cohorts ($p = 0.200$ and $p = 0.924$), consistent with the pattern observed in the XGBoost vs. LASSO comparison (Supplementary Table S2 and Supplementary Figure S4). However, as this simplified model was developed as a post-hoc exploration and was not constructed within the same cross-validation framework as the primary models, it should be interpreted as a sensitivity analysis rather than a formal comparator. Importantly, the primary contribution of this study lies not in the identification of novel individual risk factors, but in the precise prediction of a highly specific composite endpoint — severe immediate postoperative complications within 24 h of cardiac surgery — which has been substantially underrepresented in prior predictive modeling literature.

The XGBoost model demonstrated superior discrimination compared with the LASSO model in the internal validation cohort. However, there was no statistical difference between XGBoost model and the LASSO model in external validation cohorts, rather the point estimates of AUROC numerically favored LASSO model. Prior meta-analyses comparing ML with logistic regression have reported conflicting results in the context of postoperative mortality prediction in cardiac surgery [45, 46]. Attenuated discrimination of XGBoost model in external settings is consistent with prior reports showing that ML models provide only marginal improvements over logistic regression for predicting cardiac surgery outcomes across heterogeneous institutions [46–48].

We additionally constructed the equal weighted ensemble model as a supplementary analysis. The ensemble model achieved intermediate discrimination in internal validation and demonstrated performance comparable to XGBoost in external validation. No statistically significant differences were observed among the three models in external validation, suggesting that the performance gap observed internally does not translate to a meaningful clinical distinction across institutions.

We acknowledge that AUPRC and F1-scores declined substantially in external validation, reflecting reduced positive predictive ability in low-prevalence settings — a known mathematical consequence of applying models developed in high-acuity centers to populations with markedly lower event rates, rather than a failure of

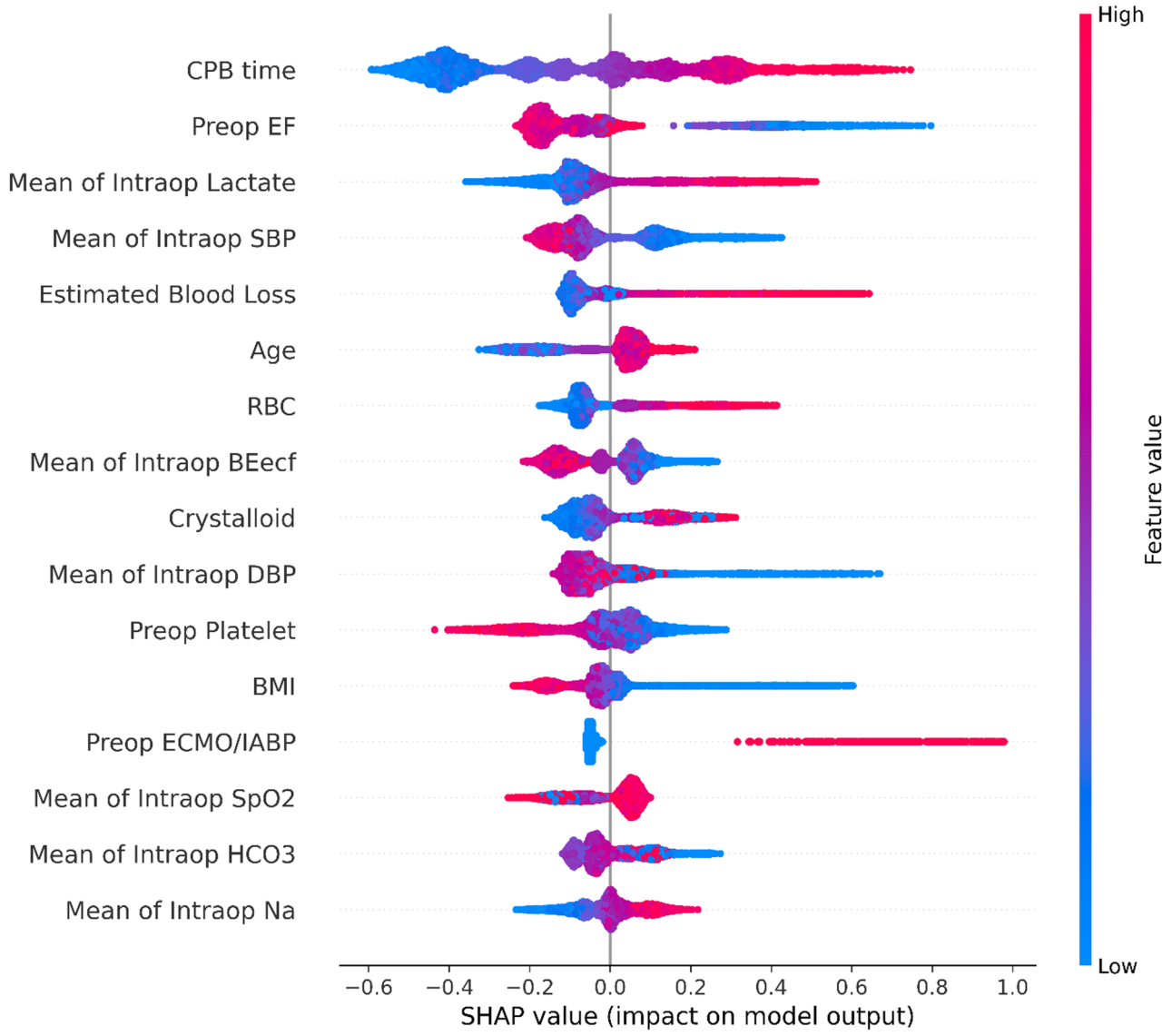

Fig. 3 Shapley additive explanations dependency plot demonstrating top 16 features of the model. SHAP, Shapley additive explanations; CPB, Cardiopulmonary bypass; preop, Preoperative; EF, Ejection fraction; intraop, Intraoperative; SBP, Systolic blood pressure; RBC, Red blood cell; BEecf, Base excess in the extracellular fluid; DBP, Diastolic blood pressure; BMI, Body mass index; ECMO, Extracorporeal membrane oxygenation; IABP, Intra-aortic balloon pump; SpO_2, Pulse oxygen saturation; HCO_3, Bicarbonate; Na, Sodium

discrimination per se. Accordingly, the model's clinical utility in external settings remains constrained, and reliable deployment would require local recalibration and prospective validation prior to implementation.

This study highlights the potential of ML approaches for predicting critical and immediate postoperative complications after cardiac surgery. With SHAP-based interpretability, individualized risk factors can be identified even within complex nonlinear relationships, thereby facilitating personalized management strategies. The modifiable predictors revealed by the model may guide patient-specific perioperative interventions, whereas non-modifiable factors can enhance preoperative risk stratification. Prediction models may help support clinical decision-making, such as determining the appropriateness of fast-track recovery protocols or the early initiation of mechanical support. A critical next step is translating these models from development to the clinical bedside. Recently, *Pan et al.* demonstrated the real-world value of a multicenter-validated ML model for diagnosing early infection after cardiovascular surgery through prospective clinical implementation, showing significant benefits in antibiotic management, ICU length of stay, and cost control [49]. Ultimately, prospective studies are needed to determine whether prediction-based decision support utilizing our model can meaningfully improve clinical outcomes.

Limitations

Our intentional 95:5 dataset split resulted in a small internal test set with wide 95% confidence intervals for internal validation metrics; this limitation was mitigated by evaluating the model on two large, independent external validation cohorts. Although external validation confirmed the model's baseline discriminative capacity, predictive accuracy varied substantially across institutions, reflecting differences in surgical practice, data acquisition protocols, and clinical management. The internal cohort spanned 17 years (2004 to 2021), whereas external cohorts comprised predominantly recent data, introducing temporal drift that could not be fully accounted for in the analysis.

Furthermore, institutional drift was evident in the markedly lower event rates of external cohorts (1.37% and 2.49%) compared with the internal cohort (5.93%). This disparity directly affected precision-oriented metrics: AUPRC declined substantially in external validation (0.076 and 0.204) compared with internal validation (0.360), and F1-scores approached zero (0.036 and 0.104), reflecting a near-complete loss of positive predictive ability in low-prevalence settings. Because such an incidence mismatch can distort clinical utility assessments, we recalibrated predicted probabilities to local event rates prior to DCA. Although net benefit was positive across all cohorts, it was substantially lower in external cohorts than in the internal validation cohort. These findings underscore that continuous updating and rigorous local recalibration are prerequisites for reliable clinical implementation.

In addition, hemodynamic parameters such as blood pressure, cardiac index, and the administration of inotropic or vasopressor agents were not incorporated into the definition of postoperative outcomes in this study. Given the current lack of consensus on the definition of postoperative hypotension in cardiac surgery, future research is warranted to establish standardized criteria for hemodynamic instability in this population.

This study focused solely on immediate postoperative complications that occurred within 24 h after surgery. Long-term outcomes, such as extended ICU stay, organ dysfunction, and late mortality, were not assessed. Further investigations integrating both short- and long-term outcomes are necessary to achieve a more comprehensive evaluation of ML-based prediction of postoperative severe complications after cardiac surgery.

Furthermore, regarding our feature selection methodology, the SHAP-based forward selection was performed on the entire development dataset, not on the inner cross-validation loop. We acknowledge that performing feature selection outside of resampling can introduce optimism bias, potentially leading to an overestimation of the model's performance in the internal evaluation. However, to mitigate this risk and objectively assess the true generalizability of our model, the final selected feature set and model were subsequently evaluated on two strictly independent external validation cohorts. Because these external datasets were completely isolated from the feature selection process, they provide a rigorous and unbiased estimate of the model's real-world predictive performance.

Finally, several potentially relevant variables were not included in our model, such as continuous intraoperative mean arterial blood pressure, cardiac index, mixed venous oxygen saturation, complexity of surgery, and detailed preoperative medical history. Incorporating these physiological and clinical variables could further enhance the model performance. Future studies integrating richer physiological data and more comprehensive comorbidity profiles may yield more accurate and clinically valuable prediction models.

Conclusion

In conclusion, we developed an XGBoost model and an equal-weighted ensemble of XGBoost and LASSO as a supplementary model for predicting severe immediate postoperative complications in patients undergoing cardiac surgery. While the XGBoost model achieved acceptable discrimination internally, external validation revealed a performance attenuation across institutions, largely attributable to variations in baseline event rates. Nevertheless, this study demonstrates that severe immediate postoperative complications after cardiac surgery are predictable using routinely available perioperative variables and provides a multi-institutional benchmark for future studies targeting this clinically critical endpoint. Further prospective validation and institution-specific recalibration are essential for translating these findings into practical decision-support tools that enhance individualized perioperative care.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12871-026-03841-9.

Supplementary Material 1.

Acknowledgements

This study used clinical data retrieved from the Seoul National University Hospital Patients Research Environment (SUPREME) system.

Authors' contributions

H.C. conceived and designed the study, and drafted the manuscript. JH. A. performed data analysis, developed the prediction model and prepared the figures and tables. JW. P. and SH. K. contributed to data collection and curation. H. L. contributed additional data analyses and performed statistical validation. JK. S., YJ. C., K. N., JW. J., J. C., and IJ. K. contributed to data acquisition and critically revised the manuscript. HC. L. and Y. J. supervised the study design and provided critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding
This research was supported by a grant of the Boston-Korea Innovative Research Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant no. : RS-2024-00403047, NTIS No. : 2460003034). The funding sources and sponsors had no role in the current study design; in the collection, analysis, and interpretation of data; in the writing of this paper; and in the decision to submit the paper for publication.

Data availability
The datasets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the institutional review boards of all participating institutions (SNUH, IRB no. 2210-145-1373, approved November 8, 2022; SNUBH, IRB no. B-2309-852-405, approved August 25, 2023; and Severance Hospital, IRB no. 4-2023-0881, approved August 31, 2023). All research procedures were conducted in accordance with the Declaration of Helsinki and relevant institutional guidelines and regulations. The requirement for informed consent was waived by each institutional review board due to the retrospective nature of the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Anesthesiology and Pain Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[2]Department of Anesthesiology and Pain Medicine, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea
[3]Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[4]Healthcare AI Research Institute, Seoul National University Hospital, Seoul, Republic of Korea
[5]Department of Anesthesiology and Pain Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[6]Department of Anesthesiology and Pain Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
[7]Anesthesia and Pain Research Institute, Yonsei University College of Medicine, Seoul, Republic of Korea
[8]Department of Transdisciplinary Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[9]Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

Received: 9 February 2026 / Accepted: 13 April 2026
Published online: 17 April 2026

References

1. Lomivorotov VV, Efremov SM, Kirov MY, Fominskiy EV, Karaskov AM. Low-Cardiac-Output Syndrome After Cardiac Surgery. J Cardiothorac Vasc Anesth. 2017;31(1):291–308.
2. Vincent JL, De Backer D. Circulatory shock. N Engl J Med. 2013;369(18):1726–34.
3. Coaston TN, Vadlakonda A, Mallick S, Aguayo E, Hallare J, Sanaiha Y, et al. Decade-long variation in the development of complications after cardiac surgery across the United States. Surgery. 2025;183:109387.
4. Anthi A, Tzelepis GE, Alivizatos P, Michalis A, Palatianos GM, Geroulanos S. Unexpected cardiac arrest after cardiac surgery: incidence, predisposing causes, and outcome of open chest cardiopulmonary resuscitation. Chest. 1998;113(1):15–9.
5. Kristensen KL, Rauer LJ, Mortensen PE, Kjeldsen BJ. Reoperation for bleeding in cardiac surgery. Interact Cardiovasc Thorac Surg. 2012;14(6):709–13.
6. Mazzeffi M, Zivot J, Buchman T, Halkos M. In-hospital mortality after cardiac surgery: patient characteristics, timing, and association with postoperative length of intensive care unit and hospital stay. Ann Thorac Surg. 2014;97(4):1220–5.
7. Khorsandi M, Shaikhrezai K, Prasad S, Pessotto R, Walker W, Berg G, et al. Advanced mechanical circulatory support for post-cardiotomy cardiogenic shock: a 20-year outcome analysis in a non-transplant unit. J Cardiothorac Surg. 2016;11:29.
8. St Andre AC, DelRossi A. Hemodynamic management of patients in the first 24 hours after cardiac surgery. Crit Care Med. 2005;33(9):2082–93.
9. Kim KM, Arghami A, Habib R, Daneshmand MA, Parsons N, Elhalabi Z, et al. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2022 Update on Outcomes and Research. Ann Thorac Surg. 2023;115(3):566–74.
10. Roques F, Nashef SA, Michel P, Gauducheau E, de Vincentiis C, Baudet E, et al. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. Eur J Cardiothorac Surg. 1999;15(6):816–22. discussion 22 – 3.
11. Shahian DM, O'Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1–coronary artery bypass grafting surgery. Ann Thorac Surg. 2009;88(1 Suppl):S2–22.
12. Kilic A, Goyal A, Miller JK, Gleason TG, Dubrawksi A. Performance of a Machine Learning Algorithm in Predicting Outcomes of Aortic Valve Replacement. Ann Thorac Surg. 2021;111(2):503–10.
13. Kilic A, Habib RH, Miller JK, Shahian DM, Dearani JA, Dubrawski AW. Supplementing Existing Societal Risk Models for Surgical Aortic Valve Replacement With Machine Learning for Improved Prediction. J Am Heart Assoc. 2021;10(22):e019697.
14. Mori M, Durant TJS, Huang C, Mortazavi BJ, Coppi A, Jean RA, et al. Toward Dynamic Risk Prediction of Outcomes After Coronary Artery Bypass Graft: Improving Risk Prediction With Intraoperative Events Using Gradient Boosting. Circ Cardiovasc Qual Outcomes. 2021;14(6):e007363.
15. Allyn J, Allou N, Augustin P, Philip I, Martinet O, Belghiti M, et al. A Comparison of a Machine Learning Model with EuroSCORE II in Predicting Mortality after Elective Cardiac Surgery: A Decision Curve Analysis. PLoS ONE. 2017;12(1):e0169772.
16. Sinha S, Dong T, Dimagli A, Vohra HA, Holmes C, Benedetto U, et al. Comparison of machine learning techniques in prediction of mortality following cardiac surgery: analysis of over 220 000 patients from a large national database. Eur J Cardiothorac Surg. 2023;63(6):ezad183. https://doi.org/10.1093/ejcts/ezad183.
17. TRIPOD + AI statement. updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:q902.
18. Chen T, Guestrin C, XGBoost:. A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA: Association for Computing Machinery; 2016. pp. 785–94.
19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al. Scikit-learn: Machine Learning in Python. 2011;12(null %J J. Mach. Learn. Res.):2825–30.
20. Akiba T, Sano S, Yanase T, Ohta T, Koyama M, Optuna. A Next-generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; Anchorage, AK, USA: Association for Computing Machinery; 2019. pp. 2623–31.
21. Ozenne B, Subtil F, Maucort-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol. 2015;68(8):855–9.
22. Sokolova M, Japkowicz N, Szpakowicz S, editors. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. 2006. Vol. 43042006. p. 1015–21.
23. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–5.
24. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167–76.
25. BRIER GW, VERIFICATION OF FORECASTS EXPRESSED, IN TERMS OF PROBABILITY %J. Mon Weather Rev. 1950;78(1):1–3.
26. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. Stat Med. 2019;38(21):4051–65.

27. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell. 2020;2(1):56–67.
28. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.
29. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. 1996;58(1):267–88.
30. Rahman A, Chang Y, Dong J, Conroy B, Natarajan A, Kinoshita T, et al. Early prediction of hemodynamic interventions in the intensive care unit using machine learning. Crit Care. 2021;25(1):388.
31. Dung-Hung C, Cong T, Zeyu J, Yu-Shan OY, Yung-Yan L. External validation of a machine learning model to predict hemodynamic instability in intensive care unit. Crit Care. 2022;26(1):215.
32. Mathis MR, Engoren MC, Williams AM, Biesterveld BE, Croteau AJ, Cai L, et al. Prediction of Postoperative Deterioration in Cardiac Surgery Patients Using Electronic Health Record and Physiologic Waveform Data. Anesthesiology. 2022;137(5):586–601.
33. Salis S, Mazzanti VV, Merli G, Salvi L, Tedesco CC, Veglia F, et al. Cardiopulmonary bypass duration is an independent predictor of morbidity and mortality after cardiac surgery. J Cardiothorac Vasc Anesth. 2008;22(6):814–22.
34. Pieri M, Belletti A, Monaco F, Pisano A, Musu M, Dalessandro V, et al. Outcome of cardiac surgery in patients with low preoperative ejection fraction. BMC Anesthesiol. 2016;16(1):97.
35. Demers P, Elkouri S, Martineau R, Couturier A, Cartier R. Outcome with high blood lactate levels during cardiopulmonary bypass in adult cardiac operation. Ann Thorac Surg. 2000;70(6):2082–6.
36. Dayan V, Cal R, Giangrossi F. Risk factors for vasoplegia after cardiac surgery: a meta-analysis. Interact Cardiovasc Thorac Surg. 2019;28(6):838–44.
37. Peigh PS, Swartz MT, Vaca KJ, Lohmann DP, Naunheim KS. Effect of advancing age on cost and outcome of coronary artery bypass grafting. Ann Thorac Surg. 1994;58(5):1362–6. discussion 6–7.
38. de la Hoz MA, Rangasamy V, Bastos AB, Xu X, Novack V, Saugel B, et al. Intraoperative Hypotension and Acute Kidney Injury, Stroke, and Mortality during and outside Cardiopulmonary Bypass: A Retrospective Observational Cohort Study. Anesthesiology. 2022;136(6):927–39.
39. Sun LY, Chung AM, Farkouh ME, van Diepen S, Weinberger J, Bourke M, et al. Defining an Intraoperative Hypotension Threshold in Association with Stroke in Cardiac Surgery. Anesthesiology. 2018;129(3):440–7.
40. Ngu JMC, Jabagi H, Chung AM, Boodhwani M, Ruel M, Bourke M, et al. Defining an Intraoperative Hypotension Threshold in Association with De Novo Renal Replacement Therapy after Cardiac Surgery. Anesthesiology. 2020;132(6):1447–57.
41. Ristovic V, de Roock S, Mesana TG, van Diepen S, Sun LY. The Impact of Preoperative Risk on the Association between Hypotension and Mortality after Cardiac Surgery: An Observational Study. J Clin Med. 2020;9(7).
42. Aronson S, Phillips-Bute B, Stafford-Smith M, Fontes M, Gaca J, Mathew JP, et al. The association of postcardiac surgery acute kidney injury with intraoperative systolic blood pressure hypotension. Anesthesiol Res Pract. 2013;2013:174091.
43. Aronson S, Stafford-Smith M, Phillips-Bute B, Shaw A, Gaca J, Newman M, et al. Intraoperative systolic blood pressure variability predicts 30-day mortality in aortocoronary bypass surgery patients. Anesthesiology. 2010;113(2):305–12.
44. Sessler DI, Bloomstone JA, Aronson S, Berry C, Gan TJ, Kellum JA, et al. Perioperative Quality Initiative consensus statement on intraoperative blood pressure, risk and outcomes for elective surgery. Br J Anaesth. 2019;122(5):563–74.
45. Benedetto U, Dimagli A, Sinha S, Cocomello L, Gibbison B, Caputo M, et al. Machine learning improves mortality risk prediction after cardiac surgery: Systematic review and meta-analysis. J Thorac Cardiovasc Surg. 2022;163(6):2075–e879.
46. Penny-Dimri JC, Bergmeir C, Perry L, Hayes L, Bellomo R, Smith JA. Machine learning to predict adverse outcomes after cardiac surgery: A systematic review and meta-analysis. J Card Surg. 2022;37(11):3838–45.
47. Li Q, Lv H, Chen Y, Shen J, Shi J, Zhou C. Development and validation of a machine learning predictive model for perioperative myocardial injury in cardiac surgery with cardiopulmonary bypass. J Cardiothorac Surg. 2024;19(1):384.
48. Zeng J, Zhang D, Lin S, Su X, Wang P, Zhao Y, et al. Comparative analysis of machine learning vs. traditional modeling approaches for predicting in-hospital mortality after cardiac surgery: temporal and spatial external validation based on a nationwide cardiac surgery registry. Eur Heart J Qual Care Clin Outcomes. 2024;10(2):121–31.
49. Pan T, Zhang H, Wang C, Wang H, Matniyaz Y, Lv ZK, et al. Development, validation, and clinical evaluation of a machine-learning based model for diagnosing early infection after cardiovascular surgery (DEICS): a multi-center cohort study. Int J Surg. 2025;111(4):2862–71.

Publisher's note