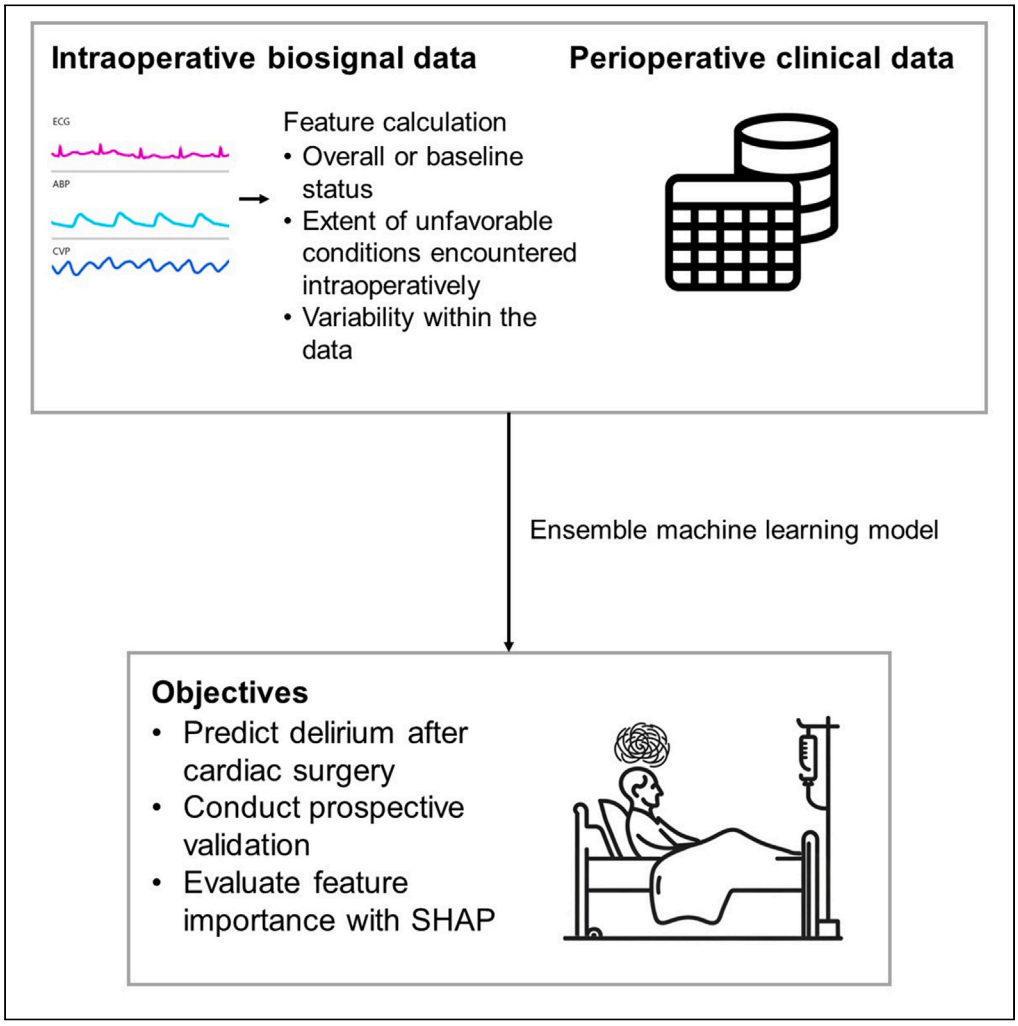
iScience

CellPress
OPEN ACCESS

Article

Machine learning with clinical and intraoperative biosignal data for predicting postoperative delirium after cardiac surgery



Changho Han, Hyun Il Kim, Sarah Soh, Ja Woo Choi, Jong Wook Song, Dukyong Yoon

sjw72331@yuhs.ac (J.W.S.)
dukyong.yoon@yonsei.ac.kr (D.Y.)

Highlights

Extracted features from biosignals to model hemodynamic fluctuations during surgery

Intraoperative biosignal features had high feature importance in the ML models

Highlighted the importance of intraoperative patient management

Highlighted the importance of high-resolution biosignal data in predictive modeling

Han et al., iScience 27, 109932
June 21, 2024
https://doi.org/10.1016/j.isci.2024.109932

iScience

CellPress
OPEN ACCESS

Article

Machine learning with clinical and intraoperative biosignal data for predicting postoperative delirium after cardiac surgery

Changho Han,[1,5] Hyun Il Kim,[2,5] Sarah Soh,[2] Ja Woo Choi,[2] Jong Wook Song,[2,*] and Dukyong Yoon[1,3,4,6,*]

SUMMARY

Early identification of patients at high risk of delirium is crucial for its prevention. Our study aimed to develop machine learning models to predict delirium after cardiac surgery using intraoperative biosignals and clinical data. We introduced a novel approach to extract relevant features from continuously measured intraoperative biosignals. These features reflect the patient's overall or baseline status, the extent of unfavorable conditions encountered intraoperatively, and beat-to-beat variability within the data. We developed a soft voting ensemble machine learning model using retrospective data from 1,912 patients. The model was then prospectively validated with data from 202 additional patients, achieving a high performance with an area under the receiver operating characteristic curve of 0.887 and an accuracy of 0.881. According to the SHapley Additive exPlanation method, several intraoperative biosignal features had high feature importance, suggesting that intraoperative patient management plays a crucial role in preventing delirium after cardiac surgery.

INTRODUCTION

Delirium is an acute neurocognitive disorder characterized by fluctuating disturbances in attention, awareness, or cognition.[1–3] Postoperative delirium is a highly prevalent and serious complication of cardiac surgery.[4] Delirium is associated with significant functional decline, higher postoperative morbidity and mortality risks, prolonged hospital stay, and healthcare costs.[5,6]

However, owing to the lack of specific treatment, current management strategies for delirium only focus on its prevention and early detection.[7–9] As such, risk stratification and identification of vulnerable patients are crucial. The pathogenesis of postoperative delirium is multifactorial, with risk factors including advanced age and preexisting cognitive impairment.[10] Furthermore, intraoperative variables, including cerebral perfusion and depth of anesthesia, are associated with postoperative delirium.[11–13] However, the impact of perioperative factors remains unclear, and a comprehensive predictive model encompassing both preoperative clinical data and intraoperative variables has not been developed.

Several recent studies have attempted to predict postoperative complications using AI.[14–17] The recent emergence of software such as VitalRecorder has made it possible to store and analyze whole high-resolution biosignal data.[18] AI enables the extraction of hidden information and nonlinear relationships and provides efficient analysis of complicated, large-sized data.[19] Using extensive perioperative clinical information and intraoperative biosignal data, this study aimed to develop machine learning models for predicting postoperative delirium after cardiac surgery and prospectively validate the performance of these models.

RESULTS

Dataset characteristics

Patients aged ≥19 years who underwent cardiac surgery were included. Data used for developing machine learning models were collected from December 14, 2018 to December 22, 2021, whereas those for prospective validation were collected from March 28, 2022 to June 28, 2022 (Figure 1). Extensive perioperative clinical data and intraoperative biosignal data were utilized for machine learning model development and validation. Specific clinical variables, which were considered to be potential predictors of postoperative delirium and are used as machine learning input, are listed in Table S1. For the continuously monitored intraoperative biosignals, we calculated features from these parameters for use as inputs for the machine learning models (Table S2). These features were designed and selected to reflect the patients' overall or

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Yongin, Republic of Korea
[2]Department of Anesthesiology and Pain Medicine, Anesthesia and Pain Research Institute Yonsei University College of Medicine, Seoul, Republic of Korea
[3]Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, Republic of Korea
[4]Institute for Innovation in Digital Healthcare (IIDH), Severance Hospital, Seoul, Republic of Korea
[5]These authors contributed equally
[6]Lead contact
*Correspondence: sjw72331@yuhs.ac (J.W.S.), dukyong.yoon@yonsei.ac.kr (D.Y.)
https://doi.org/10.1016/j.isci.2024.109932

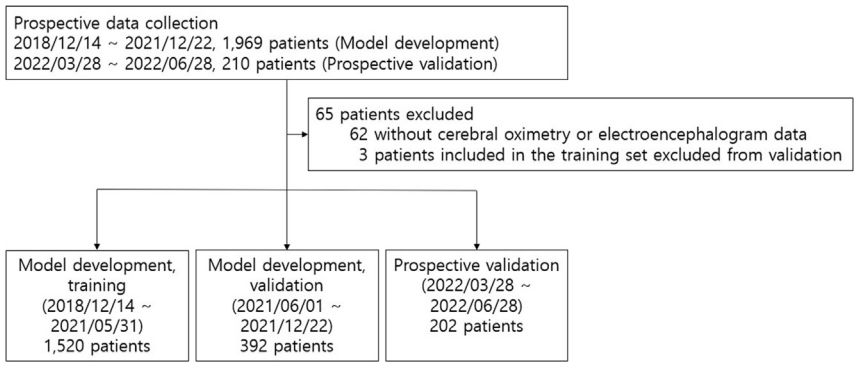


Figure 1. Patient flow diagram

baseline status (e.g., average, baseline, or lowest values), extent of unfavorable conditions encountered during surgery (e.g., duration and area under the curve falling below or above a certain value), and beat-to-beat variability within the data (coefficient of variation [CV] and average real variability [ARV]).[20]

Figure 1 shows the flow diagram of patient selection. In total, 2,179 adult patients were included: among them, 1,969 and 210 patients belonged to the training and validation sets and to the prospective validation cohort, respectively. After excluding 62 patients (55 patients from the training and validation sets and 7 patients from the prospective validation cohort) with missing cerebral oximetry or electroencephalogram (EEG) data and 3 patients who were already included in the training set (2 patients from the validation set and 1 patient from the prospective validation cohort), 2,114 patients were included. Table 1 shows the baseline clinicodemographic characteristics (demographic data, medical history, cerebral oximetry and EEG) of the entire cohort. Overall, 260 patients had postoperative delirium, and these patients were older (median [interquartile range (IQR)]: 72 · 0 [65 · 0, 76 · 0] years vs. 65 · 0 [56 · 0, 72 · 0] years, $p < 0 \cdot 001$) and included a higher proportion of males (68 · 5% vs. 61 · 7%, $p = 0 \cdot 041$) than those who did not have delirium.

Moreover, the delirium group had a higher proportion of surgeries without cardiopulmonary bypass (CPB) (42 · 3% vs. 27 · 5%, $p < 0 \cdot 001$) and with both CPB and total circulatory arrest (TCA) (11 · 5% vs. 9 · 8%, $p < 0 \cdot 001$). This group also had more comorbidities than the non-delirium group, including hypertension (76 · 5% vs. 57 · 3%, $p < 0 \cdot 001$), diabetes mellitus (40 · 0% vs. 27 · 9%, $p < 0 \cdot 001$), chronic kidney disease (30 · 4% vs. 10 · 4%, $p < 0 \cdot 001$), and previous cerebrovascular accidents (20 · 4% vs. 12 · 1%, $p < 0 \cdot 001$). Further, the delirium group had a longer operation time (median [IQR]: 231 · 0 [201 · 0, 267 · 2] min vs. 207 · 0 [164 · 0, 246 · 0] min, $p < 0 \cdot 001$), anesthesia time (median [IQR]: 305 · 0 [270 · 0, 345 · 0] min vs. 275 · 0 [230 · 0, 320 · 0] min, $p < 0 \cdot 001$), CPB duration (mean [standard deviation, SD]: 132 · 7 [64 · 4] min vs. 104 · 6 [53 · 1] min, $p < 0 \cdot 001$), and aortic cross clamp (ACC) duration (mean [SD]: 86 · 5 [42 · 6] min vs. 70 · 3 [37 · 9] min, $p < 0 \cdot 001$).

In addition, the delirium group had lower regional cerebral oxygen saturation (rSO2) (median [IQR]: 55 · 9% [49 · 4%, 60 · 4%] vs. 58 · 7% [54 · 2%, 62 · 9%], $p < 0 \cdot 001$), higher anesthesia depth (bispectral index [BIS] duration <40 or patient state index [PSI] < 25; median [IQR]: 1025 · 0 [333 · 0, 2049 · 2]/5 s vs. 535 · 5 [169 · 0, 1174 · 0]/5 s, $p < 0 \cdot 001$), and higher suppression ratio (SR) (SR duration >1%; median [IQR]: 236 · 0 [38 · 8, 763 · 0]/5 s vs. 102 · 0 [21 · 0, 332 · 0]/5 s, $p < 0 \cdot 001$) than the non-delirium group. Other clinical and biosignal characteristics of the entire cohort are presented Table S3. Notably, the delirium group had lower perfusion pressure (PP) (mean [SD]: 61 · 2 [9 · 6] mmHg vs. 64 · 0 [6 · 9] mmHg, $p < 0 \cdot 001$), higher mean pulmonary arterial pressure (mPAP) (mPAP duration >20 mmHg; mean [SD]: 1307 · 6 [858 · 5]/5 s vs. 1079 · 7 [770 · 8]/5 s, $p < 0 \cdot 001$), and lower cardiac index (CI) (CI < 2; mean [SD]: 783 · 1 [817 · 2]/5 s vs. 581 · 2 [680 · 9]/5 s, $p < 0 \cdot 001$). The characteristics of the patients in the training, validation, and prospective validation cohorts are shown in Tables S4 and S5. In the prospective validation cohort, 29 of the 202 patients had postoperative delirium.

Model performance

Figures 2 and S1 show the receiver operating characteristics (ROC) and precision-recall (PR) curves of the machine learning models in the prospective validation cohort and validation set, respectively. Eight distinct machine learning models were developed, namely XGBoost (XGB), extra trees classifier (ET), light gradient boosting machine (LGBM), random forest (RF), gradient boosting classifier (GBC), logistic regression (LR), artificial neural network (ANN), and support vector machine (SVM). Subsequently, a soft-voting ensemble (ENS) classifier was constructed by averaging the outputs of the highest-performing individual models in the validation set. To ascertain the most effective configuration for the ENS classifier, we varied the number of top-performing individual models included in the ENS, ranging from one to eight, selecting the ensemble that achieved the highest area under the ROC curve (AUROC) in the validation set. As a result, the five top-performing models, in the order of XGB, ET, LGBM, RF, and GBC, were incorporated into the soft-voting ENS classifier. To reduce the complexity of reporting, Figure 2, Table 2, Figure S1, and Table S5 only report the performances of this ENS model and the five top-performing individual models included in the ENS. The ENS model had the highest performance (prospective validation cohort: AUROC, 0 · 887; area under the PR curve [AUPRC], 0 · 499; validation set: AUROC, 0 · 782; AUPRC, 0 · 290). The individual machine learning models had AUROCs of 0 · 851–0·877 and

Table 1. Baseline clinical characteristics (demographic data and medical history), intraoperative durations and intraoperative biosignal characteristics (cerebral oximetry and electroencephalogram) of the entire cohort

	Missing	Delirium negative (*n* = 1854)	Delirium positive (*n* = 260)	*p*-Value
Sex, male, n (%)	0	1144 (61.7%)	178 (68.5%)	0.041
Age, median [Q1-Q3]	0	65 [56–72]	72.0 [65–76]	<0.001
Operation_type, n (%)	0			<0.001
No CPB		510 (27.5%)	110 (42.3%)	<0.001
CPB, no TCA		1163 (62.7%)	120 (46.2%)	
CPB, TCA		181 (9.8%)	30 (11.5%)	
Height, cm, median [Q1-Q3]	0	164.0 [156.8–170.0]	163.6 [157.0–170.0]	0.702
Weight, kg, median [Q1-Q3]	0	64.3 [56.7–72.8]	63.6 [56.1–70.6]	0.149
Body surface area, m^2, median [Q1-Q3]	0	1.71 [1.58–1.84]	1.70 [1.58–1.81]	0.177
Systolic blood pressure, mmHg, mean ± SD	3	125.0 ± 15.7	125.9 ± 16.5	0.391
Diastolic blood pressure, mmHg, mean ± SD	3	73.4 ± 11.1	70.3 ± 11.3	<0.001
Heart rate, bpm, median [Q1-Q3]	3	71.0 [63.5–79.0]	70.5 [63.0–80.0]	0.837
Emergency surgery, n (%)	0	14 (0.8%)	7 (2.7%)	0.01
Hypertension, n (%)	0	1063 (57.3%)	199 (76.5%)	<0.001
Diabetes mellitus, n (%)	0			<0.001
No diabetes		1337 (72.1%)	156 (60.0%)	
Diabetes, on oral medication		448 (24.2%)	85 (32.7%)	
Diabetes, on insulin		69 (3.7%)	19 (7.3%)	
Chronic kidney disease, n (%)	0	192 (10.4%)	79 (30.4%)	<0.001
Old cerebrovascular accident, n (%)	0	225 (12.1%)	53 (20.4%)	<0.001
Atrial fibrillation, n (%)	0	459 (24.8%)	62 (23.8%)	0.808
Liver cirrhosis, n (%)	0	42 (2.3%)	12 (4.6%)	0.041
Congestive heart failure, n (%)	0	263 (14.2%)	61 (23.5%)	<0.001
COPD, n (%)	0	56 (3.0)	22 (8.5)	<0.001
Acute MI (1 week), n (%)	0	55 (3.0%)	12 (4.6%)	0.218
Recent MI (3 months), n (%)	0	62 (3.3%)	16 (6.2%)	0.038
Old MI, n (%)	0	99 (5.3%)	32 (12.3%)	<0.001
History of cognitive impairment, n (%)	0	20 (1.1%)	5 (1.9%)	0.222
History of alcohol abuse, n (%)	0	8 (0.4%)	4 (1.5%)	0.05
Functional capacity (≥ 4 METs)[4], n (%)	4	499 (27.0%)	115 (44.2%)	<0.001
NYHA classification[5], n (%)	8			
Class I		630 (34.1%)	59 (22.7%)	<0.001
Class II		825 (44.7%)	110 (42.3%)	
Class III		345 (18.7%)	70 (26.9%)	
Class IV		46 (2.5%)	21 (8.1%)	
EuroSCORE II[6], median [Q1-Q3]	5	1.25 [0.82–2.29]	2.27 [1.25–4.34]	<0.001
Anesthesia time, minutes, median [Q1-Q3]	0	275.0 [230.0–320.0]	305.0 [270.0–345.0]	<0.001
Operation time, minutes, median [Q1-Q3]	0	207.0 [164.0.246.0]	231.0 [201.0.267.2]	<0.001
CPB duration, minutes, mean ± SD	620	104.6 ± 53.1	132.7 ± 64.4	<0.001
ACC duration, minutes, mean ± SD	642	70.3 ± 37.9	86.5 ± 42.6	<0.001
TCA duration, minutes, mean ± SD	1903	23.5 ± 12.6	28.0 ± 17.2	0.183
Baseline rSO_2, %, median [Q1-Q3]	0	63.1 [57.2–68.3]	59.2 [52.7–64.3]	<0.001
Lowest rSO_2, %, median [Q1-Q3]	0	47.0 [40.0–52.5]	42.5 [34.4–49.5]	<0.001
Average rSO_2, %, median [Q1-Q3]	0	58.7 [54.2–62.9]	55.9 [49.4–60.4]	<0.001

(*Continued on next page*)

Table 1. *Continued*

	Missing	Delirium negative (*n* = 1854)	Delirium positive (*n* = 260)	*p*-Value
CV of rSO_2, median [Q1-Q3]	0	0.080 [0.061–0.108]	0.084 [0.062–0.123]	0.127
ARV of rSO_2, %, median [Q1-Q3]	0	0.301 [0.237–0.376]	0.288 [0.236–0.350]	0.052
Duration of rSO_2 <50%, × 5 s, median [Q1-Q3]	0	32.0 [0.0–658.0]	359.0 [1.8–1569.2]	<0.001
Duration of rSO_2 < 75% from baseline value (relative decrease), × 5 s, median [Q1-Q3]	0	0.0 [0.0–120.8]	2.5 [0.0–205.8]	0.041
AUC of rSO_2 < 50%, × 5 s*%, median [Q1-Q3]	0	47.0 [0.0–2039.8]	830.8 [1.0–8556.8]	<0.001
AUC of rSO_2 < 75% from baseline value (relative decrease), × 5 s*%, median [Q1-Q3]	0	0.0 [0.0–1004.3]	16.6 [0.0–1846.6]	0.059
Anesthetic depth measurement index, n (%)	0			0.009
BIS		1080 (58.3%)	174 (66.9%)	
PSI		774 (41.7%)	86 (33.1%)	
Duration of BIS<40 or PSI<25, × 5 s, median [Q1-Q3]	0	535.5 [169.0–1174.0]	1025.0 [333.0–2049.2]	<0.001
Duration of SR>1%, × 5 s, median [Q1-Q3]	0	102.0 [21.0–332.0]	236.000 [38.8–763.0]	<0.001
AUC of SR>1%, × 5 s*%, median [Q1-Q3]	0	317.9 [22.0–2334.0]	768.8 [65.0–4435.5]	<0.001

CPB, cardiopulmonary bypass; TCA, total circulatory arrest; COPD, chronic obstructive pulmonary disease; MI, myocardial infarction; METs, metabolic equivalents; NYHA, New York Heart Association; ACC, aortic cross clamp; rSO2, cerebral regional oxygen saturation; CV, coefficient of variation; ARV, average real variability; AUC, area under the curve; BIS, bispectral index; PSI, patient state index; SR, suppression ratio; Q1, first quartile; Q3, third quartile; SD, standard deviation.

AUPRCs of 0 · 433–0·470 in the prospective validation cohort and AUROCs of 0 · 751–0·769 and AUPRCs of 0 · 265–0·281 in the validation set. Tables 2 and S6 present the performances of the models at the optimal cutoff point in the prospective validation cohort and validation set, respectively. Overall, the ENS model performed the best, achieving the highest accuracy (0 · 881) and positive predictive value (PPV) (0 · 609) and the second-highest F1 score (0 · 538) and specificity (0 · 948) in the prospective validation cohort. The performances of the remaining individual models (LR, ANN, and SVM) are separately reported in Table S7. Additionally, to provide a deeper insight into the models' learning capabilities and to reflect further on the clinical benefits of the models, their performances in the training set are also documented separately in Table S8. In the training set, the ENS model achieved high performance with an AUROC of 0.977, an AUPRC of 0.881, an accuracy of 0.907, a PPV of 0.580, and an F1 score of 0.723.

The AUROCs of the Early PREdiction of DELIRium in ICu patients (E-PRE-DELIRIC) model, a widely used tool developed to predict delirium development during intensive care unit (ICU) admission, for predicting postoperative delirium after cardiac surgery were 0 · 831 and 0 · 726 in the prospective validation cohort and validation set, respectively.[21,22] Although the AUROC of the ENS model was higher than that of the E-PRE-DELIRIC model, the DeLong test showed no significant difference between them in the prospective validation cohort and the validation set (p = 0 · 269 and 0 · 143, respectively).

Figures 3 and S2 show the top 30 variables with the highest feature importance according to the SHapley Additive exPlanations (SHAP) method. The top five variables were estimated glomerular filtration rate (eGFR), age, T3, duration of BIS <40 or PSI <25, and Katz grade 4. Patients with lower eGFR, older age, lower T3, higher duration of BIS <40 or PSI <25, and Katz grade 4 tended to have higher SHAP values. Among these variables, several biosignal features contributed significantly to model prediction. These features included duration of BIS <40 or PSI <25, average CI, AUC of PP < 60 mmHg, duration of SR > 1%, CV of rSO2, average rSO2, and ARV of mean arterial pressure (MAP).

DISCUSSION

In this study, we developed machine learning models to predict postoperative delirium after cardiac surgery using clinical and intraoperative biosignal data. Our ENS model has the best performance for predicting postoperative delirium. The novelty of our study lies in the use of features extracted from intraoperative biosignals to predict postoperative delirium, which, to the best of our knowledge, is unprecedented. In the feature importance analysis using SHAP values, the duration of BIS <40 or PSI <25, calculated from the intraoperative electroencephalogram, was among the top five important features that contributed significantly to the model's predictive ability. Moreover, several other intraoperative biosignal features, namely, average CI, AUC of PP < 60 mmHg, duration of SR > 1%, CV of rSO2, average rSO2, and ARV of MAP, were among the top 30 important features.

Numerous studies have attempted to develop machine learning models for predicting postoperative delirium. Koster et al., 2008, developed a machine learning model to predict postoperative delirium in 300 patients who underwent elective cardiac surgery.[23] Their model used only preoperative variables and showed an AUROC of 0 · 75 (95% CI, 0 · 66-0·85). Katznelson et al., 2009, developed a model using data collected from 1,059 patients who underwent cardiac surgery with CPB, and the model's C-statistic was 0 · 774.[24] Song et al., 2023, developed and compared machine learning models for predicting postoperative delirium in elderly patients using perioperative medical data, achieving

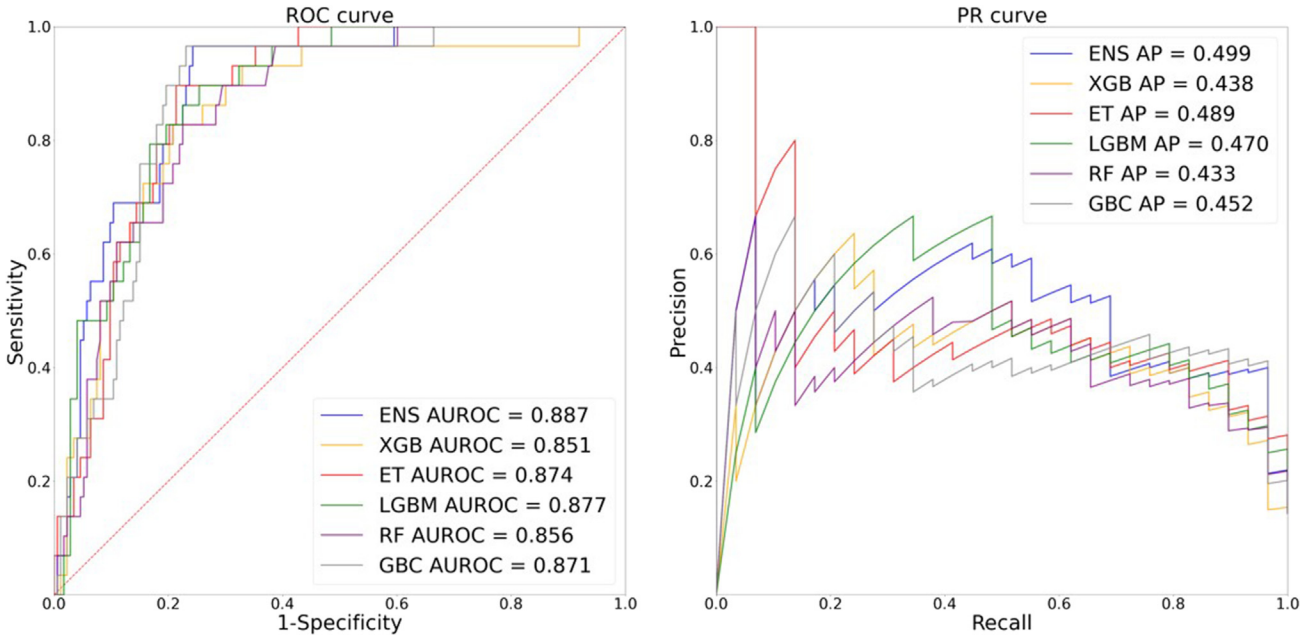


Figure 2. ROC and PR curves of the machine learning models in the prospective validation cohort

ROC: receiver operating characteristic, PR: precision recall, ENS: ensemble classifier, XGB: XGBoost, ET: extra trees classifier, LGBM: light gradient boosting machine, RF: random forest, GBC: gradient boosting classifier.

an AUROC of 0.783.[25] However, intraoperative vital signs and biosignals were not included in these predictive modeling. Rapid hemodynamic fluctuations can occur during surgery, especially cardiac surgery, which may contribute to the development of delirium.[26] To our best knowledge, our study is the first to use machine learning algorithms to predict postoperative delirium after cardiac surgery by modeling hemodynamic fluctuations reflected in intraoperative biosignals.

Intraoperative hypotension and fluctuations in blood pressure are known risk factors for delirium. In a study by Hirsch et al., 2015, increased blood pressure fluctuations, rather than absolute or relative hypotension, were predictive of postoperative delirium in non-cardiac surgery patients.[12] Ushio et al., 2022, also showed that the longer the duration of hypotension after CPB, the higher the incidence of postoperative delirium.[27] Zhang et al., 2023, also showed that increased intraoperative MAP variability may be a predictor of postoperative delirium after hip fracture surgery.[28] Extensive measurements of intraoperative biosignals may best reflect comprehensive hemodynamic fluctuations during surgery; however, including these biosignals as variables in predictive models is difficult because they require the collection of vast amounts of intraoperative time series data with conventional monitoring equipment.[26] In the present study, data were collected using VitalRecorder, a system that automatically collects time-series data from monitoring equipment, and a model that included intraoperative vital signs as variables was developed. Particularly, intraoperative vital signs, such as blood pressure, are highly likely to be modifiable risk factors that can be controlled through interventions during cardiac surgery. Therefore, the results showing the importance of intraoperative vital signs indicated that intraoperative patient management is crucial in preventing delirium after cardiac surgery.

Our study included intraoperative time-series data such as rSO2, EEG values, and MAP. Extracted at a sampling rate of 0 · 2 Hz, these variables generated high-resolution data. Consequently, there was a need to reduce the dimensionality when these variables were incorporated as inputs into the machine learning models. A substantial amount of information was lost during this process. Previous studies that used vital signs established features based on simple criteria such as the duration by which values fell outside a specific target range.[13,15] However, our study aimed to minimize information loss by defining features using various methods. These features were designed and selected to reflect the patients' overall or baseline status (e.g., average, baseline, or lowest values), the extent of unfavorable conditions encountered during surgery (e.g., duration and area under the curve falling below or above a certain value), and beat-to-beat variability within the data (CV and ARV).[20] Notably, this study used ARV as a model feature to quantify blood pressure variability. Many studies revealed that blood pressure variability as well as simple blood pressure values were associated with the occurrence of cardiovascular complication.[20,29,30] In a feature analysis based on SHAP values, the ARV of the MAP was among the top 30 variables, indicating that intraoperative blood pressure variability influences delirium development following cardiac surgery.

BIS and PSI are dimensionless numbers derived from the frontal lobe EEG signal that reflects the level of consciousness, with BIS <40 or PSI <25 indicating excessively deep anesthesia.[31,32] Burst suppression refers to an EEG pattern in which the EEG signal alternates between isoelectric patterns and bursting slow waves that occur in a comatose mental state or during overly deep anesthesia.[33] In this study, the SHAP value analysis indicated that patients with a higher depth of anesthesia (measured via EEG [BIS or PSI]) or a higher burst SR (measured via EEG) during surgery had a higher risk of postoperative delirium occurrence. This finding is consistent with those of previous studies suggesting an

Table 2. Model performances at maximum Youden J Index in the prospective validation cohort

	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
ENS	0.881	0.483	0.948	0.609	0.916	0.538
XGB	0.851	0.621	0.890	0.486	0.933	0.545
ET	0.812	0.690	0.832	0.408	0.941	0.513
LGBM	0.871	0.345	0.960	0.588	0.897	0.435
RF	0.663	0.931	0.618	0.290	0.982	0.443
GBC	0.827	0.448	0.890	0.406	0.906	0.426

PPV, positive predictive value; NPV, negative predictive value; ENS, ensemble classifier; XGB, XGBoost; ET, extra trees classifier; RF, random forest; LGBM, light gradient boosting machine; GBC, gradient boosting classifier.

association between intraoperative BIS and postoperative delirium or cognitive dysfunction.[34–36] Although the exact pathophysiology has not yet been elucidated, it is likely due to the deterioration of physical brain function and the impairment of brain network connectivity resulting from deep anesthesia.[37]

Furthermore, features related to rSO2 were among the top 30 most important features in our model. rSO2 indicates regional oxygen saturation in the brain tissue, and low rSO2 levels are associated with increased mortality and several neurological complications, including postoperative cognitive dysfunction.[38–40] Evidence regarding the association between perioperative rSO2 monitoring and postoperative delirium has been inconsistent. Several studies have demonstrated an association between preoperative rSO2 levels and postoperative delirium.[41,42] Meanwhile, meta-analyses showed no correlation between intraoperative rSO2 levels and postoperative delirium. However, these results may have been influenced by the small number of included trials.[40,43] In a prospective study conducted by Wang et al., 2019, intraoperative rSO2 desaturation was associated with postoperative delirium.[44] Our findings suggest that intraoperative rSO2 monitoring may have an impact on the prediction of postoperative delirium. More clinical trials are needed to draw definitive conclusions.

The variables with high feature importance, obtained using the SHAP method, include both modifiable and non-modifiable risk factors. Although most variables related to preoperative patient characteristics, such as age, preoperative laboratory results, and other comorbidities, are non-modifiable, the comprehensive incorporation of these variables could enhance the predictive performance of our model. This, in turn, promotes the identification of patients at high risk and the implementation of a multimodal, multidisciplinary approach for the prevention of postoperative delirium.[45] Furthermore, several intraoperative variables, especially those related to the depth of anesthesia or hemodynamics, are highly controllable. The identification of these modifiable variables, particularly those associated with intraoperative hemodynamic fluctuations and depths of anesthesia, is of paramount importance in clinical practice, given their potential for modification by physicians to prevent the occurrence of postoperative delirium. For example, an anesthesiologist can avoid inadequately deep anesthesia (indicated by low BIS and PSI, and high SR) through vigilant monitoring and titration of anesthetics. Additionally, rSO2 can be improved by optimizing hemodynamics, blood oxygen content, and cerebral blood flow.

In conclusion, our study introduces a novel approach that uses intraoperative biosignals to predict postoperative delirium. Our best model was the ENS model, with an AUROC of 0 · 887, AUPRC of 0 · 499, accuracy of 0 · 881, and an F1 score of 0 · 538. This model will enable prediction of delirium after cardiac surgery and initiation of preventive treatment, thereby improving patient prognosis and outcomes and reducing medical costs.

Limitations of the study

This study had some limitations. First, it was conducted at a single institution. Specific clinical practice protocols can vary from one institution to another. At Severance Cardiovascular Hospital, where the research was conducted, all isolated coronary artery bypass surgeries were performed off-pump, which resulted in a higher incidence of delirium in patients without CPB. In contrast, at other institutions, coronary artery bypass surgery may be performed either on-pump or off-pump. Additionally, at Severance Cardiovascular Hospital, delirium was assessed using the Intensive Care Delirium Screening Checklist (ICDSC).[46] In contrast, a significant number of institutions employ the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) for delirium screening.[47] Although both tools are effective, well-validated, and widely used, due to the retrospective nature of our model development, the results may not be directly interchangeable. Therefore, caution is warranted in generalizing our results, and further validation in other institutions is necessary to ensure the validity and robustness of our model in external environments. Second, owing to the inherent characteristics of our model development process, the model training in this study utilized retrospectively collected data. However, we attempted to offset this limitation by prospectively collecting additional data for model validation and using these data to evaluate the final performance of the model. Finally, the inclusion of a large number of features as model inputs, specifically over 100 features from pre- and intraoperative clinical data and 38 features from intraoperative biosignal data, could hinder the real-world application of our model. A supplementary platform that can automatically extract input variables, including clinical and intraoperative biosignal data; calculate predefined features; and perform risk calculation using a machine learning model is necessary for real-world implementation.

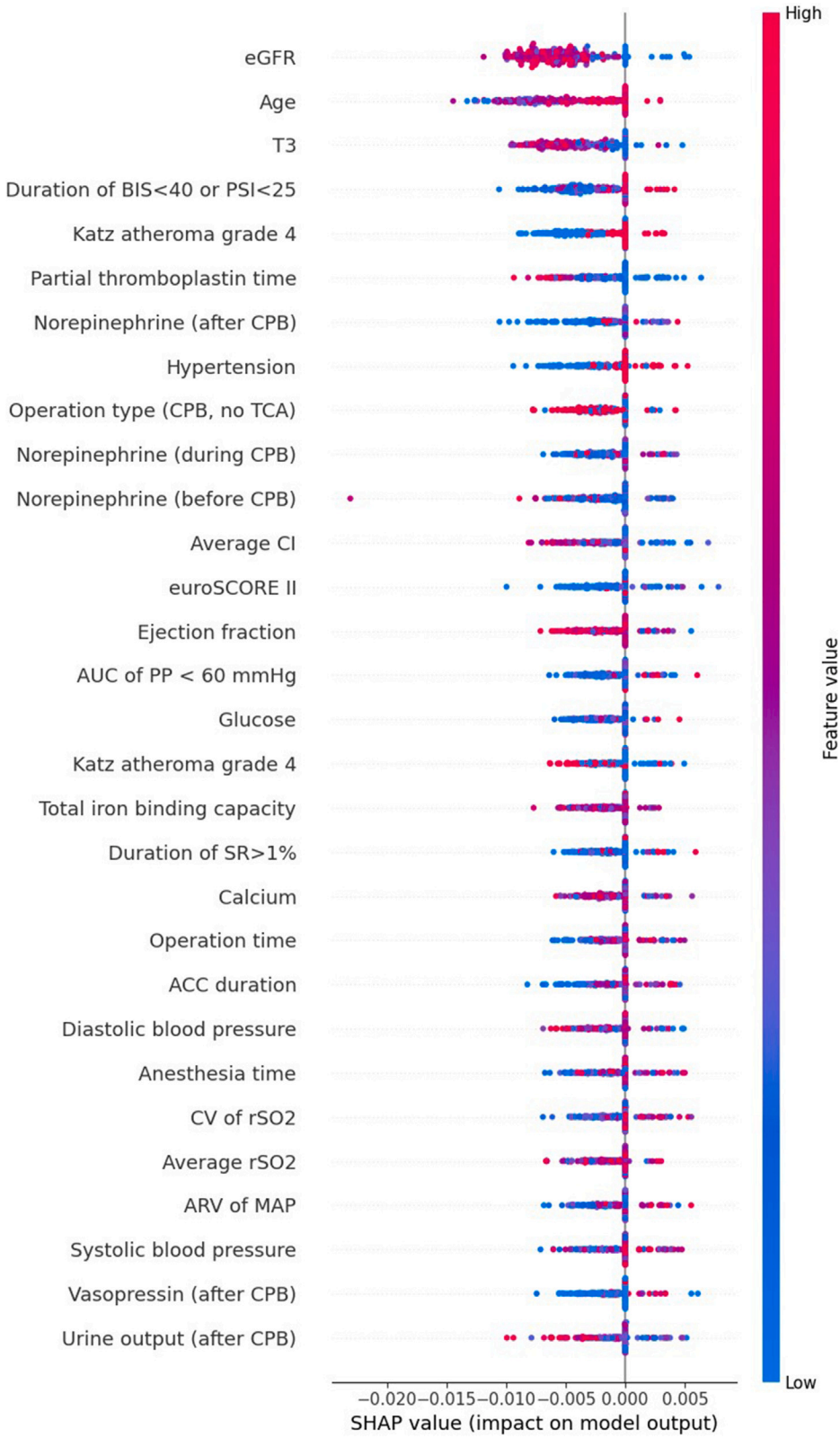

eGFR
Age
T3
Duration of BIS<40 or PSI<25
Katz atheroma grade 4
Partial thromboplastin time
Norepinephrine (after CPB)
Hypertension
Operation type (CPB, no TCA)
Norepinephrine (during CPB)
Norepinephrine (before CPB)
Average CI
euroSCORE II
Ejection fraction
AUC of PP < 60 mmHg
Glucose
Katz atheroma grade 4
Total iron binding capacity
Duration of SR>1%
Calcium
Operation time
ACC duration
Diastolic blood pressure
Anesthesia time
CV of rSO2
Average rSO2
ARV of MAP
Systolic blood pressure
Vasopressin (after CPB)
Urine output (after CPB)
−0.020 −0.015 −0.010 −0.005 0.000 0.005
SHAP value (impact on model output)
High
Low
Feature value

Figure 3. SHAP summary plot (dot)

eGFR: estimated glomerular filtration rate, BIS: bispectral index, PSI: patient state index, CPB: cardiopulmonary bypass, TCA: total circulatory arrest, CI: cardiac index, AUC: area under the curve, PP: perfusion pressure, SR: suppression ratio, ACC: aortic cross clamp, CV: coefficient of variation, rSO2: cerebral regional oxygen saturation, ARV: average real variability, MAP: mean arterial pressure, SHAP: SHapley Additive exPlanations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participant information
- METHOD DETAILS
  - Study design and ethics approval
  - Data sources and labeling
  - Intraoperative biosignal data extraction and feature calculation
  - Data preprocessing
  - Machine learning model training and performance evaluation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109932.

ACKNOWLEDGMENTS

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711196067, RS-2020-KD000095). This study was supported by a Severance Hospital Research fund for Clinical excellence(SHRC) (C-2020-0023).

AUTHOR CONTRIBUTIONS

J.W.S. and S.S. conceived and designed the overall study. H.I.K. and J.W.C. acquired the initial data. C.H. implemented the machine learning models and carried out the statistical analyses. C.H. and H.I.K. analyzed the data, interpreted the results, and drafted the manuscript. J.W.S., D.Y., and S.S. provided critical revision of the manuscript. C.H. and D.Y. searched the literature related to the prediction of postoperative delirium after cardiac surgery using machine learning. H.I.K., J.W.S., and S.S. searched the literature to identify evidence related to the association between intraoperative factors and postoperative delirium. D.Y. and J.W.S. obtained funding and provided supervision. C.H. and H.I.K. equally contributed to this work and should be considered as co-first authors. All authors approved the final version of the manuscript. All authors had full access to the data and had final responsibility for the decision to submit for publication.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 24, 2023
Revised: February 25, 2024
Accepted: May 5, 2024
Published: May 8, 2024

REFERENCES

1. O'Neal, J.B., and Shaw, A.D. (2016). Predicting, preventing, and identifying delirium after cardiac surgery. Perioperat. Med. *5*, 7. https://doi.org/10.1186/s13741-016-0032-5.
2. Rieck, K.M., Pagali, S., and Miller, D.M. (2020). Delirium in hospitalized older adults. Hosp. Pract. *48*, 3–16. https://doi.org/10.1080/21548331.2019.1709359.
3. Rengel, K.F., Pandharipande, P.P., and Hughes, C.G. (2018). Postoperative delirium. Presse Med. *47*, e53–e64. https://doi.org/10.1016/j.lpm.2018.03.012.
4. Vasilevskis, E.E., Han, J.H., Hughes, C.G., and Ely, E.W. (2012). Epidemiology and risk factors for delirium across hospital settings. Best Pract. Res. Clin. Anaesthesiol. *26*, 277–287. https://doi.org/10.1016/j.bpa.2012.07.003.
5. Ely, E.W., Shintani, A., Truman, B., Speroff, T., Gordon, S.M., Harrell, F.E., Jr., Inouye, S.K., Bernard, G.R., and Dittus, R.S. (2004). Delirium as a predictor of mortality in

mechanically ventilated patients in the intensive care unit. JAMA *291*, 1753–1762. https://doi.org/10.1001/jama.291.14.1753.
6. Neufeld, K.J., Leoutsakos, J.-M.S., Sieber, F.E., Wanamaker, B.L., Gibson Chambers, J.J., Rao, V., Schretlen, D.J., and Needham, D.M. (2013). Outcomes of early delirium diagnosis after general anesthesia in the elderly. Anesth. Analg. *117*, 471–478. https://doi.org/10.1213/ANE.0b013e3182973650.
7. Maldonado, J.R. (2017). Acute brain failure. Crit. Care Clin. *33*, 461–519. https://doi.org/10.1016/j.ccc.2017.03.013.
8. Jin, Z., Hu, J., and Ma, D. (2020). Postoperative delirium: perioperative assessment, risk reduction, and management. Br. J. Anaesth. *125*, 492–504. https://doi.org/10.1016/j.bja.2020.06.063.
9. Cai, S., Li, J., Gao, J., Pan, W., and Zhang, Y. (2022). Prediction models for postoperative delirium after cardiac surgery: Systematic review and critical appraisal. Int. J. Nurs. Stud. *136*, 104340. https://doi.org/10.1016/j.bja.2020.06.063.
10. van der Mast, R.C. (1998). Pathophysiology of delirium. J. Geriatr. Psychiatr. Neurol. *11*, 138–158. , discussion 157–158. https://doi.org/10.1177/089198879801100304.
11. Soehle, M., Dittmann, A., Ellerkmann, R.K., Baumgarten, G., Putensen, C., and Guenther, U. (2015). Intraoperative burst suppression is associated with postoperative delirium following cardiac surgery: a prospective, observational study. BMC Anesthesiol. *15*, 61. https://doi.org/10.1186/s12871-015-0051-7.
12. Hirsch, J., DePalma, G., Tsai, T.T., Sands, L.P., and Leung, J.M. (2015). Impact of intraoperative hypotension and blood pressure fluctuations on early postoperative delirium after non-cardiac surgery. Br. J. Anaesth. *115*, 418–426. https://doi.org/10.1093/bja/aeu458.
13. Maheshwari, K., Ahuja, S., Khanna, A.K., Mao, G., Perez-Protto, S., Farag, E., Turan, A., Kurz, A., and Sessler, D.I. (2020). Association Between Perioperative Hypotension and Delirium in Postoperative Critically Ill Patients: A Retrospective Cohort Analysis. Anesth. Analg. *130*, 636–643. https://doi.org/10.1213/ANE.0000000000004517.
14. Tseng, P.-Y., Chen, Y.-T., Wang, C.-H., Chiu, K.-M., Peng, Y.-S., Hsu, S.-P., Chen, K.-L., Yang, C.-Y., and Lee, O.K.-S. (2020). Prediction of the development of acute kidney injury following cardiac surgery by machine learning. Crit. Care *24*, 478. https://doi.org/10.1186/s13054-020-03179-9.
15. Xue, B., Li, D., Lu, C., King, C.R., Wildes, T., Avidan, M.S., Kannampallil, T., and Abraham, J. (2021). Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. JAMA Netw. Open *4*, e212240. https://doi.org/10.1001/jamanetworkopen.2021.2240.
16. Xie, Q., Wang, X., Pei, J., Wu, Y., Guo, Q., Su, Y., Yan, H., Nan, R., Chen, H., and Dou, X. (2022). Machine Learning-Based Prediction Models for Delirium: A Systematic Review and Meta-Analysis. J. Am. Med. Dir. Assoc. *23*, 1655–1668.e6. https://doi.org/10.1016/j.jamda.2022.06.020.
17. Peng, X., Zhu, T., Wang, T., Wang, F., Li, K., and Hao, X. (2022). Machine learning prediction of postoperative major adverse cardiovascular events in geriatric patients: a prospective cohort study. BMC Anesthesiol. *22*, 284. https://doi.org/10.1186/s12871-022-01827-x.
18. Lee, H.-C., and Jung, C.-W. (2018). Vital Recorder-a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. Sci. Rep. *8*, 1527. https://doi.org/10.1038/s41598-018-20062-4.
19. Yoon, D., Jang, J.-H., Choi, B.J., Kim, T.Y., and Han, C.H. (2020). Discovering hidden information in biosignals from patients using artificial intelligence. Korean J. Anesthesiol. *73*, 275–284. https://doi.org/10.4097/kja.19475.
20. Mena, L.J., Felix, V.G., Melgarejo, J.D., and Maestre, G.E. (2017). 24-Hour Blood Pressure Variability Assessed by Average Real Variability: A Systematic Review and Meta-Analysis. J. Am. Heart Assoc. *6*, e006895. https://doi.org/10.1161/JAHA.117.006895.
21. Wassenaar, A., van den Boogaard, M., van Achterberg, T., Slooter, A.J.C., Kuiper, M.A., Hoogendoorn, M.E., Simons, K.S., Maseda, E., Pinto, N., Jones, C., et al. (2015). Multinational development and validation of an early prediction model for delirium in ICU patients. Intensive Care Med. *41*, 1048–1056. https://doi.org/10.1007/s00134-015-3777-2.
22. van den Boogaard, M., Pickkers, P., Slooter, A.J.C., Kuiper, M.A., Spronk, P.E., van der Voort, P.H.J., van der Hoeven, J.G., Donders, R., van Achterberg, T., and Schoonhoven, L. (2012). Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. BMJ *344*, e420. https://doi.org/10.1136/bmj.e420.
23. Koster, S., Oosterveld, F.G.J., Hensens, A.G., Wijma, A., and van der Palen, J. (2008). Delirium after cardiac surgery and predictive validity of a risk checklist. Ann. Thorac. Surg. *86*, 1883–1887. https://doi.org/10.1016/j.athoracsur.2008.08.020.
24. Katznelson, R., Djaiani, G.N., Borger, M.A., Friedman, Z., Abbey, S.E., Fedorko, L., Karski, J., Mitsakakis, N., Carroll, J., and Beattie, W.S. (2009). Preoperative use of statins is associated with reduced early delirium rates after cardiac surgery. Anesthesiology *110*, 67–73. https://doi.org/10.1097/ALN.0b013e318190b4d9.
25. Song, Y.-X., Yang, X.-D., Luo, Y.-G., Ouyang, C.-L., Yu, Y., Ma, Y.-L., Li, H., Lou, J.-S., Liu, Y.-H., Chen, Y.-Q., et al. (2023). Comparison of logistic regression and machine learning methods for predicting postoperative delirium in elderly patients: A retrospective study. CNS Neurosci. Therapy *29*, 158–167. https://doi.org/10.1111/cns.13991.
26. Berger, M., Terrando, N., Smith, S.K., Browndyke, J.N., Newman, M.F., and Mathew, J.P. (2018). Neurocognitive function after cardiac surgery. Anesthesiology *129*, 829–851. https://doi.org/10.1097/ALN.0000000000002194.
27. Ushio, M., Egi, M., Fujimoto, D., Obata, N., and Mizobuchi, S. (2022). Timing, Threshold, and Duration of Intraoperative Hypotension in Cardiac Surgery: Their Associations With Postoperative Delirium. J. Cardiothorac. Vasc. Anesth. *36*, 4062–4069. https://doi.org/10.1053/j.jvca.2022.06.013.
28. Zhang, C., Song, Y., Wu, X., Miao, R., Lou, J., Ma, Y., Li, M., Mi, W., and Cao, J. (2023). Association between intraoperative mean arterial pressure variability and postoperative delirium after hip fracture surgery: a retrospective cohort study. BMC Geriatr. *23*, 735. https://doi.org/10.1186/s12877-023-04425-9.
29. Kikuya, M., Hozawa, A., Ohokubo, T., Tsuji, I., Michimata, M., Matsubara, M., Ota, M., Nagai, K., Araki, T., Satoh, H., et al. (2000). Prognostic significance of blood pressure and heart rate variabilities: the Ohasama study. Hypertension *36*, 901–906. https://doi.org/10.1161/01.hyp.36.5.901.
30. Pringle, E., Phillips, C., Thijs, L., Davidson, C., Staessen, J.A., de Leeuw, P.W., Jaaskivi, M., Nachev, C., Parati, G., O'Brien, E.T., et al. (2003). Systolic blood pressure variability as a risk factor for stroke and cardiovascular mortality in the elderly hypertensive population. J. Hypertens. *21*, 2251–2257. https://doi.org/10.1097/00004872-200312000-00012.
31. Schneider, G., Gelb, A.W., Schmeller, B., Tschakert, R., and Kochs, E. (2003). Detection of awareness in surgical patients with EEG-based indices–bispectral index and patient state index. Br. J. Anaesth. *91*, 329–335. https://doi.org/10.1093/bja/aeg188.
32. Soehle, M., Ellerkmann, R.K., Grube, M., Kuech, M., Wirz, S., Hoeft, A., and Bruhn, J. (2008). Comparison between bispectral index and patient state index as measures of the electroencephalographic effects of sevoflurane. Anesthesiology *109*, 799–805. https://doi.org/10.1097/ALN.0b013e3181895fd0.
33. Amzica, F. (2015). What does burst suppression really mean? Epilepsy Behav. *49*, 234–237. https://doi.org/10.1016/j.yebeh.2015.06.012.
34. Oliveira, C.R.D., Bernardo, W.M., and Nunes, V.M. (2017). Benefit of general anesthesia monitored by bispectral index compared with monitoring guided only by clinical parameters. Systematic review and meta-analysis. Braz. J. Anesthesiol. *67*, 72–84. https://doi.org/10.1016/j.bjane.2015.09.001.
35. Fritz, B.A., Kalarickal, P.L., Maybrier, H.R., Muench, M.R., Dearth, D., Chen, Y., Escallier, K.E., Ben Abdallah, A., Lin, N., and Avidan, M.S. (2016). Intraoperative Electroencephalogram Suppression Predicts Postoperative Delirium. Anesth. Analg. *122*, 234–242. https://doi.org/10.1213/ANE.0000000000000989.
36. Lu, X., Jin, X., Yang, S., and Xia, Y. (2018). The correlation of the depth of anesthesia and postoperative cognitive impairment: A meta-analysis based on randomized controlled trials. J. Clin. Anesth. *45*, 55–59. https://doi.org/10.1016/j.jclinane.2017.12.002.
37. Pedemonte, J.C., Plummer, G.S., Chamadia, S., Locascio, J.J., Hahm, E., Ethridge, B., Gitlin, J., Ibala, R., Mekonnen, J., Colon, K.M., et al. (2020). Electroencephalogram Burst-suppression during Cardiopulmonary Bypass in Elderly Patients Mediates Postoperative Delirium. Anesthesiology *133*, 280–292. https://doi.org/10.1097/ALN.0000000000003328.
38. Ogasawara, K., Konno, H., Yukawa, H., Endo, H., Inoue, T., and Ogawa, A. (2003). Transcranial regional cerebral oxygen saturation monitoring during carotid endarterectomy as a predictor of postoperative hyperperfusion. Neurosurgery *53*, 309–315. , discussion 314–315. https://doi.org/10.1227/01.neu.0000073547.86747.f3.
39. Kazan, R., Bracco, D., and Hemmerling, T.M. (2009). Reduced cerebral oxygen saturation measured by absolute cerebral oximetry during thoracic surgery correlates with postoperative complications. Br. J. Anaesth.

103, 811–816. https://doi.org/10.1093/bja/aep309.

40. Ding, L., Chen, D.X., and Li, Q. (2020). Effects of electroencephalography and regional cerebral oxygen saturation monitoring on perioperative neurocognitive disorders: a systematic review and meta-analysis. BMC Anesthesiol. *20*, 254. https://doi.org/10.1186/s12871-020-01163-y.
41. Susano, M.J., Dias, M., Seixas, F.S., Vide, S., Grasfield, R., Abelha, F.J., Crosby, G., Culley, D.J., and Amorim, P. (2021). Association Among Preoperative Cognitive Performance, Regional Cerebral Oxygen Saturation, and Postoperative Delirium in Older Portuguese Patients. Anesth. Analg. *132*, 846–855. https://doi.org/10.1213/ANE.0000000000005159.
42. Schoen, J., Meyerrose, J., Paarmann, H., Heringlake, M., Hueppe, M., and Berger, K.-U. (2011). Preoperative regional cerebral oxygen saturation is a predictor of postoperative delirium in on-pump cardiac surgery patients: a prospective observational trial. Crit. Care *15*, R218. https://doi.org/10.1186/cc10454.
43. He, K.-Q., Wang, S., Zhang, W., Liu, Q., and Chai, X.-Q. (2022). What is the impact of perioperative cerebral oxygen desaturation on postoperative delirium in old population: a systemic review and meta-analysis. Aging Clin. Exp. Res. *34*, 1761–1770. https://doi.org/10.1007/s40520-022-02128-6.
44. Wang, X., Feng, K., Liu, H., Liu, Y., Ye, M., Zhao, G., and Wang, T. (2019). Regional cerebral oxygen saturation and postoperative delirium in endovascular surgery: a prospective cohort study. Trials *20*, 504. https://doi.org/10.1186/s13063-019-3586-y.
45. Zhang, H., Lu, Y., Liu, M., Zou, Z., Wang, L., Xu, F.-Y., and Shi, X.-Y. (2013). Strategies for prevention of postoperative delirium: a systematic review and meta-analysis of randomized trials. Crit. Care *17*, R47. https://doi.org/10.1186/cc12566.
46. Bergeron, N., Dubois, M.J., Dumont, M., Dial, S., and Skrobik, Y. (2001). Intensive Care Delirium Screening Checklist: evaluation of a new screening tool. Intensive Care Med. *27*, 859–864. https://doi.org/10.1007/s001340100909.
47. Ely, E.W., Margolin, R., Francis, J., May, L., Truman, B., Dittus, R., Speroff, T., Gautam, S., Bernard, G.R., and Inouye, S.K. (2001). Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit. Care Med. *29*, 1370–1379. https://doi.org/10.1097/00003246-200107000-00012.
48. Mathur, S., Patel, J., Goldstein, S., and Jain, A. (2022). Bispectral Index. In StatPearls (StatPearls Publishing).
49. Drover, D., and Ortega, H.R. (2006). Patient state index. Best Pract. Res. Clin. Anaesthesiol. *20*, 121–128. https://doi.org/10.1016/j.bpa.2005.07.008.
50. Mailhot, T., Cossette, S., Lambert, J., Cournoyer, A., and Denault, A.Y. (2016). Cerebral oximetry as a biomarker of postoperative delirium in cardiac surgery patients. J. Crit. Care *34*, 17–23. https://doi.org/10.1016/j.jcrc.2016.02.024.
51. Shan, W., Chen, B., Huang, L., and Zhou, Y. (2021). The Effects of Bispectral Index-Guided Anesthesia on Postoperative Delirium in Elderly Patients: A Systematic Review and Meta-Analysis. World Neurosurg. *147*, e57–e62. https://doi.org/10.1016/j.wneu.2020.11.110.
52. Hu, Q., Deng, X., Liu, X., Wang, A., and Yang, C. (2020). A robust beat-to-beat artifact detection algorithm for pulse wave. Math. Probl Eng. *2020*, 1–8. https://doi.org/10.1155/2020/5691805.
53. Kher, R. (2019). Signal processing techniques for removing noise from ECG signals. J. Biomed. Eng. Res. *1*, 1–19. https://doi.org/10.17303/jber.2019.3.101.
54. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. https://doi.org/10.1038/nature14539.
55. Rusk, N. (2016). Deep learning. Nat. Methods *13*, 35. https://doi.org/10.1038/nmeth.3707.
56. Azur, M.J., Stuart, E.A., Frangakis, C., and Leaf, P.J. (2011). Multiple imputation by chained equations: what is it and how does it work? Int. J. Methods Psychiatr. Res. *20*, 40–49. https://doi.org/10.1002/mpr.329.
57. Sherazi, S.W.A., Bae, J.-W., and Lee, J.Y. (2021). A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome. PLoS One *16*, e0249338. https://doi.org/10.1371/journal.pone.0249338.
58. Agnihotri, D., Verma, K., Tripathi, P., and Singh, B.K. (2019). Soft voting technique to improve the performance of global filter based feature selection in text corpus. Appl. Intell. *49*, 1597–1619. https://doi.org/10.1007/s10489-018-1349-1.
59. Fluss, R., Faraggi, D., and Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. Biom. J. *47*, 458–472. https://doi.org/10.1002/bimj.200410135.
60. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From Local Explanations to Global Understanding with Explainable AI for Trees. Nat. Mach. Intell. *2*, 56–67. https://doi.org/10.1038/s42256-019-0138-9.
61. DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics *44*, 837–845. https://doi.org/10.2307/2531595.

STAR★METHODS

KEY RESOURCES TABLE

RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
python	https://www.python.org/	Version 3.10.6
scikit-learn	https://scikit-learn.org/	Version 1.3.0
matplotlib	https://matplotlib.org/	Version 3.7.2
xgboost	https://xgboost.readthedocs.io/en/stable/	Version 2.0.0
lightgbm	https://lightgbm.readthedocs.io/en/stable/	Version 4.1.0
shap	https://shap.readthedocs.io/en/latest/	Version 0.43.0
miceforest	https://pypi.org/project/miceforest/	Version 5.6.3
R	https://www.r-project.org/	Version 4.2.0
pROC	https://github.com/cran/pROC/	Version 1.18.4
VitalRecorder	https://vitaldb.net/	Version 1.8.16.8
VitalUtils	https://vitaldb.net/	Version 1.0.1
Original code of this study	https://github.com/CMI-Laboratory/PODEC_ML/	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Dukyong Yoon (dukyong.yoon@yonsei.ac.kr).

Materials availability

This study did not generate any new materials.

Data and code availability

- The data cannot be made publicly accessible due to hospital regulations. Distributing these data without the necessary consent could potentially breach patient confidentiality and contravene the approval granted by the Institutional Review Board for this study. Any additional information required to reanalyze the data reported in this paper is available from the lead contact on request.
- All original code has been deposited at https://github.com/CMI-Laboratory/PODEC_ML/ and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participant information

Patients aged ≥19 years who underwent cardiac surgery at the Severance Cardiovascular Hospital, Yonsei University Health System, Seoul, South Korea with or without cardiopulmonary bypass. Data used for developing machine learning models were collected from December 14, 2018 to December 22, 2021, whereas those for prospective validation were collected from March 28, 2022 to June 28, 2022. All participants in this study were of Korean ethnicity. In total, 2,179 adult patients were included: among them, 1,969 and 210 patients belonged to the training and validation sets and to the prospective validation cohort, respectively. After excluding 62 patients (55 patients from the training and validation sets and 7 patients from the prospective validation cohort) with missing cerebral oximetry or electroencephalogram data and 3 patients who were already included in the training set (2 patients from the validation set and 1 patient from the prospective validation cohort), 2,114 patients were included. The median age of these patients was 66 years (interquartile range 57-73), and 1,322 (62.5%) were male.

METHOD DETAILS

Study design and ethics approval

This study was conducted at the Severance Cardiovascular Hospital, Yonsei University Health System, Seoul, South Korea. The Department of Anesthesiology at Severance Cardiovascular Hospital has been developing an institutional cardiac surgery registry that collects perioperative clinical data and intraoperative biosignal data measured by VitalRecorder, a software that stores high-resolution biosignal waveforms and

vital signs.[18] The Institutional Review Board (IRB) of Severance Hospital and Yonsei University Health System approved the collection of prospective data for the establishment of this institutional cardiac surgery registry upon informed consent from cardiac surgery patients (approval number: 4-2018-1002). The machine learning models were developed using data gathered during the approved collection period. The IRB approved the retrospective data analysis and the development of machine learning models using this data (Approval number: 4-2021-0799). Additional approval was obtained from the IRB for prospective data collection upon informed consent from cardiac surgery patients (Approval number: 4-2022-0093), and the model was prospectively validated using this data.

Data sources and labeling

Patients aged ≥19 years who underwent cardiac surgery with or without CPB were included. Patients who had undergone congenital heart surgery were excluded. Clinical data including preoperative (comorbidities, blood test results, medications), intraoperative (medications, surgical procedure times, input and output during surgery), and postoperative (medications, postoperative complications) information were collected from an institutional cardiac surgery registry (Figure 1). In the current study, data used for developing machine learning models were collected from December 14, 2018 to December 22, 2021, whereas those for prospective validation were collected from March 28, 2022 to June 28, 2022. As inputs for machine learning models, preoperative or intraoperative variables reflective of patient status and possible predictors of postoperative delirium were included. The specific variables are listed in Table S1. For patients who had undergone multiple surgeries, only data from the first surgery was included to avoid data leakage and performance overestimation.

Vital signs recorded via VitalRecorder and stored in the registry include (1) rSO2 on cerebral oximetry, (2) depth of anesthesia as evaluated according to the BIS (BIS Quarto Sensor, Medtronic Corp, Minneapolis, MN, USA) or PSI (RD SedLine EEG sensor, Masimo Corp, Irvine, CA, USA) on the EEG monitor, (3) arterial blood pressure, etc.[18,48,49]

We extracted the intraoperative biosignals stored in the registry. The specific parameters used as inputs for the machine learning models are listed in Table S2. Patients with missing cerebral oximetry or EEG data in the VitalRecorder files were excluded because these two parameters were directly related to the brain and would therefore have a strong association with postoperative delirium (Figure 1).[50,51]

Patients who were diagnosed with delirium within 7 days following cardiac surgery were labeled as positive for postoperative delirium. Delirium was identified using either the ICDSC, which was assessed three times daily during the ICU stay, or through a multidisciplinary psychiatry consultation.[46]

Intraoperative biosignal data extraction and feature calculation

The original sampling frequency for each parameter stored in the VitalRecorder file varied, with the lowest sampling frequency of 0·2 Hz for rSO2. Thus, the remaining study parameters were also extracted at 0·2 Hz into CSV (comma-separated value) files via VitalUtils. Briefly, VitalUtils is a utility program designed for convenient handling of VitalRecorder files, wherein the files can be extracted as a CSV file according to a desired sampling frequency. VitalUtils was used for two reasons. First, we hypothesized that extracting values every 5 s would sufficiently reflect the patient's physiological status given the considerably longer operation time. Second, we aimed to reduce the computation time required for the calculation of the study features.

Raw VitalRecorder waveform data may be prone to unprocessed artifacts, which could have a detrimental effect on data analysis because they are not indicative of the patient's true physiological status and should be eliminated.[18,52,53] We extensively reviewed the VitalRecorder files to identify patterns of unwanted artifacts (e.g., those caused by sensor detachment or disconnection) and found that critical artifacts could be removed using a rule-based approach. Figure S3 shows examples of the artifacts, and Table S9 lists the rules adopted to remove these artifacts. After extraction, we calculated features from these parameters for use as inputs for the machine learning models (Table S2). These features were designed and selected to reflect the patients' overall or baseline status (e.g., average, baseline, or lowest values), extent of unfavorable conditions encountered during surgery (e.g., duration and area under the curve falling below or above a certain value), and beat-to-beat variability within the data (CV and ARV).[20]

Feature calculation from intraoperative biosignal parameters is further described in supplemental methods in Supplemental Information. The features were calculated in a predetermined method instead of adopting an end-to-end algorithm (e.g., using the raw data to be trained with an artificial neural network-based algorithm) owing to the following reasons. First, the raw data of these parameters were too heterogeneous (e.g., the operation time and type varied according to the application of CPB or TCA). Second, the sample size of our cohort was not sufficiently large to handle heterogeneous data using an end-to-end algorithm. A significantly large dataset is required for ANN-based end-to-end algorithms to learn meaningful features when the dataset is exceedingly complex.[54,55]

During CPB, normal physiological functions are taken over by the CPB machine as extracorporeal circulation; thus, VitalRecorder parameters related to circulation do not reflect the physiological status during that period. Therefore, for patients who underwent cardiac surgery with CPB, data for all parameters, except for cerebral oximetry and EEG (as these two parameters were not directly related to circulation), were excluded during the duration of the CPB (Figure 1).

Data preprocessing

Variables with missing values exceeding 50% of the training set were excluded. We one-hot encoded the categorical variables; for variables with missing values, a dummy variable (coded as 1 if the value was missing and 0 otherwise), replacing the original one-hot encoded vector with a vector filled with zeros, was created. Continuous variables with missing values were managed using Multiple Imputation by Chained Equations with the random Forest method.[56] The imputation model was trained in the training set and applied in the validation set during

both the model development phase and prospective validation. Z-score normalization, scaling values to obtain a mean of 0 and standard deviation of 1, was used for continuous variables. The z-score normalization formulas were determined based on the training set and subsequently applied to the validation set. For variables quantifying the depth of anesthesia, we created a separate dummy variable (0 for BIS and 1 for PSI) owing to potential measurement differences arising from vendor differences.[48,49]

Machine learning model training and performance evaluation

For the model development, the collected data were temporally split into the training and validation sets. Data of patients who underwent cardiac surgery between December 14, 2018 and May 31, 2021 and between June 1, 2021 and December 22, 2021 were assigned to the training and validation sets, respectively. We prospectively validated the developed model using data collected from patients who underwent cardiac surgery between March 28, 2022 and June 28, 2022.

Eight machine learning models were used, namely, the RF, XGB, ET, GBC, LGBM, SVM, LR, and ANN. For each model, an extensive grid of hyperparameter values were created, and the model were trained on the training set for each hyperparameter combination, thereby validating the model against a validation set. The model with the hyperparameter combination yielding the highest performance based on the AUROC was selected for final validation with the prospective validation cohort.

Ensemble models such as soft-voting ensemble classifiers often outperform individual classifiers by aggregating individual results, thereby addressing the individual classifiers' weaknesses.[57,58] Accordingly, we constructed a soft-voting ENS classifier by averaging the outputs of the highest-performing models in the validation set. To identify the most effective ENS classifier, the number of top-performing individual models (from one to eight) included in the ENS were varied, selecting the ensemble that achieved the highest AUROC in the validation set. We calculated the model's accuracy, sensitivity, specificity, PPV, negative predictive value (NPV), and F1 score at the optimal cutoff point, defined as the highest Youden J index.[59] The SHAP method was used to interpret individual predictions and quantify the contribution of each variable to the model's predictive ability.[60] We then compared our model's performance with that of E-PRE-DELIRIC, a widely used tool developed to predict delirium development during ICU admission.[21,22]

QUANTIFICATION AND STATISTICAL ANALYSIS

For continuous variables, the normality of distribution was assessed using the Shapiro–Wilk test, and normally and non-normally distributed variables were compared between the groups using the independent samples t-test and the Mann–Whitney U test, respectively. For comparisons among three or more groups, normally and non-normally distributed continuous variables were compared using the analysis of variance (ANOVA) test and the Kruskal–Wallis test, respectively. Meanwhile, categorical variables were compared using the chi-square test or Fisher's exact test, as appropriate. Normally distributed variables are presented as mean $\pm$ standard deviation, whereas non-normally distributed variables are reported as median [IQR of Q1–Q3]. AUROCs were compared using the Delong test.[61] Statistical significance was set at $p < 0{\cdot}05$ for all tests. The Delong test was conducted using the pROC library (version 1.18.4) in R (version 4.2.0). All other statistical tests were conducted using the scikit-learn library (version 1.3.0) in Python (version 3.10.6).