

Original Paper

Early Prediction of Delirium in Postcardiac Surgery Patients: Machine Learning Model Development and External Validation

Huixiu Hu^{1,2,3*}, MMed; Yuxiang Wang^{4*}, BSc; Houfeng Li^{5*}, MMed; Qinglai Zang^{6,7*}, MSc; Jing Huang⁸, MMed; Ying Zhang³, MMed; Jinjing Wu⁴, MSc; Long Liu⁴, MSc; Zhen Xing^{3*}, MMed; Yaohua Yu^{1,2*}, BMed

¹The Graduate School of Fujian Medical University, Fuzhou, Fujian, China

²Department of Anesthesiology, The First Hospital of Putian City, Putian, China

³Department of Anesthesiology, The First Affiliated Hospital of Hebei North University, Zhangjiakou, Hebei, China

⁴School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, Shanghai, China

⁵Department of Anesthesiology, Changhai Hospital, the Second Military Medical University, Shanghai, China

⁶School of Anesthesiology, Second Military Medical University, Shanghai, China

⁷Information Center, The Second Affiliated Hospital of Naval Medical University, Shanghai, China

⁸Graduate School, Wannan Medical College, Wuhu, Anhui, China

*these authors contributed equally

Corresponding Author:

Yaohua Yu, BMed

The Graduate School of Fujian Medical University

No.1 Xuefu North Road, Shangjie Town, Minhou County

Fuzhou, Fujian, 351022

China

Phone: 86 139 5956 1579

Email: yyh.8@163.com

Abstract

Background: Delirium is a frequent postoperative complication among patients who have undergone cardiac surgery and is associated with prolonged hospitalization, cognitive decline, and increased mortality. Early prediction of delirium is therefore critical for initiating timely interventions.

Objective: This study proposes the development and validation of a machine learning-based model to predict postoperative delirium in patients undergoing cardiac surgery during intensive care unit (ICU) care, facilitating the early detection of individuals at high risk of delirium and supporting clinicians in the deployment of targeted preventive strategies.

Methods: This study extracted data on postoperative cardiac surgery patients who remained in the ICU for more than 24 hours from the Medical Information Mart for Intensive Care IV version 2.0 (MIMIC-IV 2.0) database and the eICU Collaborative Research Database (eICU-CRD). The MIMIC-IV 2.0 cohort was randomly divided into a training set and an internal validation set in a 7:3 ratio, whereas the eICU-CRD functioned as an independent validation cohort. We used data from the first 24 hours of ICU monitoring to model the likelihood of delirium over the entire ICU admission period. Delirium was identified by a positive Confusion Assessment Method for the Intensive Care Unit evaluation (ie, score ≥ 4). We built predictive models by using logistic regression, support vector classifier, extreme gradient boosting (XGB), and random forest classifiers. Their performance was assessed via the area under the receiver operating characteristic curve, accuracy, sensitivity, positive predictive value, negative predictive value, and F_1 -score.

Results: The analysis involved 2124 patients from the MIMIC-IV 2.0 database and 2406 from the eICU-CRD. A set of 57 variables was selected to construct the predictive models. Among the various machine learning models tested, the XGB model demonstrated the best performance for delirium prediction during internal validation. As for external validation, the model achieved an area under the receiver operating characteristic curve of 0.75, indicating strong discriminatory ability. The most important predictive features identified by the model included hospital length of stay, minimum Glasgow Coma Scale score, mean blood pressure, Sequential Organ Failure Assessment score, weight, urine output, heart rate, and age.

Conclusions: The XGB model with strong predictive capability for ICU delirium after cardiac surgery was developed and externally validated. This model offers essential technical support for building real-time delirium alert systems and enables ongoing risk stratification and evidence-based decision-making within the ICU environment.

KEYWORDS

prediction model; machine learning; cardiac surgery; intensive care unit; delirium; Medical Information Mart for Intensive Care IV version 2.0 database; MIMIC-IV 2.0 database; eICU Collaborative Research Database; eICU-CRD

Introduction

Delirium is an acute neuropsychiatric syndrome commonly associated with encephalopathy, acute cerebral dysfunction, and states of confusion, particularly following surgical procedures [1-4]. In patients undergoing cardiac surgery, the incidence of postoperative delirium has been reported to range from 10% to 40% [5-7]. This condition is linked to a variety of adverse outcomes, including heightened pain perception, depression, cognitive impairment, and increased mortality [8,9]. Currently, the assessment of patients' arousal can be conducted using standardized tools such as the Richmond Agitation-Sedation Scale and the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) to identify different types of delirium with distinct characteristics [10-12]. Despite the existence of standardized instruments, the diagnosis of delirium frequently depends on the patient's subjective assessment of their condition. If the occurrence of delirium in patients could be predicted within a short period, it will substantially reduce the aforementioned risks.

Machine learning (ML), a branch of artificial intelligence, has driven notable progress across numerous domains of health care [13,14]. One such area where ML has shown its potential is in the postoperative surveillance with cardiac surgery, offering more information to predict delirium [15,16]. It has the potential to improve patient health care outcomes [17,18]. Compared to traditional data analysis techniques, ML models can provide more intricate predictions and perform real-time monitoring using objective data from all patients [19-21]. Furthermore, recent research has also used ML to forecast the near-term mortality rates of patients after cardiac operations. For example, Nistal-Nuño [22] constructed an extreme gradient boosting (XGB)-based predictive model to estimate 24-hour postoperative mortality following cardiac surgery. The outcome demonstrated that XGB attained an area under the receiver operating characteristic curve (AUC) of 87.5%, signifying the model's exceptional performance in forecasting intensive care unit (ICU) mortality, notably surpassing other models. Zhang et al [23] assessed an ML model, comparing it to existing severity-of-illness systems to develop a real-time tool for predicting death. However, no correlation-predictive models using ML have been developed for patients who experience delirium after cardiac surgery.

In this study, we created and verified 4 models using ML techniques to anticipate the occurrence of delirium and facilitate identification. Furthermore, we enhanced the interpretability of

the results by prioritizing the independent variables according to their predictive significance.

Methods

Ethical Considerations

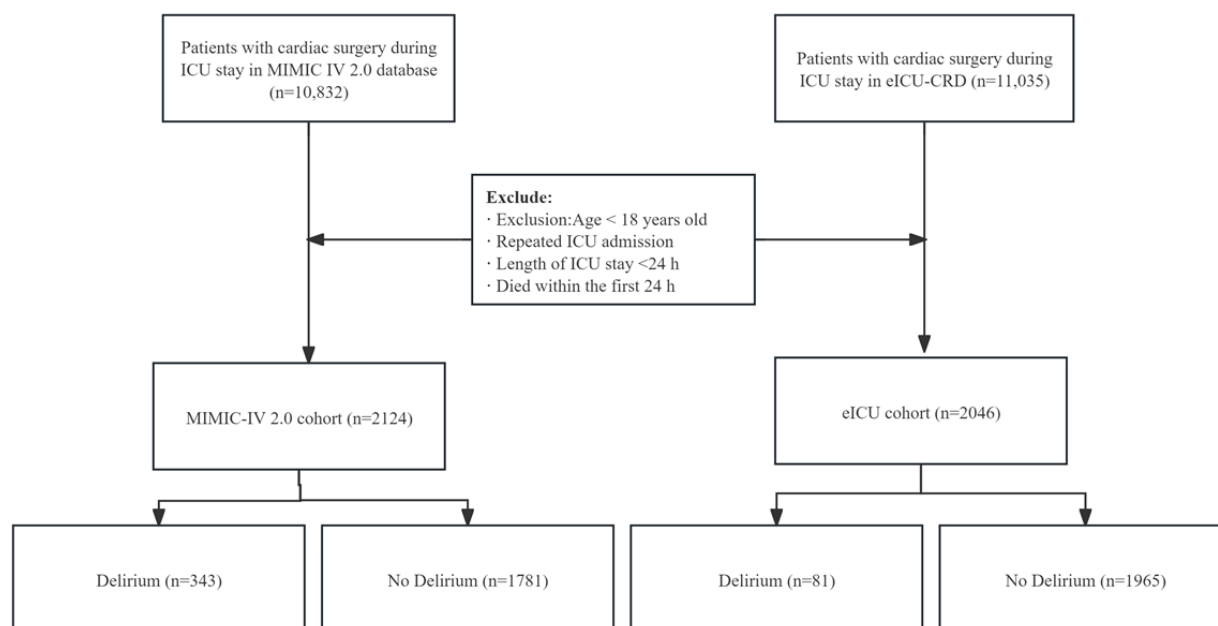
This study used data from 2 publicly available critical care databases: Medical Information Mart for Intensive Care IV version 2.0 (MIMIC-IV 2.0) database and eICU Collaborative Research Database (eICU-CRD). Both databases were approved by the institutional review boards of the Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology. As all data were fully deidentified before release, the requirement for individual informed consent was waived in accordance with the Declaration of Helsinki and applicable regulations. All team members underwent certified training in "Data or Specimens Only Research" to comply with ethical regulations governing dataset access.

Study Population

This study used 2 publicly available critical care databases to develop and validate a predictive model. The training dataset was sourced from MIMIC-IV 2.0, which includes 76,943 ICU admissions recorded at Beth Israel Deaconess Medical Center (Boston, Massachusetts, United States) between 2008 and 2019 [24]. For external validation, we used the eICU-CRD, which contains deidentified data for over 200,000 patients admitted to 208 US hospitals between 2014 and 2015 [25]. Both databases contain structured clinical data, including demographics, vital signs, laboratory test results, procedures, medications, and outcomes. We included adult patients (aged ≥ 18 years) who underwent major cardiovascular surgeries, such as coronary artery bypass grafting, heart valve repair or replacement, combined procedures, or other surgeries involving cardiopulmonary bypass.

We applied consistent inclusion and exclusion criteria across both datasets to ensure cohort comparability and data quality. Inclusion criteria required patients to meet the age threshold and have documented cardiovascular surgery. We excluded patients who had ICU stays shorter than 24 hours (to ensure sufficient observation data), missing essential demographic or outcome variables, or delirium recorded within the first 24 hours of ICU admission (to preserve the prediction time window). Although the 2 datasets differ in time range and hospital coverage, we aligned the study population by applying uniform definitions for surgical type and using a standardized 24-hour observation window. Figure 1 summarizes the baseline characteristics of the 2 cohorts.

Figure 1. Schematic representation of the study design. eICU-CRD: eICU Collaborative Research Database; ICU: intensive care unit; MIMIC-IV 2.0: Medical Information Mart for Intensive Care IV version 2.0.



Delirium Assessment

Delirium served as the primary outcome, identified on the basis of a positive CAM-ICU assessment (score ≥ 4) and consistent diagnostic coding. The observation window was the first 24 hours after ICU admission, during which patient data were collected for modeling. If at least one positive delirium assessment occurred at the time of prediction, the patient was considered delirious.

Data Extraction and Processing

The clinical data were retrieved and extracted using the structured query language, with pgAdmin 4 serving as the administrative platform for PostgreSQL. The prediction model included just clinical and laboratory characteristics that were accessible on the initial day of admission to the ICU, with patients being recognized by their unique ID numbers. The predictors consisted of the following variables— (1) demographics: age, gender, ethnicity and weight; (2) vital signs: heart rate, mean blood pressure, respiratory rate, systolic blood pressure, and temperature; (3) laboratory analysis: hemoglobin level, platelet count, white blood cell count, lactate and urine output; (4) severity scoring: Glasgow Coma Scale (GCS) and Sequential Organ Failure Assessment (SOFA) scores; (5) comorbidities: myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, diabetes (with control, without control), paraplegia, renal disease, malignant cancer, severe liver disease, and AIDS; (6) medications: opioids, barbiturates, benzodiazepines, acetaminophen, antipsychotics, anticoagulant, antihistamines, diuretics, anesthesia, and anticholinergics; (7) treatment measures: emergency admission, first care unit, last care unit, renal replacement therapy, invasive ventilation, length of ICU stay.

Missing Data Management

Variables exhibiting a missing value rate exceeding 10% were omitted to prevent potential bias. Variables with less than 10% missing values were subjected to multivariable imputations [26].

Data Balance

The dataset showed a marked imbalance, with notably fewer positive delirium cases compared with negative ones, which caused the model to lean toward predicting the majority (negative) class. To mitigate this issue, we used the Synthetic Minority Oversampling Technique to artificially augment the number of positive samples, thereby achieving a more balanced class distribution and enhancing the model's ability to generalize.

Feature Selection

The feature selection process involved using the recursive elimination of features approach of the random forest [27]. This method was used to identify the most optimal combinations of predictive variables. By examining the weight of features and their correlation and after stratified 10-fold cross-validation, final features were selected based on importance scores, correlation analysis, and cross-validation results. This process reduced dimensionality while preserving predictive power.

Model Development and Hyperparameter Tuning

The MIMIC-IV 2.0 dataset (N=2124) was randomly split into a training set (1487/2124, 70%) and a testing set (637/2124, 30%), whereas the eICU-CRD dataset was used as an external validation cohort. We developed prediction models using 4 widely adopted ML algorithms. Logistic regression was implemented for binary classification using maximum likelihood estimation [28]. Random forest, an ensemble learning method, combined multiple decision trees through majority voting to

enhance predictive performance [29]. XGB used a gradient boosting framework to iteratively build strong learners from weak ones [30]. Support vector classifier aimed to find the optimal hyperplane in a high-dimensional space for classification [31]. Bayesian optimization was used to identify optimal hyperparameters for each model, improving training efficiency and performance [32].

Model Performance Evaluation

To comprehensively assess the discriminatory performance of the prediction models, we used the receiver operating characteristic (ROC) as the primary evaluation metric. Additional metrics included accuracy, positive predictive value, negative predictive value, and sensitivity. We also reported the F_1 -score, the harmonic mean of precision and sensitivity, to reflect the balance between these two metrics. Together, these indicators were used to evaluate the clinical applicability of each model in stratifying the risk of postoperative delirium among cardiac surgery patients. Shapley Additive Explanations (SHAP) was used to investigate the interpretability of the final predictive model.

Statistical Analysis

All statistical analyses were conducted using Stata 17.0 and SPSS (version 27.0; IBM Corp). Frequencies and percentages were used to summarize categorical variables, with comparisons made via the chi-square test. The distribution of continuous variables was assessed using the Shapiro-Wilk test. Normally distributed data were reported as mean (SD) and compared using independent 2-tailed t tests. Skewed data were summarized as

median and IQR and analyzed using the Mann-Whitney U or Kruskal-Wallis test, based on group composition. Statistical significance was determined using 2-sided tests, with a threshold of $P < .05$.

Results

Baseline Characteristics

A total of 2124 patients from the MIMIC-IV 2.0 database were included in the final analysis. Among them, 16.1% (343/2124) of cardiac surgery patients were diagnosed with delirium during their hospital stay, occurring after the first day of ICU admission. In the external validation cohort, an analysis was conducted on 2046 cases obtained from the eICU-CRD, of whom 3.81% (81/2046) developed delirium during the same postoperative period, also defined as after the first ICU day. In these patients with delirium, maximum heart rate, minimum mean blood pressure, minimum hemoglobin level, minimum platelet count, maximum white blood cell count, urine output, minimum GCS score, SOFA, and length of ICU stay showed a notable disparity between the two different groups. Tables 1-3 present a concise summary of the comparison of fundamental traits, vital signs, and laboratory analysis between patients with and without delirium. According to the data, patients with delirium are predominantly male, typically older, and have longer hospital stays with higher severity scores upon admission. Additionally, factors such as weight loss, decreased urine output, and decreased mean arterial pressure may all exacerbate the likelihood of delirium in patients.

Table 1. Characteristics of patients and controls from the development dataset for the first 24-hour model cohort: demographics and vital signs.

Patient characteristics	MIMIC-IV 2.0 ^a cohort			eICU-CRD ^b cohort		
	No delirium (n=1781)	Delirium (n=343)	<i>P</i> value	No delirium (n=1965)	Delirium (n=81)	<i>P</i> value
Gender, n (%)			.10			.09
Male	1144.0 (64.2)	195.0 (56.9)		937.0 (49.5)	46 (57)	
Female	637.0 (35.8)	148.0 (43.1)		992.0 (50.5)	35 (43)	
Race, n (%)			.80			.62
Asian	47.0 (2.6)	6.0 (1.7)		29.0 (1.5)	6.0 (1.4)	
Black	80.0 (4.5)	19.0 (5.5)		475.0 (24.2)	18.0 (22.5)	
Hispanic	126.0 (7.1)	42.0 (12.2)		1227.0 (62.4)	53.0 (65)	
White	1391.0 (78.1)	219.0 (63.8)		132.0 (6.7)	12.0 (14.8)	
Unknown	29.0 (1.6)	47.0 (13.7)		96.0 (4.9)	10.0 (2.3)	
Other	108.0 (6.1)	10.0 (2.)		6.0 (0.3)	2.0 (2.5)	
Age (y), median (IQR)	70.0 (61.0-79.0)	75.0 (64.0-81.0)	<.001	68.0 (57.0-77.0)	71.0 (61.0-79.8)	<.001
Weight (kg), median (IQR)	83.8 (70.8-96.7)	80.0 (60.0-95.4)	.02	85.2 (70.0-102.0)	84.5 (67.9-103.4)	.31
Vital signs, median (IQR)						
Heart rate min (bpm)	67.0 (60.0-74.0)	68.0 (59.0-77.0)	.12	69.0 (59.0-80.0)	70.0 (60.0-80.0)	.29
Heart rate max (bpm)	94.0 (86.0-106.0)	97.0 (88.0-110.0)	<.001	101.0 (88.0-116.0)	107.0 (92.2-124.5)	<.001
Heart rate mean (bpm)	80.2 (73.3-87.5)	82.7 (75.6-89.3)	<.001	83.6 (75.7-94.4)	86.3 (75.7-97.9)	.007
Mean blood pressure min (mm Hg)	57.0 (53.0-62.0)	55.0 (49.5-60.0)	<.001	62.0 (54.0-72.0)	60.0 (52.0-68.0)	<.001
Mean blood pressure max (mm Hg)	97.0 (89.0-107.0)	95.0 (88.0-108.0)	.33	103.0 (99.0-118.0)	104.0 (91.0-122.0)	.31
Mean blood pressure mean (mm Hg)	74.6 (70.3-79.4)	73.3 (69.3-78.8)	.02	80.0 (76.6-90.4)	78.2 (70.3-88.2)	.02
Respiratory rate min (bpm)	12.0 (10.0-14.0)	12.0 (9.0-14.0)	.75	13.0 (11.0-16.0)	13.0 (11.0-16.0)	.63
Respiratory rate max (bpm)	26.0 (23.0-29.0)	26.0 (23.00-30.0)	.69	27.0 (24.0-32.00)	28.0 (24.0-33.00)	.42
Respiratory rate mean (bpm)	17.8 (16.2-19.6)	18.1 (16.2-20.2)	.23	19.2 (17.1-21.9)	19.2 (16.9-22.9)	.76
Systolic blood pressure min (mm Hg)	93.0 (91.0-95.0)	93.0 (91.00-96.0)	.85	97.0 (95.6-98.4)	97.1 (95.34-98.63)	.001
Systolic blood pressure max (mm Hg)	91.0 (93.0-95.0)	93.0 (91.0-96.0)	.003	92.0 (89.0-94.0)	91.0 (86.3-94.0)	.15
Systolic blood pressure mean (mm Hg)	97.7 (96.5-98.7)	97.9 (96.7-99.0)	.14	100.0 (99.0-100.0)	100.0 (100.0-100.0)	.92
Temperature mean (°C)	36.7 (36.5-36.9)	36.72 (36.5-37.0)	.48	36.7 (36.6-37.0)	36.8 (36.9-36.5)	.007

^aMIMIC-IV Medical Information Mart for Intensive Care IV version 2.0.^beICU-CRD: eICU Collaborative Research Database.

Table 2. Characteristics of patients and controls from the development dataset for the first 24-hour model cohort: laboratory test results and comorbidities.

Patient characteristics	MIMIC-IV 2.0 ^a cohort			eICU-CRD ^b cohort		
	No delirium (n=1781)	Delirium (n=343)	<i>P</i> value	No delirium (n=1965)	Delirium (n=81)	<i>P</i> value
Laboratory results, median (IQR)						
Hemoglobin min (g/100 mL)	8.4 (7.5-9.6)	9.0 (7.9-10.3)	<.001	8.4 (7.3-9.4)	9.0 (7.9-10.3)	<.001
Hemoglobin max (g/100 mL)	11.5 (10.0-12.9)	11.4 (10.4-14.2)	.64	11.4 (10.4-12.9)	11.4 (10.4-20.1)	.62
Platelet min (10 ⁹ /L)	147.0 (111.0-198.0)	130.0 (95.0-179.0)	<.001	145.0 (109.0-197.0)	128.0 (96.0-178.0)	<.001
Platelet count max (10 ⁹ /L)	125.0 (151.0-245.0)	188.0 (149.0-242.0)	.57	128.0 (97.0-235.0)	186.0 (153.0-245.0)	.55
White blood cell min (10 ⁹ /L)	8.8 (6.3-11.7)	9.5 (6.9-12.3)	.12	9.2 (6.3-12.0)	9.6 (7.0-12.4)	.14
White blood cell max (10 ⁹ /L)	13.1 (10.0-17.3)	14.8 (11.3-19.6)	<.001	14.0 (10.0-17.3)	14.7 (11.4-20.0)	<.001
Lactate min (mmol/L)	1.2 (0.9-1.5)	1.2 (0.9-1.6)	.21	1.4 (1.0-2.1)	1.3 (0.9-1.9)	.18
Urine output (mL)	1832.0 (1290.0-2617.0)	1575.0 (1002-2321.0)	<.001	1496.0 (796.3-2600.0)	1229.5 (600.0-2056.0)	<.001
Comorbidity, n (%)						
Myocardial infarction	30 (1.7)	3 (0.9)	.37	181 (9.2)	30 (6.8)	.11
Congestive heart failure	724 (40.7)	184 (53.6)	<.001	1251 (63.7)	51 (63)	.78
Peripheral vascular disease	281 (15.8)	84 (24.5)	<.001	14 (0.7)	1.0 (1.1)	.36
Cerebrovascular disease	186 (10.4)	65 (19.0)	<.001	90 (4.6)	6 (7.3)	.02
Chronic pulmonary disease	498 (28)	125 (36.4)	.002	303 (15.4)	11 (13.6)	.35
Renal disease	387 (21.7)	96 (28.0)	.01	321 (16.3)	13 (15.7)	.74
Diabetes with control	1538 (7.7)	30 (8.7)	.53	155 (7.9)	7 (8.5)	.58
Diabetes without control	495 (27.8)	97 (28.3)	.85	503 (30.2)	24 (29.5)	.84
Rheumatic disease	78 (4.4)	23 (6.7)	.06	82.0 (4.6)	5 (6.5)	.06
Peptic ulcer disease	13 (0.7)	7 (2)	.02	9.0 (0.5)	1 (1.5)	.02
Severe liver disease	16 (0.9)	7 (2)	.06	18.0 (1.0)	2 (2.3)	.07
Dementia	11 (0.6)	3 (0.9)	.59	13.0 (0.7)	1 (0.9)	.61
Paraplegia	15 (0.8)	12 (3.5)	<.001	16.0 (0.9)	3 (3.7)	<.001
Malignant cancer	103 (5.8)	16 (4.7)	.41	181.0 (9.2)	5 (6.8)	.11
AIDS	4 (0.2)	5 (3)	.38	1964.00 (99.0)	1 (1.2)	.04

^aMIMIC-IV Medical Information Mart for Intensive Care IV version 2.0.^beICU-CRD: eICU Collaborative Research Database.

Table 3. Characteristics of patients and controls from the development dataset for the first 24-hour model cohort: score, drug, and treatment measures.

Patient characteristics	MIMIC-IV 2.0 ^a cohort			eICU-CRD ^b cohort		
	No delirium (n=1781)	Delirium (n=343)	<i>P</i> value	No delirium (n=1965)	Delirium (n=81)	<i>P</i> value
Score, median (IQR)						
GCS ^c (min)	14.0 (14.0-15.0)	11.0 (6.0-14.0)	<.001	15.0 (14.0-15.00)	12.0 (8.0-14.0)	<.001
SOFA ^d	4.0 (2.0-7.0)	8.0 (6.0-11.0)	<.001	5.0 (3.0-7.0)	7.0 (5.0-10.0)	<.001
Drug, n (%)						
Acetaminophen	1005 (56.4)	164 (47.8)	.003	1104 (56.2)	41 (50.6)	.054
Anesthesia	38 (2.1)	5 (1.5)	.002	35 (1.9)	1 (1.2)	.02
Anticholinergics	1062 (59.6)	221 (64.4)	.10	387 (19.7)	19 (23.5)	.03
Anticoagulant	519 (29.1)	92 (26.8)	.39	539 (27.4)	36 (44.4)	.41
Antipsychotics	18 (1.0)	11 (3.2)	.001	21 (1.1)	17 (3.9)	<.001
Barbiturates	2.0 (0.1)	343 (100)	.54	3 (0.1)	81 (100)	.54
Benzodiazepines	138 (7.7)	32 (9.3)	.32	142 (8)	35 (9.5)	.34
Diuretics	694 (39)	131 (38.2)	.79	854 (43.5)	29 (35.5)	.008
Opioids	1040 (58.4)	201 (58.6)	.94	981 (49.9)	38 (46.9)	.28
Treatment measures, n (%)						
Emergency admission	1121 (62.9)	231 (67.3)	.12	380 (19.3)	17 (20.9)	.52
First care unit	1724 (96.8)	321 (93.6)	.004	1779 (90.5)	75 (92.6)	.04
Last care unit	42 (2.4)	16 (4.7)	.02	453 (23)	36 (44.4)	<.001
Renal Replacement Therapy	15 (0.8)	11 (3.2)	<.001	719 (36.6)	46 (56.8)	.047
Invasive ventilation	1179 (66.2)	298 (86.9)	<.001	1699 (86.5)	77 (95.1)	<.001
Length of ICU ^e stay, median (IQR)	2.4 (2.0-3.4)	5.9 (4.0-10.3)	<.001	3.20 (3.0-5.5)	6.7 (5.3-12.5)	<.001

^aMIMIC-IV Medical Information Mart for Intensive Care IV version 2.0.^beICU-CRD: eICU Collaborative Research Database.^cGCS: Glasgow Coma Scale.^dSOFA: Sequential Organ Failure Assessment.^eICU: intensive care unit.

Model Performance Evaluation

Using 4 ML algorithms, we developed predictive models to assess the risk of postoperative delirium in cardiac surgery patients, leveraging electronic health record data for early identification of high-risk individuals. Figure 2 presents ROC curves of all models, allowing a systematic comparison of their discriminative performance. Among the models, XGB demonstrated the best overall predictive performance, achieving the highest AUC for identifying patients at risk of delirium. The random forest classifiers also exhibited strong performance, although slightly lower than that of the XGB model. Notably, both XGB and random forest classifiers maintained high predictive accuracy, indicating good model generalizability. In contrast, support vector classifier and logistic regression models

showed substantially lower discriminative power. To further validate the clinical applicability of the XGB model, model performance was assessed using several evaluation metrics, including accuracy, sensitivity, positive predictive value, and negative predictive value (Table 4). Furthermore, the corresponding confusion matrices illustrating these metrics are displayed in Figure 3.

The model was externally validated using the eICU-CRD, a large-scale critical care dataset incorporating records from 208 hospitals, to assess its performance on independent data. The XGB model maintained strong discriminative performance in ROC analysis, with high AUC values confirming its reliability across institutions (Figure 4). The model was further rigorously validated through precision-recall analysis and calibration curves (Figures 5A and B).

Figure 2. Receiver operating characteristic curves of different machine learning algorithms evaluated on the internal validation set. LR: logistic regression; SVC: support vector classifier; RFC: random forest classifier; XGB: extreme gradient boosting.

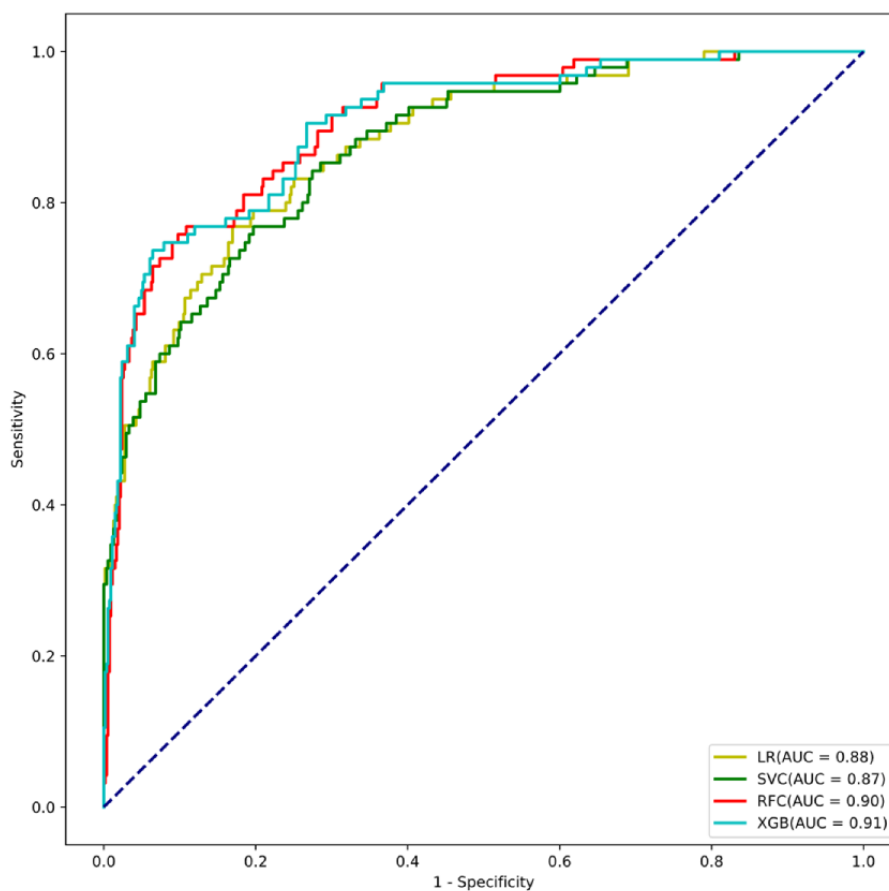


Table 4. Test set evaluation of machine learning model performance.

Model	Accuracy	Sensitivity	AUC ^a	PPV ^b	NPV ^c	F_1 -score
LR ^d	0.82	0.73	0.88	0.44	0.95	0.55
XGB ^e	0.83	0.77	0.91	0.47	0.95	0.58
SVC ^f	0.83	0.67	0.87	0.45	0.94	0.54
RFC ^g	0.88	0.75	0.90	0.58	0.95	0.65

^aAUC: area under the receiver operating characteristic curve.

^bPPV: positive predictive value.

^cNPV: negative predictive value.

^dLR: logistic regression.

^eXGB: extreme boosting gradient.

^fSVC: support vector classifier.

^gRFC: random forest classifier.

Figure 3. Confusion matrix for binary classification. (A) Logistic regression (LR), (B) extreme gradient boosting (XGB), (C) support vector classifier (SVC), (D) random forest classifier (RFC).

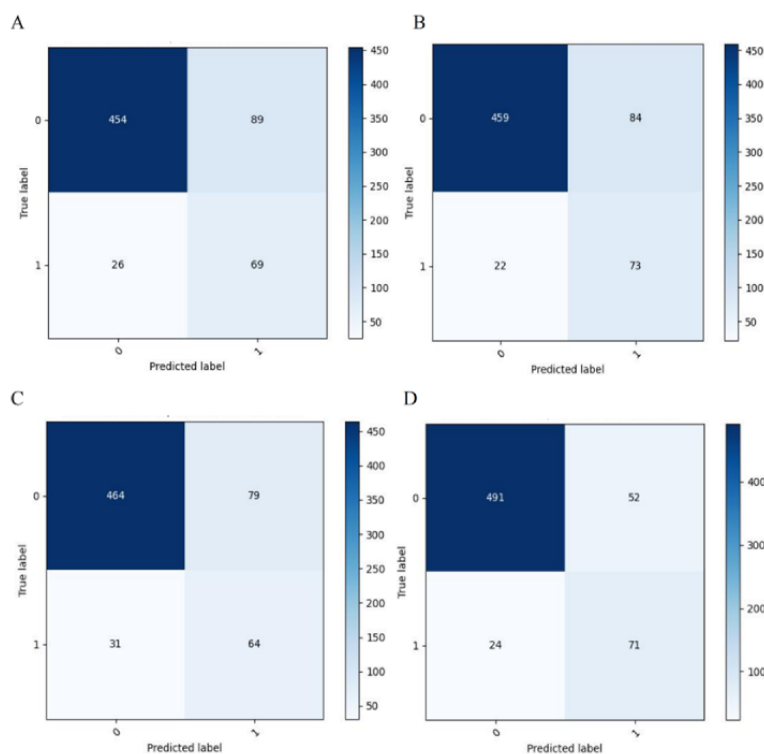


Figure 4. Receiver operating characteristic curves of various machine learning models evaluated on the external validation cohort. LR: logistic regression; RFC: random forest classifier; SVC: support vector classifier; XGB: extreme gradient boosting.

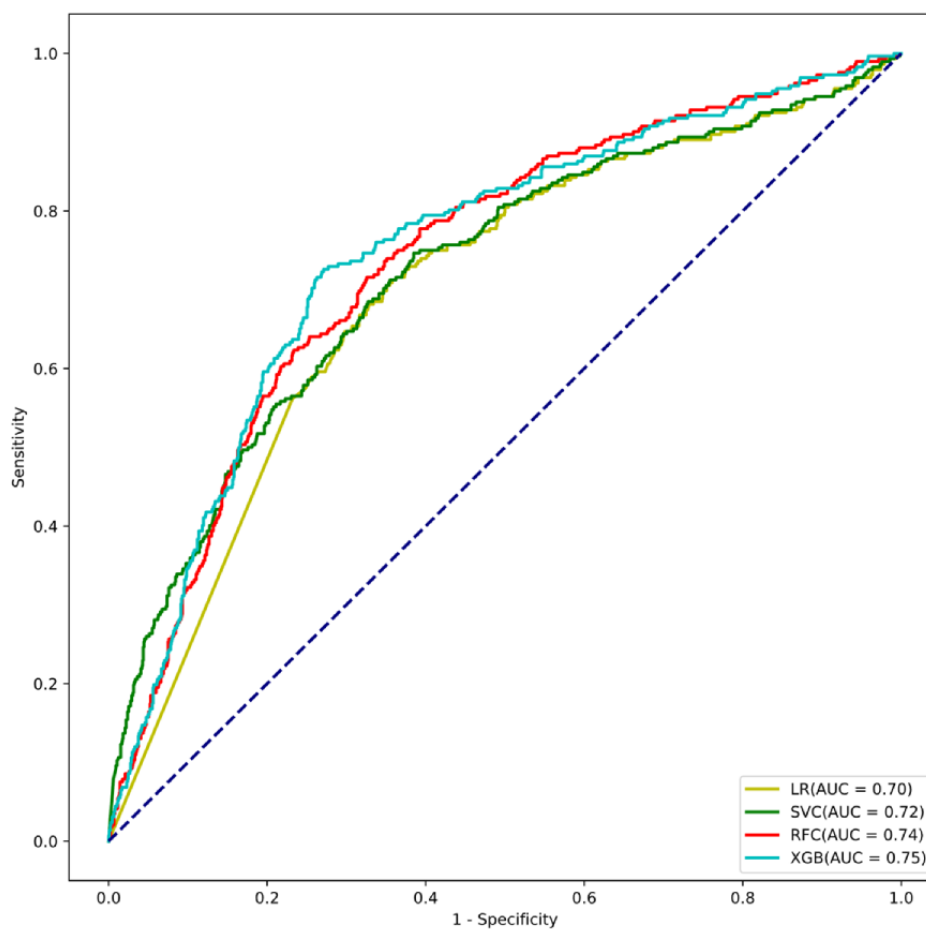
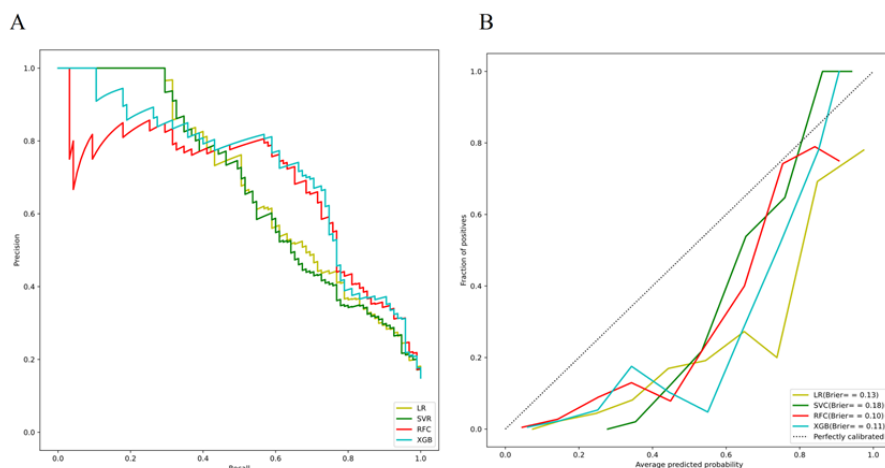


Figure 5. (A) Precision-recall curve. (B) Calibration curves and calculated Brier scores. LR: logistic regression; RFC: random forest classifier; SVC: support vector classifier; XGB: extreme gradient boosting.



Variable Importance

The predictive contribution of each variable to postoperative delirium varied among cardiac surgery ICU patients. The XGB algorithm was used to estimate variable importance and highlight the top predictors influencing model accuracy. The top predictors included length of stay in ICU, lowest GCS score,

age, mean blood pressure, SOFA score, weight, heart rate and urine output (Figure 6). To further interpret the influence and directionality of these features on model predictions, we generated SHAP summary plots (Figure 7). These visualizations provide a detailed explanation of how individual features contributed to the model output, offering insights into their relative impact and potential clinical relevance.

Figure 6. Variable contribution rankings estimated using the XGB model. GCS: Glasgow Coma Scale; ICU_los: the length time of ICU; MBP: mean blood pressure; SOFA: Sequential Organ Failure Assessment.

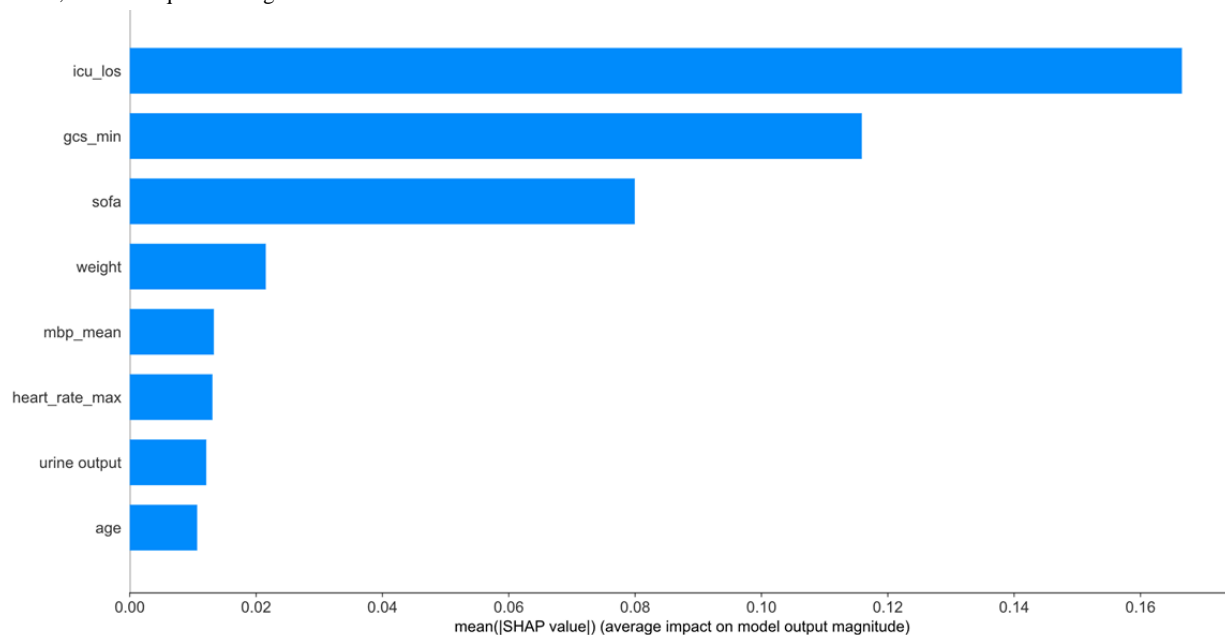
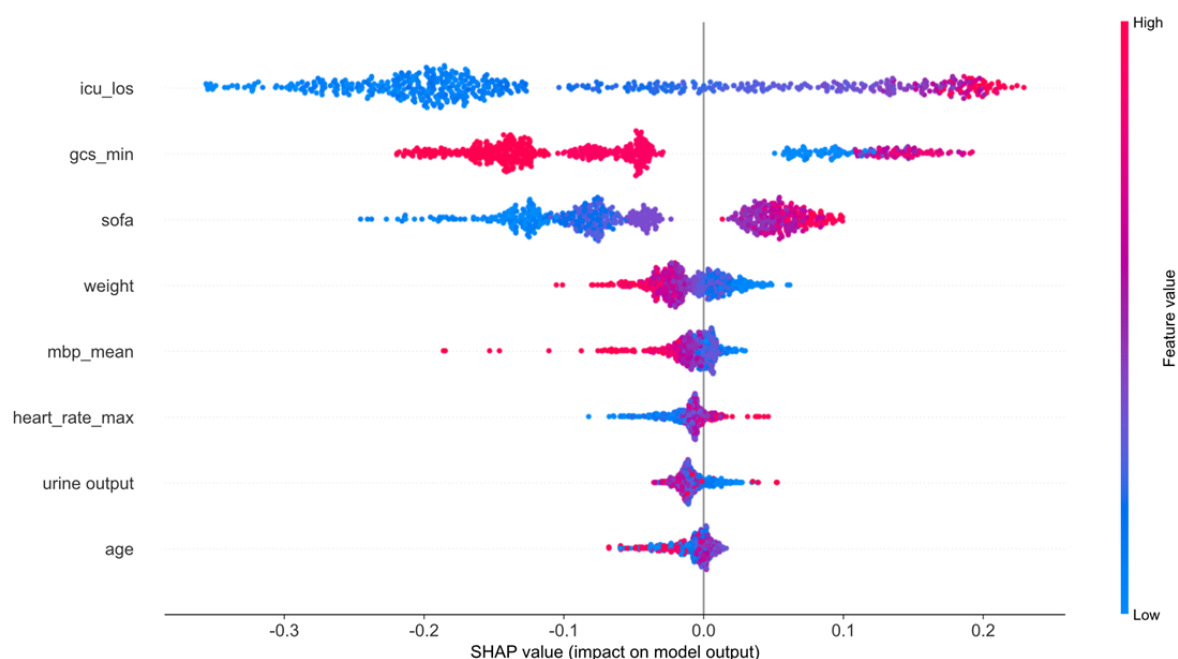


Figure 7. Shapley Additive Explanations (SHAP) summary plots illustrating feature contributions in the extreme gradient boosting (XGB) model. GCS: Glasgow Coma Scale; ICU_los: length of stay in the intensive care unit; MBP: mean blood pressure; SOFA: Sequential Organ Failure Assessment.



Discussion

Principal Findings

In this large-scale retrospective study, we developed and validated an ML model to predict postoperative delirium, a common and burdensome complication after cardiac surgery, with reported incidence rates ranging from 18% to 52% [33-35]. The XGB-based model, trained on the MIMIC-IV 2.0 dataset, demonstrated excellent performance in internal validation (AUC=0.91) and maintained strong discriminative power in external validation using the eICU-CRD dataset (AUC=0.75). Key predictors identified by the model included length of stay in ICU, min GCS score, mean blood pressure, SOFA score, weight, urine output, heart rate, and age. As far as we are aware, this study presents one of the earliest ML models that are developed to predict delirium in a diverse population of cardiac surgery patients. Importantly, our model relies on routinely available clinical data from the first 24 hours of ICU admission, enabling early identification of high-risk patients throughout their ICU stay. This early-warning capability offers a critical window for timely intervention, allowing clinicians to implement targeted strategies at the earliest stages of delirium onset. The model's high temporal relevance and wide applicability support its potential use in guiding personalized care plans and improving patient outcomes.

Comparison With Previous Work

Among existing tools for predicting postoperative cognitive complications, the CAM-ICU score remains the most widely used [36-38]. However, its applicability may be limited in certain surgical contexts or patient subgroups. Some individual populations require more tailored analysis, and it is important to note that CAM-ICU may not capture the full spectrum of delirium severity, potentially compromising its construct validity

[12,39,40]. Therefore, balancing feasibility and validity in specific clinical settings is essential. Our ML approach addresses these limitations by offering a data-driven, individualized risk assessment that adapts to patient heterogeneity and enables broader clinical applicability.

Furthermore, this study incorporated SHAP to enhance the interpretability of the XGB model. SHAP provides a transparent and traceable explanation framework for individualized risk prediction of delirium, allowing visualization of how each feature influences the model's output in both direction and magnitude [41,42]. This helps clinicians better understand the rationale behind specific predictions. Through SHAP analysis, we identified a set of clinically relevant predictors closely associated with patient deterioration, including length of stay in ICU, lowest GCS score, age, mean blood pressure, SOFA score, weight, urine output, and heart rate.

The XGB model demonstrated superior predictive performance, its clinical value extends beyond accuracy—it lies in its ability to inform real-time decision-making. Future research should focus on integrating these key features into dynamic clinical monitoring and intervention workflows. For instance, early abnormalities in high-impact variables, such as lowest GCS score, mean blood pressure, or urine output could trigger automated alerts within electronic health record systems, enabling real-time risk stratification. Additionally, individualized SHAP-based explanations could support patient-centered clinical strategies, facilitating a closed-loop “early warning–targeted intervention” model. This approach may ultimately enhance early prevention and improve outcomes in postoperative delirium management.

Strengths and Limitations

This study has several notable strengths. First, model development was based on large scale, real-world data from 2

publicly available critical care databases (MIMIC-IV 2.0 and eICU-CRD), which offer extensive clinical information and robust sample sizes, thereby increasing the trustworthiness and general applicability of our findings. Second, as far as we are aware, this is the earliest known study to apply ML techniques for predicting postoperative delirium in a broad cardiac surgery cohort. This tailored approach enables more accurate risk stratification by accounting for individual variability across surgical subtypes. One notable strength of our model lies in its exclusive use of routinely available clinical data obtained within the first day of ICU stay, eliminating the need for additional or specialized testing. The use of early time-series features to quantify dynamic changes in key physiological indicators allows real-time risk estimation with high clinical practicality and timeliness. From a translational perspective, the model demonstrated a sensitivity of 77% in identifying high-risk patients early, providing a valuable window for initiating preventive interventions. These proactive measures have the potential to reduce ICU length of stay, lower per-patient hospitalization costs, and minimize readmission risk, ultimately interrupting the cascade of adverse outcomes often associated with postoperative delirium. Finally, by focusing on the first day of ICU data, the model creates a strategic opportunity for timely clinical action, supporting the implementation of early, targeted strategies aimed at mitigating symptom progression and reducing complication rates.

Our study has several limitations. First, we acknowledge that our findings may be racially biased against White patients and the applicability to other populations may be limited due to the database being derived from Western countries. Second, the use of open public databases to obtain data may introduce missing data bias, which is unavoidable. Third, as this study is

retrospective and observational in nature, selection bias is inevitable. Specifically, the collection of retrospective data depends on existing medical records that may vary in completeness and accuracy due to differences in clinical documentation practices and data entry habits. Such inconsistencies can result in missing or inaccurate key clinical information, which in turn affects the quality of data used for model training. Moreover, retrospective studies lack the ability to actively intervene in or control the research process, making it difficult to eliminate the influence of confounding variables. This limitation weakens the strength of causal inferences drawn from the analysis. Additionally, our model has not yet undergone prospective clinical validation, which limits its current applicability in real-world clinical practice. Prospective validation, ideally conducted under well-controlled trial conditions and standardized workflows, is essential for evaluating the predictive performance and robustness of the model in diverse clinical settings. Without this step, there remains a risk that the model's actual performance may deviate from the retrospective findings, particularly when applied to different health care systems or patient populations.

Conclusions

We constructed and assessed an effective predictive model targeting postoperative delirium in patients admitted to the ICU after cardiac surgery. This model leverages routinely available patient information from the initial 24-hour ICU stay to estimate the risk of delirium throughout the hospitalization period. Given its reliance on readily available early-stage variables, the model has the potential to serve as a practical and accessible risk stratification tool for health care professionals and patients when selecting optimal cardiac treatment strategies.

Acknowledgments

The authors gratefully acknowledge the institutional review boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology for providing authorization to use data from the Medical Information Mart for Intensive Care IV version 2.0 database and eICU Collaborative Research Database.

Funding

The authors have not declared a specific grant for this research from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HH contributed to study design, data extraction, methods, data analysis, and manuscript preparation. YW, QZ, JW, and LL contributed to data interpretation and data analysis. HL, JH, and YZ contributed to data extraction and data interpretation. YY and ZX directed the project. YY is the guarantor of this study.

Conflicts of Interest

None declared.

References

1. Li T, Li J, Yuan L, Wu J, Jiang C, Daniels J, et al. RAGA Study Investigators. Effect of regional vs general anesthesia on incidence of postoperative delirium in older patients undergoing hip fracture surgery: the RAGA randomized trial. *JAMA*. Jan 04, 2022;327(1):50-58. [FREE Full text] [doi: [10.1001/jama.2021.22647](https://doi.org/10.1001/jama.2021.22647)] [Medline: [34928310](https://pubmed.ncbi.nlm.nih.gov/34928310/)]
2. Chen M, Zhang L, Shao M, Du J, Xiao Y, Zhang F, et al. E4BP4 coordinates circadian control of cognition in delirium. *Adv Sci (Weinh)*. Jun 17, 2022;9(23):e2200559. [FREE Full text] [doi: [10.1002/advs.202200559](https://doi.org/10.1002/advs.202200559)] [Medline: [35713240](https://pubmed.ncbi.nlm.nih.gov/35713240/)]
3. Wu YC, Tseng PT, Tu YK, Hsu CY, Liang CS, Yeh TC, et al. Association of delirium response and safety of pharmacological interventions for the management and prevention of delirium: a network meta-analysis. *JAMA Psychiatry*. May 01, 2019;76(5):526-535. [FREE Full text] [doi: [10.1001/jamapsychiatry.2018.4365](https://doi.org/10.1001/jamapsychiatry.2018.4365)] [Medline: [30810723](https://pubmed.ncbi.nlm.nih.gov/30810723/)]
4. Brown KN, Soo A, Faris P, Patten SB, Fiest KM, Stelfox HT. Association between delirium in the intensive care unit and subsequent neuropsychiatric disorders. *Crit Care*. Jul 31, 2020;24(1):476. [FREE Full text] [doi: [10.1186/s13054-020-03193-x](https://doi.org/10.1186/s13054-020-03193-x)] [Medline: [32736572](https://pubmed.ncbi.nlm.nih.gov/32736572/)]
5. Tschernatsch M, Juenemann M, Alhaidar F, El Shazly J, Butz M, Meyer M, et al. Epileptic seizure discharges in patients after open chamber cardiac surgery-a prospective prevalence pilot study using continuous electroencephalography. *Intensive Care Med*. Jul 2020;46(7):1418-1424. [FREE Full text] [doi: [10.1007/s00134-020-06073-8](https://doi.org/10.1007/s00134-020-06073-8)] [Medline: [32405742](https://pubmed.ncbi.nlm.nih.gov/32405742/)]
6. Li Q, Shen J, Lv H, Liu Y, Chen Y, Zhou C, et al. Incidence, risk factors, and outcomes in electroencephalographic seizures after mechanical circulatory support: a systematic review and meta-analysis. *Front Cardiovasc Med*. Aug 03, 2022;9:872005. [FREE Full text] [doi: [10.3389/fcvm.2022.872005](https://doi.org/10.3389/fcvm.2022.872005)] [Medline: [35990978](https://pubmed.ncbi.nlm.nih.gov/35990978/)]
7. Hawkins RB, Mehaffey JH, Guo A, Charles EJ, Speir AM, Rich JB, et al. Virginia Cardiac Services Quality Initiative. Postoperative atrial fibrillation is associated with increased morbidity and resource utilization after left ventricular assist device placement. *J Thorac Cardiovasc Surg*. Oct 2018;156(4):1543-9.e4. [FREE Full text] [doi: [10.1016/j.jtcvs.2018.03.169](https://doi.org/10.1016/j.jtcvs.2018.03.169)] [Medline: [29801690](https://pubmed.ncbi.nlm.nih.gov/29801690/)]
8. Sáez de Asteasu ML, Martínez-Velilla N, Zambom-Ferraresi F, Casas-Herrero Á, Cadore EL, Galbete A, et al. Assessing the impact of physical exercise on cognitive function in older medical patients during acute hospitalization: secondary analysis of a randomized trial. *PLoS Med*. Jul 05, 2019;16(7):e1002852. [FREE Full text] [doi: [10.1371/journal.pmed.1002852](https://doi.org/10.1371/journal.pmed.1002852)] [Medline: [31276501](https://pubmed.ncbi.nlm.nih.gov/31276501/)]
9. Fan G, Zhong M, Su W, An Z, Zhu Y, Chen C, et al. Effect of different anesthetic modalities on postoperative delirium in elderly hip fractures: a meta-analysis. *Medicine (Baltimore)*. Jun 07, 2024;103(23):e38418. [FREE Full text] [doi: [10.1097/MD.00000000000038418](https://doi.org/10.1097/MD.00000000000038418)] [Medline: [38847680](https://pubmed.ncbi.nlm.nih.gov/38847680/)]
10. Pun BT, Badenes R, Heras La Calle G, Orun OM, Chen W, Raman R, et al. COVID-19 Intensive Care International Study Group. Prevalence and risk factors for delirium in critically ill patients with COVID-19 (COVID-D): a multicentre cohort study. *Lancet Respir Med*. Mar 2021;9(3):239-250. [FREE Full text] [doi: [10.1016/S2213-2600\(20\)30552-X](https://doi.org/10.1016/S2213-2600(20)30552-X)] [Medline: [33428871](https://pubmed.ncbi.nlm.nih.gov/33428871/)]
11. Berger M, Terrando N, Smith SK, Browndyke JN, Newman MF, Mathew JP. Neurocognitive function after cardiac surgery: from phenotypes to mechanisms. *Anesthesiology*. Oct 2018;129(4):829-851. [FREE Full text] [doi: [10.1097/ALN.0000000000002194](https://doi.org/10.1097/ALN.0000000000002194)] [Medline: [29621031](https://pubmed.ncbi.nlm.nih.gov/29621031/)]
12. Young M, Holmes N, Kishore K, Marhoon N, Amjad S, Serpa-Neto A, et al. Natural language processing diagnosed behavioral disturbance vs confusion assessment method for the intensive care unit: prevalence, patient characteristics, overlap, and association with treatment and outcome. *Intensive Care Med*. May 2022;48(5):559-569. [FREE Full text] [doi: [10.1007/s00134-022-06650-z](https://doi.org/10.1007/s00134-022-06650-z)] [Medline: [35322288](https://pubmed.ncbi.nlm.nih.gov/35322288/)]
13. Li RC, Asch SM, Shah NH. Developing a delivery science for artificial intelligence in healthcare. *NPJ Digit Med*. Aug 21, 2020;3:107. [FREE Full text] [doi: [10.1038/s41746-020-00318-y](https://doi.org/10.1038/s41746-020-00318-y)] [Medline: [32885053](https://pubmed.ncbi.nlm.nih.gov/32885053/)]
14. Talpur N, Abdulkadir SJ, Alhussian H, Hasan MH, Aziz N, Bamhdi A. Deep Neuro-Fuzzy System application trends, challenges, and future perspectives: a systematic survey. *Artif Intell Rev*. 2023;56(2):865-913. [FREE Full text] [doi: [10.1007/s10462-022-10188-3](https://doi.org/10.1007/s10462-022-10188-3)] [Medline: [35431395](https://pubmed.ncbi.nlm.nih.gov/35431395/)]
15. Deeken F, Sánchez A, Rapp MA, Denkinger M, Brefka S, Spank J, et al. PAWEL Study Group. Outcomes of a delirium prevention program in older persons after elective surgery: a stepped-wedge cluster randomized clinical trial. *JAMA Surg*. Feb 01, 2022;157(2):e216370. [FREE Full text] [doi: [10.1001/jamasurg.2021.6370](https://doi.org/10.1001/jamasurg.2021.6370)] [Medline: [34910080](https://pubmed.ncbi.nlm.nih.gov/34910080/)]
16. Humeidan ML, Reyes JC, Mavarez-Martinez A, Roeth C, Nguyen CM, Sheridan E, et al. Effect of cognitive prehabilitation on the incidence of postoperative delirium among older adults undergoing major noncardiac surgery: the neurobics randomized clinical trial. *JAMA Surg*. Feb 01, 2021;156(2):148-156. [FREE Full text] [doi: [10.1001/jamasurg.2020.4371](https://doi.org/10.1001/jamasurg.2020.4371)] [Medline: [33175114](https://pubmed.ncbi.nlm.nih.gov/33175114/)]
17. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open*. Mar 01, 2021;4(3):e212240. [FREE Full text] [doi: [10.1001/jamanetworkopen.2021.2240](https://doi.org/10.1001/jamanetworkopen.2021.2240)] [Medline: [33783520](https://pubmed.ncbi.nlm.nih.gov/33783520/)]
18. Heilbroner SP, Few R, Mueller J, Chalwa J, Charest F, Suryadevara S, et al. Predicting cardiac adverse events in patients receiving immune checkpoint inhibitors: a machine learning approach. *J Immunother Cancer*. Oct 2021;9(10):e002545. [FREE Full text] [doi: [10.1136/jitc-2021-002545](https://doi.org/10.1136/jitc-2021-002545)] [Medline: [34607896](https://pubmed.ncbi.nlm.nih.gov/34607896/)]
19. Dara S, Dhamercherla S, Jadav SS, Babu CM, Ahsan MJ. Machine learning in drug discovery: a review. *Artif Intell Rev*. 2022;55(3):1947-1999. [FREE Full text] [doi: [10.1007/s10462-021-10058-4](https://doi.org/10.1007/s10462-021-10058-4)] [Medline: [34393317](https://pubmed.ncbi.nlm.nih.gov/34393317/)]

20. Zhong Y, Brooks MM, Kennedy EH, Bodnar LM, Naimi AI. Use of machine learning to estimate the per-protocol effect of low-dose aspirin on pregnancy outcomes: a secondary analysis of a randomized clinical trial. *JAMA Netw Open*. Mar 01, 2022;5(3):e2143414. [FREE Full text] [doi: [10.1001/jamanetworkopen.2021.43414](https://doi.org/10.1001/jamanetworkopen.2021.43414)] [Medline: [35262718](https://pubmed.ncbi.nlm.nih.gov/35262718/)]
21. Sumner SA, Bowen D, Holland K, Zwald ML, Vivolo-Kantor A, Guy GP, et al. Estimating weekly national opioid overdose deaths in near real time using multiple proxy data sources. *JAMA Netw Open*. Jul 01, 2022;5(7):e2223033. [FREE Full text] [doi: [10.1001/jamanetworkopen.2022.23033](https://doi.org/10.1001/jamanetworkopen.2022.23033)] [Medline: [35862045](https://pubmed.ncbi.nlm.nih.gov/35862045/)]
22. Nistal-Nuño B. Machine learning applied to a cardiac surgery recovery unit and to a coronary care unit for mortality prediction. *J Clin Monit Comput*. Jun 2022;36(3):751-763. [doi: [10.1007/s10877-021-00703-2](https://doi.org/10.1007/s10877-021-00703-2)] [Medline: [33860407](https://pubmed.ncbi.nlm.nih.gov/33860407/)]
23. Zhang L, Xu F, Han D, Huang T, Li S, Yin H, et al. Influence of the trajectory of the urine output for 24 h on the occurrence of AKI in patients with sepsis in intensive care unit. *J Transl Med*. Dec 20, 2021;19(1):518. [FREE Full text] [doi: [10.1186/s12967-021-03190-w](https://doi.org/10.1186/s12967-021-03190-w)] [Medline: [34930308](https://pubmed.ncbi.nlm.nih.gov/34930308/)]
24. Shi Y, Stornetta DS, Reklow RJ, Sahu A, Wabara Y, Nguyen A, et al. A brainstem peptide system activated at birth protects postnatal breathing. *Nature*. Jan 2021;589(7842):426-430. [FREE Full text] [doi: [10.1038/s41586-020-2991-4](https://doi.org/10.1038/s41586-020-2991-4)] [Medline: [33268898](https://pubmed.ncbi.nlm.nih.gov/33268898/)]
25. Davidson S, Villarroel M, Harford M, Finnegan E, Jorge J, Young D, et al. Day-to-day progression of vital-sign circadian rhythms in the intensive care unit. *Crit Care*. Apr 22, 2021;25(1):156. [FREE Full text] [doi: [10.1186/s13054-021-03574-w](https://doi.org/10.1186/s13054-021-03574-w)] [Medline: [33888129](https://pubmed.ncbi.nlm.nih.gov/33888129/)]
26. Murphy MS, Ducharme R, Hawken S, Corsi DJ, Petreich W, El-Chaâr D, et al. Exposure to intrapartum epidural analgesia and risk of autism spectrum disorder in offspring. *JAMA Netw Open*. May 02, 2022;5(5):e2214273. [FREE Full text] [doi: [10.1001/jamanetworkopen.2022.14273](https://doi.org/10.1001/jamanetworkopen.2022.14273)] [Medline: [35616937](https://pubmed.ncbi.nlm.nih.gov/35616937/)]
27. Proix T, Delgado Saa J, Christen A, Martin S, Pasley BN, Knight RT, et al. Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nat Commun*. Jan 10, 2022;13(1):48. [FREE Full text] [doi: [10.1038/s41467-021-27725-3](https://doi.org/10.1038/s41467-021-27725-3)] [Medline: [35013268](https://pubmed.ncbi.nlm.nih.gov/35013268/)]
28. Schober P, Vetter TR. Logistic regression in medical research. *Anesth Analg*. Jan 14, 2021;132(2):365-366. [FREE Full text] [doi: [10.1213/ANE.0000000000005247](https://doi.org/10.1213/ANE.0000000000005247)] [Medline: [33449558](https://pubmed.ncbi.nlm.nih.gov/33449558/)]
29. Rhodes JS, Cutler A, Moon KR. Geometry- and accuracy-preserving random forest proximities. *IEEE Trans Pattern Anal Mach Intell*. Sep 2023;45(9):10947-10959. [doi: [10.1109/TPAMI.2023.3263774](https://doi.org/10.1109/TPAMI.2023.3263774)] [Medline: [37015125](https://pubmed.ncbi.nlm.nih.gov/37015125/)]
30. Li Q, He Y, Pan J. CrossFuse-XGBoost: accurate prediction of the maximum recommended daily dose through multi-feature fusion, cross-validation screening and extreme gradient boosting. *Brief Bioinform*. Nov 22, 2023;25(1):bbad511. [FREE Full text] [doi: [10.1093/bib/bbad511](https://doi.org/10.1093/bib/bbad511)] [Medline: [38216539](https://pubmed.ncbi.nlm.nih.gov/38216539/)]
31. Que Z, Lin CJ. One-class SVM probabilistic outputs. *IEEE Trans Neural Netw Learn Syst*. Apr 2025;36(4):6244-6256. [doi: [10.1109/TNNLS.2024.3395148](https://doi.org/10.1109/TNNLS.2024.3395148)] [Medline: [38717884](https://pubmed.ncbi.nlm.nih.gov/38717884/)]
32. Chen Y, Ou Z, Zhou D, Wu X. Advancements and challenges of artificial intelligence-assisted electroencephalography in epilepsy management. *J Clin Med*. Jun 16, 2025;14(12):4270. [FREE Full text] [doi: [10.3390/jcm14124270](https://doi.org/10.3390/jcm14124270)] [Medline: [40566015](https://pubmed.ncbi.nlm.nih.gov/40566015/)]
33. Wang YY, Yue JR, Xie DM, Carter P, Li QL, Gartaganis SL, et al. Effect of the tailored, family-involved hospital elder life program on postoperative delirium and function in older adults: a randomized clinical trial. *JAMA Intern Med*. Jan 01, 2020;180(1):17-25. [FREE Full text] [doi: [10.1001/jamainternmed.2019.4446](https://doi.org/10.1001/jamainternmed.2019.4446)] [Medline: [31633738](https://pubmed.ncbi.nlm.nih.gov/31633738/)]
34. Brown CH, Neufeld KJ, Tian J, Probert J, LaFlam A, Max L, Cerebral Autoregulation Study Group, et al. Effect of targeting mean arterial pressure during cardiopulmonary bypass by monitoring cerebral autoregulation on postsurgical delirium among older patients: a nested randomized clinical trial. *JAMA Surg*. Sep 01, 2019;154(9):819-826. [FREE Full text] [doi: [10.1001/jamasurg.2019.1163](https://doi.org/10.1001/jamasurg.2019.1163)] [Medline: [31116358](https://pubmed.ncbi.nlm.nih.gov/31116358/)]
35. Leone M, Einav S, Chiumello D, Constantin JM, De Robertis E, De Abreu MG, et al. Guideline contributors. Noninvasive respiratory support in the hypoxaemic peri-operative/periprocedural patient: a joint ESA/ESICM guideline. *Intensive Care Med*. Apr 2020;46(4):697-713. [FREE Full text] [doi: [10.1007/s00134-020-05948-0](https://doi.org/10.1007/s00134-020-05948-0)] [Medline: [32157356](https://pubmed.ncbi.nlm.nih.gov/32157356/)]
36. Gélinas C, Bérubé M, Puntillo KA, Boitor M, Richard-Lalonde M, Bernard F, et al. Validation of the critical-care pain observation tool-neuro in brain-injured adults in the intensive care unit: a prospective cohort study. *Crit Care*. Apr 13, 2021;25(1):142. [FREE Full text] [doi: [10.1186/s13054-021-03561-1](https://doi.org/10.1186/s13054-021-03561-1)] [Medline: [33849619](https://pubmed.ncbi.nlm.nih.gov/33849619/)]
37. Honarmand K, Lalli RS, Priestap F, Chen JL, McIntyre CW, Owen AM, et al. Natural history of cognitive impairment in critical illness survivors. A systematic review. *Am J Respir Crit Care Med*. Jul 15, 2020;202(2):193-201. [FREE Full text] [doi: [10.1164/rccm.201904-0816CI](https://doi.org/10.1164/rccm.201904-0816CI)] [Medline: [32078780](https://pubmed.ncbi.nlm.nih.gov/32078780/)]
38. Rosas IO, Bräu N, Waters M, Go RC, Malhotra A, Hunter BD, et al. Tocilizumab in patients hospitalised with COVID-19 pneumonia: efficacy, safety, viral clearance, and antibody response from a randomised controlled trial (COVACTA). *EClinicalMedicine*. May 2022;47:101409. [FREE Full text] [doi: [10.1016/j.eclinm.2022.101409](https://doi.org/10.1016/j.eclinm.2022.101409)] [Medline: [35475258](https://pubmed.ncbi.nlm.nih.gov/35475258/)]
39. Madrigal C, Kim J, Jiang L, Lafo J, Bozzay M, Primack J, et al. Delirium and functional recovery in patients discharged to skilled nursing facilities after hospitalization for heart failure. *JAMA Netw Open*. Mar 01, 2021;4(3):e2037968. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.37968](https://doi.org/10.1001/jamanetworkopen.2020.37968)] [Medline: [33724390](https://pubmed.ncbi.nlm.nih.gov/33724390/)]

40. Saczynski JS, Koethe B, Fick DM, Vo QT, Devlin JW, Marcantonio ER, et al. Cognitive and functional change in skilled nursing facilities: differences by delirium and Alzheimer's disease and related dementias. *J Am Geriatr Soc*. Nov 2024;72(11):3501-3509. [doi: [10.1111/jgs.19112](https://doi.org/10.1111/jgs.19112)] [Medline: [39171670](https://pubmed.ncbi.nlm.nih.gov/39171670/)]
41. Huang HL, Lee JY, Lo YS, Liu IH, Huang SH, Huang YW, et al. Whole-blood 3-gene signature as a decision aid for rifapentine-based tuberculosis preventive therapy. *Clin Infect Dis*. Sep 14, 2022;75(5):743-752. [FREE Full text] [doi: [10.1093/cid/ciac003](https://doi.org/10.1093/cid/ciac003)] [Medline: [34989801](https://pubmed.ncbi.nlm.nih.gov/34989801/)]
42. McKenna M, Shackelford D, Pontes C, Ball B, Nance E. Multiple particle tracking detects changes in brain extracellular matrix and predicts neurodevelopmental age. *ACS Nano*. May 25, 2021;15(5):8559-8573. [FREE Full text] [doi: [10.1021/acsnano.1c00394](https://doi.org/10.1021/acsnano.1c00394)] [Medline: [33969999](https://pubmed.ncbi.nlm.nih.gov/33969999/)]

Abbreviations

AUC: area under the receiver operating characteristic curve
CAM-ICU: Confusion Assessment Method for the Intensive Care Unit
eICU-CRD: eICU Collaborative Research Database
GCS: Glasgow Coma Scale
ICU: intensive care unit
MIMIC-IV 2.0: Medical Information Mart for Intensive Care IV version 2.0
ML: machine learning
ROC: receiver operating characteristic
SHAP: Shapley Additive Explanations
SOFA: Sequential Organ Failure Assessment
XGB: extreme gradient boosting

Edited by J Klann; submitted 01.Mar.2025; peer-reviewed by P Tekkali, A Özel; comments to author 30.Apr.2025; accepted 26.Sep.2025; published 11.Feb.2026

Please cite as:

Hu H, Wang Y, Li H, Zang Q, Huang J, Zhang Y, Wu J, Liu L, Xing Z, Yu Y

Early Prediction of Delirium in Postcardiac Surgery Patients: Machine Learning Model Development and External Validation

JMIR Med Inform 2026;14:e73283

URL: <https://medinform.jmir.org/2026/1/e73283>

doi: [10.2196/73283](https://doi.org/10.2196/73283)

PMID:

©Huixiu Hu, Yuxiang Wang, Houfeng Li, Qinglai Zang, Jing Huang, Ying Zhang, Jinjing Wu, Long Liu, Zhen Xing, Yaohua Yu. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 11.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org>, as well as this copyright and license information must be included.