

Development and validation of an interpretable machine learning model for predicting in-hospital mortality in patients with ventricular fibrillation

Chengdi Chen

Affiliated Hospital of Guangdong Medical College Hospital

Kaixiang Zhang

Jinjiang Municipal Hospital

Tongchun Zhong

Affiliated Hospital of Guangdong Medical College Hospital

Haochun Li

Affiliated Hospital of Guangdong Medical College Hospital

Zibei Feng

Affiliated Hospital of Guangdong Medical College Hospital

Zhijian Guo

Lanzhou University Second Hospital

Zhixiong Yang

Affiliated Hospital of Guangdong Medical College Hospital

Shian Huang

Affiliated Hospital of Guangdong Medical College Hospital

Lingpin Pang

panglingpin@126.com

Affiliated Hospital of Guangdong Medical College Hospital

Research Article

Keywords: Ventricular Fibrillation, Hospital Mortality, Machine Learning, MIMIC-IV database, SHAP

Posted Date: March 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9119180/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: Timely and accurate outcome prediction is essential for clinical decision-making in patients with ventricular fibrillation. However, the interpretation of these predictions and the translation of predictive models into clinical practice are equally crucial. This study aims to develop an interpretable machine learning (IML) model that effectively predicts in-hospital mortality for ventricular fibrillation patients.

Methods: In this study, 879 patients with ventricular fibrillation from the Medical Information Mart for Intensive Care-IV (MIMIC-IV) database were randomly assigned to a training set and a test set at a ratio of 7:3. After least absolute shrinkage and selection operator (LASSO) regression analysis and the Boruta method determined the modeling variables, seven machine learning (ML) algorithms were developed to predict in-hospital mortality for patients with ventricular fibrillation using these data and externally validated in two hospitals. The area under the ROC curve (AUC) of the receiver operating characteristic (ROC) curve was calculated to assess the performance of the seven models. Decision curve analysis (DCA) was conducted to evaluate the clinical utility of the three models by estimating the net benefit at a range of threshold probabilities. Based on performance, The SHapley Additive exPlanation (SHAP) algorithm attributes interpretability to the optimal prediction model.

Results: lactate, use of beta-blockers, anion gap, red blood cell count, blood urea nitrogen, hemoglobin, heart rate, respiratory rate, and albumin were selected as the nine most influential variables. The LR model demonstrated the most robust predictive performance, achieving AUROC values of 0.845 and 0.836 in the training set and test set, respectively. Furthermore, it achieved AUROC values of 0.794 and 0.903 in the two external validation sets, respectively. DCA showed that the model had the greatest net benefit rate when the prediction probability threshold is 0.15–0.45. The use of beta-blockers is the most important feature in the prediction process. The SHAP force plot provided a visualization of the direction and degree of influence of each feature on the predicting results of the model.

Conclusion: ML is a reliable tool for predicting in-hospital mortality in patients with ventricular fibrillation. SHAP methods were used to explain intrinsic information of the LR model, which may prove clinically useful and help clinicians tailor precise management.

1 Introduction

Ventricular fibrillation (VF) represents the most life-threatening cardiac arrhythmia and is a primary mechanism underlying sudden cardiac arrest[1]. Sudden cardiac death (SCD) is one of the leading causes of death worldwide. Similar to the proportion of sudden deaths in Western countries, including the United States, approximately 544,000 people die suddenly each year in China, with the vast majority of these deaths caused by ventricular fibrillation[2–4]. Despite advancements in early defibrillation, advanced life support, and emergency percutaneous coronary intervention, the in-hospital mortality rate for VF patients admitted to the intensive care unit (ICU) remains unacceptably high, ranging from 30% to 50% [5]. VF frequently results in multi-organ dysfunction, electrolyte disturbances, acid-base imbalances,

and severe neurological sequelae, collectively elevating mortality risk[6]. Therefore, early identification of high-risk patients and timely implementation of targeted interventions are crucial for improving prognosis and optimizing resource allocation.

The in-hospital mortality of patients with ventricular fibrillation (VF) is influenced by complex, nonlinear interactions among multiple physiological systems. To accurately quantify such risks, a modeling approach capable of effectively capturing and analyzing this highly heterogeneous interplay is required[7, 8]. The widespread adoption of hospital information systems and the exponential growth in computational power have accelerated the integration of machine-learning (ML) approaches into critical-care medicine. By exploiting high-dimensional data, ML algorithms can model non-linear relationships and latent interactions, frequently achieving superior discriminative performance compared with traditional scoring systems[9, 10]. Previous studies leveraging the Medical Information Mart for Intensive Care (MIMIC) database have developed ML models that outperform conventional scores in predicting cardiac arrest, shock, and in-hospital mortality[11, 12].

Sumeet S. Chugh[13] and colleagues have prospectively developed a model to predict the risk of out-of-hospital cardiac arrest (first-time ventricular fibrillation/pulseless ventricular tachycardia) in patients with coronary heart disease; there are also studies using EMS on-site indicators and machine learning to predict out-of-hospital refractory ventricular fibrillation[14]. However, predictive studies specifically addressing in-hospital ventricular fibrillation (VF) and its fatal outcomes remain scarce. Current research has primarily focused on generic cardiac arrest or composite arrhythmia cohorts, thereby leaving this high-risk VF subpopulation inadequately investigated. Moreover, most existing models operate as "black boxes," lacking the interpretability required to gain clinical trust and satisfy regulatory standards [15, 16], leaving this high-risk VF subpopulation inadequately investigated with interpretable and tailored predictive tools. Consequently, the development of an accurate, interpretable, and externally validated machine learning model tailored to VF patients is of critical clinical importance.

Although machine learning models have shown potential in predicting the prognosis of patients with ventricular fibrillation (VF), their 'black-box' nature severely limits clinical interpretability and practical value[17]. To address this critical issue, this study introduces the SHapley Additive exPlanations (SHAP) method to reveal the prediction logic by quantifying variable contributions, thereby enhancing model transparency[18]. This study used retrospective cohort data, training models based on the MIMIC-IV (v3.1) database, and externally validating them with independent data from intensive care units of two tertiary hospitals in Guangdong, China. By systematically comparing multiple machine learning algorithms, the optimal predictive model was selected, and the SHAP method was integrated to achieve visual interpretation of risk factors. This study aims to develop a decision-support tool for managing VF patients that integrates predictive accuracy with clinical interpretability, thereby facilitating early risk stratification and enabling personalized interventions.

2 Materials and methods

2.1 Data source

The primary data utilized in this study were obtained from the MIMIC-IV database (<https://physionet.org/content/mimiciv/0.3/>), a large-scale publicly accessible dataset developed and maintained by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology (MIT). The MIMIC-IV database contains comprehensive clinical information from patients admitted to the ICU of Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, USA, between 2008 and 2019. This multidimensional dataset includes demographic characteristics, medical histories, laboratory test results, medication administrations, and detailed clinical treatment processes, encompassing a heterogeneous population of ICU patients. Access to the database requires completion of the National Institutes of Health (NIH)-certified “Protecting Human Research Participants” online training course to ensure ethical compliance. The study author, Zhijian Guo, has successfully passed the Collaborative Institutional Training Initiative (CITI) examination and obtained database access authorization (Certification ID: 13754573). As all data in MIMIC-IV have undergone de-identification and the study protocol was approved by the Institutional Review Boards of both BIDMC and MIT, no additional patient informed consent was required. Additionally, this study incorporated de-identified clinical data from ICU patients hospitalized between June 2019 and January 2025 at two first-class tertiary hospitals in China: External validation cohorts were approved by the Ethics Committees of Affiliated Hospital of Guangdong Medical University (approval No. PJKT-2025-226) and the Second Affiliated Hospital of Guangdong Medical University (approval No. PJKT2023-0549). These datasets comprehensively document various clinical parameters during ICU hospitalization, with all data undergoing rigorous de-identification procedures in accordance with ethical standards to ensure confidentiality and information security. This study adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [19]. The MIMIC-IV dataset was de-identified in accordance with the Health Insurance Portability and Accountability Act (HIPAA), balancing patient privacy with research utility[20].

2.2 Participants

Following previous studies, we used SQL queries to identify adult patients with any ICU stay who had a recorded diagnosis of ventricular fibrillation (VF) coded as ICD-9 427.41 or ICD-10 I49.0/I49.01. Inclusion and Exclusion Criteria: Inclusion: Adults with any ICU admission and a confirmed VF diagnosis. Exclusion: No confirmed VF; missing hospital data exceeding 20%; multiple ICU admissions, retaining only the first admission. The main outcome is the mortality from admission to discharge. This study primarily focuses on in-hospital mortality. Internal Validation Cohort from MIMIC-IV database: Eligible patients were randomly divided in a 7:3 ratio into a training set ($n = 615$, for modeling) and an internal test set ($n = 264$, for performance evaluation). External Validation Cohort: From January 2019 to December 2024, a total of 328 patients were consecutively included from Guangdong Medical University First Affiliated Hospital ($n = 272$) and Second Affiliated Hospital ($n = 56$), meeting the same inclusion and exclusion criteria, forming the external validation set. In total, our study included 1,207 patients. The processes of data acquisition, screening, analysis, and statistical procedures for both cohorts are summarized in Fig. 1.

2.3 Selection of variables

To ensure the model's clinical relevance and parsimony, we performed variable selection based on a comprehensive pool of candidate variables. This initial variable grouping considered three aspects: the complexity of the MIMIC-IV database[21], the extent of missing relevant data, and the prognostic risk factors for ventricular fibrillation[22]. The baseline variables uniformly extracted from the MIMIC database can be summarized into six major categories, encompassing over 52 items, all extracted within 24 hours after initial ICU admission. These cover demographics, lifestyle, comorbidities, medication, vital signs, laboratory tests, and survival information from ICU admission to 28 days post-discharge, ensuring the completeness and comparability for subsequent modeling: Demographics and body metrics: age, gender, height, weight. Lifestyle and chronic disease background: smoking history, alcohol use, hypertension, diabetes, ischemic heart disease. Evidence-based medications before discharge: statins, β -blocker use. Vital signs: respiratory rate, heart rate, systolic/diastolic blood pressure, body temperature. Routine laboratory (venous blood) metabolism: blood glucose, creatinine, BUN, uric acid, Na^+ , K^+ , HCO_3^- , anion gap. Blood gas/perfusion: lactate. Hematology: WBC, RBC, hemoglobin, platelets, NEU#, LYM#, MONO#, EOS#. Coagulation: PT, PTT, INR, TT, D-dimer, fibrinogen. Liver function: ALT, AST, GGT, albumin. Lipid profile: total cholesterol, HDL-C, triglycerides. Myocardial injury: cTnT, CK, CK-MB, LDH, NT-proBNP. All variables were collected according to the 'first valid value' principle, and those with > 20% missing data were excluded in advance to ensure that subsequent machine learning or traditional regression models use the same variable pool during derivation, internal validation, and external validation. To develop a parsimonious and generalizable prediction model, this study implemented a dual feature selection strategy integrating LASSO regression and the Boruta algorithm[23–24].

First, Least Absolute Shrinkage and Selection Operator (LASSO) regression was applied to condense the initial predictor set. By incorporating an L1 penalty term into the loss function, this method drives the coefficients of non-informative variables to zero. The optimal regularization parameter (λ) was determined via 10-fold cross-validation. Following the one standard error rule, we selected the most penalizing λ within one standard error of the minimum cross-validated error, thereby prioritizing a sparser model with stronger generalizability. Concurrently, we employed the Boruta algorithm, a wrapper method rooted in random forest to perform an all-relevant feature selection. This algorithm iteratively compares the importance of original features against randomized shadow features, retaining only those variables that consistently demonstrate significance beyond random noise. The outcomes from both selection approaches were visualized using a Venn diagram to elucidate their consensus. The final feature set for predictive modeling was defined by the intersection of variables identified by both LASSO and Boruta. This consensus-based approach enhances the robustness of the selected predictors and strengthens the interpretability of the resulting model, while maintaining an optimal balance between simplicity and predictive power.

2.4 Data pre-processing

Data from MIMIC-IV and the two external hospitals were pre-processed for outliers and missingness. Outliers were flagged by the inter-quartile range (IQR) rule: values $> Q3 + 1.5 \times IQR$ or $< Q1 - 1.5 \times IQR$. Variables with $> 20\%$ missing values were discarded. For the remainder, multivariate imputation by chained equations (MICE) was applied separately to training and test sets to avoid information leakage. MICE used linear regression for continuous variables and logistic regression for binary variables, preserving distributional properties and minimising bias. After imputation, all continuous predictors were min–max normalised to the 0–1 range[26].

2.5 Construction of the machine learning model

Seven algorithms were evaluated: logistic regression (LR), decision tree (DT), support-vector machine (SVM), random forest (RF), XGBoost, LightGBM and artificial neural network (ANN). Nested cross-validation was employed to ensure generalisability and mitigate over-fitting. Optimal hyper-parameters were identified by grid or random search with five-fold cross-validation on the augmented training data.

2.6 Interpretation analysis

Shapley Additive exPlanations (SHAP) were used to quantify each variable's contribution to the predicted probability of in-hospital mortality[27]. By computing SHAP values for every patient, we ranked predictors according to their average absolute impact on model output. This approach yields a transparent, clinically interpretable overview of which factors most strongly drive the risk of death among ICU patients with ventricular fibrillation.

2.7 Statistical analyses

Quantitative data were summarized as mean $\pm$ SD, and independent samples t-test was utilized for comparison between two groups. For quantitative data that did not follow a normal distribution, the information was expressed as median (lower quartile, upper quartile), and Wilcoxon's rank sum test was used for comparison between groups. Qualitative data were presented as frequencies and proportions, and group comparisons were performed using the chi-square test. The primary metric in this study was the AUROC. The 95% confidence intervals for the AUROC values were calculated using the Bootstrap method with 1000 repetitions. Additionally, a receiver operating characteristic curve (ROC) was plotted, and accuracy, sensitivity, specificity, and F1-score were selected to assess the discrimination of the model. The model's calibration was evaluated by calibration curve [28]. Furthermore, decision curve analysis (DCA) was employed to reflect the net clinical benefit of the model [29]. Statistical analyses were conducted using R software (version 4.2.3) and Python (version 3.8.10). All statistical tests were two-tailed, and a p-value of less than 0.05 was considered statistically significant.

3 Results

3.1 Patient characteristics

This study included a total of 1,207 ventricular fibrillation (VF) ICU patients from three medical centers, including the MIMIC-IV database (n = 879), the First Affiliated Hospital of Guangdong Medical University

(n = 272), and the Second Affiliated Hospital of Guangdong Medical University (n = 56). The in-hospital mortality rates at the three centers were 36.5% (321/879) for the MIMIC-IV cohort, 32.7% (89/272) for Center A, and 23.2% (13/56) for Center B. The MIMIC-IV cohort was randomly divided into a training set (70%, n = 615) and an internal test set (30%, n = 264) with a 7:3 ratio. Baseline characteristics of the MIMIC-IV cohort are summarized in Table 1. The mean age of the patients was 65.4 ± 14.3 years, and the cohort was predominantly male (609, 69.3%). Regarding in-hospital mortality, 321 patients (36.5%) died, while 558 (63.5%) survived. In addition, the proportions of patients with a history of smoking and drinking were 9.4% (83/879) and 9.8% (86/879), respectively. No significant differences were observed between the training and test sets regarding age, sex, in-hospital outcomes, or smoking history ($p > 0.05$), indicating a balanced distribution of key baseline characteristics between the training and test sets. This justifies the subsequent model development and evaluation on the basis of comparable groups.

3.2 Selection of predictors

This study began with 52 candidate predictor variables and sequentially performed feature selection on the training set using the Boruta algorithm and LASSO regression. As shown in Figure 2, the Boruta algorithm identified 20 important variables, while LASSO regression (Figure 3) identified 11 predictor variables with non-zero coefficients. Ultimately, we took the intersection of the results from both methods, obtaining 9 common predictor variables (Figure 4), which were included in the final model: lactate, use of beta-blockers, anion gap, red blood cell count, blood urea nitrogen, hemoglobin, heart rate, respiratory rate, and albumin.

3.3 Construction of the machine learning model

Among the seven machine learning models developed, the logistic regression (LR) model exhibited excellent and stable predictive performance across both internal testing and external validation. On the internal test set, the logistic regression model exhibited outstanding discriminative ability, with an AUROC of 0.845 (95% CI: 0.801–0.889), tying with the artificial neural network for the best performance. Additionally, the model showed balanced and robust performance across other key metrics: an accuracy of 0.792, ranking first; precision and recall of 0.766 and 0.615, respectively, reflecting a good balance, with an F1 score of 0.682. In external validation, the model demonstrated excellent generalizability. In the Center A cohort, its AUROC was 0.794 (95% CI: 0.746–0.842), the highest among all models; in the Center B cohort, the AUROC reached 0.903 (95% CI: 0.826–0.981), representing the best performance among all external validation results. In comparison to more complex models, the LR model demonstrated the smallest performance degradation on the external validation datasets, underscoring its superior robustness.

Graphical analyses further supported these findings. As shown in Figure 5, the AUC values of all models on the internal test set ranged between 0.82 and 0.86, reflecting consistently strong discrimination. In external validation (Figure 6), model performance declined to AUCs between 0.72 and 0.79, suggesting

some limitations in generalizability. Notably, the LR model exhibited the smallest decrease in performance upon external validation, indicating superior robustness among the candidates.

Calibration curves (Figure 7 and Figure 8) revealed close alignment between the validation sets and the internal test set, with only a minor upward shift in external validation, implying negligible overfitting and good generalizability. Decision curve analysis confirmed the clinical utility of the LR model: in the internal test set (Figure 9), it provided a higher net benefit than the "treat-all" strategy within a threshold range of 0.15–0.45. In external validation (Figure 10), although this beneficial range narrowed to 0.20–0.40 compared to the internal test set, the model still yielded significantly positive net benefit, confirming its sustained clinical utility across different patient populations.

In summary, the logistic regression model demonstrated comprehensive advantages in discriminative ability, metric balance, cross-center generalizability, predictive stability, and clinical practicality, and was therefore selected as the optimal predictive model in this study.

3.4 Model Interpretation

The global feature importance derived from the LR model is summarized in Figure 11. In descending order of influence, the most important predictors were: β -blocker use, anion gap, lactate, blood urea nitrogen (BUN), respiratory rate, red blood cell (RBC) count, heart rate, albumin, and hemoglobin. The SHAP beeswarm plot in Figure 11 illustrates that β -blocker use, anion gap, and lactate consistently had the largest impact on the predicted mortality probability across all validation cases, while the contributions of BUN, respiratory rate, RBC count, heart rate, albumin, and hemoglobin were relatively smaller. Furthermore, the SHAP scatter plots (Figure 12) indicate that elevated levels of lactate, BUN, anion gap, respiratory rate, and heart rate, as well as the presence of anemia or low albumin, were associated with an increased risk of mortality. In contrast, β -blocker use was associated with a significantly reduced risk. To enhance local interpretability, we selected two representative patients with contrasting outcomes from the external validation set and used SHAP force plots to explain the individual predictions. As shown in Figure 13, the first case was a low-risk patient. The baseline mortality probability was 0.30, and the model's final prediction was 0.33. A modestly elevated BUN level (represented by a red bar) slightly increased the risk. However, this was largely offset by protective factors, including higher hemoglobin, hematocrit, and albumin levels, along with β -blocker use (shown as blue bars), resulting in a final risk score close to the baseline and indicating a relatively favorable prognosis. Figure 14 illustrates a high-risk case. The baseline probability was 0.20, but the final prediction was 0.62. Key risk-increasing factors (red bars) included the absence of β -blocker therapy, advanced age, elevated lactate, low bicarbonate, low hemoglobin, and low albumin. Mildly elevated BUN and a normal white blood cell count (blue bars) offered only limited protective effects. These favorable factors were insufficient to counterbalance the multiple strong risk drivers, leading to a substantially elevated mortality probability that underscores the need for prompt and intensive clinical intervention.

Table 1 A Comparison of Clinical Characteristics Between Training and Testing Sets Among ICU Patients with Ventricular Fibrillation from MIMIC-IV.

Patient characteristics	Total (n = 879)	Test Set (n = 264)	Training Set (n = 615)	<i>P</i> value
Age, mean (SD)	65.4 (14.3)	66.0 (13.8)	65.2 (14.5)	0.459
outcome, n (%)				
survival	558 (63.5)	168 (63.6)	390 (63.4)	1.000
death	321 (36.5)	96 (36.4)	225 (36.6)	
Sex n (%)				0.237
Women	270 (30.7)	89 (33.7)	181 (29.4)	
Men	609 (69.3)	175 (66.3)	434 (70.6)	
Smoking history n (%)				0.886
No	796 (90.6)	238 (90.2)	558 (90.7)	
Yes	83 (9.4)	26 (9.8)	57 (9.3)	
Drinking history n (%)				0.057
No	793 (90.2)	230 (87.1)	563 (91.5)	
Yes	86 (9.8)	34 (12.9)	52 (8.5)	
Hypertension n (%)				0.240
No	745 (84.8)	230 (87.1)	515 (83.7)	
Yes	134 (15.2)	34 (12.9)	100 (16.3)	
Diabetes n (%)				0.387
No	573 (65.2)	166 (62.9)	407 (66.2)	
Yes	306 (34.8)	98 (37.1)	208 (33.8)	
Ischemic-heart-diseases n (%)				0.912
No	442 (50.3)	134 (50.8)	308 (50.1)	
Yes	437 (49.7)	130 (49.2)	307 (49.9)	
Statin use, n (%)				0.399
No	313 (35.6)	100 (37.9)	213 (34.6)	
Yes	566 (64.4)	164 (62.1)	402 (65.4)	
Beta-blocker use, n (%)				0.707

No	286 (32.5)	83 (31.4)	203 (33.0)	
Yes	593 (67.5)	181 (68.6)	412 (67.0)	
Respiratory Rate, mean (SD)	20.0 (4.2)	19.9 (4.0)	20.0 (4.3)	0.733
Heart Rate, mean (SD)	80.1 (16.4)	80.2 (16.4)	80.0 (16.4)	0.899
DBP, mean (SD)	63.2 (10.7)	62.7 (10.8)	63.4 (10.7)	0.367
SBP, mean (SD)	110.6 (15.8)	111.2 (15.3)	110.3 (16.0)	0.436
Glucose, median [Q1,Q3]	146.0 [123.5,196.0]	145.0 [125.0,185.2]	146.0 [123.0,200.0]	0.554
Weight, mean (SD)	86.8 (21.9)	85.0 (21.1)	87.5 (22.2)	0.113
Height, mean (SD)	172.3 (8.0)	171.5 (8.3)	172.7 (7.9)	0.06
Creatinine, median [Q1,Q3]	1.0 [1.0,3.0]	1.0 [1.0,3.0]	2.0 [1.0,3.0]	0.763
BUN, median [Q1,Q3]	30.0 [18.0,50.0]	29.0 [18.8,47.2]	31.0 [18.0,50.5]	0.428
Potassium, mean (SD)	5.0 (1.0)	4.9 (1.0)	5.0 (0.9)	0.06
Aniongap, median [Q1,Q3]	18.0 [15.0,22.0]	17.5 [14.0,22.0]	18.0 [15.0,22.0]	0.62
Sodium, mean (SD)	141.9 (5.6)	142.2 (5.6)	141.8 (5.5)	0.317
Bicarbonate, mean (SD)	25.8 (5.5)	25.9 (4.9)	25.8 (5.7)	0.738
Platelet, mean (SD)	250.6 (117.2)	255.6 (130.8)	248.5 (111.0)	0.438
RBC, mean (SD)	4.0 (0.8)	3.9 (0.8)	4.0 (0.8)	0.243
WBC, median [Q1,Q3]	16.0 [12.0,21.5]	15.0 [11.0,22.0]	16.0 [12.0,21.0]	0.452
Hemoglobin, median [Q1,Q3]	12.0 [10.0,14.0]	12.0 [10.0,14.0]	12.0 [10.0,14.0]	0.269
PT, median [Q1,Q3]	17.0 [14.0,24.0]	16.0 [14.0,23.0]	17.0 [14.0,24.0]	0.231
PTT, median [Q1,Q3]	63.0 [35.0,124.5]	53.5 [33.0,108.0]	66.0 [35.5,129.0]	0.116
INR, median [Q1,Q3]	2.0 [1.0,2.0]	2.0 [1.0,2.0]	2.0 [1.0,2.0]	0.18
Temperature, median [Q1,Q3]	37.0 [36.0,37.0]	37.0 [36.0,37.0]	37.0 [36.0,37.0]	0.65
AST, median [Q1,Q3]	238.0 [75.0,794.0]	220.5 [81.8,789.5]	256.0 [72.5,796.0]	0.455
ALT, median [Q1,Q3]	139.0 [48.0,418.0]	139.0 [50.0,417.0]	137.0 [46.5,418.0]	0.691
Lactate, median [Q1,Q3]	5.0 [3.0,6.0]	5.0 [2.0,6.0]	6.0 [3.0,6.0]	0.334

CK-MB, median [Q1,Q3]	82.0 [13.5,88.0]	82.0 [13.0,90.0]	81.0 [14.0,87.0]	0.645
CK, median [Q1,Q3]	3115.0 [507.5,3710.5]	3065.5 [524.8,3706.5]	3115.0 [506.5,3715.0]	0.957
Troponin-t, median [Q1,Q3]	3.0 [0.0,3.0]	3.0 [0.8,3.0]	3.0 [0.0,3.0]	0.701
LDH, median [Q1,Q3]	1108.0 [421.5,1259.0]	895.5 [400.5,1250.2]	1184.0 [441.5,1263.5]	0.029
lymphocytes-abs, mean (SD)	1.8 (0.9)	1.8 (0.9)	1.7 (0.9)	0.309
Neutrophils-abs, mean (SD)	13.8 (5.8)	14.0 (5.9)	13.7 (5.8)	0.536
Monocytes-abs, mean (SD)	0.9 (0.6)	0.9 (0.6)	0.9 (0.6)	0.982
Albumin, mean (SD)	3.1 (0.5)	3.1 (0.5)	3.1 (0.5)	0.711
Fibrinogen, mean (SD)	390.9 (142.7)	387.1 (141.2)	392.6 (143.4)	0.603
Triglycerides, mean (SD)	176.1 (56.2)	177.0 (59.4)	175.7 (54.8)	0.761
Cholesterol-hdl, mean (SD)	41.4 (4.9)	41.1 (5.1)	41.5 (4.9)	0.298
Cholesterol, mean (SD)	75.0 (11.7)	74.8 (13.2)	75.1 (11.0)	0.804
NT-proBNP, mean (SD)	12043.8 (2674.9)	11880.4 (2711.3)	12113.9 (2658.2)	0.24
Uric acid, mean (SD)	6.0 (0.2)	6.0 (0.2)	6.0 (0.2)	0.966
D-dimer, mean (SD)	7098.0 (332.2)	7098.1 (342.0)	7098.0 (328.1)	0.996
GGT, mean (SD)	224.2 (6.0)	224.0 (6.0)	224.3 (6.1)	0.422
TT, mean (SD)	52.2 (1.3)	52.2 (1.3)	52.2 (1.3)	0.762

Note: ICU, intensive care units; DBP, diastolic blood pressure; SBP, systolic blood pressure; BUN, blood urea nitrogen; PT, prothrombin time; PTT, partial thromboplastin time; INR, international normalized ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; CK-MB, creatine kinase-MB; CK, creatine kinase; LDH, lactate dehydrogenase; NT-proBNP, N-terminal pro-B-type natriuretic peptide; GGT, gamma-glutamyl transferase; TT, thrombin time; statistically significant at $p < 0.05$.

Table 2 Comparison of Model Performance on the Development Set from the MIMIC-IV Database and External Validation Sets from Two Independent Medical Centers.

	AUROC (95% CI)	Accuracy	precision	recall	f1-score	specificity
Test set						
LR	0.845 (0.801 - 0.889)	0.792	0.766	0.615	0.682	0.893
DT	0.755 (0.703 - 0.807)	0.750	0.647	0.688	0.667	0.786
RF	0.833 (0.788 - 0.879)	0.777	0.718	0.635	0.674	0.857
LightGBM	0.824 (0.778- 0.870)	0.731	0.903	0.292	0.441	0.982
XGBoost	0.793 (0.745 - 0.842)	0.708	0.581	0.708	0.638	0.708
SVM	0.828 (0.782- 0.873)	0.784	0.729	0.646	0.685	0.863
ANN	0.845 (0.802 - 0.889)	0.784	0.705	0.698	0.702	0.833
Training set						
LR	0.836 (0.807 - 0.866)	0.766	0.724	0.582	0.645	0.872
DT	0.832 (0.778 - 0.886)	0.784	0.671	0.809	0.733	0.769
RF	0.856 (0.806 - 0.907)	0.800	0.725	0.735	0.730	0.838
LightGBM	0.844 (0.792 - 0.896)	0.730	0.875	0.309	0.457	0.974
XGBoost	0.855 (0.804 - 0.906)	0.805	0.7162	0.779	0.746	0.821
SVM	0.857 (0.806 - 0.907)	0.805	0.735	0.735	0.735	0.846
ANN	0.853 (0.802 - 0.904)	0.784	0.700	0.721	0.710	0.821
Validation set(centerA)						
LR	0.794 (0.746 - 0.842)	0.750	0.662	0.483	0.558	0.880
DT	0.765 (0.673 - 0.856)	0.659	0.484	0.556	0.517	0.709

RF	0.725 (0.629 - 0.822)	0.744	0.667	0.444	0.533	0.891
LightGBM	0.746 (0.652 - 0.840)	0.756	0.684	0.481	0.565	0.891
XGBoost	0.754 (0.661 - 0.847)	0.744	0.636	0.519	0.571	0.855
SVM	0.720 (0.623 - 0.813)	0.683	1.000	0.037	0.071	1.000
ANN	0.747 (0.653 - 0.841)	0.744	0.615	0.593	0.604	0.818
Validation set(centerB)						
LR	0.903 (0.826 - 0.981)	0.821	0.636	0.538	0.583	0.908
DT	0.827 (0.647 - 1.000)	0.824	0.667	0.500	0.571	0.923
RF	0.788 (0.594 - 0.983)	0.765	0.500	0.500	0.500	0.846

Note: The internal development set was derived from the MIMIC-IV clinical database. External validation was performed to rigorously assess the models' generalizability: Validation set (Center A) was sourced from Affiliated Hospital of Guangdong Medical University; Validation set (Center B) was sourced from The Second Affiliated Hospital of Guangdong Medical University. AUROC, area under the receiver operating characteristic; LR, logistic regression; DT, decision tree; RF, random forest; LightGBM, Light Gradient Boosting Machine; XGBoost, eXtreme Gradient Boosting; SVM, support vector machine; ANN, artificial neural network.

4 Discussion

This study, based on the MIMIC-IV database and data from two tertiary hospitals in Guangdong Province, constructed seven machine learning models to predict in-hospital mortality risk for ICU patients with ventricular fibrillation. Among the seven candidate models, the logistic regression (LR) model demonstrated the best overall performance, with an AUROC of 0.845 on the internal test set and AUROCs of 0.794 and 0.903 in the two external validation cohorts. The superior performance of the relatively simple LR model, which matched or exceeded that of more complex counterparts like the artificial neural network (ANN) and XGBoost, can be attributed to several factors. The rigorously selected, clinically coherent set of features likely establishes a relationship with the outcome that is predominantly linear or monotonic, a scenario in which LR excels. Furthermore, the inherent simplicity of LR confers a significant advantage in settings with limited sample sizes by effectively mitigating overfitting, thereby enhancing generalizability to external data, as evidenced by its robust validation

performance. The linearity assumption of LR aligns with the predominantly monotonic relationships between selected predictors and mortality, reducing overfitting risks in moderate-sized cohorts. It maintained excellent discrimination, good calibration, and outstanding clinical net benefit on both the internal test set and external validation cohorts. More importantly, the LR model itself offers transparent probability outputs and interpretable parameters, providing a foundation for clinical trust and application [26, 30, 31]. Through a robust feature selection process, we identified nine readily available clinical features—lactate, β -blocker use, anion gap, red blood cell count, blood urea nitrogen, hemoglobin, heart rate, respiratory rate, and albumin—highlighting the model's clinical applicability. Furthermore, the subsequent SHAP analysis confirmed the model's clinical plausibility, as the ranking of variable importance highly coincided with previously established pathophysiology. Considering the need for clinical decisions to be communicated to patients, we further applied SHAP for interpretability analysis of the LR model. Combined with relevant literature, the resulting ranking of variable importance highly coincided with previously established predictors of mortality risk in ventricular fibrillation, enhancing the model's credibility and clinical acceptability.

The interpretability analysis based on the SHAP method revealed the key pathophysiological mechanisms driving the model's predictions. Firstly, the non-use of β -blockers was identified as the most consistent risk factor, which aligns with previous research findings. Its mechanism may be related to the inhibition of excessive sympathetic nervous activation and the stabilization of myocardial electrophysiology [32, 33]. Secondly, metabolic disorders are the core risk-driving factors: elevated lactate and anion gap together indicate tissue hypoperfusion and acidosis [35–39]. Moreover, signs of sympathetic excitation (such as increased heart rate and respiratory rate) directly elevate the predicted risk of death [40–43]. In addition, anemia and hypoalbuminemia jointly increase patient risk by affecting oxygen transport and triggering inflammation and malnutrition. [44, 45]. Finally, the elevation of renal function indicators such as BUN has also been confirmed as an independent risk factor [46–48].

Individual SHAP force plots confirmed that the model can quantify the main risk and protective factors for each patient. In low-risk cases, beta-blocker use and normal hematocrit offset the slight increase in risk due to elevated BUN, keeping the predicted probability close to baseline. In contrast, high-risk cases showed a sharp rise in predicted probability (from 0.20 to 0.62) due to multiple metabolic derangements combined with the absence of β -blocker, providing a clear visual cue for immediate interventions such as acidosis correction, oxygenation optimization, volume resuscitation, and consideration of β -blocker initiation. By quantifying feature contributions, SHAP interprets model predictions and provides actionable insights to support clinical decision-making and facilitate personalized treatment planning [49, 50].

This study has several limitations. First, external validation was conducted only in two tertiary hospitals in Guangdong Province (sample size < 400), which limits the model's generalizability across different regions, healthcare levels, and populations. Second, although β -blockers showed a protective effect in the SHAP analysis, this finding is based on observational data and may be subject to indication bias. Future randomized controlled trials are warranted to confirm the causal protective effect of β -blockers in

this specific population. Third, the model only utilized static indicators within the first 24 hours of ICU admission, failing to capture the dynamic evolution of relevant variables. Furthermore, due to the limitations of the MIMIC-IV database, this study could not include important indicators such as specific electrocardiogram parameters for ventricular fibrillation and echocardiographic data, which may be valuable for revealing arrhythmia mechanisms and assessing cardiac structure and function. To address these limitations, we plan to conduct multicenter prospective studies, promote randomized controlled trials to verify drug effects, and incorporate temporal dynamic information and more comprehensive cardiac specialty parameters in future modeling.

5 Conclusion

This multicenter study developed and validated an interpretable machine learning-based early warning model for predicting in-hospital mortality among 1,207 ICU patients with ventricular fibrillation. A curated set of nine routinely available clinical predictors was identified. The logistic regression model exhibited robust generalizability across external validation cohorts. Based on SHAP analysis, the model reveals that the use of β -blockers is the primary protective factor, while metabolic acidosis, sympathetic activation, and anemia/hypoalbuminemia are key drivers of risk. Thereby, the model enables clear interpretation of its predictions at the individual patient level.

Abbreviations

Artificial Neural Network (ANN); Decision Tree (DT); Least Absolute Shrinkage and Selection Operator (LASSO); Light Gradient Boosting Machine (LightGBM); Logistic Regression (LR); Multivariate Imputation by Chained Equations (MICE); Machine Learning (ML); Random Forest (RF); Shapley Additive exPlanations (SHAP); Support Vector Machine (SVM); eXtreme Gradient Boosting (XGBoost); Alanine Aminotransferase (ALT); Aspartate Aminotransferase (AST); Blood Urea Nitrogen (BUN); Creatine Kinase (CK); Creatine Kinase-MB (CK-MB); Diastolic Blood Pressure (DBP); Gamma-Glutamyl Transferase (GGT); High-Density Lipoprotein Cholesterol (HDL-C); International Normalized Ratio (INR); Lactate Dehydrogenase (LDH); N-terminal pro-B-type Natriuretic Peptide (NT-proBNP); Prothrombin Time (PT); Partial Thromboplastin Time (PTT); Red Blood Cell Count (RBC); Systolic Blood Pressure (SBP); Thrombin Time (TT); White Blood Cell Count (WBC); Intensive Care Unit (ICU); Sudden Cardiac Death (SCD); Ventricular Fibrillation (VF); Area Under the Receiver Operating Characteristic Curve (AUROC); Decision Curve Analysis (DCA); Health Insurance Portability and Accountability Act (HIPAA); Interquartile Range (IQR); Medical Information Mart for Intensive Care IV (MIMIC-IV); Receiver Operating Characteristic (ROC); Structured Query Language (SQL); Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD).

Declarations

6 Acknowledgements

Not applicable.

Funding

This work was supported by the National Health Commission Center for Medical and Health Technology Development Research Grant (WKZX2024GM0206), the Guangdong Medical University Key Project of the First Batch of Clinical and Basic Science Innovation Programme (GDMULCJC2024005), and the Guangdong Medical University Affiliated Hospital Intramural Funded Project (LCYJ2023B009).

Competing interests

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

Ethics approval and consent to participate

The use of the de-identified MIMIC-IV database was approved by the Institutional Review Boards of MIT and BIDMC, and the requirement for informed consent was waived,an author (Zhijian Guo) complied with the requirements for accessing the database and was responsible for data extraction.External validation cohorts were approved by the Ethics Committees of Affiliated Hospital of Guangdong Medical University (approval No. PJKT-2025-226) and the Second Affiliated Hospital of Guangdong Medical University (approval No. PJKT2023-0549).

Data availability

The MIMIC-IV database we used in the study is publicly available. Reasonable requests regarding Affiliated Hospital of Guangdong Medical University data access should be addressed to Chengdi Chen.This study was supported by the Big Data Platform of Affiliated Hospital of Guangdong Medical University.

Author Statement

C.C., K.X.Z., and L.P. contributed to the conception and design of the study, and were responsible for data collection, integrity, and accuracy. C.C. drafted the manuscript. Z.J.G. and C.C. participated in data analysis and interpretation. H.S. contributed to manuscript revision and approved the final version.

Declaration of competing interest

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

References

1. Ravikumar V, Kong X, Tan NY, Christopolous G, Ladas TP, Jiang Z, et al. Complexity analysis of electrical activity during endocardial and epicardial biventricular mapping of ventricular fibrillation. *J Interv Card Electrophysiol*. 2023;
2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. *Circulation*. 2021;143:e254–743.
3. Zhang S. Sudden cardiac death in China: current status and future perspectives. *Europace*. 2015;17 Suppl 2:ii14-18.
4. Chugh SS, Reinier K, Uy-Evanado A, Chugh HS, Elashoff D, Young C, et al. Prediction of Sudden Cardiac Death Manifesting with Documented Ventricular Fibrillation or Pulseless Ventricular Tachycardia. *JACC Clin Electrophysiol*. 2022;8:411–23.
5. Wiberg S, Hassager C, Thomsen JH, Frydland M, Høfsten DE, Engstrøm T, et al. GLP-1 analogues for neuroprotection after out-of-hospital cardiac arrest: study protocol for a randomized controlled trial. *Trials*. 2016;17:304.
6. Brown SR, Brogan TV, McMullan DM, Roberts JS. Outcomes After In-Hospital Pediatric Recurrent Cardiac Arrests. *Pediatr Crit Care Med*. 2020;21:e981–7.
7. Lee JH, Hwang SY, Kim HR, Kim YW, Kang MJ, Cho KW, et al. Effectiveness of the sequential organ failure assessment, acute physiology and chronic health evaluation II, and simplified acute physiology score II prognostic scoring systems in paraquat-poisoned patients in the intensive care unit. *Hum Exp Toxicol*. 2017;36:431–7.
8. Sun D, Ding H, Zhao C, Li Y, Wang J, Yan J, et al. Value of SOFA, APACHE IV and SAPS II scoring systems in predicting short-term mortality in patients with acute myocarditis. *Oncotarget*. 2017;8:63073–83.
9. Kumar R, Ruhel R, van Wijnen AJ. Unlocking biological complexity: the role of machine learning in integrative multi-omics. *Acad Biol*. 2024;2:10.20935/acadbiol7428.
10. Brandao-de-Resende C, Melo M, Lee E, Jindal A, Neo YN, Sanghi P, et al. A machine learning system to optimise triage in an adult ophthalmic emergency department: a model development and validation study. *EClinicalMedicine*. 2023;66:102331.
11. Liu C, Elmer J, Arefan D, Pease M, Wu S. Interpretable machine learning model for imaging-based outcome prediction after cardiac arrest. *Resuscitation*. 2023;191:109894.
12. Kim YK, Koo JH, Lee SJ, Song HS, Lee M. Explainable Artificial Intelligence Warning Model Using an Ensemble Approach for In-Hospital Cardiac Arrest Prediction: Retrospective Cohort Study. *J Med Internet Res*. 2023;25:e48244.
13. Chugh SS, Reinier K, Uy-Evanado A, Chugh HS, Elashoff D, Young C, et al. Prediction of Sudden Cardiac Death Manifesting with Documented Ventricular Fibrillation or Pulseless Ventricular Tachycardia. *JACC Clin Electrophysiol*. 2022;8:411–23.
14. Rahadian RE, Okada Y, Shahidah N, Hong D, Ng YY, Chia MYC, et al. Machine learning prediction of refractory ventricular fibrillation in out-of-hospital cardiac arrest using features available to EMS.

Resusc Plus. 2024;18:100606.

15. Basubrin O. Current Status and Future of Artificial Intelligence in Medicine. *Cureus*. 17:e77561.
16. Alemu R, Sharew NT, Arsano YY, Ahmed M, Tekola-Ayele F, Mersha TB, et al. Multi-omics approaches for understanding gene-environment interactions in noncommunicable diseases: techniques, translation, and equity issues. *Hum Genomics*. 2025;19:8.
17. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–38.
18. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. 2022;214:106584.
19. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
20. Banks DL. The Health Insurance Portability and Accountability Act: does it live up to the promise? *J Med Syst*. 2006;30:45–50.
21. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10:1.
22. Zhou H, Zhu Y, Zhang Z, Mei Y, Lv J, Zhang G, et al. Ventricular fibrillation/ventricular tachycardia within 72 h of VA-ECMO: incidence, outcomes, risk factors, and management. *ESC Heart Fail*. 2024;11:524–32.
23. Austin PC, van Buuren S. Logistic regression vs. predictive mean matching for imputing binary covariates. *Stat Methods Med Res*. 2023;32:2172–83.
24. Gao J, Kwan PW, Shi D. Sparse kernel learning with LASSO and Bayesian inference algorithm. *Neural Netw*. 2010;23:257–64.
25. Yang JR, Chen Q, Wang H, Hu XY, Guo YM, Chen JZ. Reliable CA-(Q)SAR generation based on entropy weight optimized by grid search and correction factors. *Comput Biol Med*. 2022;146:105573.
26. Austin PC, van Buuren S. Logistic regression vs. predictive mean matching for imputing binary covariates. *Stat Methods Med Res*. 2023;32:2172–83.
27. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. 2022;214:106584.
28. Austin PC, Harrell FE Jr, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. *Stat Med*. 2020;39(21):2714-2742.
29. Tsalatsanis A, Hozo I, Vickers A, Djulbegovic B. A regret theory approach to decision curve analysis: a novel method for eliciting decision makers' preferences and decision-making. *BMC Med Inform Decis Mak*. 2010;10:51.

30. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med*. 2011;18:1099–104.
31. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774–81.
32. Koracevic G, Stojanovic M, Lovic D, Zdravkovic M, Sakac D. Certain beta blockers (e.g., bisoprolol) may be reevaluated in hypertension guidelines for patients with left ventricular hypertrophy to diminish the ventricular arrhythmic risk. *J Hum Hypertens*. 2021;35:564–76.
33. Min B, Kluger J, Kalus JS, Guertin D, McBride BF, White CM. The impact of catecholamines in patients with or without beta-blockers on the ventricular fibrillation cycle length and ventricular fibrillation cycle length variability. *Ann Noninvasive Electrocardiol*. 2005;10:305–11.
34. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J*. 2021;21:1643–8.
35. Qian X, Sheng Y, Jiang Y, Xu Y. Associations of serum lactate and lactate clearance with delirium in the early stage of ICU: a retrospective cohort study of the MIMIC-IV database. *Front Neurol*. 2024;15:1371827.
36. Sheng B, Li S, Wang H, Guo Z, Zhang Z. Association between serum lactate, lactate clearance rate, and 30-day mortality among patients undergoing mechanical ventilation: A retrospective cohort study of MIMIC-III database. *Medicine (Baltimore)*. 2024;103:e35818.
37. Kraut JA, Madias NE. Metabolic acidosis: pathophysiology, diagnosis and management. *Nat Rev Nephrol*. 2010;6:274–85.
38. Achanti A, Szerlip HM. Acid-Base Disorders in the Critically Ill Patient. *Clin J Am Soc Nephrol*. 2023;18:102–12.
39. Ahn SY, Ryu J, Baek SH, Han JW, Lee JH, Ahn S, et al. Serum anion gap is predictive of mortality in an elderly population. *Exp Gerontol*. 2014;50:122–7.
40. Yu X, Jiao Z, Yang F, Xin Q. Construction and Validation of an Early Warning Model for Predicting the 28-Day Mortality in Sepsis Patients with Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis*. 2025;20:1373–85.
41. Richard J-C, Brochard L, Breton L, Aboab J, Vandelet P, Tamion F, et al. Influence of respiratory rate on gas trapping during low volume ventilation of patients with acute lung injury. *Intensive Care Med*. 2002;28:1078–83.
42. Bruning R, Dykes H, Jones TW, Wayne NB, Sikora Newsome A. Beta-Adrenergic Blockade in Critical Illness. *Front Pharmacol*. 2021;12:735841.
43. Richalet J-P, Hermand E, Lhuissier FJ. Cardiovascular physiology and pathophysiology at high altitude. *Nat Rev Cardiol*. 2024;21:75–88.
44. Arnau-Barrés I, Pascual-Dapena A, López-Montesinos I, Gómez-Zorrilla S, Sorlí L, Herrero M, et al. Severe Hypoalbuminemia at Admission Is Strongly Associated with Worse Prognosis in Older Adults with SARS-CoV-2 Infection. *J Clin Med*. 2021;10:5134.

45. van den Brink DP, Kleinveld DJB, Sloos PH, Thomas KA, Stensballe J, Johansson PI, et al. Plasma as a resuscitation fluid for volume-depleted shock: Potential benefits and risks. *Transfusion*. 2021;61 Suppl 1:S301–12.
46. Beier K, Eppanapally S, Bazick HS, Chang D, Mahadevappa K, Gibbons FK, et al. Elevation of blood urea nitrogen is predictive of long-term mortality in critically ill patients independent of “normal” creatinine. *Crit Care Med*. 2011;39:305–13.
47. Deng T, Wu D, Liu S-S, Chen X-L, Zhao Z-W, Zhang L-L. Association of blood urea nitrogen with 28-day mortality in critically ill patients: A multi-center retrospective study based on the eICU collaborative research database. *PLoS One*. 2025;20:e0317315.
48. Zhang J, Qin Y, Zhou C, Luo Y, Wei H, Ge H, et al. Elevated BUN Upon Admission as a Predictor of in-Hospital Mortality Among Patients with Acute Exacerbation of COPD: A Secondary Analysis of Multicenter Cohort Study. *Int J Chron Obstruct Pulmon Dis*. 2023;18:1445–55.
49. Cordella C, Marte MJ, Liu H, Kiran S. An Introduction to Machine Learning for Speech-Language Pathologists: Concepts, Terminology, and Emerging Applications. *Perspect ASHA Spec Interest Groups*. 2025;10:432–50.
50. Takefuji Y. Beyond XGBoost and SHAP: Unveiling true feature importance. *J Hazard Mater*. 2025;488:13738.

Figures

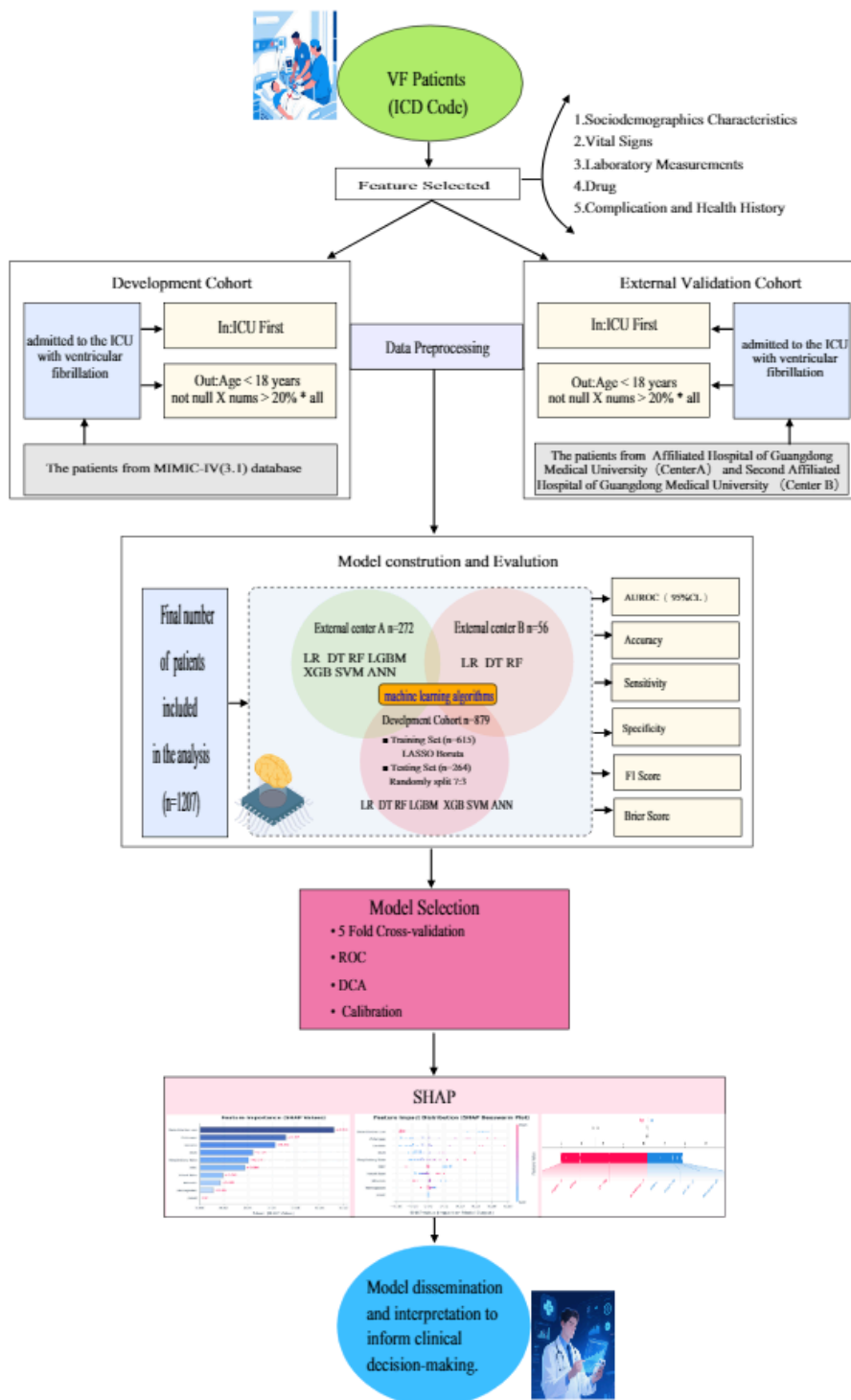


Figure 1

Development and validation workflow for the ventricular fibrillation early-warning model.

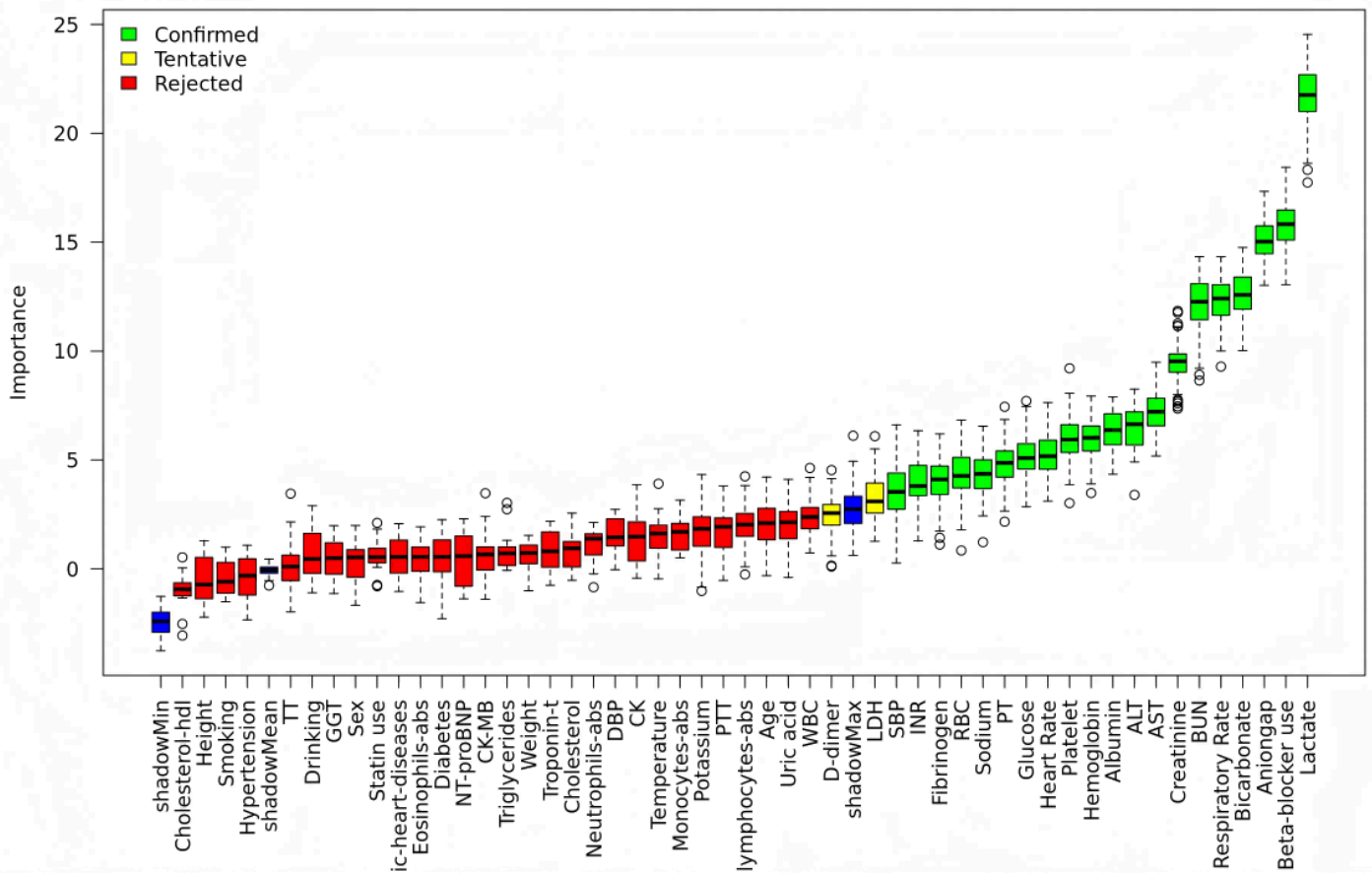


Figure 2

Feature Selection using the Boruta Algorithm.

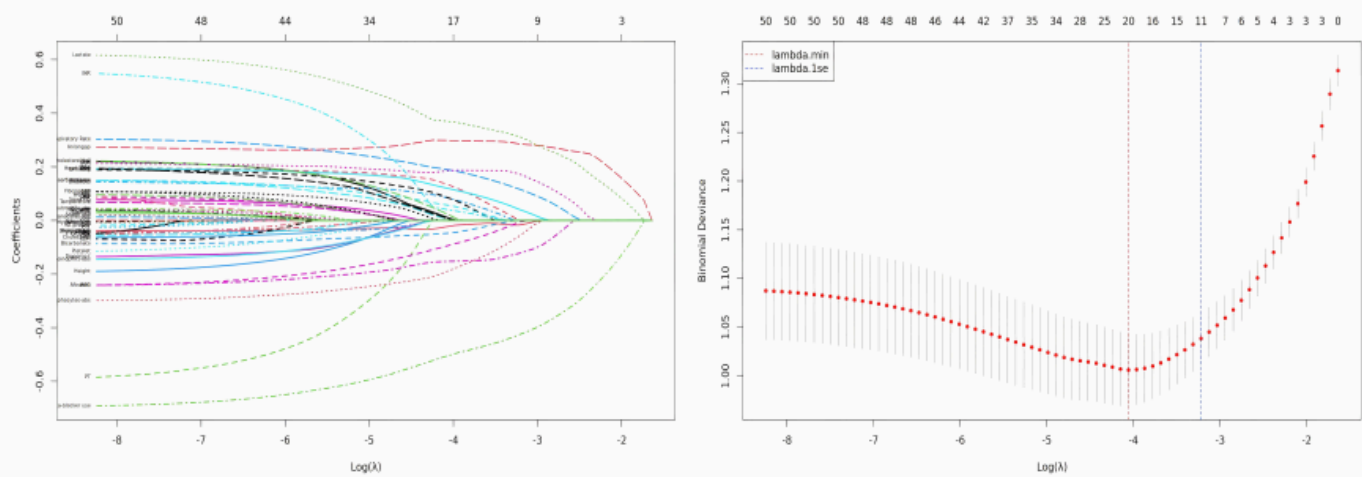


Figure 3

Feature Selection using the LASSO Algorithm.

Feature Selection by Boruta and Lasso

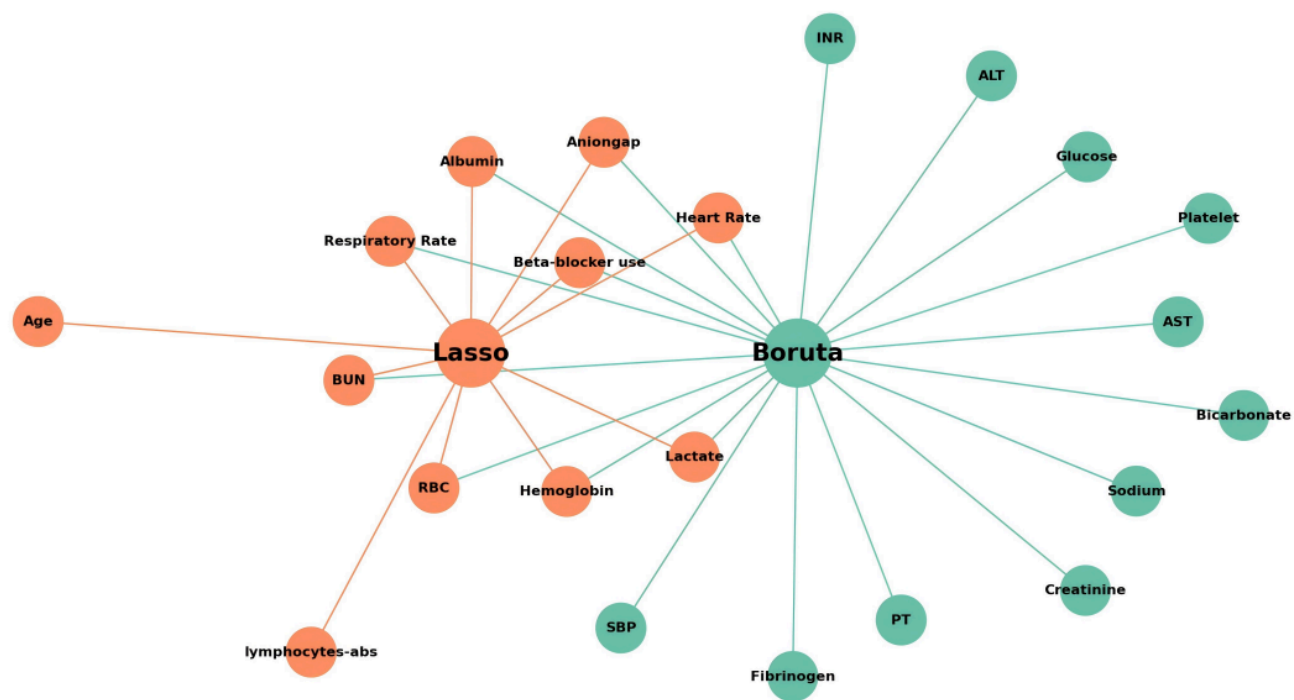


Figure 4

Overlap of Features Selected by Boruta and LASSO Algorithms.

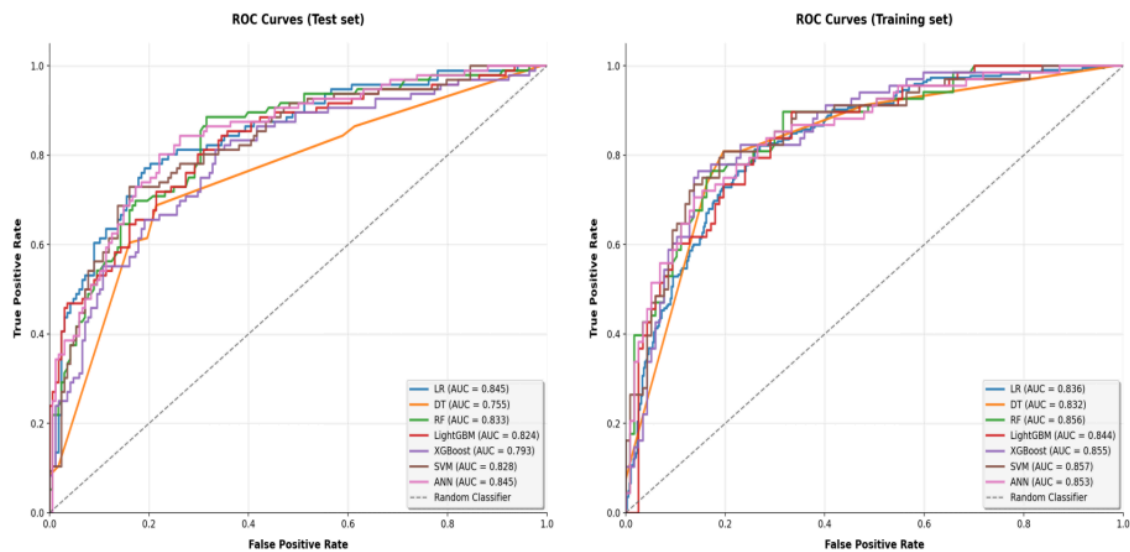


Figure 5

The ROC curves for the internal MIMIC-IV training and test sets.

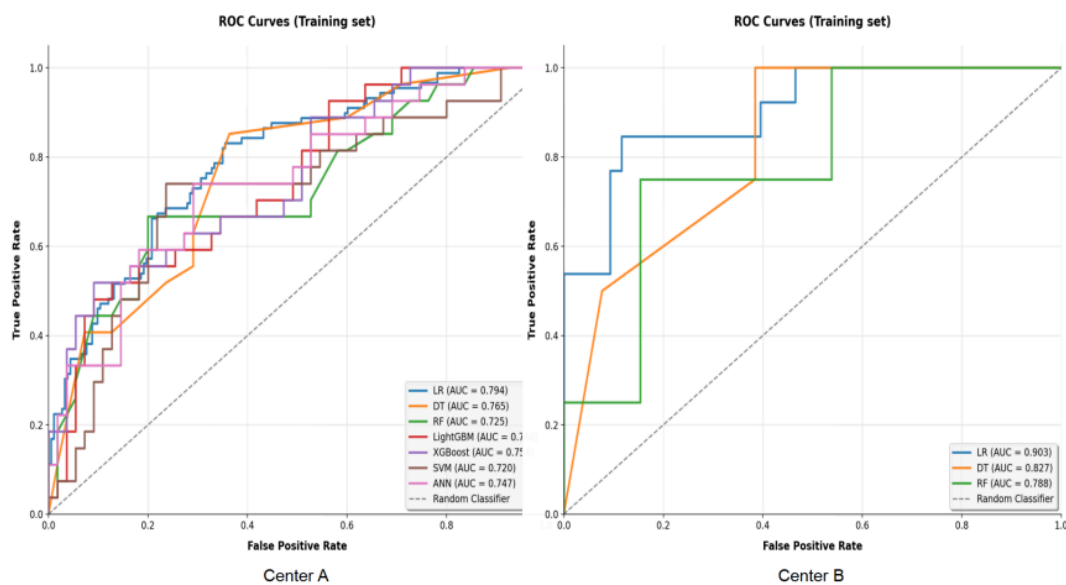


Figure 6

The ROC of the external-validation cohorts from Center A and Center B.

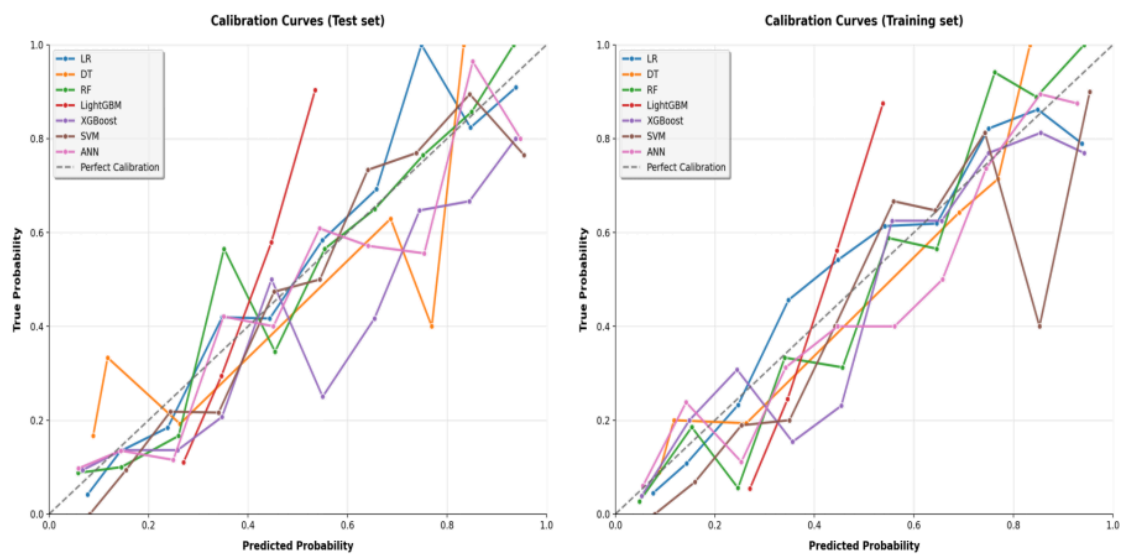


Figure 7

The calibration curves (CCs) of the internal MIMIC-IV test and training sets.

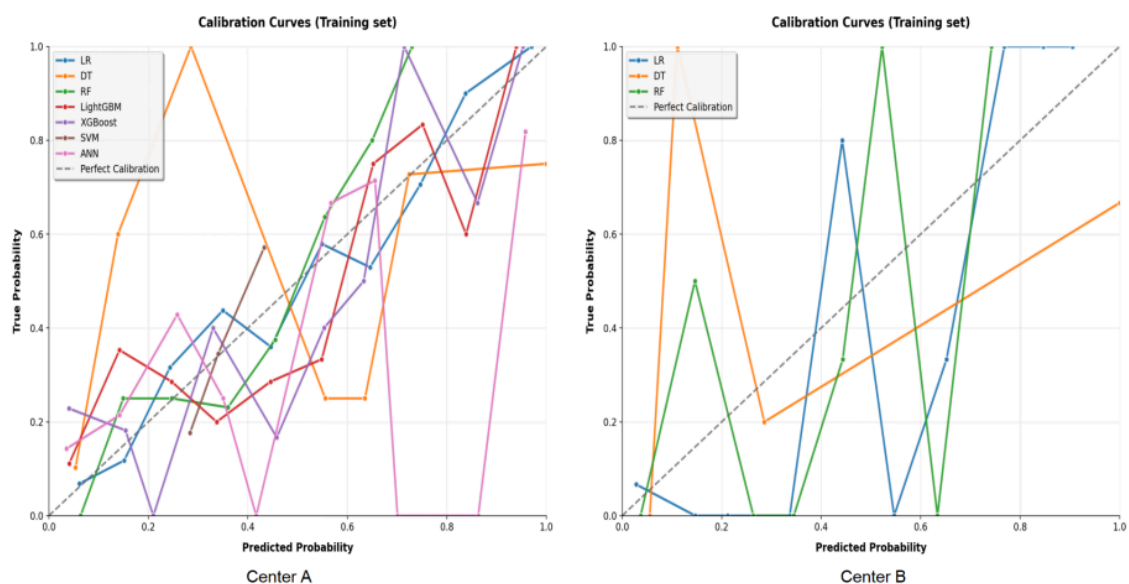


Figure 8

The calibration curves (CCs) of the external-validation cohorts from Center A and Center B.

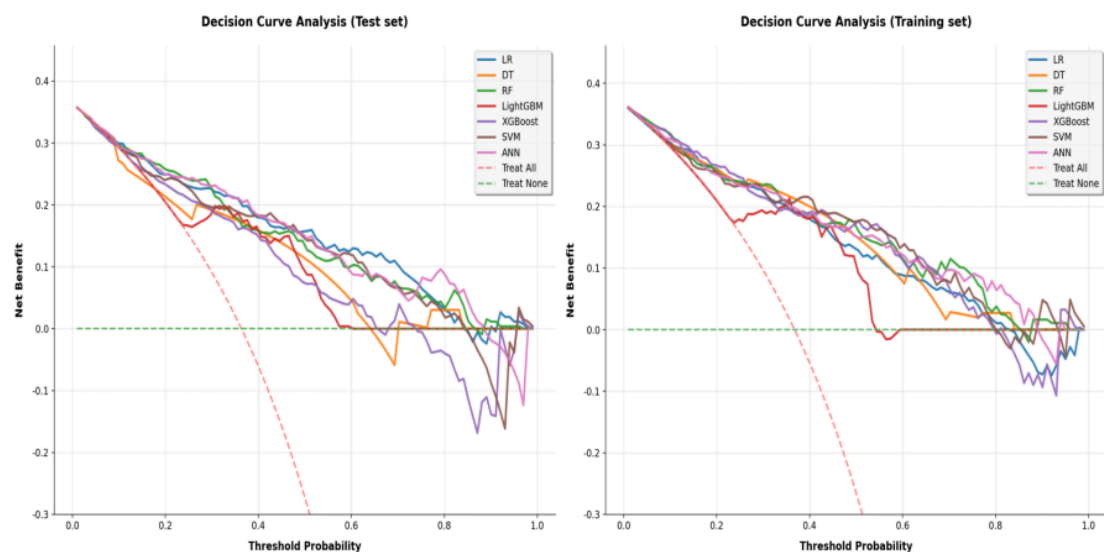


Figure 9

Decision curve analysis (DCA) of the internal MIMIC-IV test and training sets.

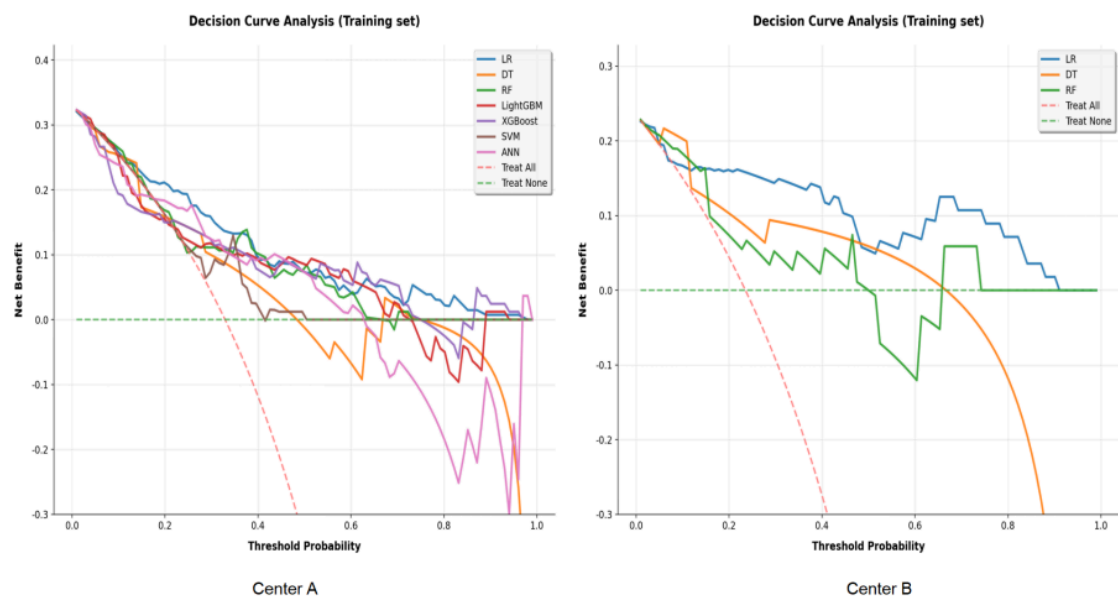


Figure 10

Decision curves of the external-validation cohorts from Center A and Center B.

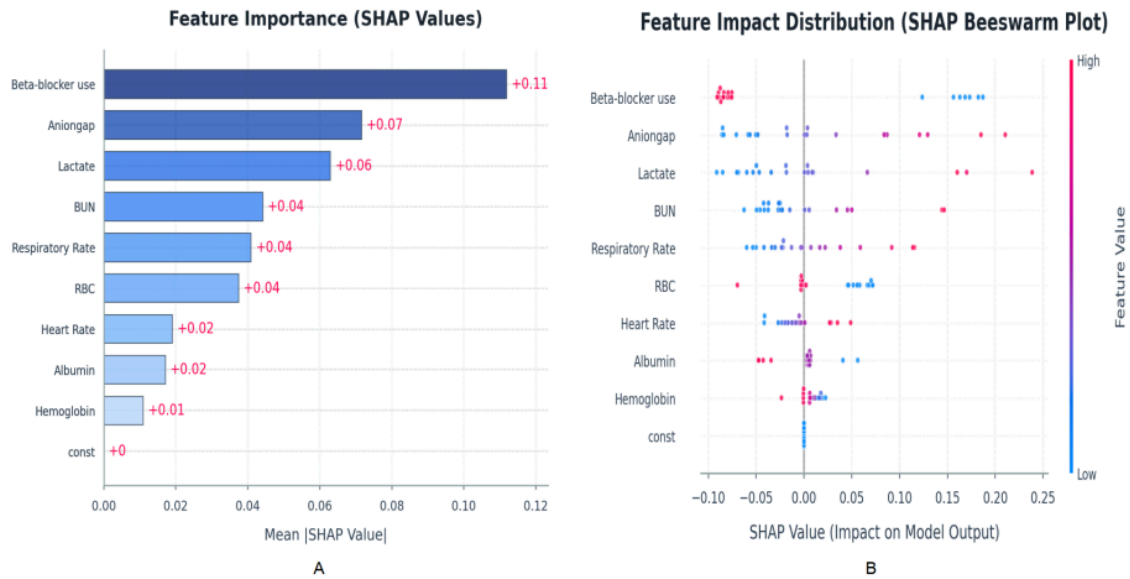


Figure 11

Feature Importance Analysis for the Prediction Model. (A) Mean SHAP values for the top predictive features; (B) Beeswarm plot of SHAP value distributions for each feature.

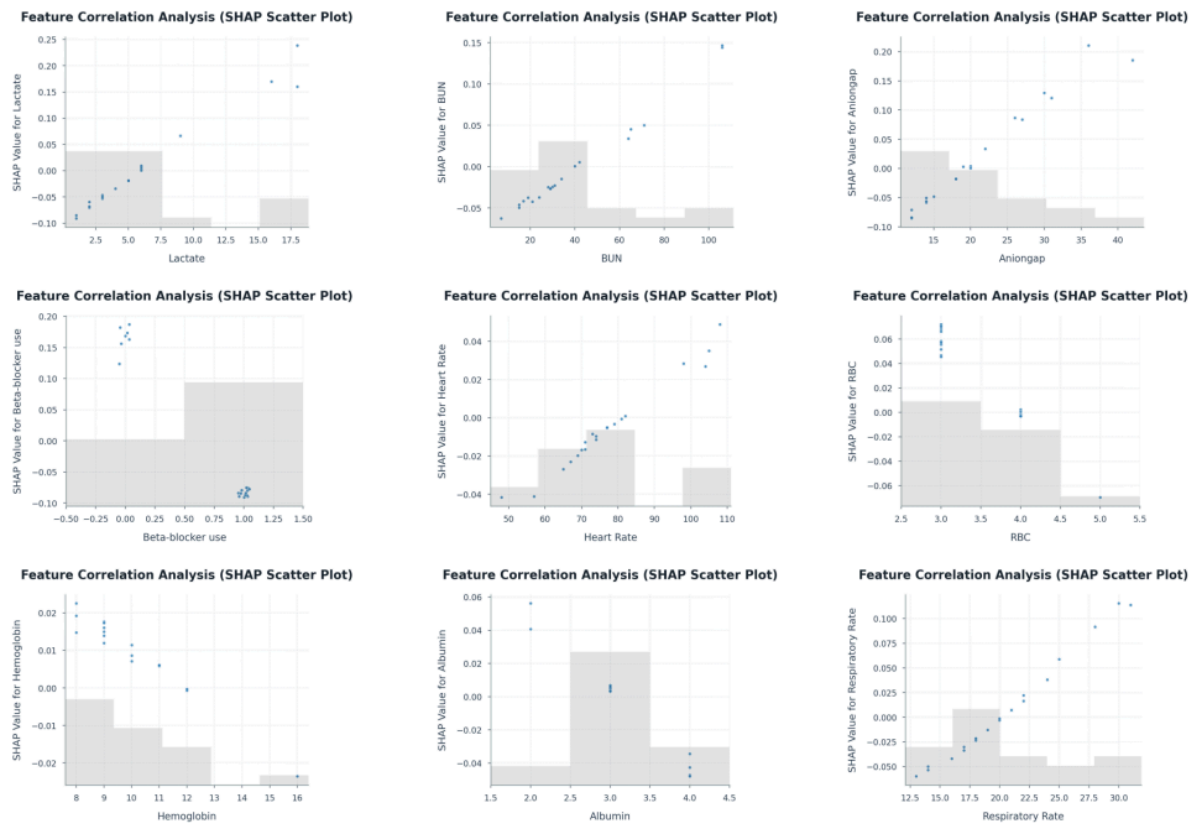


Figure 12

The main effect of variables on in-hospital mortality. Higher SHAP value leads to higher mortality risk.

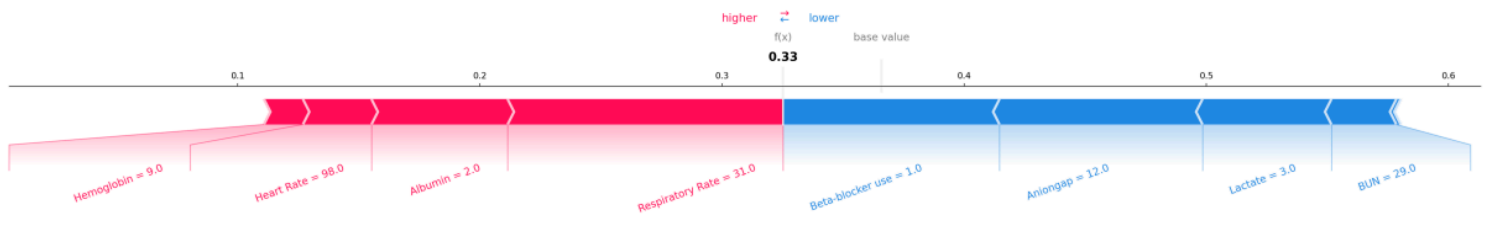


Figure 13

Prediction of individual mortality risk based on the SHAP.

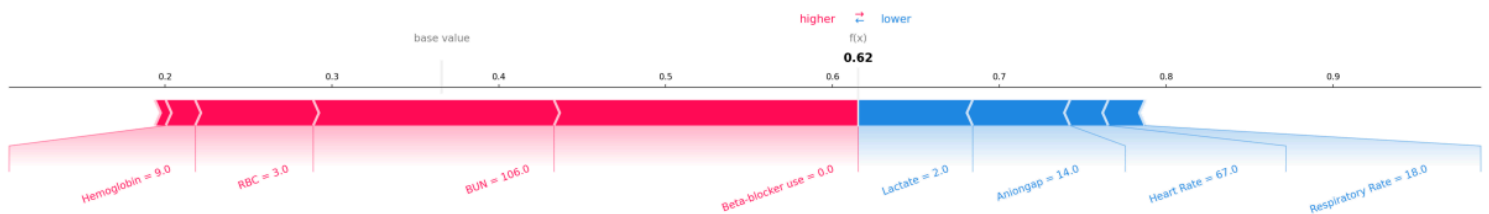


Figure 14

Prediction of individual mortality risk based on the SHAP.