Interpretable Machine Learning Models for Predicting In-Hospital Mortality in Patients with Chronic Critical Illness and Heart Failure: A Multicenter Study

Min He, Yongqi Lin, Siyu Ren, Pengzhan Li, Guoqing Liu, Liangbo Hu, Lingyan Lei, Yue Wang, Xiaocong Zeng

Submitted to: Journal of Medical Internet Research
on: June 01, 2024

Table of Contents

Interpretable Machine Learning Models for Predicting In-Hospital Mortality in Patients with Chronic Critical Illness and Heart Failure: A Multicenter Study

Min He[1*]; Yongqi Lin[1*]; Siyu Ren[1*]; Pengzhan Li[1]; Guoqing Liu[1]; Liangbo Hu[1]; Lingyan Lei[1]; Yue Wang[1]; Xiaocong Zeng[1, 2, 3]

[1]Department of Cardiology, The First Affiliated Hospital of Guangxi Medical University Nanning CN
[2]Guangxi Key Laboratory Base of Precision Medicine in Cardiocerebrovascular Diseases Control and Prevention & Guangxi Clinical Research Center for Cardio-cerebrovascular Diseases Nanning CN
[3]School of Basic Medical Sciences Nanning CN
[*]these authors contributed equally

Corresponding Author:
Xiaocong Zeng
Department of Cardiology, The First Affiliated Hospital of Guangxi Medical University
6 Shuangyong Road
Nanning
CN

Abstract

Background: Heart failure (HF) is a leading cause of morbidity and mortality among patients in intensive care units (ICUs), particularly those with chronic critical illness (CCI).

Objective: We aimed to develop and validate a machine learning (ML) model to predict in-hospital mortality for in CCI patients with CCI and HF.

Methods: Retrospective data encompassing medical records from over 200 hospitals were obtained from the Medical Information Mart for Intensive Care III (MIMIC-III), MIMIC-IV, and eICU Collaborative Research Database (eICU-CRD). Patients diagnosed with CCI and HF at their first ICU admission were included. The MIMIC-III and -IV datasets were used as a derivation cohort, while that from eICU-CRD was employed as a validation cohort. Key predictive features were identified utilizing the recursive feature elimination with 10-fold cross-validation method. Subsequently, multiple ML algorithms were evaluated, including Random Forest, K-Nearest Neighbors, Support Vector Machine (SVM), Extreme Gradient Boosting, Naive Bayes, Light Gradient Boosting Machine, and Adaptive Boosting. The performance of the models was assessed via metrics such as area under the receiver operating characteristic curve (AUROC), decision curve analysis, accuracy, sensitivity, specificity, and F1 score. Furthermore, model interpretability was enhanced by applying the SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) methods, providing insights into the contribution of individual features to the predictive outcomes.

Results: A total of 780 (males: 451 [57.8%]) and 610 (males: 343 [56.2%]) patients with CCI and HF were allocated to the derivation and validation cohorts, respectively. Eleven features were selected to develop the prediction models. Among all models, the SVM algorithm-based model demonstrated high predictive accuracy (derivation cohort: AUROC, 0.781; sensitivity, 0.739; specificity, 0.691; and F1 score, 0.613; validation cohort: AUROC, 0.683; accuracy, 0.645; sensitivity, 0.607; specificity, 0.656; and F1 score, 0.44). The SHAP and LIME analyses evaluated the feature contributions, highlighting Sequential Organ Failure Assessment score, oxyhemoglobin saturation, diastolic blood pressure, and systolic blood pressure as significant predictors of in-hospital mortality.

Conclusions: The SVM model developed in this study effectively predicts in-hospital mortality in patients with CCI and HF and can serve as a reliable tool for early intervention and improved patient management. Furthermore, this ML model combines high accuracy with interpretability, thereby substantially contributing to clinical predictive analytics.

(JMIR Preprints 01/06/2024:62815)
DOI: https://doi.org/10.2196/preprints.62815

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Interpretable Machine Learning Models for Predicting In-Hospital Mortality in Patients with Chronic Critical Illness and Heart Failure: A Multicenter Study

Min He[1], Yongqi Lin[1], Siyu Ren[1], Pengzhan Li[1], Guoqing Liu[1], Liangbo Hu[1], Lingyan Lei[1], Yue Wang[1] and Xiaocong Zeng[1,2,3*]

[1] Department of Cardiology, The First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning 530021, Guangxi, China;

[2] Guangxi Key Laboratory Base of Precision Medicine in Cardiocerebrovascular Diseases Control and Prevention & Guangxi Clinical Research Center for Cardio-cerebrovascular Diseases, Nanning, Guangxi, China;

[3] School of Basic Medical Sciences, Guangxi Medical University, Nanning, Guangxi, China;

***** Correspondence to**:** Xiaocong Zeng, Department of Cardiology, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China. Email: aban829@163.com

Abstract

Background: Heart failure (HF) is a leading cause of morbidity and mortality among patients in intensive care units (ICUs), particularly those with chronic critical illness (CCI).
Objective: We aimed to develop and validate a machine learning (ML) model to predict in-hospital mortality for in CCI patients with CCI and HF.

Methods: Retrospective data encompassing medical records from over 200 hospitals were obtained from the Medical Information Mart for Intensive Care III (MIMIC-III), MIMIC-IV, and eICU Collaborative Research Database (eICU-CRD). Patients diagnosed with CCI and HF at their first ICU admission were included. The MIMIC-III and -IV datasets were used as a derivation cohort, while that from eICU-CRD was employed as a validation cohort. Key predictive features were identified utilizing the recursive feature elimination with 10-fold cross-validation method. Subsequently, multiple ML algorithms were evaluated, including Random Forest, K-Nearest Neighbors, Support Vector Machine (SVM), Extreme Gradient Boosting, Naive Bayes, Light Gradient Boosting Machine, and Adaptive Boosting. The performance of the models was assessed via metrics such as area under the receiver operating characteristic curve (AUROC), decision curve analysis, accuracy, sensitivity, specificity, and F1 score. Furthermore, model interpretability was enhanced by applying the SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) methods, providing insights into the contribution of individual features to the predictive outcomes.
Results: A total of 780 (males: 451 [57.8%]) and 610 (males: 343 [56.2%]) patients with CCI and HF were allocated to the derivation and validation cohorts, respectively. Eleven features were selected to develop the prediction models. Among all models, the SVM algorithm-based model demonstrated high predictive accuracy (derivation cohort: AUROC, 0.781; sensitivity, 0.739; specificity, 0.691; and F1 score, 0.613; validation cohort: AUROC, 0.683; accuracy, 0.645; sensitivity, 0.607; specificity, 0.656; and F1 score, 0.44). The SHAP and LIME analyses evaluated the feature contributions, highlighting Sequential Organ Failure Assessment score, oxyhemoglobin saturation, diastolic blood pressure, and systolic blood pressure as significant predictors of in-hospital mortality.
Conclusions: The SVM model developed in this study effectively predicts in-hospital mortality in patients with CCI and HF and can serve as a reliable tool for early intervention and improved patient management. Furthermore, this ML model combines high accuracy with interpretability, thereby substantially contributing to clinical predictive analytics.

Keywords: Machine Learning; Chronic Critical Illness; Heart Failure; In-Hospital Mortality; SHAP; LIME

Introduction

Heart failure (HF), characterized by impaired cardiac function, is a critical cardiovascular health issue of the 21st century, exhibiting high mortality and extensive morbidity[1]. HF is one of the major causes of cardiovascular deaths worldwide, and over 64 million people globally and approximately 9 million in China are affected by this severe condition[2,3]. Moreover, population aging and diagnostic advancements are expected to increase the number of HF cases[4]. Although progress has been achieved in the treatment procedures, patients with HF, particularly those in intensive care units (ICUs), continue to experience prominent adversities such as prolonged ICU admissions, high healthcare expenditures, and elevated mortality[5].
In 1985, Girard coined the term chronic critical illness (CCI) to refer to patients requiring sustained ICU care[6]. Over the last 30 years, ICU care has advanced in the US. However, CCI occurrences have risen to approximately 35 cases per 100,000 people, with a total of 380,000 reported cases and a cost of $35 billion to the healthcare system in 2009[7]. Patients with CCI often experience swift and severe clinical deterioration, exhibiting 1-year mortalities of 50[8]. Therefore, the consequences of CCI not only affect the patients and their families but also have a pronounced impact on the healthcare infrastructure and societal structures.
Additionally, the interplay between HF and CCI is complex. HF commonly leads to CCI, notably worsening patient outcomes[9]. Furthermore, the families of patients with CCI face prognostic

uncertainties, eventually placing a heavy strain on medical resources[10]. Consequently, improving the prediction accuracy of in-hospital mortality can enhance prognostic consultations, facilitate collaborative decision-making, and optimize patient management.

Machine learning (ML) algorithms have been pivotal in advancing mortality reduction strategies for critically ill patients. These algorithms, which utilize extensive datasets encompassing demographics, clinical profiles, and therapeutic details, are increasingly being applied to detect, diagnose, and predict cardiovascular disease outcomes, ultimately facilitating the development of precise, tailored treatment plans[11]. However, the inherent complexity of ML models poses challenges to their interpretability, a crucial step for effectively integrating them into medical practice[12]. The SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) methods have been developed to mitigate these issues by providing clear insight into the influence of features on predictions, an essential aspect for refining patient-specific treatment strategies[13,14]. Although the potential of ML models for predicting HF mortality is well-established, its practical application is still emerging and requires further validation across diverse clinical cohorts[15,16]. ML algorithms can be used to identify individuals at an elevated risk, anticipate disease trajectory, and facilitate timely medical interventions, thus proving especially beneficial in managing HF and CCI[17]. However, the current predictive models rarely address the unique needs of patients with comorbid CCI and HF.

Therefore, this study aims to develop an ML model that precisely predicts in-hospital mortality in this patient group. This constructed prediction model will provide substantial support to healthcare professionals in assessing disease severity and prognosis.

Methods

Data sources and study population

In this study, the primary data sources were the Medical Information Mart for Intensive Care (MIMIC) and the eICU Collaborative Research Database (eICU-CRD). The MIMIC datasets, including MIMIC-III "CareVue" (version 1.4) and MIMIC-IV (version v2.2), provide extensive clinical data sourced from patients admitted to the Beth Israel Deaconess Medical Center. The MIMIC-III "CareVue" subset contains health information from 2001 to 2008, whereas MIMIC-IV covers data from 2008 to 2019 and does not overlap with MIMIC-III "CareVue"[18,19]. These two databases form the derivation cohort of this study, presenting diverse patient details such as demographics, vital signs, and laboratory test results. The eICU-CRD serves as the validation cohort, and it features anonymized records from over 200,000 ICU admissions in more than 200 US hospitals from 2014 to 2015[20].

This study enrolled patients >18 years of age who were diagnosed with CCI and HF in the ICU and were experiencing their first ICU admission. CCI was defined as an ICU stay of >14 days and persistent organ dysfunction, as demonstrated by a cardiovascular Sequential Organ Failure Assessment (SOFA) score of ≥1 or a SOFA score of ≥2 for any other organ system assessed using the most critical measurements from day 14[21]. Figure 1 illustrates the study flowchart. The researcher, Min He, was authorized to access these databases (record ID: 57369428). Since all protected health information was anonymized, obtaining ethical approval and patient consent was not required.

Data extraction

Structured Query Language was utilized to gather data on patients diagnosed with CCI and HF during their ICU stay, specifically retrieving information from the 14th day post-admission. The collected data included demographic information, SOFA scores, laboratory examination results, and comorbid conditions. All diagnoses were classified according to the International Classification of Diseases, versions 9 and 10. The variables in the dataset comprised age, gender, body mass index (BMI), heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood

pressure (MBP), oxyhemoglobin saturation (SpO_2), urine output, and laboratory tests covering routine blood counts, electrolyte levels, and liver and kidney function, along with coagulation and arterial blood gas measurements. Missing data were handled by applying the K-Nearest Neighbors (KNN) imputation technique, wherein the nearest complete data points were used to estimate the missing values. Variables with over 50% of missing data were excluded to ensure data quality and analytical precision. Patients of >89 years of age were excluded. The primary study outcome was mortality during the hospital stay.

Statistical Analysis

Standard descriptive statistics were employed to describe the baseline characteristics of the included patients. The Shapiro–Wilk test was utilized to determine the normality of continuous variables. Normally distributed variables were expressed as mean ± standard deviation and compared using t-tests. Non-normally distributed variables were presented as medians and interquartile ranges, and the Wilcoxon rank-sum test was employed for comparative analysis. Categorical variables were depicted as frequencies and percentages, with the Pearson chi-square test for evaluating differences. All statistical tests were two-tailed, with the statistical significance set at a *P*-value of <.05. Data analysis was performed using R software (version 4.3.3).

Study design

The study population was divided into training (70%) and testing (30%) sets. Continuous variables were standardized to a mean of 0 and a standard deviation of 1 to enhance the stability and comparability of the models. Variables exhibiting no variance or high multicollinearity were removed to refine the analytical process further. Subsequently, feature selection was performed using the recursive feature elimination with 10-fold cross-validation (RFECV) method with the Random Forest (RF) algorithm, which iteratively ranked and pruned features to enhance model performance.
Various ML algorithms, including RF, KNN, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Naive Bayes (NB), Light Gradient Boosting Machine (LGBM), and Adaptive Boosting (AdaBoost), were evaluated for the development of predictive models. Hyperparameters were optimized via GridSearchCV, which conducted a systematic search within a 10-fold cross-validation (CV) framework to enhance accuracy and prevent model overfitting. Model performance was quantitatively assessed using metrics such as area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, and F1 score. The classification threshold was determined using Youden's index to achieve an optimal balance between sensitivity and specificity. Additionally, decision curve analysis (DCA) was utilized to determine the clinical utility of the models by evaluating the net benefits across a range of threshold probabilities, thus providing insight into the practical implications of model deployment in clinical settings. Calibration curves were also generated to compare the predicted probabilities with actual outcomes, ensuring the accuracy of model predictions. Moreover, the models underwent external validation with an independent cohort to confirm the generalizability and repeatability of their predictions. Lastly, SHAP and LIME techniques were implemented to achieve a detailed analysis of the ML outputs. The SHAP method, which is grounded in game theory, assigns feature importance and clarifies their influence on model predictions, thereby providing an understanding of the effects of the features via intuitive visualizations. The LIME approach offers localized explanations for complex model predictions by approximating a simpler model near the prediction point, thus enhancing model transparency crucial for clinical applicability.

Results

Baseline characteristics

A total of 780 patients with CCI and HF were included in the derivation cohort (males: 451 [57.8%]), whereas 610 patients with CCI and HF were allocated to the validation cohort (males: 343 [56.2%]). Furthermore, the patients were classified into survival and non-survival groups based on their survival outcomes at discharge. Table 1 presents the baseline characteristics of the derivation cohort after grouping according to survival outcome, with the in-hospital mortality rate of this cohort being 31.5% (n = 230). Key variables, such as age, BMI, SOFA score, SBP, DBP, MBP, SpO2, white and red blood cell counts, red blood cell distribution width (RDW), serum calcium and creatinine levels, the anion gap, partial pressure of oxygen, prothrombin time, international normalized ratio, and blood urea nitrogen (BUN) level, along with urine output showed significant differences between surviving and non-surviving patients (all P<.05). However, no significant differences were found in comorbidities between the patients in these two groups (all P>.05). The baseline characteristics of the validation cohort are provided with in-hospital mortality of 29.5% (n = 139) (Multimedia Appendix 1).

Table 1. Baseline characteristics of the derivation cohort.

Variables	Survival group (N=550)	Non-survival group (N=230)	Total (N=780)	*P*
Gender, male, n (%)	321 (58.4%)	130 (56.5%)	451 (57.8%)	.692
Age, years	70.5 (59.0–78.0)	74.0 (65.0–81.0)	71.0 (61.0–79.0)	<.001
[a]BMI, kg/m2	29.9 (26.8–33.2)	28.7 (25.2–31.9)	29.6 (26.2–32.9)	.001
[b]SOFA score	4.0 (3.0–6.0)	6.0 (4.0–8.0)	4.0 (3.0–7.0)	<.001
Vital signs, median (IQR)				
Heart rate, bpm	85.8 (74.3–96.2)	87.6 (76.7–98.0)	86.3 (75.1–96.7)	.311
Respiratory rate, bpm	21.3 (18.5–24.5)	21.5 (18.6–25.2)	21.4 (18.5–24.9)	.733
[c]SBP, mmHg	110.9 (104.3–122.7)	106.2 (101.7–112.8)	109.1 (103.2–119.2)	<.001
[d]DBP, mmHg	57.0 (52.7–63.5)	54.1 (51.0–57.3)	56.0 (51.8–61.6)	<.001
[e]MBP, mmHg	70.7 (66.5–78.5)	67.0 (64.1–71.3)	69.4 (65.3–76.3)	<.001
[f]SpO2, %	97.5 (96.0–98.8)	97.1 (95.6–98.5)	97.5 (96.0–98.8)	.030
[g]pO2, mmHg	104.3 (92.5–119.5)	101.5 (85.3–117.0)	103.8 (90.4–119.0)	.030
[h]pCO2, mmHg	44.0 (39.0–49.7)	43.8 (38.2–50.3)	44.0 (39.0–49.8)	.692
Laboratory tests, median (IQR)				
Red blood cell, #/uL	3.1 (2.8–3.4)	3.1 (2.8–3.4)	3.1 (2.8–3.4)	.371
[i]MCH, pg	29.9 (28.8–31.0)	30.1 (28.9–31.2)	29.9 (28.8–31.0)	.186
[j]MCHC, g/L	32.6 (31.6–33.5)	32.7 (31.7–33.7)	32.6 (31.7–33.6)	.180
Mean corpuscular volume, fl	91.0 (88.0–95.0)	91.2 (88.0–95.5)	91.0 (88.0–95.0)	.985
[k]RDW, %	15.9 (14.8–17.3)	17.1 (15.6–18.7)	16.2 (15.1–17.9)	<.001
White blood cell, K/uL	12.8 (9.4–15.6)	13.8 (9.7–18.3)	12.9 (9.5–16.3)	.012
Platelet count, K/uL	284.0 (197.0–386.0)	226.0 (127.7–354.0)	272.0 (173.5–380.5)	<.001
Hemoglobin, g/dL	9.3 (8.3–10.2)	9.2 (8.4–10.1)	9.2 (8.4–10.1)	.535
pH	7.42 (7.39–7.45)	7.40 (7.37–7.44)	7.4 (7.4–7.4)	<.001
Anion gap, mmol/L	13.1 (11.0–16.0)	14.0 (12.0–17.0)	13.5 (11.3–16.0)	.024
Prothrombin time, s	15.0 (13.7–16.4)	15.2 (14.0–17.0)	15.0 (13.8–16.4)	.027
Partial thromboplastin time, s	47.5 (31.7–60.3)	46.8 (31.6–57.2)	47.1 (31.6–59.9)	.947
International normalized ratio	1.4 (1.2–1.5)	1.4 (1.3–1.6)	1.4 (1.2–1.5)	.012
Blood urea nitrogen, mg/dL	36.5 (24.2–55.0)	42.0 (29.0–74.7)	37.6 (25.0–58.8)	.001
Creatinine, mg/dL	1.2 (0.8–1.9)	1.4 (0.9–2.2)	1.2 (0.8–2.0)	.003
Calcium, mg/dL	8.4 (8.1–8.8)	8.3 (8.0–8.7)	8.4 (8.0–8.8)	.040

Variables	Survival group (N=550)	Non-survival group (N=230)	Total (N=780)	*P*
Chloride, mEq/L	103.0 (98.5–107.0)	103.4 (99.0–108.0)	103.0 (99.0–107.5)	.250
Magnesium, mg/dL	2.1 (2.0–2.4)	2.1 (2.0–2.3)	2.1 (2.0–2.4)	.153
Phosphate, mg/dL	3.6 (3.1–4.3)	3.7 (3.0–4.3)	3.6 (3.1–4.3)	.831
Potassium, mEq/L	4.0 (3.8–4.3)	4.0 (3.8–4.4)	4.0 (3.8–4.4)	.823
Sodium, mEq/L	140.0 (137.0–143.7)	140.0 (136.0–145.0)	140.0 (136.5–144.0)	.970
Glucose, mg/dL	131.0 (111.7–154.0)	134.2 (115.0–160.0)	131.8 (113.0–154.6)	.201
Urine output, ml/24h	3533.5 (2225.4–5266.0)	2785.0 (1055.0–4485.0)	3311.5 (1933.6–5042.5)	<.001
Comorbidities, n (%)				
Atrial fibrillation	291 (52.9%)	125 (54.3%)	416 (53.3%)	.773
Acute kidney injury	453 (82.4%)	183 (79.6%)	636 (81.5%)	.414
Arrhythmia	370 (67.3%)	159 (69.1%)	529 (67.8%)	.673
Cardiomyopathy	108 (19.6%)	40 (17.4%)	148 (19%)	.529
Cerebrovascular disease	16 (2.9%)	2 (0.9%)	18 (2.3%)	.142
Coronary heart disease	254 (46.2%)	102 (44.3%)	356 (45.6%)	.696
Chronic kidney disease	131 (23.8%)	61 (26.5%)	192 (24.6%)	.479
[l]COPD	57 (10.4%)	19 (8.3%)	76 (9.7%)	.441
Diabetes	89 (16.2%)	38 (16.5%)	127 (16.3%)	.991
Dyslipidemia	143 (26%)	54 (23.5%)	197 (25.3%)	.516
Heart arrest	54 (9.8%)	30 (13%)	84 (10.8%)	.231
Acute myocardial infarction	132 (24%)	54 (23.5%)	186 (23.8%)	.949
Valve heart disease	123 (22.4%)	47 (20.4%)	170 (21.8%)	.617
Respiratory failure	369 (67.1%)	165 (71.7%)	534 (68.5%)	.234
Mechanical ventilation	532 (96.7%)	214 (93%)	746 (95.6%)	.035

[a]BMI: body mass index.
[b]SOFA: Sequential Organ Failure Assessment.
[c]SBP: systolic blood pressure.
[d]DBP: diastolic blood pressure.
[e]MBP: mean blood pressure.
[f]SpO_2: oxyhemoglobin saturation.
[g]pO_2: partial pressure of oxygen.
[h]pCO_2: partial pressure of carbon dioxide.
[i]MCH: mean corpuscular hemoglobin.
[j]MCHC: mean corpuscular hemoglobin concentration.
[l]RDW: Red blood cell distribution width.
[l]COPD: Chronic obstructive pulmonary disease.

Features selected and model performance

The RFECV method identified 11 critical predictors from the training data, with these predictors achieving the highest accuracy (Figure 2). Further, the predictive models were developed by employing the following seven ML algorithms based on these 11 critical predictors: RF, KNN, SVM, XGBoost, NB, LGBM, and AdaBoost. Hyperparameter tuning was then conducted via GridSearchCV to optimize the models within a 10-fold CV framework, with detailed information provided in Multimedia Appendix 2. Specifically, for the SVM model, the optimized hyperparameters were: cost = 26.68, degree = 1, gamma = 23.36, kernel = polynomial, and type = C-classification.

In the derivation cohort, the SVM model demonstrated robust performance with an AUROC of 0.781, an accuracy of 0.748, a sensitivity of 0.739, a specificity of 0.691, and an F1 score of 0.613 (Table 2 and Figure 3A). In the validation cohort, the SVM model maintained its effective performance, yielding an AUROC of 0.675, an accuracy of 0.645, a sensitivity of 0.607, a specificity

of 0.656, and an F1 score of 0.443 (Table 2 and Figure 3B). The AUROCs of the other six models (RF, KNN, LGBM, NB, AdaBoost, and XGBoost) were 0.759, 0.759, 0.744, 0.751, 0.721, and 0.682 in the derivation cohort (Table 2 and Figure 3A), 0.643, 0.646, 0.615, 0.638, 0.661, and 0.616 in the validation cohort (Table 2). The SVM algorithm emerged as the superior model, consistently exhibiting the best performance metrics. The DCA of the SVM model in the two cohorts (Figure 4A-B) highlighted significant clinical benefits across various decision thresholds. Finally, calibration curves illustrated in Multimedia Appendix 3A-B established the consistency of the SVM model, with the calibration curve closely aligning with the ideal curve and thus verifying the accuracy of the model's probability predictions against actual outcomes. Given these favorable findings, the SVM model was selected for further explainability analysis.

Table 2. Performance of the seven machine learning models in the derivation and validation cohorts.

Model	Threshold	[a]AUROC	Accuracy	Sensitivity	Specificity	F 1 score
Derivation cohort						
[b]RF	0.268	0.759	0.692	0.812	0.642	0.609
[c]KNN	0.250	0.759	0.714	0.768	0.691	0.613
[d]SVM	0.316	0.781	0.748	0.739	0.751	0.633
[e]LGBM	0.237	0.744	0.705	0.739	0.691	0.596
[f]NB	0.162	0.751	0.667	0.855	0.588	0.602
[g]AdaBoost	0.253	0.721	0.624	0.797	0.552	0.556
[h]XGBoost	0.377	0.682	0.598	0.855	0.491	0.557
Validation cohort						
RF	0.273	0.643	0.550	0.752	0.488	0.438
KNN	0.317	0.646	0.678	0.483	0.738	0.412
SVM	0.310	0.675	0.645	0.607	0.656	0.443
LGBM	0.548	0.615	0.720	0.366	0.828	0.379
NB	0.322	0.638	0.650	0.552	0.679	0.423
AdaBoost	0.131	0.661	0.622	0.634	0.618	0.439
XGBoost	0.500	0.616	0.738	0.303	0.870	0.351

[a]AUROC: area under the receiver operating characteristic curve.
[b]RF: Random Forest.
[c]KNN: K-Nearest Neighbors.
[d]SVM: Support Vector Machine.
[e]XGBoost: Extreme Gradient Boosting.
[f]NB: Naive Baye.
[g]LGBM: Light Gradient Boosting Machine.
[h]AdaBoost: Adaptive Boosting.

Explainability of SVM Model

In this analysis, the testing set data were assessed to explore the SHAP values within the SVM model, with the SHAP values exceeding 0 implying a promoting effect on the outcome. Figure 5A presents a summary plot that visualizes the overall influence of each feature. Colors on the plot denote the magnitude of feature values, i.e., yellow indicates high values, while purple represents low values. Additionally, the deviation of a data point from the zero baseline SHAP value directly correlates with its impact on the model's prediction, enhancing our comprehension of outcome manipulation by the feature values. Further analysis of the directionality of effects demonstrated that the elevated SOFA score, BUN level, and RDW were associated with increased mortality risk. Conversely, higher SpO_2, DBP, and SBP were inversely correlated with mortality risks. Figure 5B is a feature importance plot that ranks features by their SHAP values, highlighting those that are most

predictive of circulatory dysfunction (SOFA score, SpO_2, DBP, and SBP) as the most influential. Partial dependence plots were also employed to display the relationships of individual features with mortality likelihood (Figure 6). In these plots, positive linear relationships were identified between mortality risk and SOFA score, BUN level, age, RDW, and platelet count, while negative linear relationships were observed between mortality risk and SpO_2, DBP, SBP, BMI, urine output, and pH. The study further examined the potential interactions between platelet count and other significant prognostic features, such as SOFA score and RDW (Multimedia Appendix 4A-B).

Next, the SHAP and LIME methods were employed to delineate the influence of clinical features on the prognostic outcomes for two patients with distinct in-hospital outcomes (i.e., survival and non-survival). The starting point for the SHAP waterfall plots was the mean prediction of the dataset for in-hospital mortality, and the influence of each feature on prediction adjustments was graphically represented from this point (Figure 6). Color-coded arrows in these plots indicated whether individual features increased (yellow arrows) or decreased (red arrows) the mortality risk. In Patient 1 who had death as the outcome, the SHAP values (Figure 7A) reflected a 47.1% mortality risk, with the contributing features to this risk comprising a BMI of 18.6 kg/m^2, DBP of 46.6 mmHg, SOFA score of 7, SpO_2 of 96.5%, SBP of 109 mmHg, urine output of 2856 ml, age of 71 years, BUN level of 50 mg/dL, and RDW of 17.1%. Conversely, a platelet count of 61 × 109/L and a pH of 7.42 were associated with reduced mortality risk. Additionally, the LIME results for Patient 1 demonstrated a mortality probability of 48% (Multimedia Appendix 5A). In the case of Patient 2 who had survival as the outcome, the SHAP results indicated a 75.5% survival probability, with an improved survival rate linked to low SOFA scores and high SpO_2 levels (Figure 7B). Correspondingly, the LIME results of Patient 2 suggested an elevated survival probability of 83% (Multimedia Appendix 5B).

Discussion

Principal Results

In this study, we developed and validated seven ML models to predict in-hospital mortality in patients with CCI and HF, leveraging multicenter data from over 200 hospitals within the MIMIC and eICU databases. Using the RFECV method, we identified 11 key predictors, ranked by their SHAP values as follows: SOFA score, SpO2, DBP, SBP, BUN level, age, RDW, BMI, urine output, pH, and platelet count. The SVM model had superior performance compared to the other models. The SHAP and LIME methods were implemented to augment the model interpretability, thereby clarifying the decision-making processes of the SVM model.

CCI is a condition in which patients who overcome an initial acute illness continue to exhibit severe organ dysfunction, requiring extended ICU care[22]. Thus, CCI can cause significant strain on healthcare resources due to elevated ICU occupancies. Studies of the RO-AHFS and ALARM-HF registries have reported in-hospital mortality rates of 17.3% and 17.8%, respectively, for patients with HF in ICUs[23,24]. Other researchers have found significantly higher mortality rates, with patients with CCI and HF demonstrating mortality rates of 28.6% and 33.6% in Japan and China, respectively[17,25]. The current study revealed an in-hospital mortality of approximately 30%, highlighting the urgent need for improved prognostic evaluations and therapeutic interventions. CCI and HF both represent formidable public health issues, necessitating comprehensive investigations to identify alterable risk factors and formulate effective management and prevention methods.

ML effectively enhances the accuracy of diagnoses, treatments, and risk assessments in HF-related CCI management. The utilization of ML in this field has profoundly influenced patient outcomes by providing more personalized and precise treatment solutions[11]. In this study, the RFECV method isolates key predictors critical to understanding mortality risk in CCI patients with CCI and HF. Consequently, our model integrates only the most vitally basic and common clinical prognostic indicators, improving the generalizability and achieving high prediction performance in both derivation and validation cohorts. The 2018 General Data Protection Regulation emphasizes the

necessity for clarity in ML decisions, thus promoting the need for such interpretative tools[26]. The implementation of the SHAP and LIME techniques in our study successfully addressed the complexities of "black box" ML algorithms, thereby enhancing model transparency. This clarity in the model is vital for improve outcomes of patients with CCI and HF.

We determined the impact of clinical features on individual patient outcomes by performing in-depth analyses of two patients from the testing set with different prognostic outcomes. The SHAP and LIME interpretative algorithms demonstrated whether the various clinical indicators raised or lowered the in-hospital mortality risk in each patient. In Patient 1 (Figure 6A), the ML model indicated a 47.1% likelihood of mortality based on the SHAP method. The influential features increasing this risk included BMI, DBP, SOFA score, SpO_2, SBP, urine output, age, BUN level, and RDW. Conversely, elevated platelet counts and improved pH levels contributed to reducing this risk. Ultimately, the patient succumbed to the disease, confirming the prediction. In Patient 2 (Figure 6B), the model estimated a 75.5% probability of survival based on the SHAP method. The features such as a lower SOFA score, higher SpO_2, higher pH level, lower RDW, increased urine output, and higher platelet count contributed to reducing the mortality risk. Conversely, increased SBP, DBP, age, BMI, and BUN level heightened the mortality risk. Further, the survival of the patient at discharge validated the model prediction. The LIME method exhibited similar findings to those of the SHAP method, clarifying the directional effects of these features and enriching clinical understanding across methodologies (Multimedia Appendix 5A-B). Moreover, the discrepancies between the model predictions and the baseline values highlight the cumulative influence of these clinical predictors and confirm the complexity of the analytical process for predicting patient outcomes. Therefore, these two interpretative methods provided a comprehensive insight into the prognostic features.

Comparison with Prior Work

Our study revealed that the increased risk of in-hospital mortality among patients with CCI and HF was correlated to the decrease in circulatory dysfunction markers, such as SOFA score, SpO_2, DBP, and SBP. The SOFA score, a metric for assessing organ dysfunction, measures the severity across six systems: respiratory, circulatory, renal, hematologic, hepatic, and central nervous systems. This tool has been established as a crucial mortality indicator in patients with HF[27]. The SOFA score was strongly predictive of mortality in this study, exhibiting the most significant weight within the SVM model for patients with CCI and HF. Furthermore, previous studies have shown that patients with HF and lower blood pressure (BP) typically have poorer outcomes than those with HF and higher BP, primarily attributed to diminished cardiac output and pump efficiency[28]. Additionally, lower SpO_2 levels, an indicator of potential HF aggravation, were found to profoundly influence patient prognosis[29].

Population aging is associated with escalated morbidity and healthcare costs, particularly among older adults who constitute a significant proportion of CCI cases. Additionally, HF continues to be the predominant cause of hospital admissions among individuals >65 years[30]. A prior study demonstrated that older age can substantially increase mortality risk for critically ill individuals with HF[31]. Our data showed a similar trend with that study, displaying a significant difference in the median age between non-surviving and surviving patients (74.0 years vs. 70.5 years, $P<.001$). These patients are often of advanced age and present with persistent organ dysfunction, two features that heighten mortality risk[32].

Our study also identified elevated BUN levels and reduced urine output as significant predictors of in-hospital mortality for patients with CCI and HF. Low urine output is common in ICU patients and is strongly associated with poor outcomes due to its implications for renal parenchymal damage and potential role in maintaining intravascular volume[33]. Moreover, the decreased renal function can lead to fluid retention, increased cardiac load, and potentially worsened HF symptoms. The interplay between progressive cardiovascular and renal impairments can frequently exacerbate CCI, contributing to a cyclical escalation in overall disease burden and ultimately causing cardiorenal

syndrome[34]. Earlier investigations have consistently shown that fluid overload is significantly correlated with higher mortality in patients with HF[35]. Previous ML analyses have recognized BUN, a marker of renal function, as a crucial predictor of mortality in individuals with HF[16].
In this study, we determined that RDW significantly influenced in-hospital mortality in patients with CCI and HF. Alterations in RDW, which are often ascribed to nutritional deficits (notably iron, folate, and vitamin B_{12} deficiencies) or bone marrow suppression, are prevalent in patients with HF and have been shown to predict adverse outcomes more reliably than other established risk features[36]. All these findings emphasize the significance of the overall evaluation of clinical profiles rather than the examination of individual metrics alone to accurately estimate mortality risks using ML models.

Limitations

However, this study has a few limitations that must be acknowledged. First, the use of retrospective data from the MIMIC and eICU databases, which do not provide cardiological parameters, may have introduced selection and information biases, potentially compromising the precision and reducing the broader applicability of the model. Hence, prospective studies are critical to confirm its utility in real-life clinical scenarios. Second, our model only utilized laboratory data collected on the 14th day of admission. This approach may have overlooked any dynamic changes in these indicators during the hospital stay. Finally, our study data predominantly represented Western demographics, thereby restricting the applicability of our model to other groups, such as Asian populations. Therefore, future investigations should aim to employ a diverse dataset and incorporate dynamic clinical variables to enhance the predictive power of their models for in-hospital mortality.

Conclusions

The SVM model developed in this study may serve as a reliable ML instrument for predicting the in-hospital mortality of patients with CCI and HF. Furthermore, global and local interpretable techniques can be used to comprehensively understand the underlying data of SVM models, thereby potentially enhancing the clinical utility. These insights will also enable clinicians to devise targeted management strategies crucial for improving the survival rates in patients with CCI and HF.

Acknowledgements

We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript. This work was supported by the National Natural Science Foundation of China (grant numbers 82260069 and 81860071).

Conflicts of Interest

None declared

Abbreviations

AdaBoost: Adaptive Boosting
AUROC: area under the receiver operating characteristic curve
BMI: body mass index
BP: blood pressure
BUN: blood urea nitrogen
CCI: chronic critical illness
CV: cross-validation
DBP: diastolic blood pressure
DCA: decision curve analysis
eICU-CRD: eICU Collaborative Research Database

HF: heart failure
HR: heart rate
ICU: intensive care unit
KNN: K-Nearest Neighbors
LGBM: Light Gradient Boosting Machine
LIME: Local Interpretable Model-Agnostic Explanations
MBP: mean blood pressure
MIMIC: Medical Information Mart for Intensive Care
ML: machine learning
NB: Naive Bayes
RF: Random Forest
RFECV: recursive feature elimination with cross-validation
RDW: red blood cell distribution width
SBP: systolic blood pressure
SHAP: SHapley Additive exPlanations
SOFA: Sequential Organ Failure Assessment
SpO2: oxyhemoglobin saturation
SVM: Support Vector Machine
XGBoost: Extreme Gradient Boosting

References

1. Ac A, T A, M J, D C, Sd W, Gh G, Hj R. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. Circ Heart Fail Circ Heart Fail; 2013 Jan 9;6(5):881-889. PMID:23888045

2. The Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on Cardiovascular Health and Diseases in China 2021: An Updated Summary. Biomed Environ Sci BES 2022 Jul 20;35(7):573–603. PMID:35945174

3. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Lond Engl 2018 Nov 10;392(10159):1789–1858. PMID:30496104

4. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272–3287. PMID:35150240

5. Komanduri S, Jadhao Y, Guduru SS, Cheriyath P, Wert Y. Prevalence and risk factors of heart failure in the USA: NHANES 2013 - 2014 epidemiological follow-up study. J Community Hosp Intern Med Perspect 2017 Jan;7(1):15–20. PMID:28634519

6. Girard K, Raffin TA. The chronically critically ill: to save or let die? Respir Care 1985 May;30(5):339–347. PMID:10315661

7. Kahn JM, Le T, Angus DC, Cox CE, Hough CL, White DB, Yende S, Carson SS, ProVent Study Group Investigators. The epidemiology of chronic critical illness in the United States. Crit Care Med 2015 Feb;43(2):282–287. PMID:25377018

8. Unroe M, Kahn JM, Carson SS, Govert JA, Martinu T, Sathy SJ, Clay AS, Chia J, Gray A, Tulsky JA, Cox CE. One-year trajectories of care and resource utilization for recipients of prolonged mechanical ventilation: a cohort study. Ann Intern Med 2010 Aug 3;153(3):167–175. PMID:20679561

9. Pérez-Anibal E, Contreras-Arrieta S, Rojas-Suárez J, Coronell-Rodríguez W, Aguilar-Schotborgh M, Borre-Naranjo D, Almanza-Hurtado A, Dueñas-Castell C. Association of Chronic Critical Illness and COVID-19 in Patients Admitted to Intensive Care Units: A Prospective Cohort Study. Arch Bronconeumol 2023 Feb;59(2):126–128. PMID:36400653

10. Je N, Af M, Sl C, N T, S W, Gi A, Rs M. Communication about chronic critical illness. Arch Intern Med Arch Intern Med; 2007 Oct 12;167(22). PMID:18071175

11. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, Ashley E, Dudley JT. Artificial Intelligence in Cardiology. J Am Coll Cardiol 2018 Jun 12;71(23):2668–2679. PMID:29880128

12. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA 2017 Aug 8;318(6):517–518. PMID:28727867

13. Ribeiro MT, Singh S, Guestrin C. " Why should i trust you?" Explaining the predictions of any classifier. Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min 2016. p. 1135–1144. doi: 10.1145/2939672.2939778

14. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan 17;2(1):56–67. doi: 10.1038/s42256-019-0138-9

15. Woolley RJ, Ceelen D, Ouwerkerk W, Tromp J, Figarska SM, Anker SD, Dickstein K, Filippatos G, Zannad F, Metra M, Ng L, Samani N, van Veldhuisen DJ, Lang C, Lam CS, Voors AA. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. Eur J Heart Fail 2021 Jun;23(6):983–991. PMID:33651430

16. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, Jacoby DL, Masoudi FA, Spertus JA, Krumholz HM. Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction. JACC Heart Fail 2020 Jan;8(1):12–21. PMID:31606361

17. Liu P, Li S, Zheng T, Wu J, Fan Y, Liu X, Gong W, Xie H, Liu J, Li Y, Jiang H, Zhao F, Zhang J, Wu L, Ren H, Hong Z, Chen J, Gu G, Wang G, Zhang Z, Wu X, Zhao Y, Ren J. Subphenotyping heterogeneous patients with chronic critical illness to guide individualised fluid balance treatment using machine learning: a retrospective cohort study. EClinicalMedicine 2023 May;59:101970. PMID:37131542

18. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, Pollard TJ, Hao S, Moody B, Gow B, Lehman L-WH, Celi LA, Mark RG. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 3;10(1):1. PMID:36596836

19. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, Moody B, Szolovits P,

Anthony Celi L, Mark RG. MIMIC-III, a freely accessible critical care database. Sci Data Nature Publishing Group; 2016 May 24;3(1):160035. doi: 10.1038/sdata.2016.35

20. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data Nature Publishing Group; 2018 Sep 11;5(1):180178. doi: 10.1038/sdata.2018.178

21. Gardner AK, Ghita GL, Wang Z, Ozrazgat-Baslanti T, Raymond SL, Mankowski RT, Brumback BA, Efron PA, Bihorac A, Moore FA, Anton SD, Brakenridge SC. The Development of Chronic Critical Illness Determines Physical Function, Quality of Life, and Long-Term Survival Among Early Survivors of Sepsis in Surgical ICUs. Crit Care Med 2019 Apr;47(4):566–573. PMID:30664526

22. Gardner AK, Ghita GL, Wang Z, Ozrazgat-Baslanti T, Raymond SL, Mankowski RT, Brumback BA, Efron PA, Bihorac A, Moore FA, Anton SD, Brakenridge SC. The Development of Chronic Critical Illness Determines Physical Function, Quality of Life, and Long-Term Survival Among Early Survivors of Sepsis in Surgical ICUs. Crit Care Med 2019 Apr;47(4):566–573. PMID:30664526

23. Chioncel O, Ambrosy AP, Filipescu D, Bubenek S, Vinereanu D, Petris A, Collins SP, Macarie C, Gheorghiade M, Romanian Acute Heart Failure Syndromes Study Investigators. Patterns of intensive care unit admissions in patients hospitalized for heart failure: insights from the RO-AHFS registry. J Cardiovasc Med Hagerstown Md 2015 May;16(5):331–340. PMID:24710424

24. Follath F, Yilmaz MB, Delgado JF, Parissis JT, Porcher R, Gayat E, Burrows N, McLean A, Vilas-Boas F, Mebazaa A. Clinical presentation, management and outcomes in the Acute Heart Failure Global Survey of Standard Treatment (ALARM-HF). Intensive Care Med 2011 Apr;37(4):619–626. PMID:21210078

25. Ohbe H, Matsui H, Fushimi K, Yasunaga H. Epidemiology of Chronic Critical Illness in Japan: A Nationwide Inpatient Database Study. Crit Care Med 2021 Jan 1;49(1):70–78. PMID:33177360

26. Ig C, R A, A S, B X, B L. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff Proj Hope Health Aff (Millwood); 2014 Jul;33(7). PMID:25006139

27. Li X, Wang Z, Zhao W, Shi R, Zhu Y, Pan H, Wang D. Machine learning algorithm for predict the in-hospital mortality in critically ill patients with congestive heart failure combined with chronic kidney disease. Ren Fail 2024 Dec;46(1):2315298. PMID:38357763

28. Messerli FH, Rimoldi SF, Bangalore S. The Transition From Hypertension to Heart Failure: Contemporary Update. JACC Heart Fail 2017 Aug;5(8):543–551. PMID:28711447

29. R de la E, M A, G M, Jc R, P M, D S, J F, E S, G N, M L, C B, E A, A V, J S, V B, Jl G, J N, P C, Aa V, A B-G, J N. Bending oxygen saturation index and risk of worsening heart failure events in chronic heart failure. Eur J Heart Fail Eur J Heart Fail; 2022 Nov;24(11). PMID:36054502

30. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev

Cardiol 2011 Jan;8(1):30–41. PMID:21060326

31. A O, Jm P, D P, Mn B. Mechanisms of Post-critical Illness Cardiovascular Disease. Front Cardiovasc Med Front Cardiovasc Med; 2022 Jul 15;9. PMID:35911546

32. Nelson JE, Cox CE, Hope AA, Carson SS. Chronic critical illness. Am J Respir Crit Care Med 2010 Aug 15;182(4):446–454. PMID:20448093

33. Lin P-C, Huang H-C, Komorowski M, Lin W-K, Chang C-M, Chen K-T, Li Y-C, Lin M-C. A machine learning approach for predicting urine output after fluid administration. Comput Methods Programs Biomed 2019 Aug;177:155–159. PMID:31319943

34. Loftus TJ, Filiberto AC, Ozrazgat-Baslanti T, Gopal S, Bihorac A. Cardiovascular and Renal Disease in Chronic Critical Illness. J Clin Med 2021 Apr 9;10(8):1601. PMID:33918938

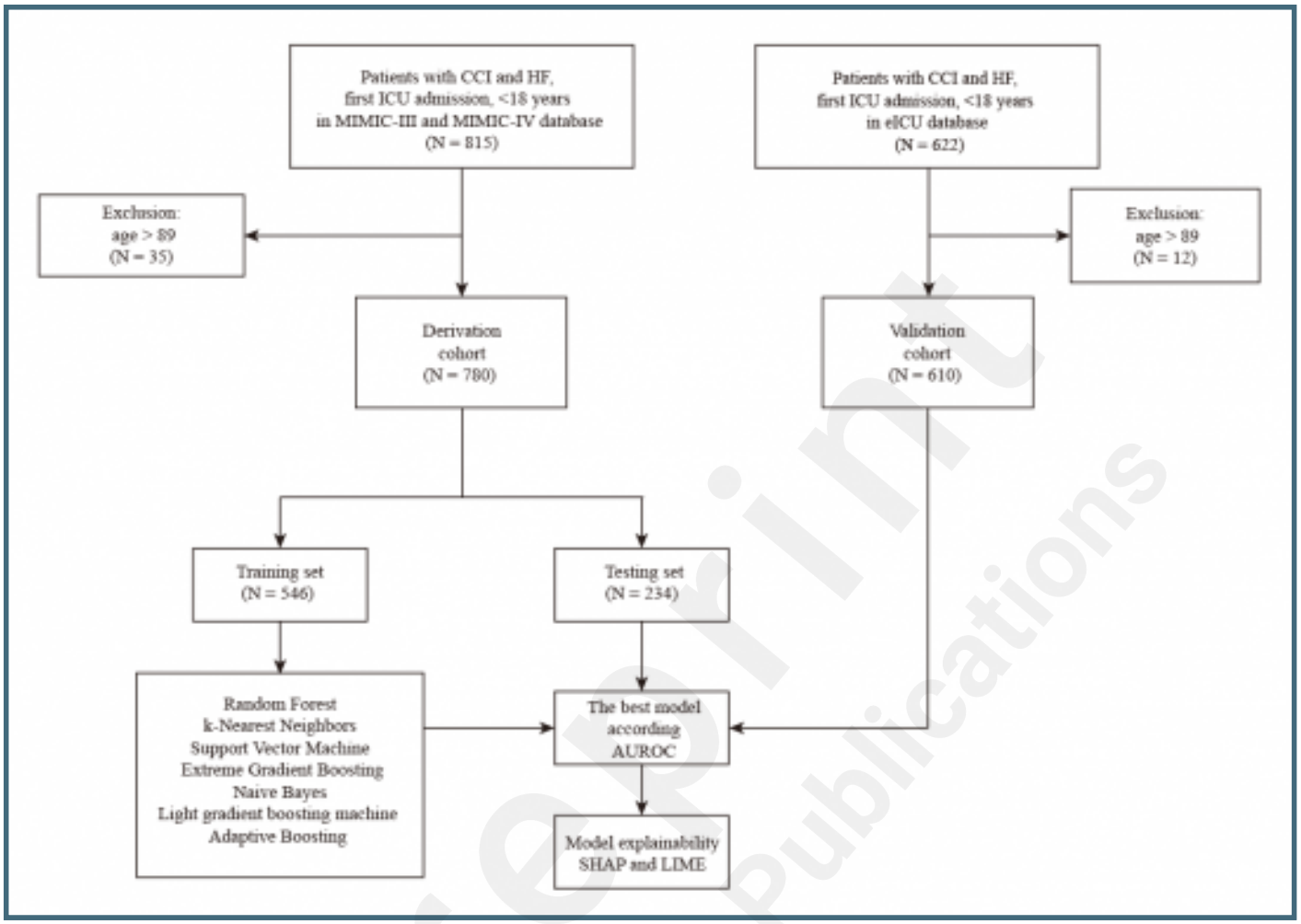
35. Schmidt M, Bailey M, Kelly J, Hodgson C, Cooper DJ, Scheinkestel C, Pellegrino V, Bellomo R, Pilcher D. Impact of fluid balance on outcome of adult patients treated with extracorporeal membrane oxygenation. Intensive Care Med 2014 Sep;40(9):1256–1266. PMID:24934814

36. Melchio R, Rinaldi G, Testa E, Giraudo A, Serraino C, Bracco C, Spadafora L, Falcetta A, Leccardi S, Silvestri A, Fenoglio L. Red cell distribution width predicts mid-term prognosis in patients hospitalized with acute heart failure: the RDW in Acute Heart Failure (RE-AHF) study. Intern Emerg Med 2019 Mar;14(2):239–247. PMID:30276661

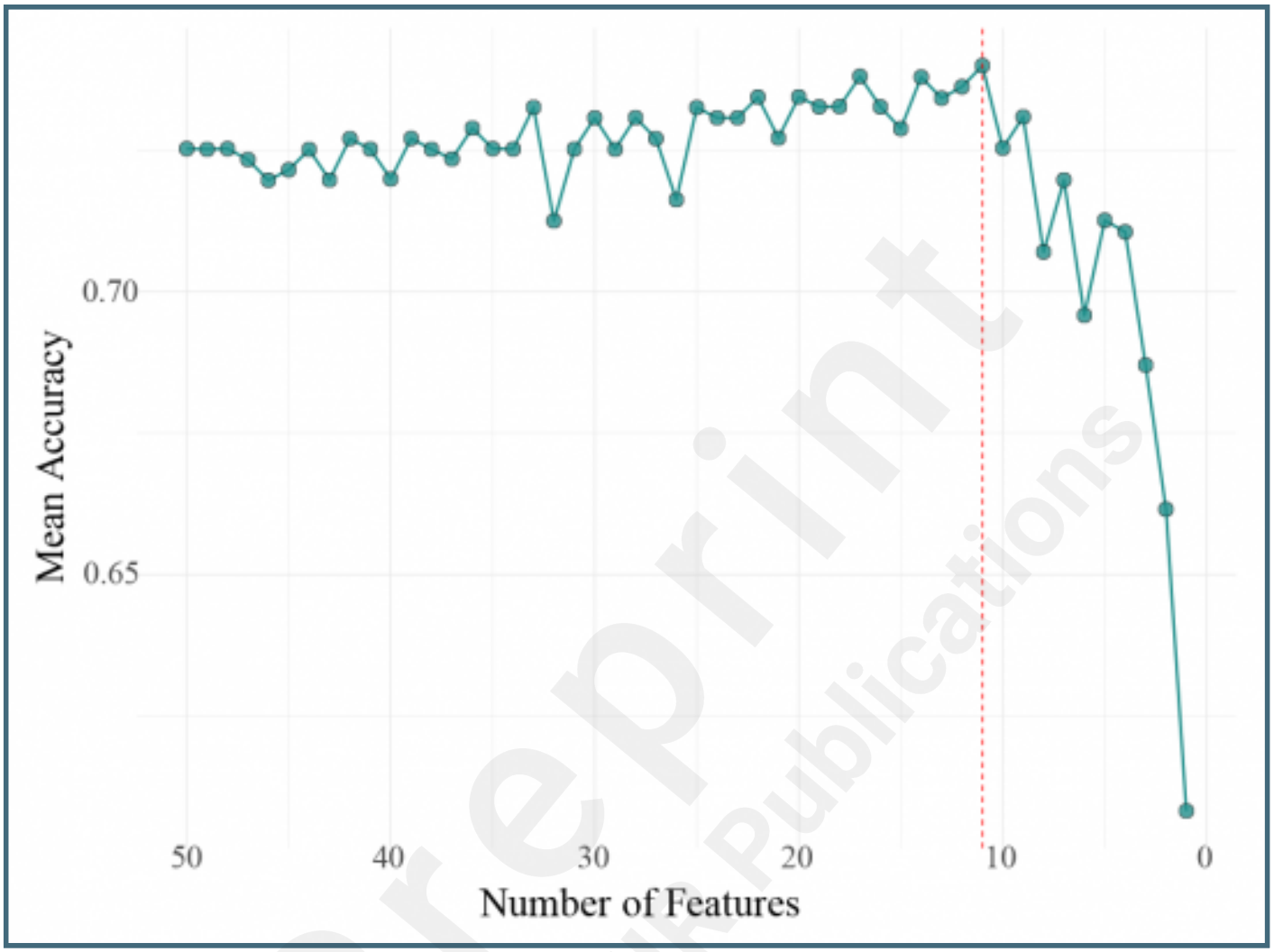
Supplementary Files

Figures

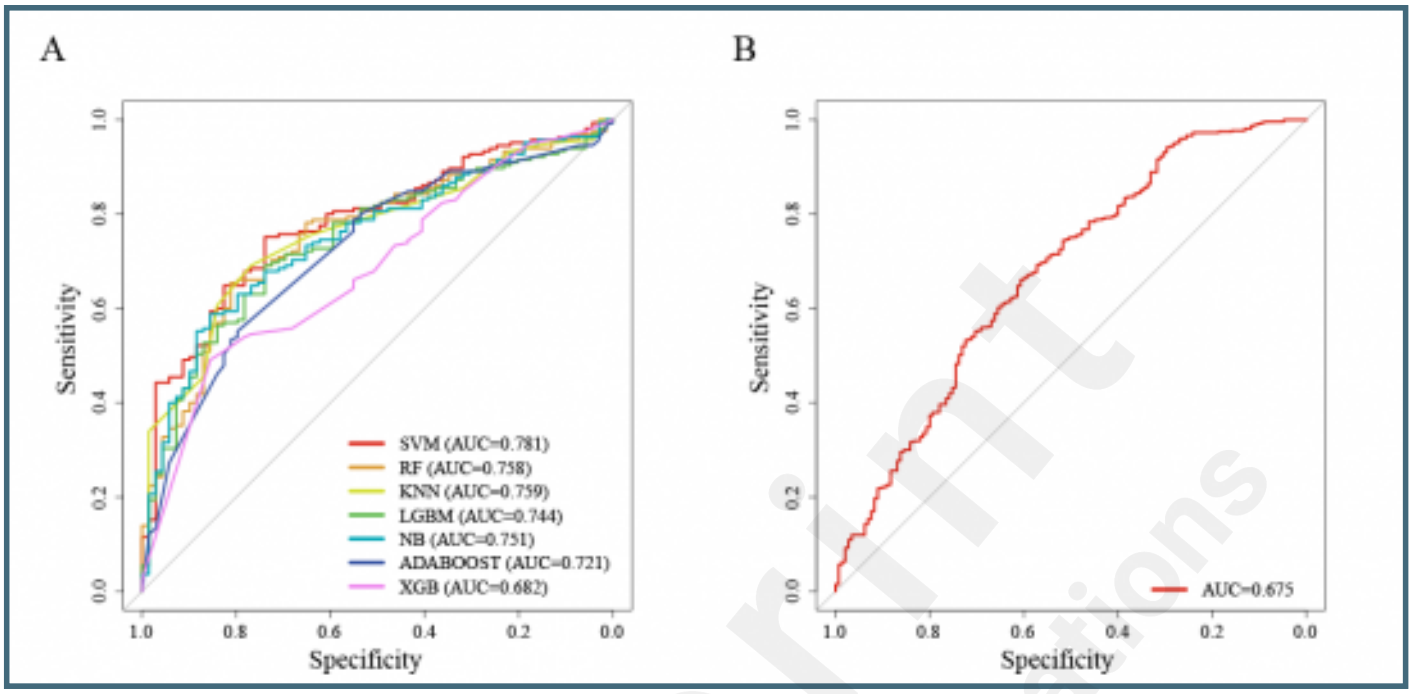
Flowchart of the study.



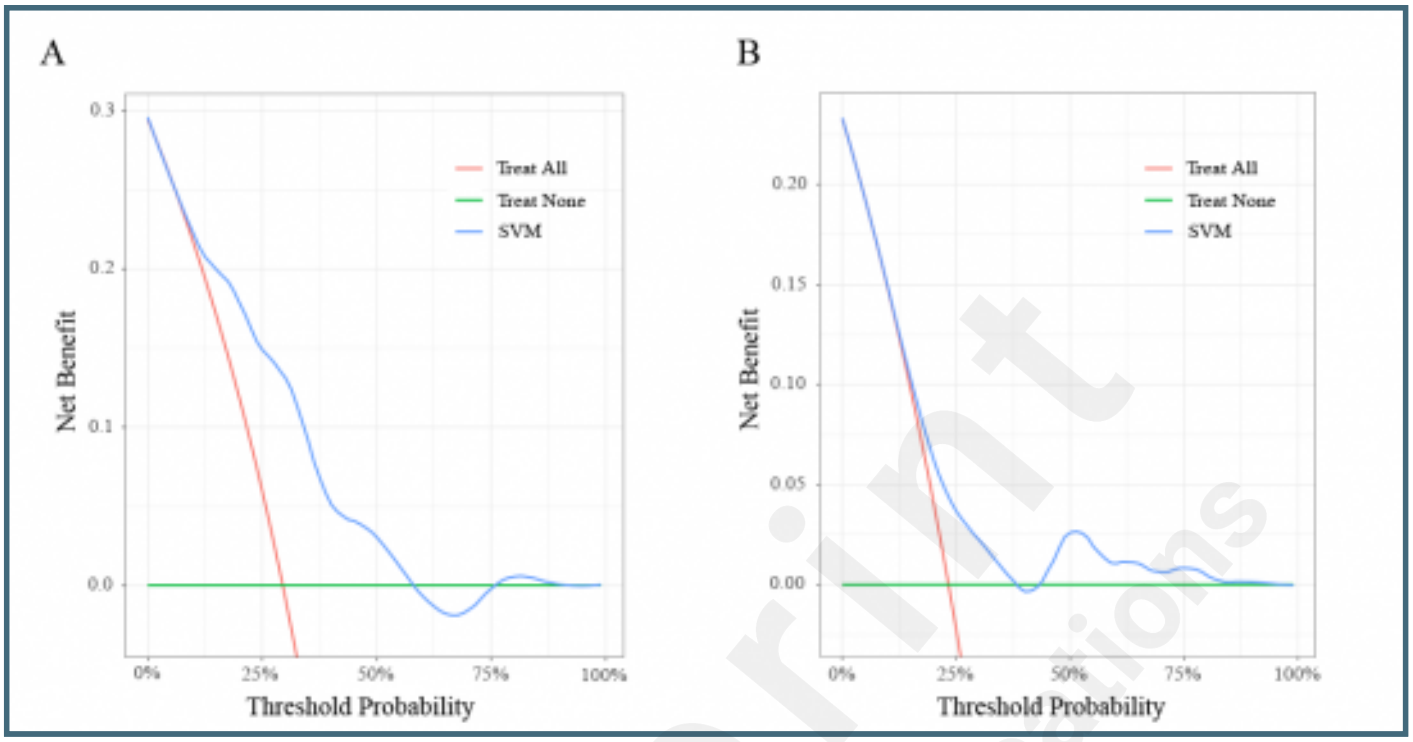
Results of the recursive feature elimination with 10-fold cross-validation method for identifying the predictive features.



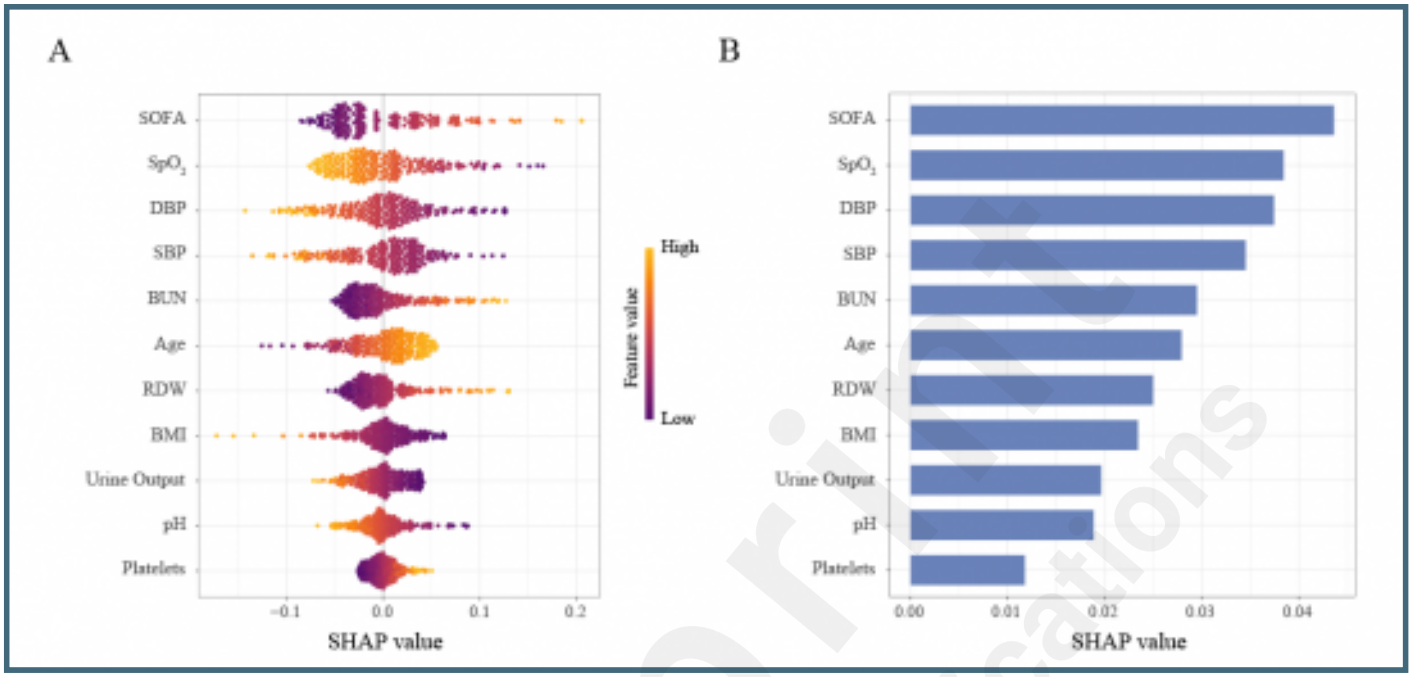
ROC curve of in-hospital mortality. (A) ROC curves comparing the seven machine learning models in the derivation cohort. (B) ROC curves of the validation cohort based on the SVM model.



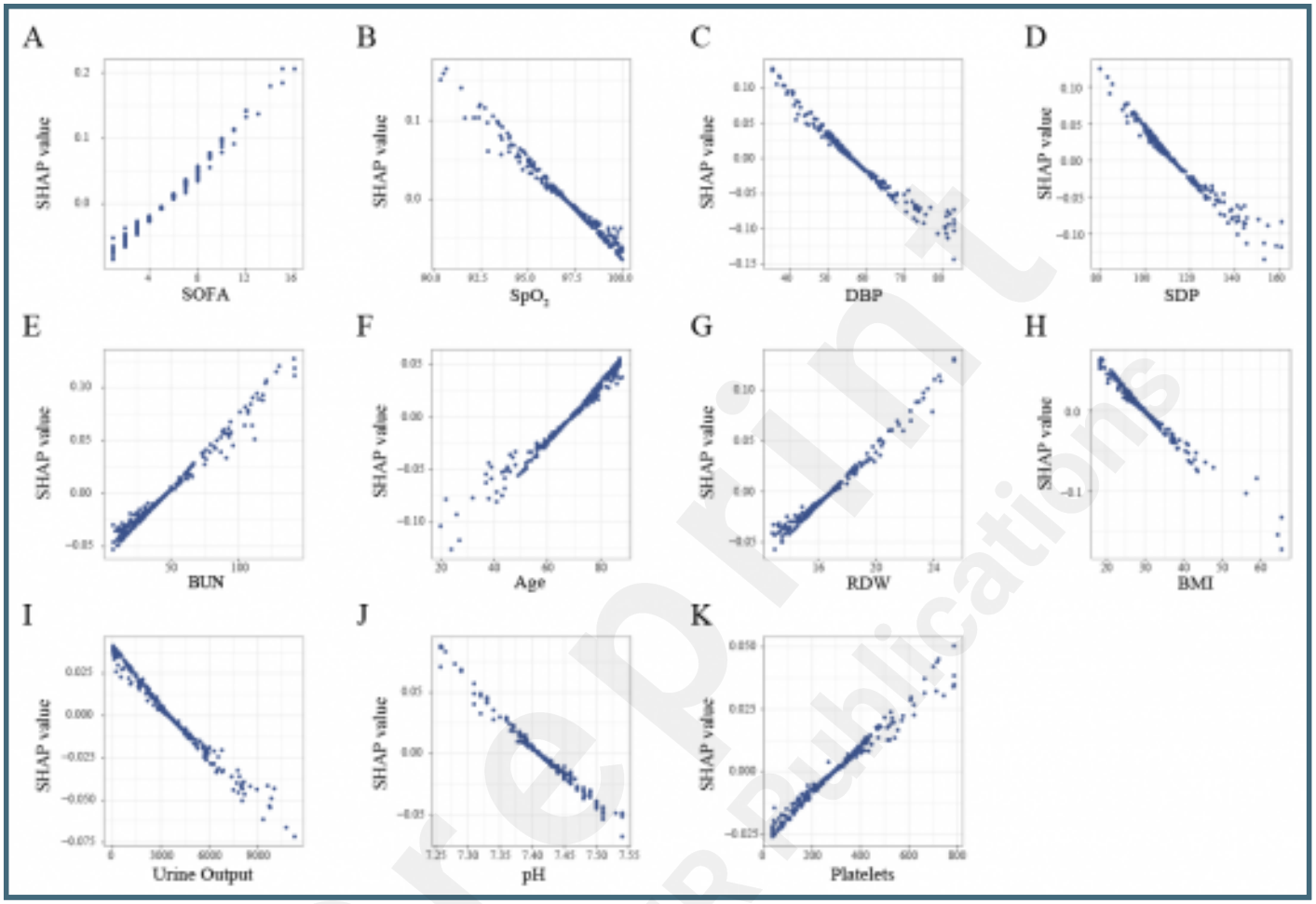
Decision curve analysis of the SVM model in the derivation (A) and validation cohorts (B).



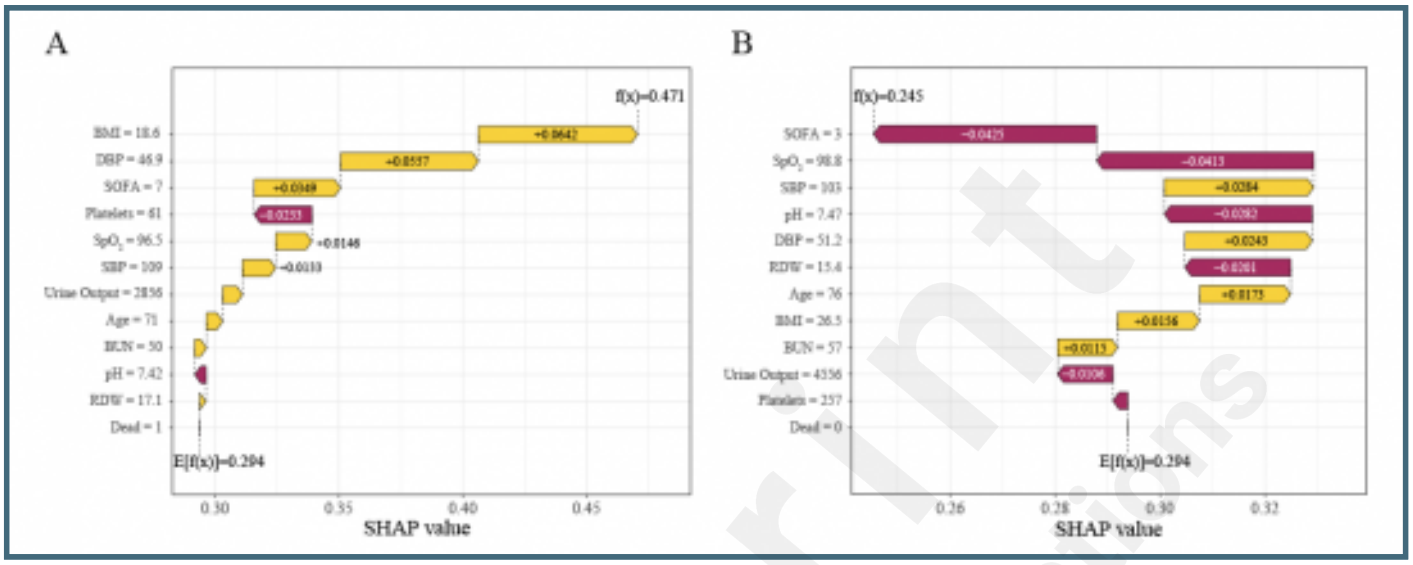
Interpretation of the SVM model. (A) Summary plot based on the SHAP values. A higher SHAP value of a feature suggests a higher risk contribution. Colors on the plot denote the magnitude of the feature values, wherein high values are indicated in yellow, while low values are shown in purple. (B) Importance ranking of the 11 identified features according to the mean (|SHAP value|).



Partial dependence plots of the SVM model based on the SHAP values. (A–K) Effects of SOFA score, SpO2, DBP, SBP, BUN level, age, RDW, BMI, urine output, pH, and platelet count on the output of the SVM prediction model. SHAP values exceeding 0 imply a promoting effect on in-hospital mortality risk.



Interpretation of the SVM model in two patients with distinct prognostic outcomes based on the SHAP method. (A) Patient 1, who had death as the outcome, was correctly predicted to experience death. (B) Patient 3, who had survival as the outcome, was accurately predicted to survive. This plot shows the significant features contributing to pushing the model output. Mortality risk is elevated (yellow arrows) or reduced (red arrows).



Multimedia Appendixes

Baseline characteristics of the validation cohort.
URL: http://asset.jmir.pub/assets/da8d43a1fe2ce5737f91f7b77e04dfc9.docx

Parameter Settings of the six machine learning models.
URL: http://asset.jmir.pub/assets/4e694d3302847bfc8388b7d0abda5122.doc

Calibration curves of the SVM model in the derivation (A) and validation cohorts (B).
URL: http://asset.jmir.pub/assets/2f8f13b9f6f73552985047e47d0ce403.png

Potential interactions of platelet count with SOFA score (A) and RDW (B).
URL: http://asset.jmir.pub/assets/1652ff7f1679c57d14904000b4a82f6f.png

Interpretation of the SVM model in two patients with different prognostic outcomes based on the LIME method. (A) Patient 1, who had death as the outcome, was correctly predicted to experience death. (B) Patient 2, who had survival as the outcome, was accurately predicted to survive. The blue box indicates that the features are risk factors for in-hospital mortality, while the orange box suggests that the features are protective factors.
URL: http://asset.jmir.pub/assets/0cb50b019c28bafca1c739e2b4a82c93.png