Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Mortality Risk Among Non-Cardiac Surgical Patients in the Surgical Intensive Care Unit: A Retrospective Study Based on the MIMIC-IV Database

Mengke Ma
Jinzhou Medical University, The People's Hospital of Liaoning Province
Jiatong Liu
General Hospital of Northern Theater Command
Caiyun Li
The People's Hospital of Liaoning Province
Yingxue Chen
China Medical University
Huishu Jia
General Hospital of Northern Theater Command
Aijie Hou
The People's Hospital of Liaoning Province
Hongzeng Xu
hongzengxu@foxmail.com

Jinzhou Medical University, The People's Hospital of Liaoning Province

Article

Keywords: Non-cardiac surgery, deep learning, machine learning, 30-day mortality, MIMIC database

Posted Date: May 15th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6231511/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 22nd, 2026. See the published version at https://doi.org/10.1038/s41598-026-72135-4.

Abstract

Background

Accurate prediction of mortality risk in non-cardiac surgical patients is critically important for informing clinical decision-making and resource allocation. This study aims to develop a predictive model utilizing deep learning and machine learning to assess mortality risk in this patient population.

Methods

Clinical indicators and electrocardiogram (ECG) signals were extracted from non-cardiac surgery patients in the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. Five machine learning models were developed to estimate 30-day mortality risk: logistic regression (LR), decision tree (DT), light gradient boosting machine (LGBM), extreme gradient boosting (XGBoost), and support vector machine (SVM). To enhance analysis scope, two additional models, backpropagation neural network (BPNN) and recurrent neural network (RNN), were constructed and their performance compared to initial models. SHAP was employed to analyze the optimal model, identifying the most influential risk factors from both global and local perspectives.

Results

Among 4843 MIMIC-IV patients, 526 (10.8%) died within 30 days after non-cardiac surgery. The LGBM model surpassed other machine learning and deep learning models, attaining the highest scores for accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), which were 0.949, 0.925, 0.983, 0.95, and 0.97, respectively. Compared to the traditional revised cardiac risk index (RCRI) score, the LGBM model significantly improved classification accuracy. SHAP analysis revealed that preoperative INR, bicarbonate, BUN, and creatinine levels were the four key variables influencing the performance of the LGBM model.

Conclusion

The LGBM model provides a new, convenient approach for the prognosis and assessment of non-cardiac surgical patients. This tool has the potential to offer effective decision support for clinicians in their risk assessment and clinical decision-making processes.

1. Background

Worldwide, approximately 250 million surgeries are performed annually, with around 200 million being non-cardiac surgeries [1]. The incidence of postoperative complications is 37%, and it has been reported that 1.8 million adults die within 30 days of non-cardiac surgery, with mortality rates ranging from 0.8–

1.5% [2]. In a 7-day cohort study, up to 8% of patients undergoing non-cardiac surgery required intensive care, with in-hospital mortality rates ranging from 1.2–21.5%[3]. Patients undergoing non-cardiac surgery may experience a range of postoperative complications, such as cardiovascular events, cerebrovascular accidents, and bleeding incidents. These risk factors can lead to patient death, especially in those with conditions such as hypertension, heart failure, and severe aortic stenosis [4]. Although the postoperative mortality rate has been reduced through improvements in preoperative management, postoperative death remains a severe complication, posing a threat to human life and safety, and placing significant burdens on society and families. It has even become an important public health issue. Therefore, identifying and predicting these risk factors is crucial for reducing patient mortality.

Currently, several preoperative risk assessment tools have been widely applied in clinical practice, including the revised cardiac risk index (RCRI)[5], the American college of surgeons national surgical quality improvement program's myocardial infarction or cardiac arrest risk calculator (ACS-NSQIP-MICA) [6], and the ACS-NSQIP surgical risk calculator[7]. However, these traditional tools have several limitations. For instance, they primarily depend on physicians' medical history records and statistical methods (such as logistic regression or cox regression), and perform poorly when dealing with complex nonlinear data. Moreover, some tools lack external validation (such as the surgical outcome risk tool, SORT) [8], or are not user-friendly for direct clinical application due to complex data and high manual processing requirements (such as the physiological and operative severity score for the enumeration of mortality and morbidity, POSSUM) [9]. These limitations underscore the urgent need to develop clinically accessible risk assessment tools specifically designed for non-cardiac surgical patients, in order to bridge the existing gap and enhance the comprehensiveness and predictive efficiency of risk assessment.

To address the aforementioned challenges, numerous researchers have leveraged deep learning (DL) for automatic feature extraction from electronic health records (EHRs), achieving remarkable success. For instance, Nguyen et al. successfully employed convolutional neural networks (CNNs) to capture feature information from EHRs, including temporal dynamics, and accurately predicted postoperative mortality using electrocardiogram (ECG) data [10]. With the advancement in computing technology, sophisticated bidirectional neural network models have emerged as state-of-the-art solutions. An et al. proposed a bidirectional long short-term memory (Bi-LSTM) approach [11], which learns temporal dependencies in data from both directions, thereby more comprehensively capturing contextual information in clinical data and facilitating more precise disease risk prediction. Although these methods have significantly improved the accuracy of risk prediction, their requirement for a large number of parameters and substantial computational resources has led to high computing power consumption, failing to meet the demand for rapid response in intensive care unit (ICU) clinical applications.

In recent years, machine learning (ML) technologies have garnered increasing attention in the medical field, particularly for predicting postoperative risks. One such technique, extreme gradient boosting (XGBoost), has demonstrated remarkable predictive performance in handling non-linear data, allowing for more accurate simulation of complex clinical data and thus providing better support for clinical

decision-making. Therefore, the application of ML techniques capable of capturing non-linear relationships holds great significance for improving postoperative management in non-cardiac surgery patients. This study aims to construct an optimal predictive model for 30-day mortality after non-cardiac surgery using ML technologies. The objective is to provide new perspectives and approaches for clinical management and treatment of non-cardiac surgery patients, ultimately reducing mortality rates, improving healthcare quality, and optimizing medical resource allocation in surgical intensive care unit (SICU).

2. Methods

This section presents the flowchart for the proposed methodology, as illustrated in Fig. 1.

2.1 Database

This study utilized data from the MIMIC-IV database[12]. The MIMIC-III database, spanning from 2001 to 2012, included over 40,000 ICU patient cases from the Beth Israel Deaconess Medical Center. MIMIC-IV expanded upon MIMIC-III and updated patient records from 2008 to 2019. This dataset adheres to the standards of the International Classification of Diseases, 9th Revision (ICD-9) coding system [13]. The MIMIC series databases provide de-identified and publicly available data, therefore ethical approval was not required for this study[14].

2.2 Participants

According to the 2022 ESC/ESA non-cardiac surgery guidelines, non-cardiac surgeries are categorized into low-risk, moderate-risk, and high-risk groups. Moderate-risk surgeries (such as peripheral artery angioplasty, abdominal, endovascular aneurysm repair, carotid endarterectomy, head and neck, gynecological, neurosurgery/major orthopedic surgery, and major urological surgery) are associated with moderate complication risk, while high-risk surgeries (such as adrenalectomy, aortic or major vessel resection, pancreaticoduodenectomy, liver resection, biliary, esophagectomy, pneumonectomy, lung transplantation, radical cystectomy, and bowel perforation repair) carry higher morbidity and mortality risks[15]. This study focuses on elective surgical patients admitted to the surgical ICU, excluding potential confounders and emergency surgery cases. The exclusion criteria were as follows: 1) only the first ICU admission was included; 2) patients with ICU stays < 24 hours were excluded[16]; 3) patients aged ≥ 18; 4) data from cardiac and other low-risk surgeries were excluded; 5) to avoid duplicate data. Ultimately, 4843 patients met the criteria and were included in the study.

2.3 Variable Selection

In this study, software tools such as PostgreSQL 13 and pgAdmin 4, along with Python 3.9.7, were utilized to screen and organize data for variable selection[17]. Through expert consultation and a comprehensive review of the literature, it was determined that 40 clinical features should be included in the model as predictive characteristics for each patient's medical record. SQL queries were executed to extract the following clinical information:

(1) Laboratory results on the first day of ICU admission: red blood cell count, white blood cell count, platelet count, hematocrit, hemoglobin, mean corpuscular hemoglobin concentration, mean corpuscular volume, red cell distribution width, bicarbonate, creatinine, blood urea nitrogen, thrombin time, activated partial thromboplastin time (APTT), calcium, sodium, potassium, international normalized ratio (INR), SAPS II score, and RCRI score.

(2) Demographic information: including patient race, age, sex, alcohol consumption, and weight.

(3) Vital signs on the first day of ICU admission: temperature, respiratory rate, mean arterial pressure, heart rate, oxygen saturation, and blood glucose levels.

(4) In this study, we obtained the electrocardiogram (ECG) machine signals recorded on the first day of the intensive care unit (ICU) for the patients. These parameters include QRS wave duration, P wave axis, QRS wave axis, T wave axis, and T wave end time.

The QRS wave duration refers to the time interval from the onset to the end of the QRS complex on an ECG. Under normal circumstances, the QRS wave duration should be less than 120 milliseconds. The P wave axis indicates the electrical axis direction of the P wave on an ECG, reflecting the direction of atrial depolarization. The QRS wave axis represents the electrical axis direction of the QRS complex on an ECG, reflecting the direction of ventricular depolarization. The T wave axis indicates the electrical axis direction of the T wave on an ECG, reflecting the direction of ventricular repolarization. The T wave end time refers to the time point at which the T wave ends on an ECG, which is used to assess the duration of ventricular repolarization. These parameters are of significant importance in ECG analysis and can be used to evaluate the cardiac electrophysiological status and potential heart diseases.

(5) Comorbidities: diabetes, congestive heart failure, myocardial infarction, hypertension, cerebrovascular disease.

2.4 Data Preprocessing

The outliers in the data were removed directly, and the multiple imputation method was used to address the issue of missing data in the predictor variables[18]. Variables with a missing rate exceeding 30% were excluded to avoid potential bias. Due to the large number of postoperative survivors compared to postoperative deaths, resulting in data imbalance, the synthetic minority oversampling technique (SMOTE)[19], an oversampling technique for imbalanced data, was used. It balanced the dataset by generating new synthetic samples through interpolation between minority class samples. A new training dataset was generated using SMOTE, which increased the ratio of in-hospital death events to non-death events from 1:8.33 to 1:1. The new data was similar to the original because they were generated based on the original features.

2.5 Feature Selection

The random forest (RF) algorithm significantly enhances classification accuracy while reducing the number of unnecessary, irrelevant, and redundant features [20]. The random forest (RF) algorithm was

used to calculate feature importance, laying the groundwork for the subsequent feature selection stage. Additionally, visualizing feature importance helped deepen the understanding and interpretation of the model's predictions. Figure 2 shows the 12 most important features calculated using the RF algorithm: INR, age, BUN, creatinine, PT, respiration, temperature, P axis, QRS wave duration, oxygen saturation, bicarbonate, and cerebrovascular disease.

2.6 Model Construction and Evaluation

In this study, two types of neural network models suitable for classification tasks were selected, namely the backpropagation neural network (Fig. 3) and the recursive neural network (Fig. 4). A three-layer shallow network structure was used to reduce the complexity of deep learning models and enhance their interpretability. Dropout was employed as an effective regularization technique to prevent overfitting [21]. Both deep neural networks used binary cross-entropy loss (BCELoss) as the loss function, which is more suitable for binary classification tasks [22]. For comparative experiments, logistic regression (LR), decision tree (DT), light gradient boosting machine (LGBM), extreme gradient boosting (XGBoost), and support vector machine (SVM) were also established. Participants were randomly divided into two cohorts: one for fitting the predictive model (the training cohort), accounting for 70% of the total samples (n = 3139), and the other for evaluating the model (the testing cohort), accounting for 30% of the participants (n = 1344). In this study, a 10-fold cross-validation method was used to estimate the performance of machine learning (ML) and deep learning (DL)[23]. This method divides the training data into 10 equal parts, trains the model on 9 parts, and validates it on the remaining part, providing 10 independent performance estimates to help accurately assess the model's predictive ability. GridSearchCV was employed to meticulously tune hyperparameters for various machine learning models[24]. Through this process, the optimal parameter combinations were determined, and the evaluation metrics from GridSearchCV were recorded, which are presented in Table 2.

To evaluate the efficacy of different machine learning models in predicting the risk of death after non-cardiac surgery, a range of evaluation metrics were used, including accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUROC). The AUC and F1 score were taken as the core indicators for measuring model performance.

2.7 Statistical Analysis

This study assumes that the collected data follows a normal distribution pattern. Categorical variables are presented as percentages, while continuous variables are described as mean ± standard deviation. The Kolmogorov-Smirnov test was applied to continuous variables, and the Chi-square test was used for categorical variables. All statistical analyses were performed using Python, with a significance level set at $P < 0.05$[25].

3. Results

3.1 Baseline Characteristics

As shown in Table 1, this study included a total of 4843 participants, with 526 (10.9%) deaths occurring within 30 days after non-cardiac surgery. Of the participants, 46.4% were female and 53.6% were male, with an average age of 62.8 ± 16.6 years for survivors and 72.1 ± 15.0 years for deceased patients. Elderly and white patients had a higher incidence of postoperative mortality. Hypertension was the most common comorbidity, affecting 2488 individuals, and it was more prevalent in patients who died after non-cardiac surgery compared to myocardial infarction, heart failure, and cerebrovascular disease. Both QRS wave duration and P-axis were significantly associated with postoperative mortality in non-cardiac surgery patients.

Table 1
Baseline characteristics of the patients

	Total	**Non-survivors**	**Survivors**	***p* value**
	N = 4843	***N = 4317***	***N = 526***	
Age	63.77 (16.69)	62.8 (16.6)	72.1 (15.0)	<0.001
Race				<0.001
ASIAN	280 (5.8)	256 (5.93%)	24 (4.56%)	
BLACK	154 (3.2)	144 (3.34%)	10 (1.90%)	
OTHER	1259 (26.0)	1053 (24.4%)	206 (39.2%)	
WHITE	3150 (65.0)	2864 (66.3%)	286 (54.4%)	
Gender				0.621
Female	2248 (46.4)	1998 (46.3%)	250 (47.5%)	
Male	2595 (53.6)	2319 (53.7%)	276 (52.5%)	
Alcohol abuse				0.439
Never	3770 (77.8)	3368 (78.0%)	402 (76.4%)	
Ofen	1073 (22.2)	949 (22.0%)	124 (23.6%)	
Weight (kg)	83.18 (25.20)	83.5 (25.0)	80.4 (26.7)	0.011
Hemoglobin(%)	11.06 (2.18)	11.1 (2.15)	10.7 (2.36)	<0.001
Bicarbonate (mmol/L)	22.86 (4.06)	23.0 (3.91)	21.6 (4.94)	<0.001
Calcium(mg/dL)	8.24 (0.82)	8.25 (0.80)	8.17 (0.96)	0.069
Chloride(mg/dL)	104.91 (5.47)	105 (5.34)	105 (6.42)	0.189
Sodium(mmol/L)	138.72 (4.42)	139 (4.26)	139 (5.56)	0.679
Potassium (mmol/L)	4.12 (0.67)	4.11 (0.65)	4.23 (0.80)	0.001
Creatinine(mg/dL)	1.13 (1.20)	1.09 (1.20)	1.43 (1.14)	<0.001
Bun(mg/dL)	21.42 (17.39)	20.3 (16.0)	30.8 (24.0)	<0.001
Platelet($\times 10^9$/L)	218.24 (103.39)	219 (102)	213 (117)	0.274
RBC($\times 10^{12}$/L)	3.68 (0.73)	3.70 (0.72)	3.54 (0.84)	<0.001
WBC($\times 10^9$/L)	12.86 (11.36)	12.6 (11.1)	14.7 (13.2)	0.001
MCV	90.59 (6.38)	90.4 (6.28)	92.2 (6.98)	<0.001

	Total	Non-survivors	Survivors	*p* value
	N = 4843	***N = 4317***	***N = 526***	
MCHC	33.35 (1.50)	33.4 (1.49)	33.0 (1.58)	<0.001
RDW	14.50 (1.95)	14.4 (1.89)	15.3 (2.25)	<0.001
PT	14.63 (6.32)	14.4 (5.82)	16.8 (9.21)	<0.001
PTT	34.23 (19.72)	33.5 (18.8)	40.3 (25.4)	<0.001
INR	1.33 (0.62)	1.30 (0.58)	1.54 (0.89)	<0.001
Heart rate	87.68 (20.61)	87.4 (20.5)	90.4 (21.3)	0.002
MDP	86.02 (18.82)	86.3 (18.7)	83.8 (19.9)	0.006
Glucose (mmol/L)	141.56 (43.43)	141 (42.8)	144 (48.2)	0.310
Respiration	18.55 (5.95)	18.4 (5.77)	20.0 (7.11)	<0.001
Oxygen saturation	97.22 (3.93)	97.3 (3.68)	96.5 (5.49)	0.001
Temperature	36.69 (0.83)	36.7 (0.79)	36.5 (1.06)	<0.001
QRS wave duration	96.28 (20.01)	95.7 (19.3)	101 (24.8)	<0.001
T wave end time	641.72 (610.96)	627 (128)	763 (1814)	0.085
P wave axis	3846.27 (10037.48)	3441 (9551)	7175 (12918)	<0.001
QRS wave axis	46.26 (862.96)	36.0 (647)	131 (1848)	0.245
T wave axis	158.92 (1956.43)	129 (1701)	404 (3384)	0.067
SAPS II score	6.23 (2.37)	6.20 (2.43)	6.48 (1.81)	0.001
RCRI score	0.95 (0.97)	0.91 (0.95)	1.29 (1.09)	<0.001
Myocardial infarct				<0.001
NO	4407 (91.0)	3959 (91.7%)	448 (85.2%)	
YES	436 (9.0)	358 (8.29%)	78 (14.8%)	
Congestive heart failure				<0.001
NO	4174 (86.2)	3785 (87.7%)	389 (74.0%)	
YES	669 (13.8)	532 (12.3%)	137 (26.0%)	
Cerebrovascular disease				<0.001
NO	3808 (78.6)	3435 (79.6%)	373 (70.9%)	

	Total	Non-survivors	Survivors	*p* value
	N = 4843	***N = 4317***	***N = 526***	
YES	1035 (21.4)	882 (20.4%)	153 (29.1%)	
Diabetes				0.656
NO	3825 (79.0)	3414 (79.1%)	411 (78.1%)	
YES	1018 (21.0)	903 (20.9%)	115 (21.9%)	
Hypertension				0.068
NO	2355 (48.6)	2079 (48.2%)	276 (52.5%)	
YES	2488 (51.4)	2238 (51.8%)	250 (47.5%)	

Table 2 GridSearchCV optimization for machine learning models.

ML Models	Ranging Value of GridSearchCV	Optimal Parameter Combination
LR	'C': [0.01, 0.1, 1, 10, 100]	'C': 0.1
	'penalty': ['l1', 'l2']	'penalty': l1
	'solver': ['liblinear', 'saga']	'solver': liblinear
	'n_estimators': [10,50, 100, 200]	'n_estimators': 100
LGBM	'learning_rate': [0.01, 0.1, 0.2]	'learning_rate': 0.2
	'num_leaves': [20, 50, 70]	'num_leaves': 70
DT DT	'max_depth': [3, 5, 10]	'max_depth': 10
	'min_samples_split': [2, 5, 10]	'min_samples_split': 5
	'min_samples_leaf': [1, 2, 4]	'min_samples_leaf': 2
SVM	'C': [0.1, 1, 10, 100]	'C': 10
	'kernel': ['linear', 'rbf', 'poly']	'kernel': rbf
	'gamma': ['scale', 'auto']	'gamma': auto
	'max_iter': [1000, 2000]	'max_iter': 1000
XGBoost CatBoost	'learning_rate': [0.01, 0.05, 0.1]	'learning_rate': 0.1
	'max_depth': [3, 5, 7]	'max_depth': 7
	'n_estimators': [100, 200, 300]	'n_estimators': 200

3.2 Model Comparison

We performed a comparative analysis of deep learning classifiers and machine learning classifiers on the clinical record dataset of non-cardiac surgery patients. Figure 5 presents the accuracy, precision, recall, and F1 score of different classifier models. For a more comprehensive evaluation of the model's performance in relation to AUC, the ROC curve is shown in Fig. 6. Compared to other models, the LGBM model demonstrated superior AUC performance, showing a higher balance between sensitivity (True Positive Rate, TPR) and specificity (True Negative Rate, TNR). This resulted in a high true positive rate while effectively reducing false positives. The results indicate that LGBM achieved the highest scores for AUC (0.96) and F1 score (0.949), demonstrating excellent performance in this task. The F1 score represents the balance between precision and recall, while AUC evaluates the model's ability to distinguish between positive and negative instances. Among the models compared, LGBM significantly

outperformed the others on these evaluation metrics, highlighting its overall superior performance and predictive accuracy in this classification task. The remaining models still demonstrated good predictive capabilities, as follows: XGBoost (AUC 0.95, F1 0.949); BPNN (AUC 0.94, F1 0.904); RNN (AUC 0.89, F1 0.858); DT (AUC 0.95, F1 0.808); SVM (AUC 0.82, F1 0.929); LR (AUC 0.74, F1 0.77). In contrast, the RCRI score had an AUC of 0.59 and an F1 score of 0.443. This further confirms that the LGBM model outperforms the traditional RCRI score in terms of both accuracy and predictive ability.

3.3 Explainability Analysis

Based on the LGBM model, which exhibited optimal predictive performance, an explainability analysis was conducted using shapley additive explanations (SHAP) values to interpret the non-cardiac surgery 30-day mortality prediction model. The shapley value, from cooperative game theory, is now widely used to identify important features that affect predictions [26]. It has been adapted as a model-agnostic approach to explain predictions. Rooted in shapley values, this approach uses a game-theoretic perspective to measure feature importance, effectively navigating the opacity often associated with machine learning models and achieving consistent interpretability. This method provides explanations from both global and local perspectives, helping ICU healthcare providers better understand the disease progression in non-cardiac surgery patients and develop more personalized and effective treatment plans.

3.3.1 Global Interpretation

In global interpretation, SHAP values can reveal the importance of features and describe their relationship with the output and their interactions. Figure 7 displays the SHAP values, showing the cumulative impact ranking of features. Each feature’s horizontal position on the plot indicates whether it is associated with an increase or decrease in the prediction trend. In this visualization, blue indicates a positive contribution to the prediction, while purple represents a negative effect. Features such as INR, age, BUN, creatinine, PT, respiration, P axis, QRS wave duration, and cerebrovascular disease positively impact the prediction of mortality risk. In contrast, temperature, bicarbonate, and oxygen saturation have a negative effect on survival prediction.

The bar chart ranks the features based on their average absolute SHAP values, from highest to lowest. The top four features are INR, BUN, bicarbonate, and creatinine. This ranking demonstrates the contribution of each feature to the model’s performance. The higher the absolute value of the SHAP values, the greater the feature's importance and its impact on the model’s output.

3.3.2 Local Interpretation

In Fig. 8, f(x) represents the probability of 30-day mortality after non-cardiac surgery for an individual patient sample, calculated by summing the SHAP values of the patient's features. The baseline corresponds to the average value of f(x) across all samples. This local interpretation plot shows how different features contribute to the deviation from the baseline in the model’s output. Red features

increase the output, while blue features reduce it. Notably, for this specific patient, our LGBM model predicts a higher mortality risk compared to the baseline. The most significant factors affecting this patient's outcome are INR, respiration, BUN, and bicarbonate. By providing explicit interpretations based on specific patient data, healthcare providers can understand which clinical indicators most significantly contribute to the patient's mortality risk, thereby enabling precision medicine.

4. Discussion

In this study, we developed and validated two deep learning algorithms and five traditional machine learning algorithms for predicting 30-day mortality after non-cardiac surgery using the MIMIC-IV database. Among these models, the LGBM algorithm outperformed BPNN, RNN, SVM, XGBoost, LR, and DT. The LGBM model demonstrated robust performance, with strong identification and calibration abilities, showing considerable net benefits in clinical practice. By analyzing the LGBM model using the SHAP method, we achieved optimal performance with relatively minimal computational resources and data requirements. This model has the potential to become an important tool for predicting mortality risk, providing valuable assistance to clinicians in decision-making.

The revised cardiac risk index (RCRI), last revised in 1999, includes six key factors: preoperative serum creatinine level > 2 mg/dL, coronary artery disease, cerebrovascular events, insulin use, persistent heart failure, and high-risk surgeries such as abdominal, thoracic, or groin vascular procedures[27]. Subsequent studies have validated this index, demonstrating its moderate predictive ability for cardiac mortality and nonfatal cardiac arrest in non-cardiac surgery patients [28]. Our study found that the LGBM model outperformed the traditional RCRI scoring system, highlighting its superior predictive accuracy. This finding suggests that advanced machine learning models can significantly enhance the prediction of postoperative outcomes compared to conventional scoring systems.

Preoperative laboratory indicators play a crucial role in the perioperative period. One study showed that highly sensitive cardiac troponin T, total cholesterol, and high-density lipoprotein (HDL) levels were highly correlated with 6-month mortality after non-cardiac surgery, and a random forest algorithm with an AUC of 0.96 was used to model these factors[29]. This underscores the importance of incorporating comprehensive laboratory data into risk prediction models. By adjusting these laboratory indicators preoperatively, the risk of 30-day mortality during the perioperative period may be reduced. These identified risk factors can serve as the basis for a comprehensive preoperative anesthetic assessment, providing essential information to assist anesthesiologists in making informed clinical decisions.

Elderly patients undergoing non-cardiac surgery have a higher likelihood of complications and poorer general conditions compared to middle-aged and younger patients[30]. Significantly higher complication rates and mortality have been reported among patients aged 80 and above compared to younger groups[31]. Similarly, Troisi determined a significant uptick in death risks for elderly individuals receiving intensive care[32]. A multicenter study showed that prothrombin (PT) is the most abundant coagulation factor in blood. Prolonged PT can lead to excessive bleeding during surgery, which may negatively affect

prognosis[33]. Since prothrombin is produced by the liver, its abnormal levels may suggest liver dysfunction. Patients on warfarin are more likely to have underlying cardiac or cerebrovascular diseases. The therapeutic window for the INR, which is an indicator of warfarin dosage, is narrow, and an INR between 5.0–9.0 increases the risk of bleeding. In such cases, the patient is likely to be in an emergency situation. Therefore, PT and INR are closely related to postoperative prognosis. Additionally, elevated levels of creatinine and BUN are commonly linked to acute kidney injury (AKI), which in turn is associated with a variety of postoperative complications, such as extended hospital stays and higher medical expenses [34]. Continuous monitoring of perioperative vital signs helps healthcare providers detect changes in a patient's condition early and intervene in a timely manner, which is crucial for reducing perioperative mortality and the incidence of major postoperative complications.

In our study, 20.2% of the patients who died had preoperative atrial fibrillation as indicated by their electrocardiogram (ECG). Atrial fibrillation is a common condition in non-cardiac surgery, potentially leading to hemodynamic instability and increasing the risk of stroke, thereby prolonging hospital stay and increasing costs [35]. The extended QRS wave duration was found to be a significant predictor of major adverse cardiac events (MACE) in elderly hypertensive patients with heart failure [36]. Patients with a QRS duration > 120ms had a higher incidence of cardiac adverse events. Prolonged QRS duration is an independent predictor of myocardial infarction or coronary heart disease-related mortality and an independent risk factor for heart failure patients' cardiovascular or all-cause mortality.

Precise risk prediction is essential for elderly patients undergoing non-cardiac surgery, as they are more susceptible to postoperative complications and mortality due to the presence of multiple comorbidities. Effective risk prediction models can significantly enhance perioperative management and improve outcomes by facilitating targeted interventions and optimizing clinical decision-making. In the preoperative phase, identifying high-risk patients through risk prediction models allows for early, tailored interventions, such as optimizing treatment plans and adjusting medications, thereby reducing the incidence of perioperative complications and enhancing surgical safety. During the perioperative period, real-time risk prediction enables dynamic monitoring and early detection of potential risks, facilitating timely interventions and reducing postoperative mortality and hospital stay duration. In the postoperative phase, risk prediction models aid in closely monitoring high-risk patients, enabling prompt management of complications. Additionally, these models provide psychological support to patients and their families, enhancing satisfaction and compliance.

In summary, comprehensive risk prediction and management strategies can significantly improve postoperative outcomes and quality of life for elderly patients undergoing non-cardiac surgery. Future work should focus on integrating advanced predictive analytics into clinical workflows to further enhance patient care.

It is crucial to acknowledge the limitations of this study. First, data with missing values exceeding 30% were excluded, which may introduce bias and result in an incomplete dataset. Second, the model was developed based on ICU admission data from the first day, lacking subsequent dynamic data. Third, the

data in this study were sourced from a specific surgical ICU population within the MIMIC-IV database, which may limit its generalizability. Therefore, larger-scale and multicenter studies are needed to validate the model's effectiveness in different settings. Fourth, the dataset used in this study only includes populations of European and American descent, without considering differences among various racial and ethnic groups. Further research should aim to expand the sample's racial diversity and data size. Lastly, the current implementation and visualization of the model rely on Python software, which may pose challenges for those unfamiliar with the Python environment.

5. Conclusion

This study developed a model aimed at predicting the 30-day mortality rate for patients undergoing non-cardiac surgery. By using SHAP values for both global and local explanations, we analyzed the risk factors contributing to mortality in non-cardiac surgery. This analysis will help physicians identify key risk factors, which, if controlled through interventions, can significantly improve survival outcomes for non-cardiac surgery patients, thereby enabling precision medicine. In conclusion, our study provides a new, non-invasive strategy for assessing the prognosis of non-cardiac surgeries.

Abbreviations

ECG
Electrocardiogram
SICU
Surgical Intensive Care Unit
LGBM
Light gradient boosting machine
XGBoost
Extreme gradient boosting
SVM
Support vector machine
DT
Decision tree
RF
Random forest
BPNN
Backpropagation neural network
RNN
Recurrent neural network
LR
Logistic regression
SHAP
Shapley additive explanations

SAPS II
Simplified acute physiology score II
RCRI
Revised cardiac risk index
PT
Prothrombin time
INR
International normalized ratio
APTT
Activated partial thromboplastin time
HDL
High-density lipoprotein
AUC
Area under the curve
ROC
Receiver operating characteristic
SMOTE
Synthetic minority over-sampling technique
BIDMC
Beth Israel Deaconess Medical Center
MIMIC-IV
Medical information mart for intensive care IV
NSQIP
National surgical quality improvement program
ACS
American college of surgeons
BUN
Blood urea nitrogen
NCS
Non-cardiac surgery

Declarations

Contributions

MKM contributed to the study conception and design. Material preparation, data collection and analysis were performed by JTL and HSJ. The draft of the manuscript was written by CYL and YXC. HZX and AJH revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Mengke Ma has completed the online test and received a certification number, which is Record ID: 59607662. However, we adhered to fundamental ethical research principles to ensure the validity and fairness of the investigation.

Consent for publication

All authors have thoroughly reviewed the full manuscript and agreed to the publication of the final version.

Declaration of Competing Interest

The authors have no competing interest to declare.

Funding

This work received funding from the Natural Science Foundation of Liaoning Province under Grant (No. 2023-MS-054).

Author Contribution

MKM contributed to the study conception and design. Material preparation, data collection and analysis were performed by JTL and HSJ. The draft of the manuscript was written by CYL and YXC. HZX and AJH revised the manuscript. All authors read and approved the final manuscript.

Data Availability

The datasets analyzed during the current study can be obtained from (https://physionet.org/content/mimiciv/2.0/).

References

1. Santangelo, G. et al. Risk of cardiovascular complications during non-cardiac surgery and preoperative cardiac evaluation. *Trends Cardiovasc. Med.* **32**, 271–284 (2022).
2. Haynes, A. B. et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N. Engl. J. Med.* **360**, 491–499 (2009).
3. Pearse, R. M. et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* **380**, 1059–1065 (2012).
4. Nuttall, G. A. et al. Perioperative Mortality: A Retrospective Cohort Study of 75,446 Noncardiac Surgery Patients. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes* 8, 435–442 (2024).

5. TH, L. et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* **100**, 1043–1049 (1999).
6. Davenport, D. L., Bowe, E. A., Henderson, W. G., Khuri, S. F. & Mentzer, R. M. National Surgical Quality Improvement Program (NSQIP) Risk Factors Can Be Used to Validate American Society of Anesthesiologists Physical Status Classification (ASA PS) Levels. *Ann. Surg.* **243**, 636–644 (2006).
7. Bilimoria, K. Y. et al. Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons. *J. Am. Coll. Surg.* **217**, 833–842e833 (2013).
8. Oakland, K. et al. External validation of the Surgical Outcome Risk Tool (SORT) in 3305 abdominal surgery patients in the independent sector in the UK. *Perioperative Med.* **10** (2021).
9. Prytherch, D. R. et al. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *Br. J. Surg.* **85**, 1217–1220 (1998).
10. Nguyen, P., Tran, T., Wickramasinghe, N. & Venkatesh, S. $\mathtt {Deepr}$: A Convolutional Net for Medical Records. *IEEE J. Biomedical Health Inf.* **21**, 22–30 (2017).
11. Pizzi, C. et al. Development and verification of prediction models for preventing cardiovascular diseases. *Plos One* **14** (2019).
12. Johnson, A. E. W. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* **10** (2023).
13. Farkas, R. & Szarvas, G. Automatic construction of rule-based ICD-9-CM coding systems. *BMC Bioinform.* **9** (2008).
14. Zhang, G. et al. Predicting sepsis in-hospital mortality with machine learning: a multi-center study using clinical and inflammatory biomarkers. *Eur. J. Med. Res.* **29** (2024).
15. Wu, X., Hu, J. & Zhang, J. Machine learning-based model for predicting major adverse cardiovascular and cerebrovascular events in patients aged 65 years and older undergoing noncardiac surgery. *BMC Geriatr.* **23**, 819 (2023).
16. Liu, S., Chen, M., Tang, L., Li, X. & Zhou, S. Association between Serum Ferritin and Prognosis in Patients with Ischemic Heart Disease in Intensive Care Units. *J. Clin. Med.* **12** (2023).
17. Dou, J. et al. Association between triglyceride glucose-body mass and one-year all-cause mortality of patients with heart failure: a retrospective study utilizing the MIMIC-IV database. *Cardiovasc. Diabetol.* **22** (2023).
18. Dhanka, S. & Maini, S. A hybridization of XGBoost machine learning model by Optuna hyperparameter tuning suite for cardiovascular disease classification with significant effect of outliers and heterogeneous training datasets. *Int. J. Cardiol.* **420** (2025).
19. Huang, J. et al. Prediction of mortality events of patients with acute heart failure in intensive care unit based on deep neural network. *Comput. Methods Programs Biomed.* **256** (2024).

20. Wang, T. Improved random forest classification model combined with C5.0 algorithm for vegetation feature analysis in non-agricultural environments. *Sci. Rep.* **14** (2024).
21. Pansambal, B. H. & Nandgaokar, A. B. Integrating Dropout Regularization Technique at Different Layers to Improve the Performance of Neural Networks. *Department Electron. Telecommunications Eng.* **14**, 716–721 (2023).
22. Mushava, J. & Murray, M. Flexible loss functions for binary classification in gradient-boosted decision trees: An application to credit scoring. *Expert Syst. Appl.* **238** (2024).
23. Allgaier, J., Pryss, R. & Cross-Validation Visualized A Narrative Guide to Advanced Methods. *Mach. Learn. Knowl. Extr.* **6**, 1378–1388 (2024).
24. Dhanka, S., Bhardwaj, V. K. & Maini, S. Comprehensive analysis of supervised algorithms for coronary artery heart disease detection. *Expert Syst.* **40** (2023).
25. Peng, X. et al. A simple machine learning model for the prediction of acute kidney injury following noncardiac surgery in geriatric patients: a prospective cohort study. *BMC Geriatr.* **24** (2024).
26. Mastropietro, A., Feldmann, C. & Bajorath, J. Calculation of exact Shapley values for explaining support vector machine models using the radial basis function kernel. *Sci. Rep.* **13** (2023).
27. Schmidt, G. et al. Comparison of preoperative NT-proBNP and simple cardiac risk scores for predicting postoperative morbidity after non-cardiac surgery with intermediate or high surgical risk. *Perioperative Med.* **13** (2024).
28. Ford, M. K., Beattie, W. S. & Wijeysundera, D. N. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann. Intern. Med.* **152**, 26–35 (2010).
29. Wu, X. D. et al. Risk factors prediction of 6-month mortality after noncardiac surgery of older patients in China: a multicentre retrospective cohort study. *Int. J. Surg. (London England)*. **110**, 219–228 (2024).
30. Adriana Nunes Machado, Maria do Carmo Sitta, Wilson Jacob Filho & & Garcez-Leme, L. E.-g. Prognostic Factors for Mortality Among Patients Above the 6th Decade Undergoing Non-Cardiac Surgery: (Cares – Clinical Assessment and Research in Elderly Surgical Patients). *Clinical Science* 63, 151–156 (2008).
31. Cicek, V. et al. A New Risk Prediction Model for the Assessment of Myocardial Injury in Elderly Patients Undergoing Non-Elective Surgery. *J. Cardiovasc. Dev. Disease* **12** (2024).
32. Troisi, F. et al. Clinical complexity of an Italian cardiovascular intensive care unit: the role of mortality and severity risk scores. *J. Cardiovasc. Med. (Hagerstown Md)*. **25**, 511–518 (2024).
33. Guo, Q. Y., Peng, J., Shan, T. C. & Xu, M. Risk Factors for Mortality in Critically Ill Patients with Coagulation Abnormalities: A Retrospective Cohort Study. *Curr. Med. Sci.* (2024).
34. Yang, H., Chen, Y., He, J., Li, Y. & Feng, Y. Advances in the diagnosis of early biomarkers for acute kidney injury: a literature review. *BMC Nephrol.* **26** (2025).

35. Wong, J. K. et al. P-Wave Characteristics on Routine Preoperative Electrocardiogram Improve Prediction of New-Onset Postoperative Atrial Fibrillation in Cardiac Surgery. *J. Cardiothorac. Vasc. Anesth.* **28**, 1497–1504 (2014).
36. Lin, C. H. et al. A multitask deep learning model utilizing electrocardiograms for major cardiovascular adverse events prediction. *npj Digit. Med.* **8** (2025).

Figures

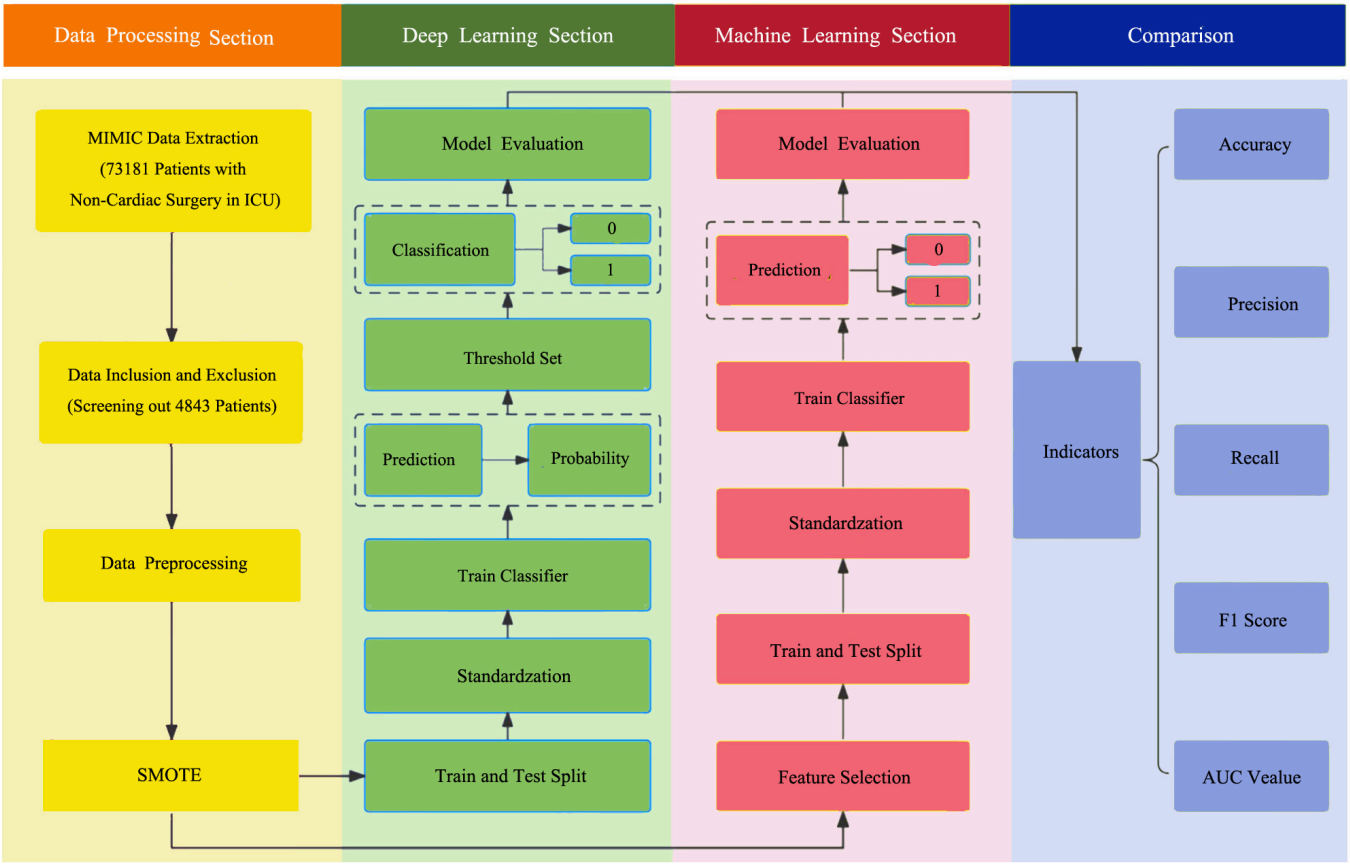

Figure 1

Flowchart of proposed non-cardiac surgical patient's survival prediction framework.

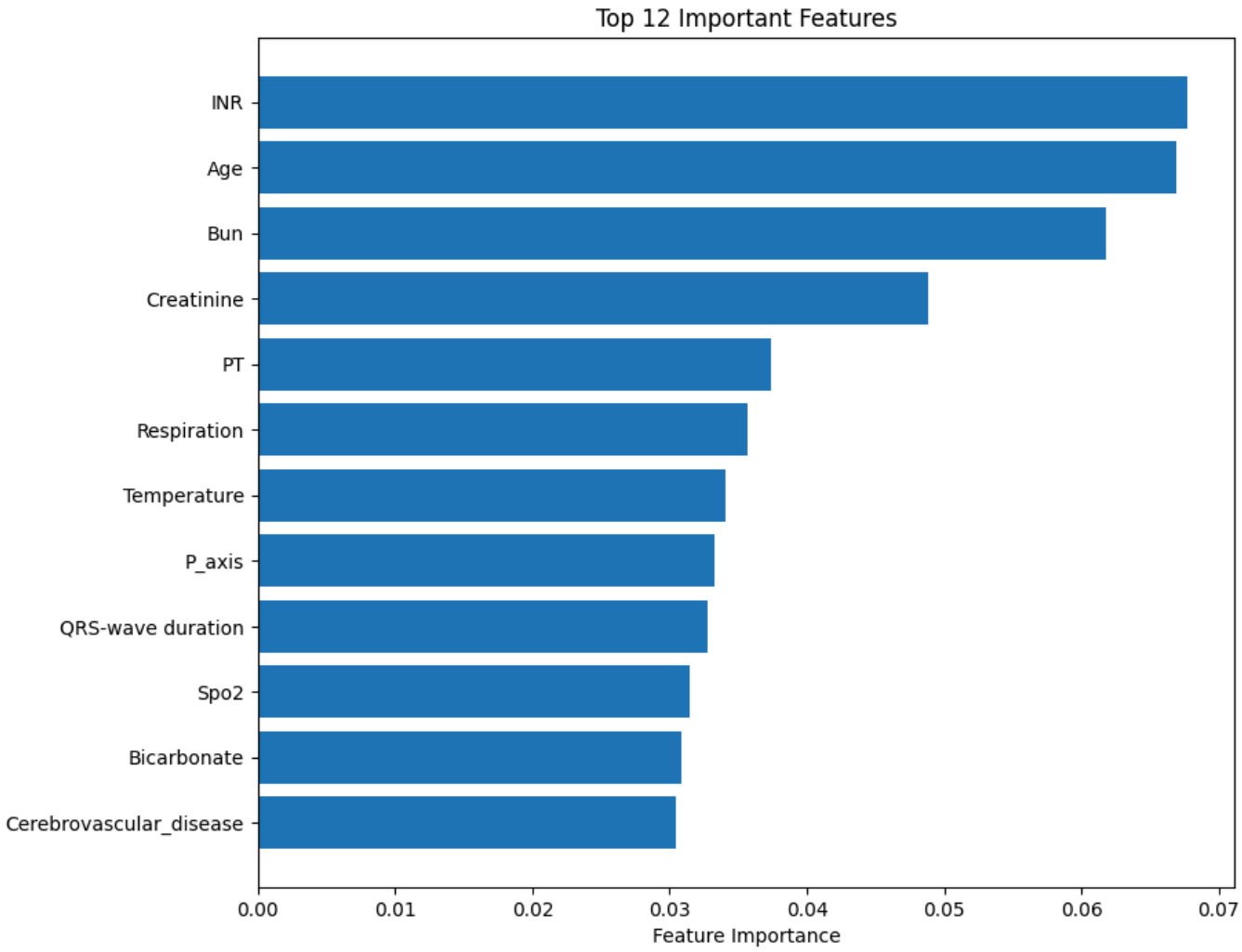

Figure 2

Top 12 important features in the random forest model.

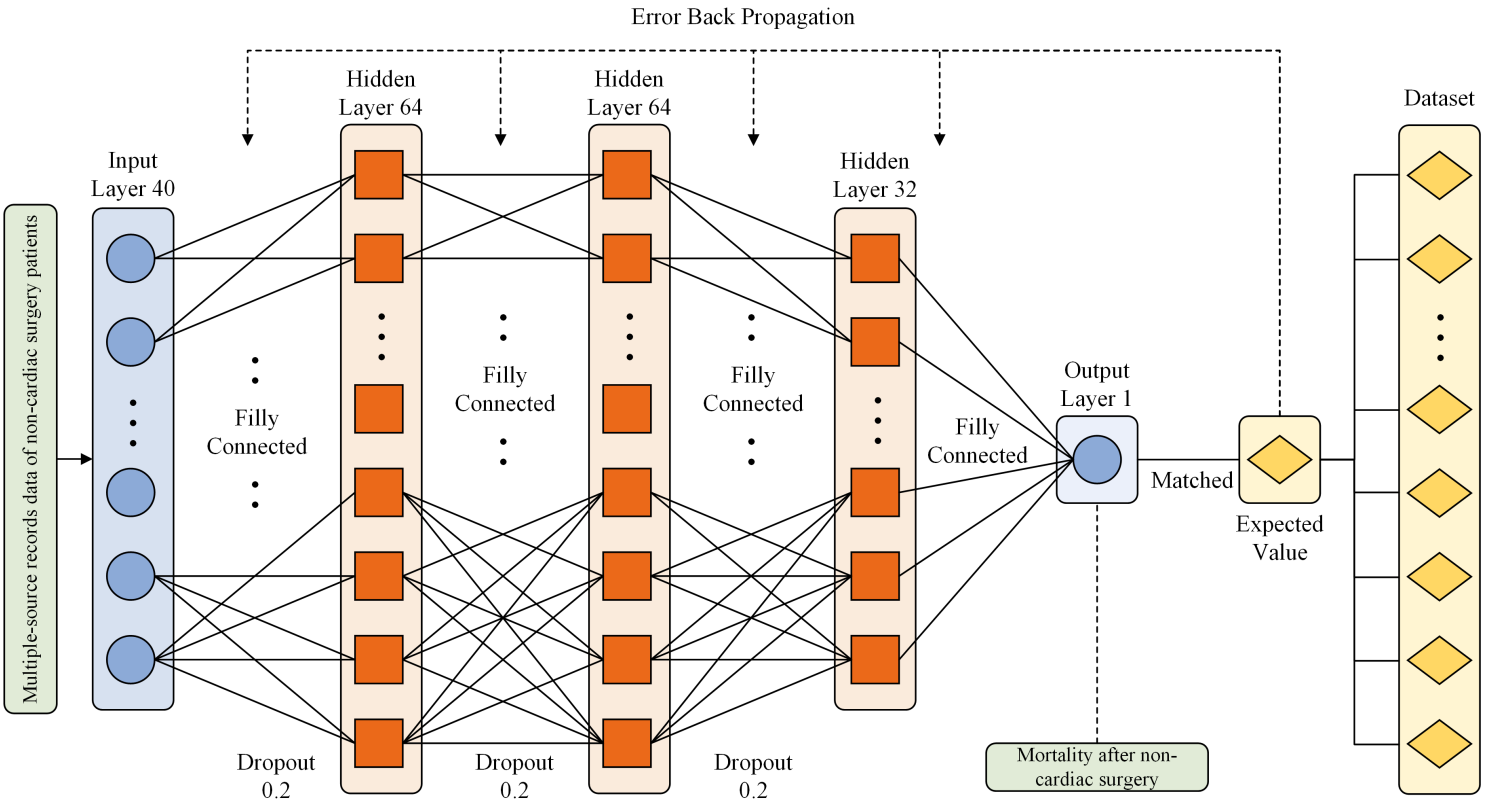

Figure 3

The architecture of the proposed BP Neural Network.

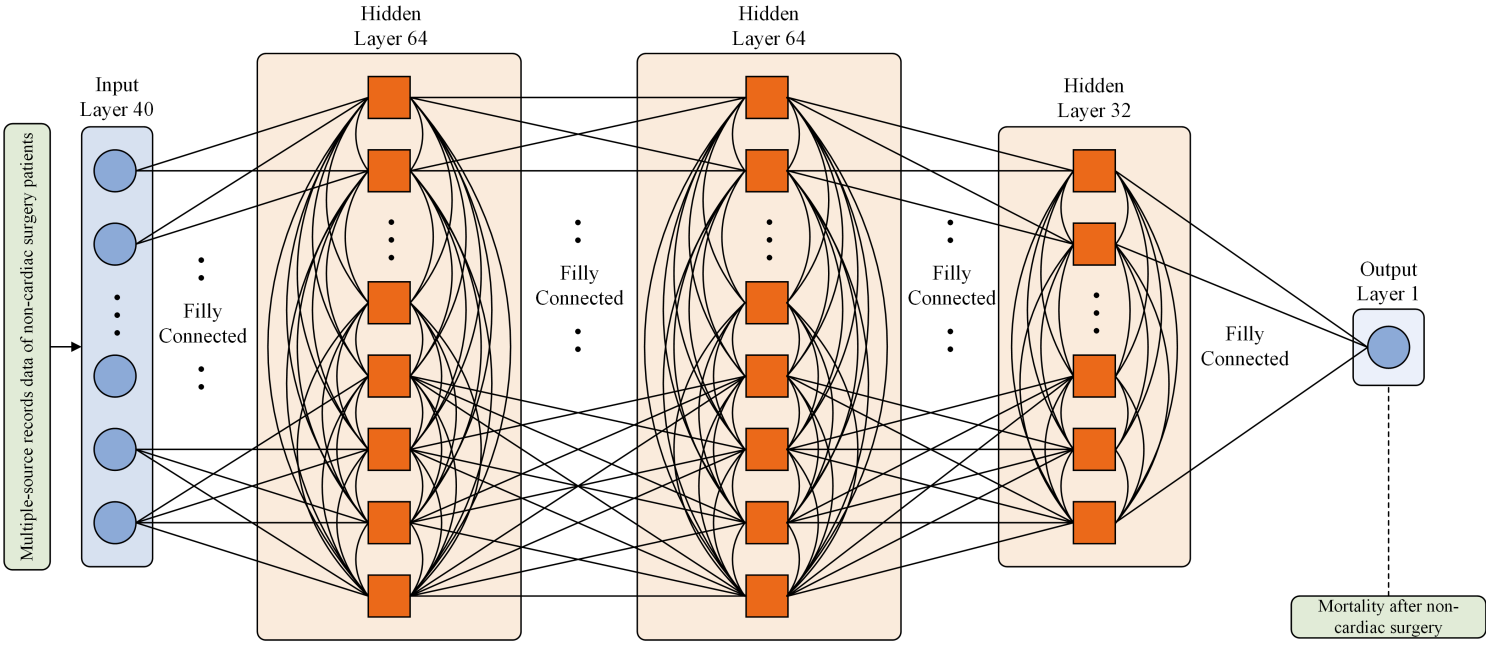

Figure 4

The architecture of the proposed recurrent neural network.

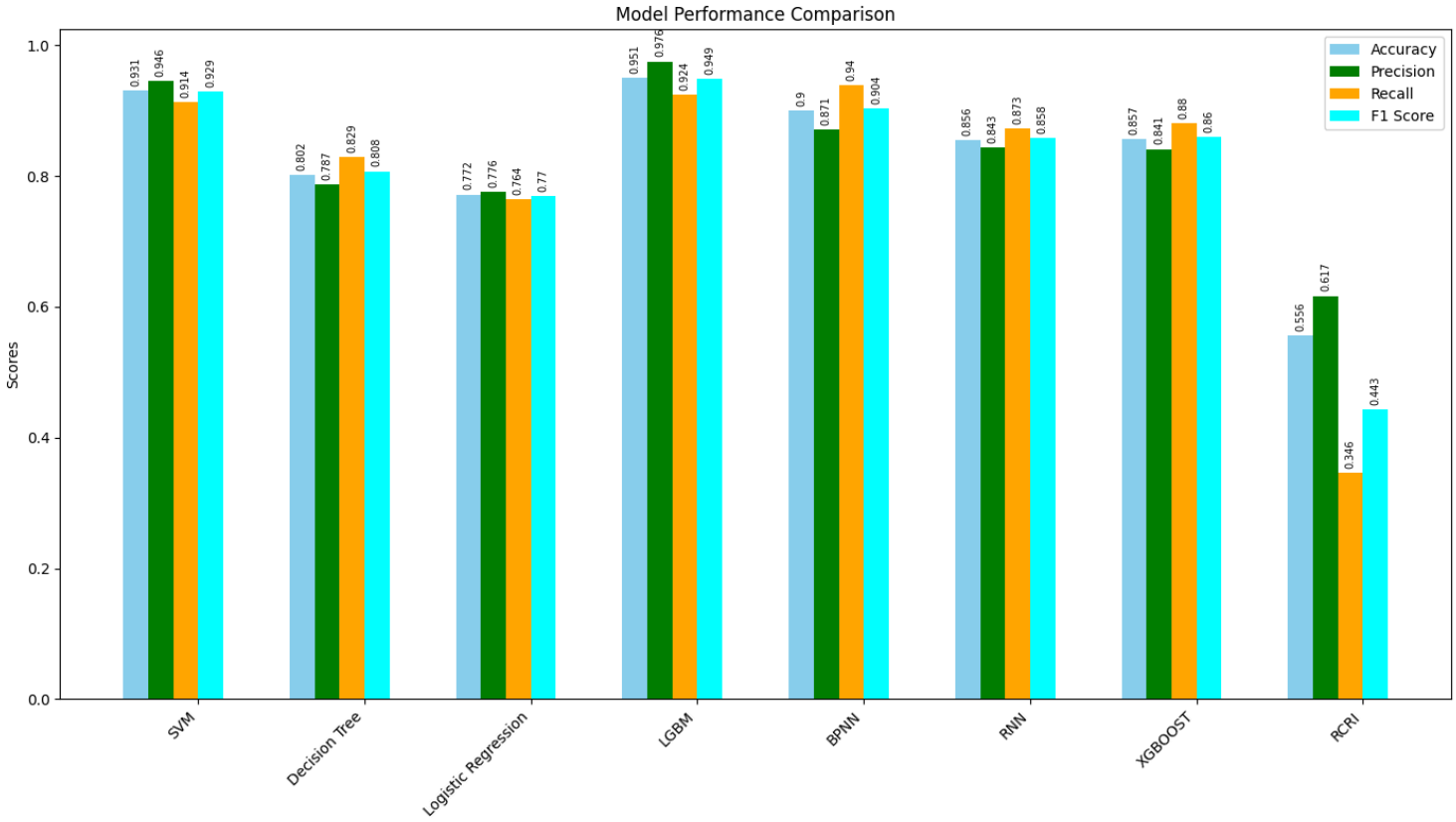

Figure 5

The classification performance of models.

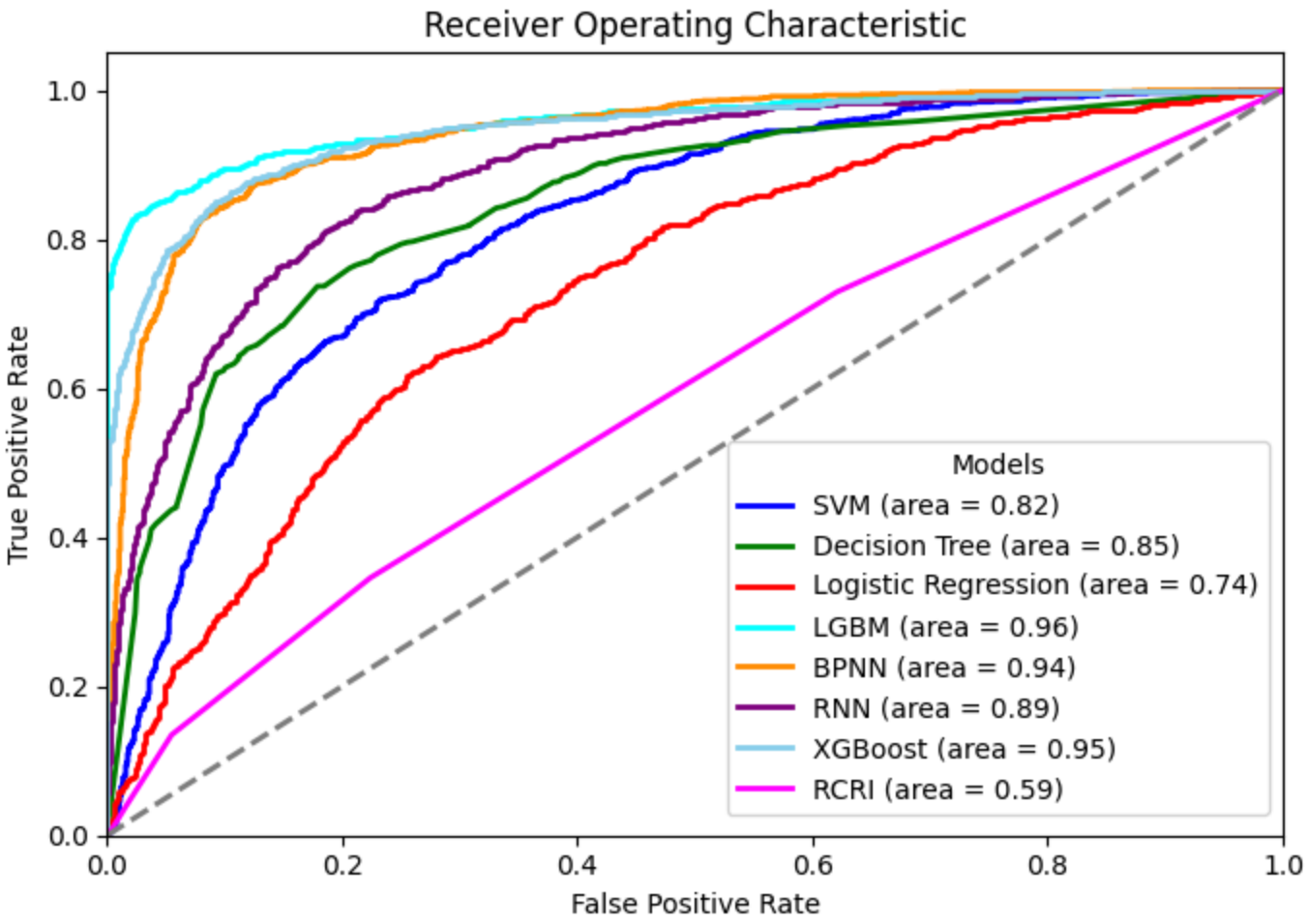

Figure 6

The ROC curves of models.

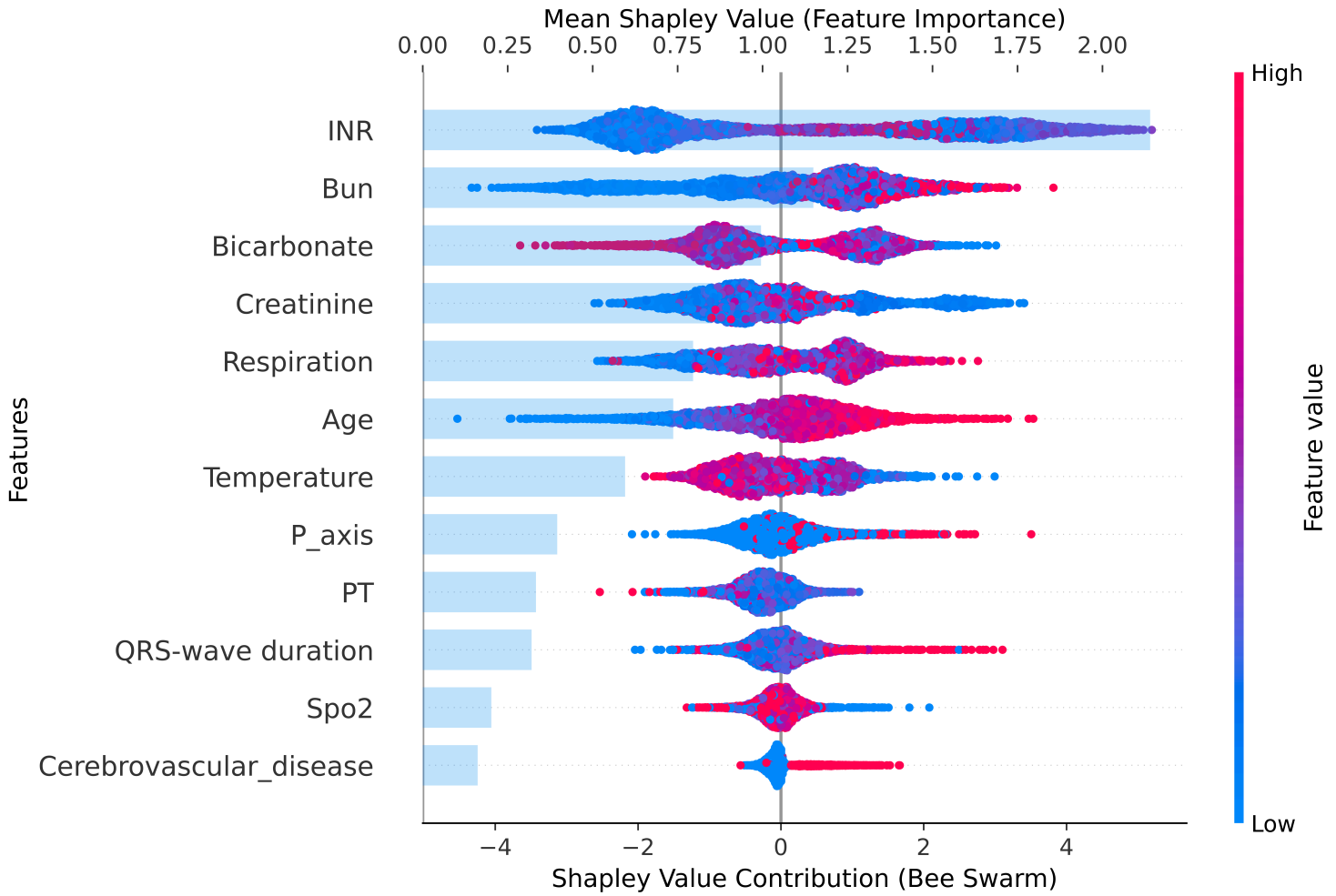

Figure 7

SHAP Values of The Top 15 Most Contributory Features on All Samples.

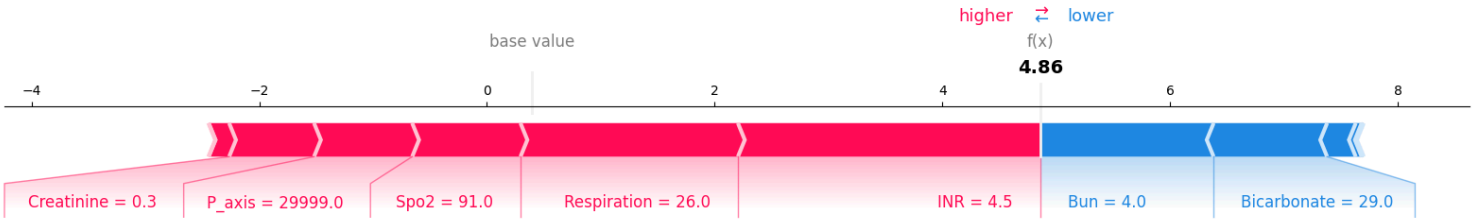

Figure 8

Indicating the local feature interpretation of a randomly selected single patient sample.