WHIOCE PUBLISHING PTE. LTD.
PROVIDING FIRST-CLASS SCIENTIFIC INFORMATION FOR TOP SCHOLARS

2026 Volume 11, Issue 5
Advances in Precision Medicine
ISSN: 2424-9106 (Online)
ISSN: 2424-8592 (Print)

An Interpretable XGBoost Model with LASSO Feature Selection for Predicting In-Hospital Mortality in Patients with Sepsis

Yaohua Tang[1], Yuexian Liu[2]*

[1]Yunnan Medical Health College, Kunming 650000, Yunnan, China
[2]Neurology Department of the 920th Hospital of the Joint Logistics Support Force, Kunming 650000, Yunnan, China

***Corresponding author:** Yuexian Liu, 15288238661@163.com

Abstract: Background: Sepsis remains a leading cause of in-hospital death in intensive care units (ICUs). This study developed and internally validated an interpretable machine-learning model for predicting in-hospital mortality in patients with sepsis. *Methods:* This retrospective cohort study used the Medical Information Mart for Intensive Care IV (MIMIC-IV, version 2.2) database. Adult patients meeting the Sepsis-3 criteria were included. Candidate predictors were extracted from the first 24 hours of ICU admission. The least absolute shrinkage and selection operator (LASSO) regression was used for feature selection. Extreme gradient boosting (XGBoost), logistic regression, and a decision tree were trained on a 70% training set and evaluated on a 30% held-out test set by the area under the receiver operating characteristic curve (AUC). SHapley Additive exPlanations (SHAP) were used to interpret the final model. *Results:* A total of 4, 262 patients with sepsis were included; the in-hospital mortality rate was 19.9%. LASSO retained 15 predictors. The XGBoost model achieved the highest discrimination (AUC 0.86; 95% confidence interval 0.84–0.88), outperforming logistic regression (0.80) and the decision tree (0.78). SHAP analysis identified lactate, mean arterial pressure, age, SOFA score, and PaO_2 as the dominant predictors, with risk rising steeply at lactate > 2 mmol/L and mean arterial pressure < 65 mmHg. *Conclusions:* An interpretable XGBoost model based on routinely collected early ICU variables accurately predicted in-hospital mortality in sepsis and may support early risk stratification.

Keywords: Sepsis; Machine learning; XGBoost; LASSO; SHAP; In-hospital mortality; MIMIC-IV

Online publication: May 26, 2026

1. Introduction

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection [1]. According to the Global Burden of Disease Study 2017, an estimated 48.9 million sepsis cases and 11 million sepsis-related deaths occurred globally in 2017, and in-hospital mortality in contemporary ICU cohorts still ranges from 25% to 30% [2,3]. Because each hour of delayed treatment increases mortality, early risk stratification is central to effective care [4]. Traditional severity scores such as the Sequential Organ Failure Assessment (SOFA) and Acute Physiology and Chronic Health Evaluation II (APACHE II) rely on linear weighting, however, and capture only a limited subset of the high-dimensional, non-linear physiology of sepsis [5,6].

Machine-learning approaches can model complex, non-linear interactions directly from data. Tree-based ensembles, in particular extreme gradient boosting (XGBoost), have shown strong performance in sepsis prediction, while Shapley

Additive exPlanations (SHAP) provide a principled framework for attributing predictions to specific features, reconciling accuracy with interpretability [7–9]. The present study aimed to develop and internally validate an interpretable XGBoost model for in-hospital mortality in sepsis after LASSO-based feature selection, and to use SHAP to identify the key drivers of mortality risk.

2. Materials and methods

2.1. Study design and participants

This retrospective cohort study used the Medical Information Mart for Intensive Care IV (MIMIC-IV, version 2.2) database, which contains de-identified data for more than 70, 000 ICU admissions at Beth Israel Deaconess Medical Center (Boston, MA, USA) between 2008 and 2019 [10]. Access was granted after the Collaborative Institutional Training Initiative training and the PhysioNet data use agreement. Adult patients (≥ 18 years) with a first ICU stay meeting the Sepsis-3 criteria-suspected infection with an acute SOFA increase of ≥ 2 points-were eligible [1]. Patients with an ICU stay < 24 hours, missing outcome, or > 30% of predictors missing were excluded. Because the data are de-identified, individual consent and additional ethics approval were not required. The study is reported per the TRIPOD statement [11].

2.2. Outcome and predictors

The primary outcome was in-hospital mortality. Candidate predictors were extracted from the first 24 hours of ICU admission and included demographics (age, sex), vital signs (mean arterial pressure [MAP], heart rate, respiratory rate, temperature, SpO_2), laboratory values (lactate, creatinine, bilirubin, platelet count, hemoglobin, white blood cell count, electrolytes, anion gap, partial pressure of oxygen [PaO_2], pH, bicarbonate, international normalized ratio), comorbidities (hypertension, diabetes, malignancy, chronic kidney disease), and the SOFA score. The most abnormal value within 24 hours was used for repeated measures. Variables with > 5% to 30% missingness were imputed by multivariate imputation by chained equations [12].

2.3. Feature selection and model development

To reduce dimensionality and multicollinearity, LASSO (L1-penalized) regression with 10-fold cross-validation was applied on the training set, retaining variables with non-zero coefficients at the one-standard-error lambda (lambda.1se) [13]. Three models were then built on the selected features: logistic regression, a cost-complexity-pruned decision tree, and XGBoost [14]. The data were split into a 70% training set and a 30% held-out test set with stratification on the outcome. XGBoost hyperparameters were tuned by 5-fold cross-validation with Bayesian optimization using the AUC as the objective, with early stopping.

2.4. Evaluation, interpretation, and statistics

Discrimination was quantified by the AUC with 95% confidence intervals (CIs) from 1, 000 bootstrap resamples; accuracy, sensitivity, specificity, and F1 score were computed at the Youden-optimal threshold, and calibration by the Brier score. The final model was interpreted using SHAP [9]. Continuous variables are presented as mean ± standard deviation or median (interquartile range) and compared by the *t*-test or Mann-Whitney U test; categorical variables by the chi-square test. A two-sided $p < 0.05$ was considered significant. Analyses used R 4.3.0 and Python 3.10 (scikit-learn, xgboost, shap).

3. Results

In total, 4, 262 patients with sepsis were included. The mean age was 65.3 ± 16.1 years, 53.3% were male, and 849 (19.9%) died in hospital.

Compared with survivors, non-survivors were older and had higher SOFA and lactate values, lower MAP, and more marked laboratory abnormalities **(Table 1)**.

Table 1. Baseline characteristics by in-hospital mortality

Variable	Survivors (n = 3, 413)	Non-survivors (n = 849)
Age, years	63.8 ± 16.0	71.4 ± 14.3**
Male sex, n (%)	1, 812 (53.1)	460 (54.2)
Mean arterial pressure, mmHg	78.6 ± 16.2	70.1 ± 17.0**
Heart rate, bpm	86.5 ± 18.3	94.2 ± 20.1**
SpO_2, %	97.1 ± 2.6	95.4 ± 4.5**
Lactate, mmol/L	1.8 (1.2-2.8)	3.6 (2.2–6.1)**
Creatinine, mg/dL	1.3 (0.9-2.1)	1.8 (1.2–2.9)**
Bilirubin, mg/dL	0.8 (0.5-1.5)	1.4 (0.8–2.9)**
Platelet count, $x10^9$/L	204 ± 95	165 ± 92**
PaO_2, mmHg	119 ± 56	104 ± 51**
pH	7.36 ± 0.08	7.30 ± 0.12**
SOFA score	5.1 ± 3.0	8.2 ± 3.7**
Hypertension, n (%)	1, 648 (48.3)	476 (56.1)**
Malignancy, n (%)	478 (14.0)	197 (23.2)**

Values are mean ± standard deviation, median (interquartile range), or n (%). ** $p < 0.001$ vs. survivors. SOFA, Sequential Organ Failure Assessment; SpO_2, peripheral oxygen saturation; PaO_2, partial pressure of oxygen.

LASSO regression with 10-fold cross-validation retained 15 predictors, spanning demographics (age), hemodynamics (MAP, heart rate), oxygenation (PaO_2, SpO_2, pH, bicarbonate), tissue perfusion (lactate), organ function (creatinine, bilirubin, platelets, INR), and severity (SOFA). On the held-out test set, XGBoost showed the best discrimination (AUC 0.86; 95% CI 0.84–0.88), followed by logistic regression (0.80) and the decision tree (0.78) (**Table 2**). XGBoost also had the highest accuracy (0.82), sensitivity (0.78), and F1 score (0.66), and the lowest Brier score (0.12). (**Table 3**)

Table 2. Performance of the predictive models on the held-out test set.

Model	AUC (95% CI)	Accuracy	Sensitivity	Specificity	F1	Brier
Logistic regression	0.80 (0.78–0.82)	0.76	0.71	0.77	0.58	0.14
Decision tree	0.78 (0.75–0.80)	0.74	0.69	0.75	0.55	0.15
XGBoost	0.86 (0.84–0.88)	0.82	0.78	0.83	0.66	0.12

AUC, area under the receiver operating characteristic curve; CI, confidence interval; XGBoost, extreme gradient boosting.

Table 3. Top predictors ranked by mean absolute SHAP value.

Rank	Predictor	Direction of effect on mortality risk
1	Lactate	Higher value → Increased risk (steep rise > 2 mmol/L)
2	Mean arterial pressure	Lower value → Increased risk (steep rise < 65 mmHg)
3	Age	Older → Increased risk
4	SOFA score	Higher score → Increased risk
5	PaO_2	Lower value → Increased risk
6	Bicarbonate	Lower value → Increased risk
7	Blood urea nitrogen	Higher value → Increased risk
8	Platelet count	Lower value → Increased risk
9	Respiratory rate	Higher value → Increased risk

Rank	Predictor	Direction of effect on mortality risk
10	pH	Lower value → Increased risk

SHAP, Shapley Additive exPlanations; SOFA, Sequential Organ Failure Assessment; PaO_2, partial pressure of oxygen.

Global SHAP analysis showed that elevated lactate, older age, and higher SOFA scores increased the predicted mortality risk, whereas higher MAP and PaO_2 were protective. Lactate and MAP exhibited clear threshold behavior-risk rose steeply above a lactate of about 2 mmol/L and below a MAP of 65 mmHg-consistent with the clinical definitions of hyperlactatemia and shock.

4. Discussion

In this retrospective study of 4, 262 patients with sepsis from MIMIC-IV, an interpretable XGBoost model predicted in-hospital mortality with an AUC of 0.86, outperforming logistic regression and a decision tree. LASSO feature selection reduced more than thirty candidate variables to fifteen interpretable predictors, lowering overfitting risk and improving clinical face validity, and SHAP analysis confirmed that the model learned a biologically coherent signal dominated by lactate, MAP, age, SOFA score, and PaO_2.

These findings align with the existing literature, in which XGBoost-based sepsis outcome models typically achieve AUC values between 0.80 and 0.88 [7,8,15]. The dominance of lactate is expected: it is a direct marker of anaerobic metabolism and tissue hypoperfusion, and its prognostic value is well established [16]. The SHAP thresholds at lactate > 2 mmol/L and MAP < 65 mmHg mirror the criteria used to define hyperlactatemia and shock, supporting the validity of the model and the Surviving Sepsis Campaign MAP target [1,3]. Age and SOFA score capture chronic reserve and acute organ dysfunction, complementing the perfusion markers.

This study has limitations. It is a single-center, retrospective analysis of a US database, so external validation on independent cohorts is required before clinical use [17]. Repeated measures were summarized as the worst value within 24 hours, which discards temporal dynamics; dynamic approaches such as long short-term memory networks may add value and warrant further study. Finally, the model predicts in-hospital mortality only; longer-term endpoints would be valuable.

In conclusion, an interpretable XGBoost model built from fifteen routinely collected early ICU variables accurately predicted in-hospital mortality in sepsis, with biologically plausible predictors. After external validation, it may support early risk stratification and individualized intervention.

Funding

Yunnan Provincial Department of Education Science Research Fund (Project No.: 2025J2315)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Singer M, Deutschman C, Seymour C, et al., 2016, The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA, 315(8): 801–810.

[2] Rudd K, Johnson S, Agesa K, et al., 2020, Global, Regional, and National Sepsis Incidence and Mortality, 1990–2017:

Analysis for the Global Burden of Disease Study. Lancet, 395(10219): 200–211.

[3] Evans L, Rhodes A, Alhazzani W, et al., 2021, Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Intensive Care Medicine, 47(11): 1181–1247.

[4] Seymour C, Gesten F, Prescott H, et al., 2017, Time to Treatment and Mortality During Mandated Emergency Care for Sepsis. New England Journal of Medicine, 376(23): 2235–2244.

[5] Vincent J, Moreno R, Takala J, et al., 1996, The SOFA (Sepsis-Related Organ Failure Assessment) Score to Describe Organ Dysfunction/Failure. Intensive Care Medicine, 22(7): 707–710.

[6] Knaus W, Draper E, Wagner D, et al., 1985, APACHE II: A Severity of Disease Classification System. Critical Care Medicine, 13(10): 818–829.

[7] Gao J, Lu Y, Ashrafi N, et al., 2024, Prediction of Sepsis Mortality in ICU Patients Using Machine Learning Methods. BMC Medical Informatics and Decision Making, 24(1): 154.

[8] Chen S, Fan J, Pishgar E, et al., 2025, XGBoost-Based Prediction of ICU Mortality in Sepsis-Associated Acute Kidney Injury Patients Using MIMIC-IV Database with Validation from eICU Database. medRxiv, 2025.02.24.25322816.

[9] Lundberg S, Lee S, 2017, A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30: 4765–4774.

[10] Johnson A, Bulgarelli L, Shen L, et al., 2023, MIMIC-IV, a Freely Accessible Electronic Health Record Dataset. Scientific Data, 10(1): 1.

[11] Collins G, Reitsma J, Altman D, et al., 2015, Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. BMJ, 350: g7594.

[12] van Buuren S, Groothuis-Oudshoorn K, 2011, mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3): 1–67.

[13] Tibshirani R, 1996, Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B, 58(1): 267–288.

[14] Chen T, Guestrin C, 2016, XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[15] Zabihi M, Kiranyaz S, Gabbouj M, 2019, Sepsis Prediction in Intensive Care Unit Using Ensemble of XGBoost Models. Computing in Cardiology, 45: 1–4.

[16] Bakker J, Nijsten M, Jansen T, 2013, Clinical Use of Lactate Monitoring in Critically Ill Patients. Annals of Intensive Care, 3: 12.

[17] Moons K, Kengne A, Grobbee D, et al., 2012, Risk Prediction Models: II. External Validation, Model Updating, and Impact Assessment. Heart, 98(9): 691–698.