

RESEARCH

Open Access



Association between the lactate dehydrogenase-to-albumin ratio and 28-day mortality in septic patients with malignancies: analysis of the MIMIC-IV database

Yongshi Shen^{1†}, Kangni Lin^{1†}, Liuxin Yang^{2†}, Peng Zheng¹, Wei Zhang¹, Jinsen Weng^{1*} and Yong Ye^{1*}

Abstract

Background Sepsis remains a leading cause of mortality in critically ill patients, particularly those with malignancies who face heightened risks due to immunosuppression and metabolic dysregulation. This study aimed to evaluate the prognostic value of the lactate dehydrogenase-to-albumin ratio (LDAR) for predicting 28-day ICU mortality in septic patients with malignancies.

Methods A retrospective cohort analysis was conducted using data from 1,635 septic patients with malignancies in the MIMIC-IV (3.1) database. Participants were stratified into quartiles based on LDAR values. The primary outcome was 28-day ICU mortality, with secondary outcomes including in-hospital and ICU mortality. Multivariable logistic regression, restricted cubic spline (RCS) analysis, and machine learning models were employed to assess associations between LDAR and outcomes. Subgroup analyses and feature importance evaluations were performed to validate robustness. The Shapley additive explanations method was used to enhance model interpretability and assess individual predictor contributions.

Results Higher LDAR is independently associated with increased 28-day ICU mortality (OR: 3.441, 95% CI: 2.497–4.741), ICU mortality (OR: 3.478, 95% CI: 2.396–5.049), and in-hospital mortality (OR: 3.747, 95% CI: 2.688–5.222), even after adjustment, highlighting its potential as a prognostic marker in ICU patients. RCS analysis revealed a nonlinear relationship, with mortality risk escalating sharply beyond $\log_2(\text{LDAR}) = 6.940$. Metastatic cancer patients had higher median LDAR (135.0 vs. 118.5, $P = 0.004$) and mortality rates (52.0% vs. 36.4%, $P < 0.001$). Boruta feature selection showed that LDAR as the top predictor of mortality. Nine machine learning model with 20 variables were built, with random forest model performing best, achieving an AUC of 0.751 (0.708–0.794) in validation and 0.727 (0.682–0.772) in text cohort.

[†]Yongshi Shen, Kangni Lin and Liuxin Yang contributed equally to this article.

*Correspondence:
Jinsen Weng
fjzyxk@163.com
Yong Ye
13489033270@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions LDAR is a robust, independent prognostic biomarker for 28-day ICU mortality in septic patients with malignancies, outperforming traditional scoring systems. The identified threshold ($\log_2(\text{LDAR}) \geq 6.940$) may aid early risk stratification and clinical decision-making. Prospective studies are warranted to validate these findings and explore dynamic LDAR monitoring in diverse populations.

Keywords Lactate dehydrogenase-to-albumin ratio, 28-day mortality, Sepsis, Malignancy, MIMIC-IV database

Introduction

Sepsis, a systemic inflammatory response syndrome caused by infection, remains a significant cause of morbidity and mortality worldwide [1–3]. To enhance and unify clinical practices, the Surviving Sepsis Campaign has developed and periodically revises guidelines for managing patients with sepsis and septic shock [4, 5]. The increasing incidence of cancer has been accompanied by an increasing risk of sepsis [5, 6]. Compared with the general population, cancer patients are ten times more likely to develop sepsis [7]. Among septic patients, those with concurrent malignancies face increased risks due to malignancy-induced immunosuppression, complex metabolic dysfunction, and treatment-related complications such as chemotherapy [8, 9]. Traditional scoring systems such as SOFA and SAPSII are widely used but may not fully capture the unique inflammatory and metabolic derangements in septic oncology patients.

Lactate dehydrogenase (LDH) and albumin (Alb) are widely used biomarkers in clinical practice, reflecting tissue damage and metabolic dysfunction, as well as nutritional and inflammatory status, respectively. Elevated LDH levels have been linked to inflammation, hypoxia, and organ dysfunction in sepsis, whereas hypoalbuminemia is associated with disease severity and poor prognosis. While each marker alone is broadly indicative of disease severity, especially in critical illnesses, they may be insufficiently sensitive or specific for the complex interplay of immunosuppression, tumor burden, and sepsis. Recently, the lactate dehydrogenase to albumin ratio (LDAR) has gained widespread attention as a composite biomarker that integrates metabolic dysregulation and nutritional status, demonstrating superior prognostic value. Studies have shown that LDAR is closely associated with mortality in various diseases, such as malignancies and sepsis [10–14]. However, investigations of LDAR in patients with malignancies complicated by sepsis remain extremely limited, and the lack of data makes it unclear whether LDAR retains prognostic validity across different tumor types (e.g., solid vs. hematologic) or sepsis etiologies (e.g., pulmonary vs. abdominal). This knowledge gap is particularly significant, as patients with sepsis and malignancies exhibit unique pathophysiological features that may alter the predictive performance of LDAR. Current research has not sufficiently explored the potential utility of LDAR in this high-risk population. Investigating the prognostic value of LDAR in septic patients with

malignancies could address this gap and provide critical evidence to guide clinical decision-making.

This study aimed to evaluate the clinical utility of the LDH/Alb ratio specifically for predicting mortality risk in septic patients with malignancies, focusing on 28-day ICU mortality endpoint.

Materials and methods

Data source

This study was designed as a retrospective cohort study based on the publicly available Medical Information Mart for Intensive Care-IV (MIMIC-IV) version 3.1 database. The MIMIC-IV contains more than 50,000 ICU admission records from Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, covering the period from 2008 to 2021 [15]. The database provides comprehensive information on patient demographics, vital signs, laboratory results, and diagnoses coded according to the Ninth (ICD-9) and Tenth (ICD-10) revisions of the international classification of diseases. To access and utilize the MIMIC-IV database, one of the study team members (Yongshi Shen) completed the required data-user certification and extracted the relevant study variables (certification number: 37021016). Because the data in the MIMIC-IV are fully deidentified, the requirement for individual patient informed consent was waived. The inclusion criteria were as follows: (1) Patients aged 18 to 100 years; (2) Patients meeting the diagnostic criteria for sepsis based on the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3); (3) Patients with a documented malignancy. Exclusion criteria: (1) ICU length of stay less than 6 h; (2) Multiple ICU admissions for sepsis—only the first admission was included; (3) Insufficient data; (4) There was no malignancy. Figure 1 shows the patient screening process.

Study outcomes

The primary outcome was all-cause mortality within 28 days of ICU admission. The secondary outcomes included in-hospital mortality and ICU mortality. In-hospital mortality refers to death occurring at any point during the entire hospital admission, and ICU mortality pertains to death during the ICU stay itself.

Variable extraction

We used structured query language via PostgreSQL (version 17.2) and Navicat Premium (version 16) to

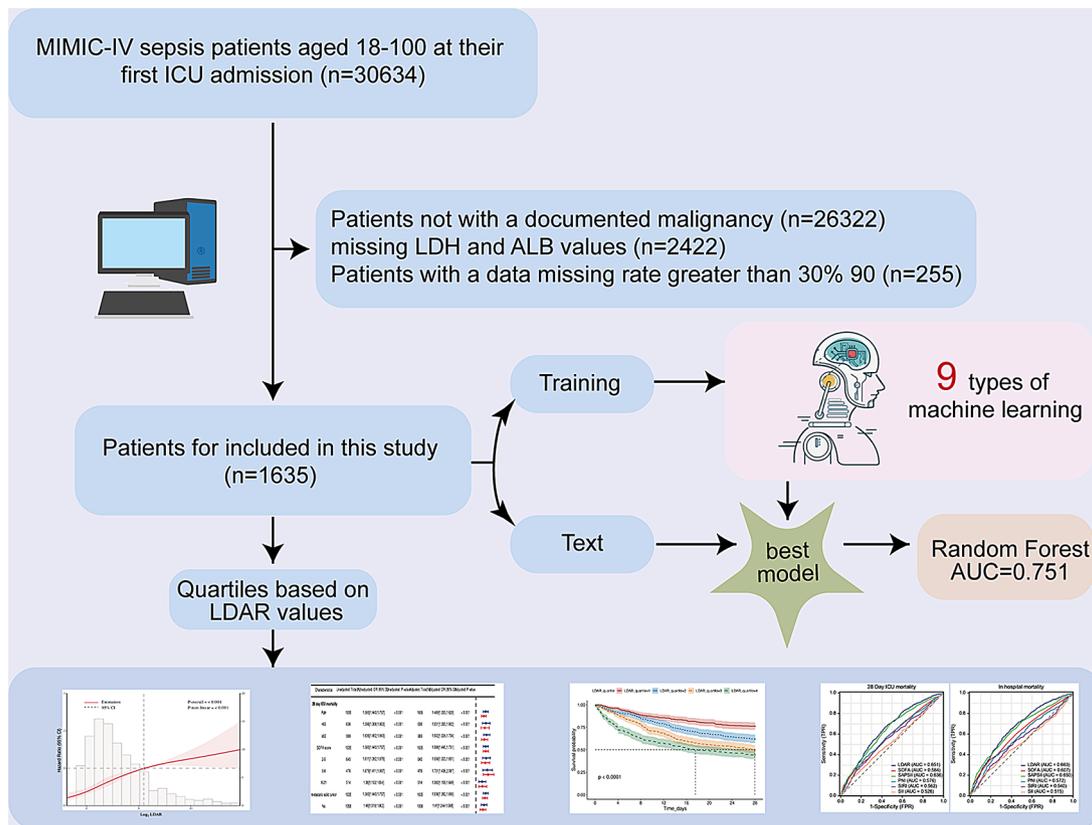


Fig. 1 Study flowchart. LDH: lactate dehydrogenase; ALB: albumin; LDAR: lactate dehydrogenase-to-albumin ratio

extract the data. The baseline characteristics collected for each patient included age, sex, body mass index (BMI), the Sequential Organ Failure Assessment (SOFA) score, and the Simplified Acute Physiology Score II (SAPSII). v , and the Charlson comorbidity index was also calculated. We additionally collected data on vital signs, clinical outcomes, and laboratory measurements, including LDH, ALB, bicarbonate, white blood cell count (WBC), hemoglobin, mean corpuscular volume (MCV), platelet count, alanine aminotransferase (ALT), aspartate aminotransferase (AST), creatinine, blood urea nitrogen (BUN), glucose, potassium, calcium, and prothrombin time (PT), among others. All laboratory parameters extracted from the MIMIC-IV (3.1) were the first measurements taken after ICU admission. The SOFA score was the maximum score recorded within 24 h of ICU admission. Follow-up started on the date of ICU admission and ended on the date of patient death.

The LDAR was defined as the LDH value (U/L) divided by the albumin value (g/ μ L). The inflammation indices were calculated as follows: prognostic nutritional index (PNI) = $10 \times$ serum albumin level (g/dL) + $0.005 \times$ total lymphocyte count (/mm³); systemic inflammation response index (SIRI) = neutrophil count \times monocyte count/lymphocyte count ($\times 10^9$ /L); and systemic immune

inflammation index (SII) = platelet count \times neutrophil count/lymphocyte count ($\times 10^9$ /L).

Data cleaning

Variables with >30% missingness were excluded (e.g., procalcitonin). We excluded a total of 34 variables due to high missingness. For remaining variables, multiple imputation was performed using the R mice package. To assess potential multicollinearity among variables, we calculated variance inflation factors (VIFs); if the VIF for any variable exceeded 5, that variable was removed from the analysis owing to collinearity. We excluded 14 variables due to multicollinearity.

Statistical analysis

Overall analysis

Because this was a retrospective study, a formal sample size calculation was not performed. Patients were divided into four groups according to the quartile of LDAR. Continuous variables were described according to their distribution. Normally distributed variables are reported as the means \pm standard deviations (SDs) and were compared via one-way analysis of variance (ANOVA). Non-normally distributed variables are presented as medians and interquartile ranges (IQRs) and were compared via the Mann-Whitney U test or the Kruskal-Wallis test.

Categorical variables are expressed as counts (n) and percentages (%) and were compared via the chi-square (χ^2) test or Fisher's exact test, as appropriate. To evaluate the associations between LDAR and the risk of in-hospital mortality, ICU mortality, and 28-day mortality, we performed multivariable logistic regression. Odds ratios (ORs) and corresponding 95% confidence intervals (CIs) were calculated to quantify the effect of the LDAR index on these outcomes. Two models were specified: Model 1: Unadjusted. Model 2: Adjusted for age, sex, BMI, vital signs (heart rate, mean arterial pressure, respiratory rate, oxygen saturation, and temperature), SOFA score, and the presence of metastatic solid tumors.

Subgroup analysis

We conducted both univariable and multivariable analyses across predefined subgroups (age, sex, SOFA score, and presence of metastatic solid tumors) to validate the relationships between the LDAR index and in-hospital, ICU, and 28-day mortality. In the multivariable analysis, we adjusted for the following covariates: age, BMI, gender, heart rate, respiratory rate, MBP, SOFA score, Comorbidities (myocardial infarction, congestive heart failure, cerebrovascular disease, chronic pulmonary disease, renal disease, severe liver disease and metastatic solid tumor) and infection site (respiratory, gastrointestinal, genitourinary). (Note: the sex subgroup analysis did not adjust for sex as a covariate.) The subgroups were further stratified by age (<65 years vs. ≥ 65 years), SOFA score (2–5 points, 5–8 points, and 8–21 points), and presence of a metastatic solid tumor. Each subgroup was analyzed via a logistic regression model, and the results are illustrated in a forest plot showing the OR and 95% CI. Kaplan–Meier analyses were also performed as sensitivity analyses to explore the associations between the LDAR index and the 28-day ICU mortality and in-hospital mortality endpoints.

Nonlinear relationship analysis

To investigate the nonlinear relationship between LDAR and 28-day ICU mortality, we employed a restricted cubic spline (RCS) analysis. Given that LDAR was skewed and had a wide range, we applied a \log_2 transformation [$\log_2(\text{LDAR})$] to improve the distribution's symmetry and reduce the impact of outliers, thereby enhancing model robustness. Following Harrell's recommendations, four knots were placed at the 10th, 35th, 65th, and 90th percentiles of the \log_2 (LDAR) distribution. In the model, 28-day ICU mortality was the dependent variable, \log_2 (LDAR) was the independent variable, and adjustments were made for age, BMI, gender, heart rate, respiratory rate, MBP, SOFA score, Comorbidities (myocardial infarction, congestive heart failure, cerebrovascular disease, chronic pulmonary disease, renal disease, severe

liver disease and metastatic solid tumor) and infection site (respiratory, gastrointestinal, genitourinary). Nonlinearity was assessed by examining the overall association (P-overall) and the nonlinear effect (P-nonlinear). The results are presented graphically, and 95% CIs are used to assess model stability and precision. The proportional hazards assumption was checked using the Schoenfeld residuals. (Figure S1)

Predictive model construction and validation

Feature selection

To identify important variables, we employed the Boruta algorithm, which compares the Z score of each real feature with that of "shadow features." If the Z score of a real feature is significantly higher than the maximum Z score of the shadow features, that variable is deemed important. The variables identified by the Boruta algorithm were used for subsequent model development.

Model construction and hyperparameter optimization

In the model development phase, we used random assignment to split the dataset into a training (70%) and a testing (30%) sets. We utilized nine machine learning algorithms—the extreme gradient boosting (XGBoost) classifier, logistic regression, light gradient boosting machine (LGBM) classifier, random forest classifier, adaptive boosting (AdaBoost) classifier, multilayer perceptron (MLP) classifier, support vector machine (SVM), Gaussian naive Bayes (GNB), and K-neighbors (KNN) classifier—to construct predictive models.

To improve model performance, we applied a grid search for hyperparameter optimization. Grid search systematically explores a predefined parameter space and identifies the combination that yields the highest performance. During this process, the training set was further divided via 10-fold cross validation to ensure model robustness. The model with the highest area under the receiver operating characteristic curve (AUC) across the 10-fold cross-validation was selected as the final model.

Model performance evaluation

The testing set was used for independent evaluation of the model's performance. The key evaluation metrics included the following: the receiver operating characteristic (ROC) curve and the area under the curve (AUC) for assessing overall predictive ability; the calibration curve for evaluating the accuracy of absolute risk predictions; and decision curve analysis (DCA) for determining clinical utility across various risk thresholds.

We further applied Shapley additive explanations (SHAPs) to depict the contribution of each feature to the prediction. Group-level SHAP plots illustrate the overall impact of each feature on the model output, and SHAP

evaluations for selected cases help clarify how specific features influence individual predictions.

All the statistical analyses were conducted in R software (version 4.3.1). A two-sided p value < 0.05 was considered statistically significant.

Results

Baseline characteristics

Data from 1,635 sepsis patients meeting the inclusion and exclusion criteria were extracted from MIMIC-IV 3.1 (Fig. 1). The VIFs for the variables are shown in Table S1 and indicate that there is no multicollinearity among the variables. The proportion of missing data for each variable is detailed in Table S2, while the baseline characteristics of the study population are presented in Table 1 (see the end of the article). Among these patients, 984 (60.18%) were male. The comorbidities included myocardial infarction in 217 patients (13.27%), congestive heart failure in 421 patients (25.75%), cerebrovascular disease in 161 patients (9.87%), peripheral vascular disease in 131 patients (8.01%), chronic pulmonary disease in 429 patients (26.24%), and type I/II diabetes in 429 patients (26.24%). Additional comorbidities were kidney disease in 350 patients (21.41%), severe liver disease in 208 patients (12.72%), and metastatic solid tumors in 577 patients (35.29%). Patients were stratified into four quartiles on the basis of LDAR values: quartile 1 ($27.941 \leq \text{LDAR} < 81.579$), quartile 2 ($81.579 \leq \text{LDAR} < 122.800$), quartile 3 ($122.800 \leq \text{LDAR} < 230.345$), and quartile 4 ($230.345 \leq \text{LDAR} \leq 10,990.909$). Quartile 3 contained 408 cases, whereas the other three quartiles each contained 409 cases. Quartile 4 patients were older and had higher heart rates, respiratory rates, SOFA scores, SAPSII scores, INR, PT, PTT, albumin levels, calcium, blood urea nitrogen, serum potassium, creatinine, alanine transaminase, alkaline phosphatase, aspartate transaminase, LDH, total bilirubin, white blood cell count, neutrophil count, monocyte count, lymphocyte count, eosinophil count, basophil count, red blood cell distribution width, platelet count, lactate level, and arterial-alveolar oxygen gradient and lower oxygenation indices. Mortality rates (ICU 28-day mortality, ICU mortality, and in-hospital mortality) were also higher in Quartile 4.

Clinical outcomes

Table 2 presents the results of the logistic regression analysis. Models I and II revealed that, compared with those in quartile 1, 28-day ICU mortality, ICU mortality, and in-hospital mortality significantly increased in quartiles 2, 3, and 4.

Table 3 shows the comparison of clinical characteristics between patients with nonmetastatic and those with metastatic solid tumors. Patients with metastatic

tumors had significantly higher LDAR levels ($p = 0.004$) and lower albumin levels ($p < 0.001$) than did those with nonmetastatic tumors, whereas LDH levels did not significantly differ between the groups ($p = 0.399$). The rates of intensive care unit (ICU) mortality (30.5% vs. 22.6%, $p < 0.001$), in-hospital mortality (37.0% vs. 34.2%, $p = 0.002$), and 28-day intensive care unit (ICU) mortality (52.0% vs. 36.4%, $p < 0.001$) were significantly greater in the metastatic group. However, patients with nonmetastatic tumors had longer ICU stays (4.88 vs. 4.09 days, $p = 0.001$) and total hospital stays (13.94 vs. 11.73 days, $p < 0.001$). These findings suggest that while metastatic tumor patients have a greater risk of mortality, nonmetastatic tumor patients may require prolonged hospitalization and intensive care support.

Restricted cubic spline (RCS)

Figure 2 illustrates the RCS analysis results. After adjusting for age, BMI, heart rate, mean arterial pressure, respiratory rate, sex, SOFA score, and the presence of cancer, the relationships between $\log_2(\text{LDAR})$ and hazard ratios (HR) for 28-day ICU mortality and all-cause ICU mortality exhibited significant nonlinear characteristics. At lower $\log_2(\text{LDAR})$ values, the risk increased slowly, but above the inflection point (6.940), the risk rose sharply. $\log_2(\text{LDAR})$ was identified as a significant predictive variable, with its impact becoming particularly pronounced beyond the threshold of 6.940.

On the basis of the results of the RCS analysis, LDAR was classified into a high-LDAR group ($\log_2(\text{LDAR}) \geq 6.940$) and a low-LDAR group ($\log_2(\text{LDAR}) < 6.940$). As shown in Supplementary Table S3, patients in the high-LDAR group had a greater incidence of metastatic solid tumors, an elevated heart rate, an increased respiratory rate, higher lactate levels, and significantly higher SOFA and SAPSII scores, indicating greater disease severity. The 28-day ICU mortality rate was greater in the high-LDAR group than in the low-LDAR group (52.93% vs. 30.85%, $P < 0.001$). The ICU all-cause mortality rate was significantly greater in the high-LDAR group than in the low-LDAR group (35.09% vs. 15.67%, $P < 0.001$), and a similar trend was observed for in-hospital all-cause mortality (48.53% vs. 25.46%, $P < 0.001$).

Subgroup analysis

To further validate the relationship between LDAR and mortality outcomes, stratified analyses were performed on the basis of age, sex, SOFA score, and the presence of metastatic solid tumors. Figure 3 shows that the associations between LDAR and 28-day ICU mortality, ICU mortality and in-hospital mortality remained significant across all subgroups, irrespective of covariate adjustments. These findings support the stability of LDAR as a

Table 1 Patient demographics and baseline characteristics

Variable	group					P value ²
	Overall N= 1,635 ¹	Quartile 1 N= 409 ¹	Quartile 2 N= 409 ¹	Quartile 3 N= 408 ¹	Quartile 4 N= 409 ¹	
Age, M (Q ₁ , Q ₃)	68.246 (59.961, 77.275)	68.640 (60.551, 77.414)	70.895 (62.767, 79.138)	67.211 (58.557, 75.981)	66.807 (59.068, 75.944)	< 0.001
Heartrate, M (Q ₁ , Q ₃)	98.000 (83.000, 113.000)	92.000 (78.000, 110.000)	97.000 (83.000, 114.000)	99.000 (85.500, 113.000)	101.000 (86.000, 116.000)	< 0.001
Weight, M (Q ₁ , Q ₃)	80.600 (70.400, 92.340)	79.600 (70.000, 88.630)	79.170 (69.100, 90.200)	81.495 (69.855, 94.230)	82.800 (72.250, 97.000)	0.002
BMI, M (Q ₁ , Q ₃)	197.363 (105.844, 290.740)	199.164 (88.389, 287.389)	208.028 (127.214, 307.055)	185.981 (91.531, 277.881)	192.894 (106.509, 283.567)	0.076
Spo ₂ , M (Q ₁ , Q ₃)	97.000 (94.000, 99.000)	97.000 (95.000, 99.000)	97.000 (94.000, 100.000)	97.000 (94.000, 100.000)	97.000 (94.000, 99.000)	0.650
MAP, M (Q ₁ , Q ₃)	78.000 (68.000, 91.000)	78.000 (68.000, 90.000)	78.000 (67.000, 91.000)	77.000 (68.000, 89.500)	79.000 (68.000, 93.000)	0.703
Respiratory rate, M (Q ₁ , Q ₃)	21.000 (17.000, 25.000)	20.000 (17.000, 24.000)	21.000 (16.000, 25.000)	22.000 (17.000, 26.000)	22.000 (18.000, 27.000)	< 0.001
Temperature, M (Q ₁ , Q ₃)	36.830 (36.500, 37.170)	36.780 (36.560, 37.170)	36.830 (36.440, 37.220)	36.830 (36.500, 37.280)	36.780 (36.440, 37.110)	0.103
Infection site, n (%)						
Respiratory	460(28.135)	110(26.895)	109(26.650)	121(29.657)	120(29.340)	
Gastrointestinal	196(11.988)	52(12.714)	43(10.513)	60(14.706)	41(10.024)	
Genitourinary	223(13.639)	58(14.181)	57(13.936)	52(12.745)	56(13.692)	
Others	882(53.945)	220(53.790)	231(56.479)	211(51.716)	220(53.790)	
Charlson comorbidity index, M (Q ₁ , Q ₃)	8.000 (6.000, 10.000)	8.000 (6.000, 10.000)	8.000 (6.000, 10.000)	8.000 (6.000, 10.000)	8.000 (6.000, 10.000)	0.759
SOFA, M (Q ₁ , Q ₃)	7.000 (4.000, 9.000)	5.000 (4.000, 8.000)	6.000 (4.000, 9.000)	7.000 (4.000, 10.000)	8.000 (5.000, 11.000)	< 0.001
SAPSII, M (Q ₁ , Q ₃)	49.000 (39.000, 60.000)	45.000 (36.000, 55.000)	50.000 (40.000, 60.000)	49.000 (40.000, 60.000)	53.000 (42.000, 65.000)	< 0.001
INR, M (Q ₁ , Q ₃)	1.400 (1.200, 1.700)	1.300 (1.200, 1.600)	1.400 (1.200, 1.700)	1.400 (1.200, 1.700)	1.500 (1.300, 1.900)	< 0.001
PT, M (Q ₁ , Q ₃)	15.600 (13.700, 18.600)	14.800 (13.000, 17.400)	15.300 (13.600, 18.500)	15.550 (13.750, 18.200)	16.900 (14.600, 20.900)	< 0.001
PTT, M (Q ₁ , Q ₃)	32.600 (28.200, 39.700)	31.300 (28.300, 36.400)	32.500 (27.700, 39.000)	33.100 (28.350, 41.200)	33.700 (28.400, 42.700)	< 0.001
Albumin, M (Q ₁ , Q ₃)	2.700 (2.300, 3.100)	3.000 (2.600, 3.400)	2.700 (2.300, 3.000)	2.600 (2.100, 3.100)	2.500 (2.100, 2.900)	< 0.001
Calcium, M (Q ₁ , Q ₃)	8.100 (7.500, 8.700)	8.200 (7.600, 8.800)	8.100 (7.600, 8.600)	8.100 (7.500, 8.700)	8.100 (7.400, 8.600)	0.004
BUN, M (Q ₁ , Q ₃)	24.000 (16.000, 42.000)	21.000 (14.000, 34.000)	24.000 (16.000, 39.000)	24.500 (16.000, 40.000)	29.000 (18.000, 52.000)	< 0.001
Potassium, M (Q ₁ , Q ₃)	4.200 (3.700, 4.700)	4.000 (3.700, 4.600)	4.100 (3.700, 4.600)	4.100 (3.700, 4.700)	4.500 (3.900, 5.100)	< 0.001
Creatinine, M (Q ₁ , Q ₃)	1.100 (0.800, 1.800)	1.000 (0.700, 1.700)	1.100 (0.800, 1.700)	1.100 (0.700, 1.800)	1.300 (0.900, 2.100)	< 0.001
ALT, M (Q ₁ , Q ₃)	30.000 (16.000, 70.000)	19.000 (12.000, 36.000)	25.000 (14.000, 52.000)	34.500 (18.000, 72.500)	67.000 (25.000, 258.000)	< 0.001
ALP, M (Q ₁ , Q ₃)	99.000 (66.000, 172.000)	84.000 (59.000, 123.000)	90.000 (64.000, 145.000)	101.000 (70.000, 171.500)	141.000 (81.000, 260.000)	< 0.001
AST, M (Q ₁ , Q ₃)	44.700 (23.000, 120.000)	24.000 (16.000, 42.000)	37.000 (21.000, 66.000)	50.500 (30.500, 122.000)	150.000 (56.000, 585.000)	< 0.001
LDH, M (Q ₁ , Q ₃)	321.000 (228.000, 574.000)	186.000 (155.000, 221.000)	270.000 (234.000, 310.000)	412.500 (342.000, 491.000)	1,027.000 (703.000, 1,918.000)	< 0.001
Total Bilirubin, M (Q ₁ , Q ₃)	0.900 (0.400, 2.000)	0.700 (0.400, 1.300)	0.800 (0.400, 1.700)	0.900 (0.450, 2.100)	1.300 (0.600, 2.600)	< 0.001
WBC, M (Q ₁ , Q ₃)	11.500 (6.300, 18.300)	10.000 (5.300, 15.400)	11.800 (6.800, 18.000)	11.250 (5.650, 18.100)	13.700 (7.500, 24.000)	< 0.001
Neutrophils, M (Q ₁ , Q ₃)	9.087 (4.700, 14.898)	7.490 (3.846, 12.228)	9.550 (5.118, 14.613)	9.375 (4.654, 15.330)	10.140 (5.730, 17.140)	< 0.001
Monocytes, M (Q ₁ , Q ₃)	0.570 (0.269, 0.979)	0.520 (0.230, 0.847)	0.580 (0.325, 0.948)	0.573 (0.248, 1.014)	0.611 (0.280, 1.120)	0.028
Lymphocytes, M (Q ₁ , Q ₃)	0.830 (0.440, 1.373)	0.780 (0.436, 1.320)	0.790 (0.425, 1.300)	0.808 (0.381, 1.300)	0.965 (0.500, 1.590)	0.004

Table 1 (continued)

Variable	group					P value ²
	Overall N= 1,635 ¹	Quartile 1 N= 409 ¹	Quartile 2 N= 409 ¹	Quartile 3 N= 408 ¹	Quartile 4 N= 409 ¹	
Eosinophils, M (Q₁, Q₃)	0.018 (0.000, 0.093)	0.026 (0.000, 0.110)	0.024 (0.000, 0.114)	0.012 (0.000, 0.081)	0.000 (0.000, 0.061)	< 0.001
Basophils, M (Q₁, Q₃)	0.010 (0.000, 0.030)	0.013 (0.000, 0.031)	0.010 (0.000, 0.028)	0.010 (0.000, 0.029)	0.000 (0.000, 0.028)	< 0.001
RDW, M (Q₁, Q₃)	16.100 (14.600, 18.200)	15.700 (14.300, 17.600)	15.900 (14.600, 17.700)	16.400 (14.600, 18.300)	16.800 (15.000, 19.100)	< 0.001
Platelet, M (Q₁, Q₃)	155.000 (84.000, 256.000)	170.000 (102.000, 256.000)	165.000 (98.000, 259.000)	156.000 (87.000, 257.500)	121.000 (61.000, 231.000)	< 0.001
Lactate, M (Q₁, Q₃)	2.100 (1.400, 3.400)	1.700 (1.200, 2.530)	1.900 (1.300, 3.000)	2.100 (1.400, 3.250)	3.080 (1.900, 6.000)	< 0.001
Po₂, M (Q₁, Q₃)	78.000 (45.000, 134.000)	81.000 (47.000, 138.000)	81.000 (43.000, 141.000)	74.850 (44.000, 129.000)	73.000 (45.000, 122.000)	0.422
Pao₂/Fio₂ratio, M (Q₁, Q₃)	154.000 (100.000, 250.000)	162.500 (110.452, 261.000)	162.000 (100.000, 275.000)	148.505 (91.714, 226.333)	142.388 (94.000, 244.250)	0.004
Aado₂, M (Q₁, Q₃)	285.970 (194.500, 395.350)	265.350 (180.450, 344.675)	267.005 (176.750, 366.850)	300.060 (210.835, 413.550)	317.500 (220.170, 460.900)	< 0.001
Glucose, M (Q₁, Q₃)	131.000 (105.000, 171.000)	127.000 (107.000, 162.000)	132.000 (105.000, 171.000)	131.000 (104.500, 165.000)	136.000 (102.000, 200.000)	0.593
Gender, n (%)						0.746
No	651 (39.8%)	159 (38.9%)	169 (41.3%)	167 (40.9%)	156 (38.1%)	
Yes	984 (60.2%)	250 (61.1%)	240 (58.7%)	241 (59.1%)	253 (61.9%)	
Myocardial infarct, n (%)						0.101
No	1,418 (86.7%)	361 (88.3%)	358 (87.5%)	359 (88.0%)	340 (83.1%)	
Yes	217 (13.3%)	48 (11.7%)	51 (12.5%)	49 (12.0%)	69 (16.9%)	
Congestive heart failure, n (%)						0.130
No	1,214 (74.3%)	307 (75.1%)	287 (70.2%)	304 (74.5%)	316 (77.3%)	
Yes	421 (25.7%)	102 (24.9%)	122 (29.8%)	104 (25.5%)	93 (22.7%)	
Peripheral vascular disease, n (%)						0.841
No	1,504 (92.0%)	380 (92.9%)	374 (91.4%)	373 (91.4%)	377 (92.2%)	
Yes	131 (8.0%)	29 (7.1%)	35 (8.6%)	35 (8.6%)	32 (7.8%)	
Cerebrovascular dis- ease, n (%)						0.778
No	1,474 (90.2%)	372 (91.0%)	368 (90.0%)	363 (89.0%)	371 (90.7%)	
Yes	161 (9.8%)	37 (9.0%)	41 (10.0%)	45 (11.0%)	38 (9.3%)	
Chronic pulmonary disease, n (%)						0.018
No	1,206 (73.8%)	284 (69.4%)	306 (74.8%)	294 (72.1%)	322 (78.7%)	
Yes	429 (26.2%)	125 (30.6%)	103 (25.2%)	114 (27.9%)	87 (21.3%)	
Diabetes, n (%)						0.883
No	1,206 (73.8%)	297 (72.6%)	300 (73.3%)	306 (75.0%)	303 (74.1%)	
Yes	429 (26.2%)	112 (27.4%)	109 (26.7%)	102 (25.0%)	106 (25.9%)	
Renal disease, n (%)						0.426
No	1,285 (78.6%)	312 (76.3%)	318 (77.8%)	329 (80.6%)	326 (79.7%)	
Yes	350 (21.4%)	97 (23.7%)	91 (22.2%)	79 (19.4%)	83 (20.3%)	
Severe liver disease, n (%)						0.206
No	1,427 (87.3%)	366 (89.5%)	362 (88.5%)	350 (85.8%)	349 (85.3%)	
Yes	208 (12.7%)	43 (10.5%)	47 (11.5%)	58 (14.2%)	60 (14.7%)	
Metastatic solid tumor, n (%)						0.075
No	1,058 (64.7%)	281 (68.7%)	273 (66.7%)	253 (62.0%)	251 (61.4%)	
Yes	577 (35.3%)	128 (31.3%)	136 (33.3%)	155 (38.0%)	158 (38.6%)	

Table 1 (continued)

Variable	group					P value ²
	Overall N= 1,635 ¹	Quartile 1 N= 409 ¹	Quartile 2 N= 409 ¹	Quartile 3 N= 408 ¹	Quartile 4 N= 409 ¹	
Hematological malignancy, n (%)						0.214
No	1360(83.180)	352(86.064)	337(82.396)	341(83.578)	330(80.685)	
Yes	275(16.820)	57(13.936)	72(17.604)	67(16.422)	79(19.315)	
Vasoactive, n (%)						0.119
No	745 (45.6%)	206 (50.4%)	175 (42.8%)	177 (43.4%)	187 (45.7%)	
Yes	890 (54.4%)	203 (49.6%)	234 (57.2%)	231 (56.6%)	222 (54.3%)	
Mechanical Ventilation, n (%)						0.138
No	192 (11.7%)	56 (13.7%)	36 (8.8%)	47 (11.5%)	53 (13.0%)	
Yes	1,443 (88.3%)	353 (86.3%)	373 (91.2%)	361 (88.5%)	356 (87.0%)	
CRRT, n (%)						<0.001
No	1,439 (88.0%)	392 (95.8%)	371 (90.7%)	363 (89.0%)	313 (76.5%)	
Yes	196 (12.0%)	17 (4.2%)	38 (9.3%)	45 (11.0%)	96 (23.5%)	
AKI, n (%)						<0.001
No	204 (12.5%)	80 (19.6%)	40 (9.8%)	41 (10.0%)	43 (10.5%)	
Yes	1,431 (87.5%)	329 (80.4%)	369 (90.2%)	367 (90.0%)	366 (89.5%)	
28 Day ICU mortality, n (%)						<0.001
No	950 (58.1%)	312 (76.3%)	253 (61.9%)	204 (50.0%)	181 (44.3%)	
Yes	685 (41.9%)	97 (23.7%)	156 (38.1%)	204 (50.0%)	228 (55.7%)	
ICU mortality, n (%)						<0.001
No	1,220 (74.6%)	357 (87.3%)	332 (81.2%)	286 (70.1%)	245 (59.9%)	
Yes	415 (25.4%)	52 (12.7%)	77 (18.8%)	122 (29.9%)	164 (40.1%)	
In hospital mortality, n (%)						<0.001
No	1,030 (63.0%)	331 (80.9%)	278 (68.0%)	228 (55.9%)	193 (47.2%)	
Yes	605 (37.0%)	78 (19.1%)	131 (32.0%)	180 (44.1%)	216 (52.8%)	
Hospital_Los_Day, M (Q₁, Q₃)	12.988 (7.008, 23.410)	12.051 (6.786, 20.547)	14.390 (8.565, 24.566)	13.929 (7.354, 23.650)	11.733 (5.438, 24.350)	0.004
ICU_Los_Day, M (Q₁, Q₃)	4.560 (2.320, 9.200)	3.940 (2.100, 7.510)	4.960 (2.850, 10.050)	5.230 (2.560, 11.020)	4.160 (1.980, 8.890)	<0.001

¹ Median (IQR) or frequency (%)² Kruskal–Wallis rank sum test; Pearson's chi-square test

LDAR: Lactate Dehydrogenase to Albumin Ratio; BMI: Body Mass Index; MAP: Mean Arterial Pressure; SOFA: Sequential Organ Failure Assessment; APACHE II: Acute Physiology Score II; SAPS II: Simplified Acute Physiology Score II; OASIS: Oxford Acute Severity of Illness Score; INR: International Normalized Ratio; PT: Prothrombin Time; PTT: Partial Thromboplastin Time; BUN: Blood Urea Nitrogen; ALT: Alanine Aminotransferase; ALP: Alkaline Phosphatase; AST: Aspartate Aminotransferase; LDH: Lactate Dehydrogenase; WBC: White Blood Cell Count; RDW: Red Cell Distribution Width; AADO₂: Alveolar-Arterial Oxygen Difference; CRRT: Continuous Renal Replacement Therapy; AKI: Acute Kidney Injury

prognostic risk indicator in sepsis patients. When 28-day survival in ICU patients was analyzed, Fig. 4 shows significant differences in survival curves among the LDAR quartiles ($p < 0.0001$). The Kaplan–Meier survival curves revealed the highest survival rates in Quartile 1 and the lowest rates in Quartile 4. By Day 28 in the ICU, Quartile 1 patients had a survival rate close to 80%, whereas Quartile 4 patients had a survival rate of approximately 50%. Dynamic changes in LDAR were significantly associated with survival.

According to the RCS analysis, the survival probability of the high-LDAR group significantly decreased compared with that of the low-LDAR group (Figure S2).

The difference between the groups was highly significant ($P < 0.0001$), reinforcing LDAR as a robust prognostic biomarker for ICU mortality in septic patients with malignancies.

Figure 5 Displays the feature selection results via the Boruta algorithm, identifying variables in green as important, those in red as unimportant, and those in yellow as tentative. LDAR was identified as the most important parameter.

The hyperparameters of the nine models are listed in Table S4, and their detailed performance metrics are shown in Table S5. Figure 6A presents the ROC curves of these models, with AUC values indicating model

Table 2 Relationships between the LDAR index groups and 28-day ICU mortality, in-hospital mortality, and ICU mortality

Exposure	model1		model2	
	OR (95%CI)	P value	OR (95%CI)	P value
28-day ICU mortality				
LDAR				
Quartile 1				
Quartile 2	1.983 (1.465,2.685)	<0.001	1.757 (1.281,2.412)	<0.001
Quartile 3	3.216 (2.385,4.339)	<0.001	2.967 (2.166,4.065)	<0.001
Quartile 4	4.052 (3.002,5.469)	<0.001	3.441 (2.497,4.741)	<0.001
ICU mortality				
LDAR				
Quartile 1				
Quartile 2	1.592(1.086,2.334)	0.017	1.323 (0.890,1.967)	0.166
Quartile 3	2.929 (2.043,4.197)	<0.001	2.465 (1.693,3.591)	<0.001
Quartile 4	4.596 (3.233,6.533)	<0.001	3.478 (2.396,5.049)	<0.001
In-hospital mortality				
LDAR				
Quartile 1				
Quartile 2	2.000 (1.448,2.761)	<0.001	1.719 (1.230,2.402)	0.002
Quartile 3	3.350 (2.446,4.589)	<0.001	2.919(2.102,4.054)	<0.001
Quartile 4	4.749 (3.470,6.501)	<0.001	3.747 (2.688,5.222)	<0.001

Model 1: Unadjusted

Model 2: Adjusted for age, sex, BMI, vital signs (heart rate, mean arterial pressure, respiratory rate, oxygen saturation, and body temperature), SOFA score, and the presence of metastatic solid tumors

LDAR: Lactate Dehydrogenase to Albumin Ratio

Table 3 Comparison of clinical characteristics between patients with nonmetastatic and metastatic solid tumors

Variable	Group			P Value
	Overall (n = 1635)	Nonmetastatic solid tumor (n = 1058)	metastatic solid tumor (n = 577)	
LDAR, M (Q ₁ , Q ₃)	122.80 (81.62, 229.46)	118.54 (80.07, 220.83)	135.00 (85.91, 244.44)	0.004
Albumin, M (Q ₁ , Q ₃)	2.70 (2.30, 3.10)	2.80 (2.40, 3.20)	2.60 (2.20, 3.00)	<0.001
LDH, M (Q ₁ , Q ₃)	321.00 (228.00, 574.00)	316.50 (230.00, 543.75)	334.00 (225.00, 634.00)	0.399
ICU-mortality, n (%)	415 (25.4)	239 (22.6)	176 (30.5)	<0.001
In-hospital mortality, n (%)	605 (37.0)	362 (34.2)	243 (42.1)	0.002
28-Day ICU mortality, n (%)	685 (41.9)	385 (36.4)	300 (52.0)	<0.001
ICU_Los_Day, M (Q ₁ , Q ₃)	4.56 (2.32, 9.19)	4.88 (2.43, 9.88)	4.09 (2.16, 7.69)	0.001
Hospital_Los_Day, M (Q ₁ , Q ₃)	12.99 (7.01, 23.35)	13.94 (7.58, 26.82)	11.73 (6.32, 19.50)	<0.001

M: Median, Q1: 1st quartile, Q3: 3rd quartile

LDAR: Lactate dehydrogenase to albumin ratio; LDH: Lactate dehydrogenase

performance: XGBoost (0.711), logistic regression (0.661), LGBM (0.612), random forest (0.751), AdaBoost (0.698), MLP (0.651), SVM (0.678), GNB (0.667), and KNN (0.621). Figure 6B shows that the random forest model had the highest net benefit within a 20-60% threshold probability range. The calibration curves (Figure S3) indicate that the random forest model closely aligned with the observed outcomes, achieving the best calibration performance (Brier score = 0.201). Thus, the random forest model was selected as the optimal model for predicting 28-day ICU mortality. Figure S4 shows its performance in the test set (AUC = 0.727, accuracy = 0.678).

Interpretability analysis

SHAP was utilized for intuitive model interpretation. Figure 7A presents a scatterplot ranking features by their cumulative SHAP value impact, highlighting LDAR, ALP, age, BUN, LDH, PO2/FiO2, and others as the top 20 predictors for 28-day ICU mortality. Figure 7B illustrates a case study of the model’s prediction process for an individual patient, with red indicating positive contributions and blue representing negative contributions. The model predicted a 65% probability of ICU mortality within 28 days for this patient.

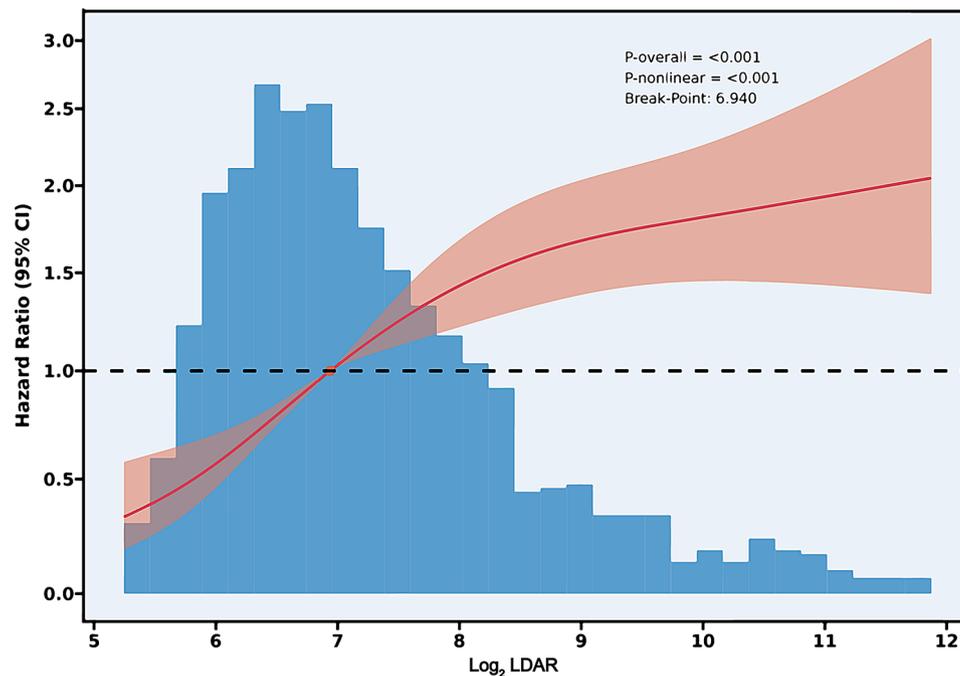


Fig. 2 Relationship between Log₂ LDAR and 28-Day ICU Mortality Risk. The solid red line represents the estimated adjusted hazard ratios (HRs), whereas the shaded area indicates the 95% CI. The solid red dot corresponds to the inflection point (log₂ LDAR = 6.940), representing the minimum hazard ratio. The horizontal dashed line indicates a hazard ratio of 1.0. Statistical significance was observed for both the overall model (P-overall < 0.001) and the non-linear relationship (P-nonlinear < 0.001). LDAR: lactate dehydrogenase-to-albumin ratio; CI: confidence interval

Relationships between LDAR and other prognostic assessment indices

The ROC curve analysis (Fig. 8) demonstrated that the LDAR score consistently outperformed the other scoring systems and indices in predicting patient outcomes, achieving the highest AUC values of 0.651 and 0.663 for the two clinical outcomes, respectively. In comparison, SAPSII (AUC = 0.636 and 0.650) and SOFA (AUC = 0.584 and 0.607) showed moderate predictive capability, whereas the indices PNI, SIRI, and SII exhibited lower predictive performance, with AUC values below 0.58 in both analyses. This highlights the superior discriminative ability of LDAR in this context.

Discussion

This study provides a comprehensive evaluation of LDAR as a prognostic biomarker in septic patients with malignancies. The findings highlight the unique value of the LDAR in predicting 28-day mortality, with a significant nonlinear relationship identified and a critical threshold of log₂(LDAR) = 6.940 established. The nonlinear pattern suggests that once lactate dehydrogenase and albumin values shift beyond a critical ratio, it may reflect irreversible organ dysfunction, severe tissue hypoxia, or profound metabolic disturbance—particularly relevant in cancer patients whose disease burden and treatment side effects accelerate these pathophysiological processes. Early recognition of high LDAR (≥ 6.940 in log₂

scale) may help clinicians identify patients at risk of deterioration. Aggressive interventions—such as optimizing hemodynamics, improving nutritional support, and close monitoring—could potentially mitigate mortality in this subgroup.

The Boruta algorithm is a popular tool for feature selection, employing a randomness-based approach to determine which variables are most pertinent to predicting a target outcome [16]. In the present study, Boruta identified LDAR as the top-ranking feature within the green zone, with the highest Z score among all selected variables. This finding suggests that LDAR could play a critical role in the study and is strongly associated with the research objectives. The analysis underscored the relevance of LDAR in predicting 28-day all-cause mortality among sepsis patients. Nonetheless, it is essential to acknowledge that LDAR might not serve as the sole determinant. First, although the Boruta algorithm is a powerful tool for feature selection, its results can be influenced by correlations between variables. Consequently, while LDAR's significance within the model is notable, it does not necessarily indicate that it is the most critical factor. Second, logistic regression analysis further demonstrated that higher LDAR levels are linked to an increased risk of 28-day mortality in sepsis patients, which aligns with Boruta's classification of LDAR as an essential feature. This finding reinforces the evidence

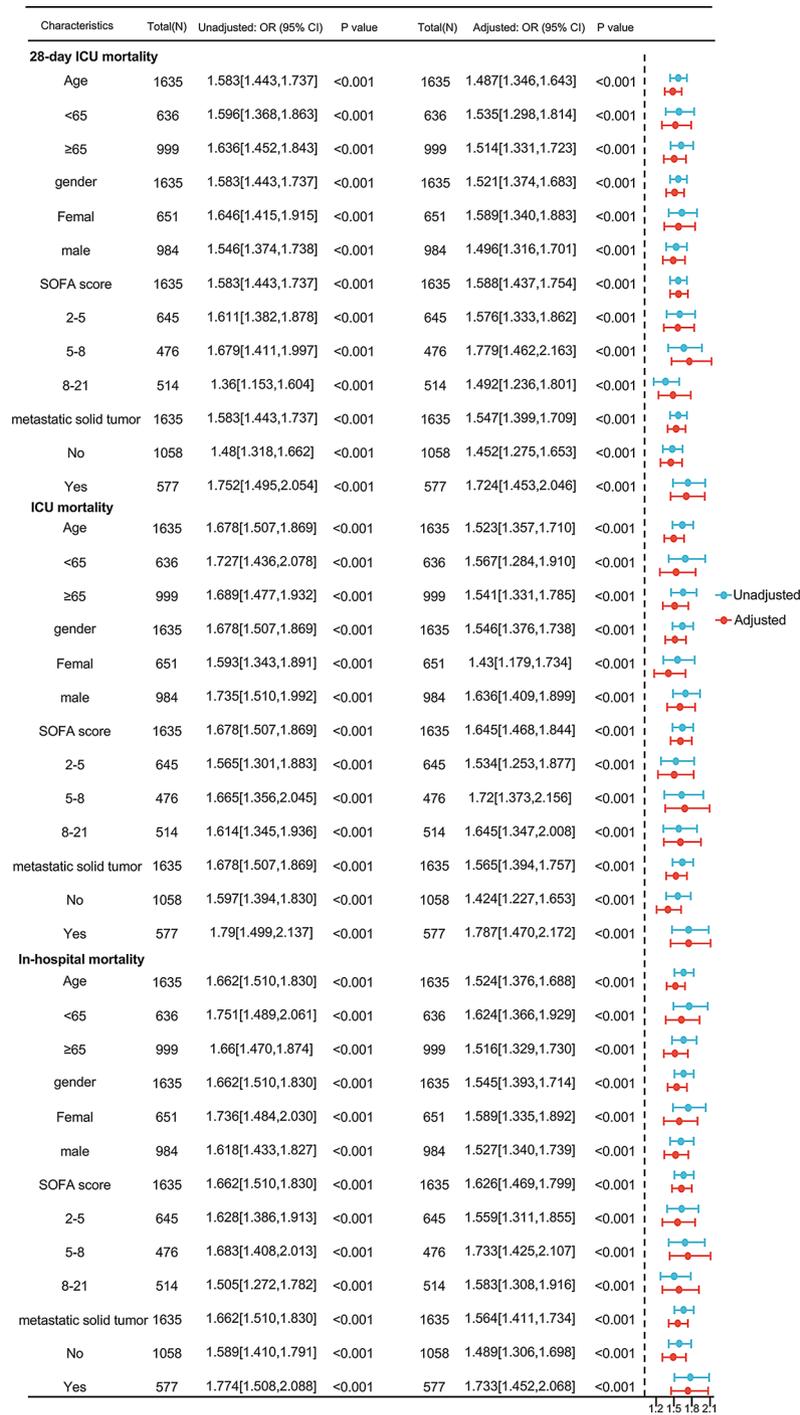


Fig. 3 Subgroup analysis of the associations between LDAR and ICU and hospital mortality outcomes. Each subgroup's hazard ratio is depicted as a dot, with horizontal lines indicating the 95% CI. Red markers represent the adjusted HR, whereas blue markers represent the unadjusted HR. LDAR: lactate dehydrogenase-to-albumin ratio; HR: hazard ratio; CI: confidence interval

supporting LDAR as a potential predictive marker for 28-day all-cause mortality in sepsis patients.

Recent studies have highlighted the lactate dehydrogenase-to-albumin ratio as a significant prognostic marker in critical care settings. Xiao et al. conducted a retrospective study on 5,784 sepsis patients from the

MIMIC-IV database and reported that elevated LDAR values were independently associated with 28-day and 90-day all-cause mortality [17], and our identification of an inflection point at $\log_2(\text{LDAR}) = 6.940$ further refines this relationship. Similarly, Guan et al. reported that an increased LDH/Alb ratio (≥ 10.57) was linked to

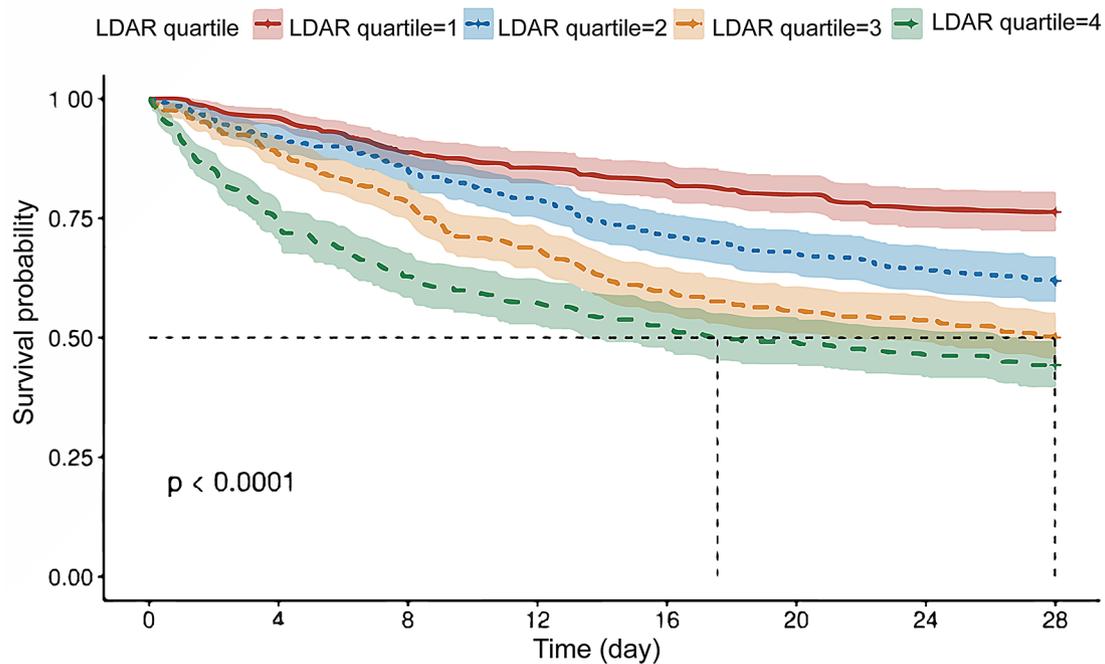


Fig. 4 Kaplan-Meier survival curves for 28-day ICU mortality stratified by LDAR quartiles. The colored lines represent the survival probabilities for quartiles 1 (red), 2 (blue), 3 (orange), and 4 (green). The shaded areas indicate 95% confidence intervals. Survival rates decrease progressively from Quartile 1 to Quartile 4, with Quartile 4 showing the lowest survival probability by Day 28. The table below displays the number of patients at risk over time for each quartile. The statistical significance of differences among quartiles is denoted by $P < 0.0001$. LDAR: Lactate dehydrogenase to albumin ratio

increased all-cause mortality in patients with severe sepsis [18]. However, while these studies primarily focused on general sepsis populations, our work specifically addresses septic patients with malignancies—a group with unique immune–metabolic challenges. Elevated LDAR levels are correlated with increased mortality across various conditions, including cardiac arrest, acute pulmonary embolism, and acute respiratory distress syndrome [14, 19, 20]. These studies highlight the prognostic value of LDAR; however, subgroup analyses specifically for patients with malignancies have not been performed. Patients with malignancies often present unique pathophysiological characteristics due to tumor progression or treatment-related factors, which may contribute to elevated LDH levels and reduced albumin levels [21–24]. These differences inherently distinguish their LDH and albumin profiles from those of the general sepsis population, potentially altering the prognostic value of LDAR. Our study supports these distinctions, demonstrating that the average LDAR in the included cohort was greater than that reported in similar studies. Furthermore, in our study, metastatic cancer patients typically exhibited marked LDH elevations and lower albumin levels, resulting in particularly high LDAR values that were directly associated with poorer outcomes. By using restricted cubic spline analysis, we captured the nonlinear relationship between LDAR and mortality—an aspect that linear models in previous studies overlooked. While individual

markers provide valuable prognostic insights—elevated serum LDH has been identified as an independent risk factor for 28-day mortality in sepsis [25], and low serum albumin levels (for instance, below 2.5 mg/dL) along with high SOFA scores at discharge have been linked to poorer long-term outcomes in sepsis survivors [26], with low albumin consistently associated with increased risk of death in severe sepsis regardless of exogenous administration [27]—these markers alone may not fully capture the complex pathophysiology of sepsis in cancer patients. In contrast, the composite LDAR leverages the strengths of both markers, underscoring its robust prognostic value for risk stratification and tailored intervention in this high-risk population.

Machine learning has been widely applied in critical care settings to enhance risk stratification and improve outcome prediction for septic patients. Previous studies have demonstrated the effectiveness of Machine learning models in classifying sepsis subtypes, identifying high-risk patient groups, and optimizing individualized treatment strategies. For instance, Qin et al. utilized ML-driven clustering to derive four distinct pediatric sepsis phenotypes based on inflammatory markers and organ dysfunction [28]. Their approach provided a refined understanding of sepsis heterogeneity, facilitating personalized therapeutic interventions. Similarly, Takkavatakarn et al. developed ML-based serum creatinine trajectory models for acute kidney injury (AKI)

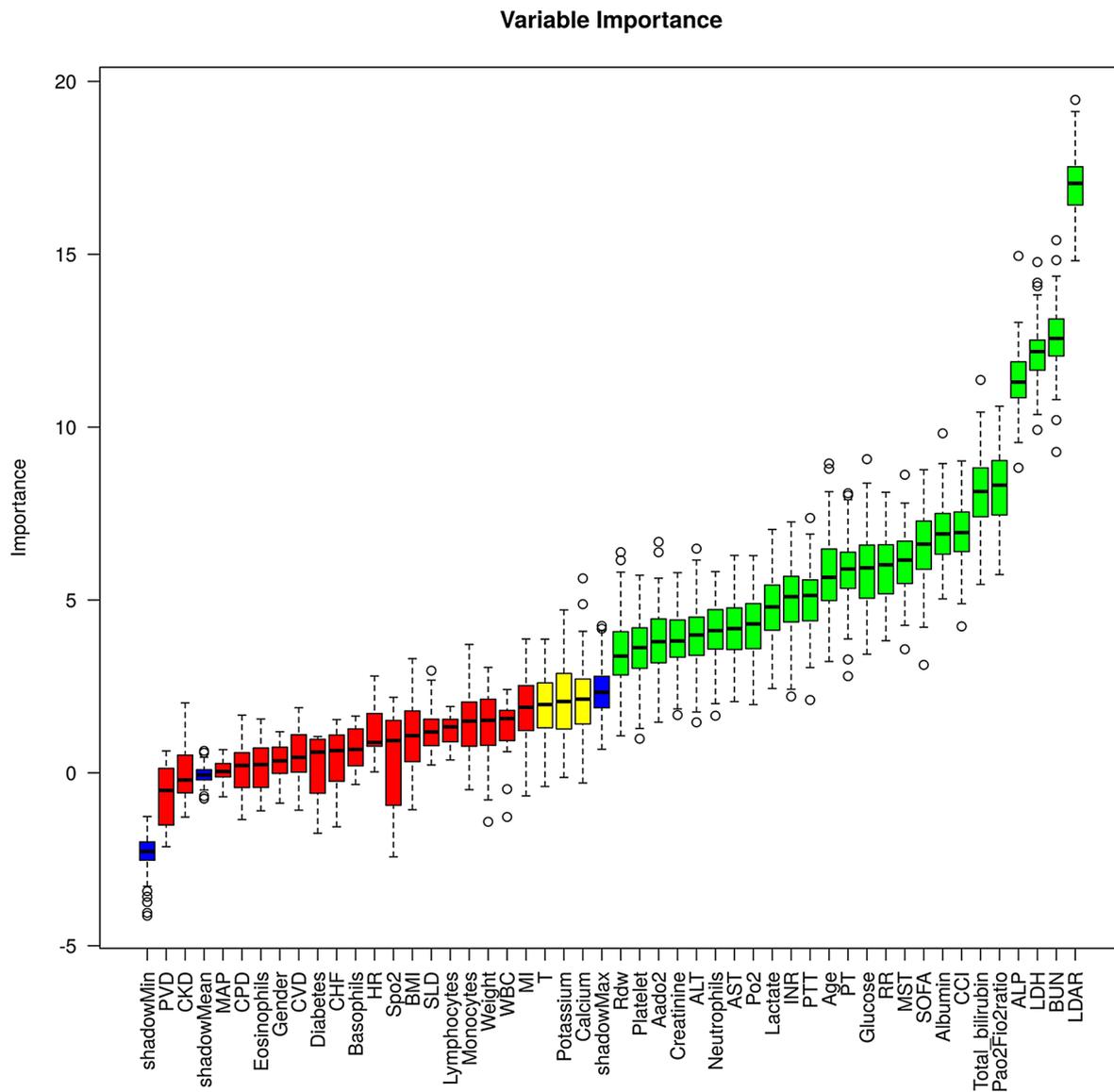


Fig. 5 Variable importance identified by the Boruta algorithm. The variables are ranked by importance from left to right, with boxplots representing the distribution of importance scores for each variable. Green indicates variables deemed important, yellow represents tentative variables, and red signifies unimportant variables. PVD: peripheral vascular disease; CKD: chronic kidney disease; MAP: mean arterial pressure; CPD: chronic pulmonary disease; CVD: cardiovascular disease; CHF: congestive heart failure; HR: heart rate; BMI: body mass index; SLD: severe liver disease; WBC: white blood cell; MI: myocardial infarction; T: temperature; RDW: red cell distribution width; ALT: alanine aminotransferase; AST: aspartate aminotransferase; INR: international normalized ratio; PTT: partial thromboplastin time; PT: prothrombin time; RR: respiratory rate; MST: metastatic solid tumor; SOFA: Sequential Organ Failure Assessment; CCI: Charlson Comorbidity Index; ALP: alkaline phosphatase; LDH: lactate dehydrogenase; BUN: blood urea nitrogen; LDAR: lactate dehydrogenase-to-albumin ratio

in septic patients [29]. Gao et al. developed a random forest model using a reduced set of clinical features to enhance interpretability [30]. In contrast, our study focuses on the prognostic value of the LDAR, a readily available biomarker for mortality risk stratification in septic patients with malignancies. Our model's AUC was relatively lower than that of prior studies, which may be attributed to several factors. First, many previous ML-based sepsis prediction models incorporated a broader range of inflammatory biomarkers (e.g., procalcitonin,

interleukins, or C-reactive protein), which were not available in our dataset because of high missing data rates, as well as the unique characteristics of septic patients with malignancies, who exhibit distinct immune-metabolic responses that may not align with those of general ICU populations.

From a mechanistic perspective, LDAR encapsulates the combined effects of LDH and albumin levels, two well-established markers of metabolic and inflammatory disturbances. Elevated LDH reflects tissue injury and

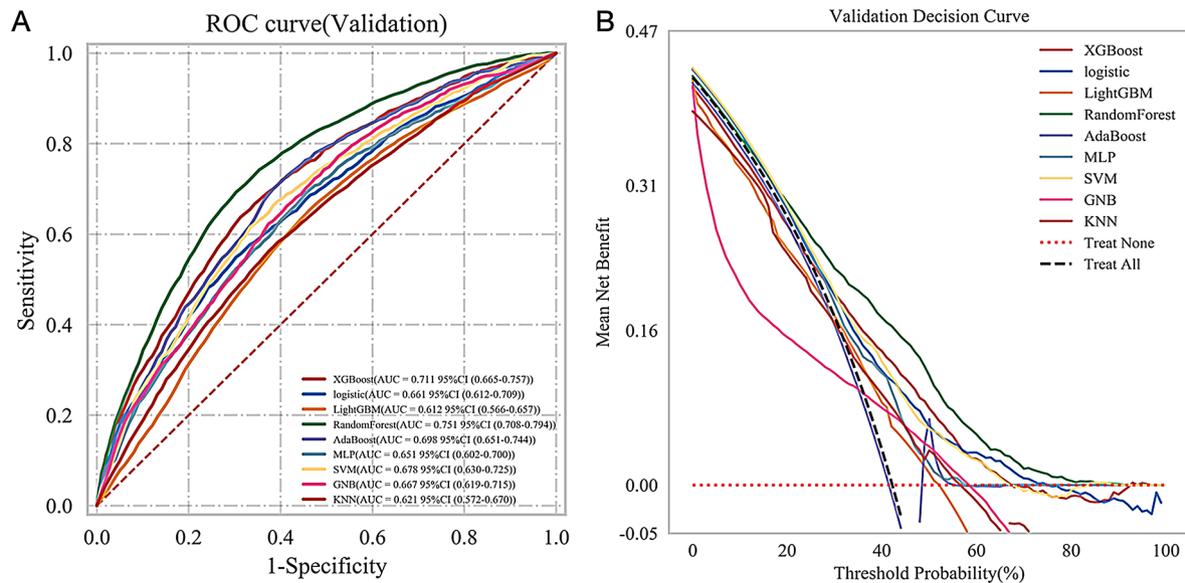


Fig. 6 Model validation: ROC and decision curve analyses. **(A)** ROC curves for various models in the validation dataset. The AUC values demonstrate the predictive performance of each model. **(B)** Decision curve analysis for the models in the validation dataset. The “Treat None” and “Treat All” strategies are represented by the red dotted and black dashed lines, respectively. LightGBM: Light Gradient Boosting Machine; MLP: Multilayer Perceptron; SVM: Support Vector Machine; GNB: Gaussian Naive Bayes; KNN: K-Nearest Neighbors; XGBoost: Extreme Gradient Boosting; AdaBoost: Adaptive Boosting. ROC: Receiver Operating Characteristics; AUC: Area under the curve

metabolic dysregulation, which are commonly observed in sepsis and malignancy due to increased oxidative stress and anaerobic glycolysis [21, 31]. Concurrently, hypoalbuminemia indicates nutritional deficits and immune dysfunction, further exacerbating systemic vulnerability [32–34]. LDAR, as a composite index, integrates these pathological processes, offering a holistic reflection of the interplay between inflammation and metabolic failure. The pronounced mortality risk at higher LDAR levels suggests that oxidative stress and immune suppression may reach critical thresholds, triggering irreversible organ failure. Our study further revealed that when the $\log_2(\text{LDAR})$ exceeded 6.940, the ICU 28-day mortality rate significantly increased, highlighting a potential threshold beyond which the risk of adverse outcomes escalated markedly. This finding underscores the critical role of LDAR in reflecting disease severity and its potential utility as a predictive marker for clinical outcomes in sepsis patients.

Our study demonstrated that LDAR outperforms traditional scoring systems (SAPSII, SOFA) and other biomarkers (PNI, SIRI, and SII) in predicting both 28-day ICU mortality and in-hospital mortality. However, relying on a single indicator for prognostic prediction may lack accuracy; therefore, LDAR can be integrated into ICU scoring systems such as SOFA or SAPSII to improve early risk prediction and resource allocation. Early recognition of high LDAR (≥ 6.940 in \log_2 scale) may help clinicians identify patients at risk of deterioration. Aggressive interventions—such as optimizing hemodynamics, improving

nutritional support, and close monitoring—could potentially mitigate mortality in this subgroup. Monitoring LDAR trends may offer insights into disease progression and treatment response; for example, a sustained increase could indicate worsening conditions, suggesting the need for intensified interventions, whereas a decrease might signal recovery, supporting de-escalation decisions. Thus, LDAR serves as both a prognostic marker and a dynamic indicator of the clinical course.

While this study establishes the significance of LDAR, several limitations warrant discussion. The retrospective, single-center design limits the generalizability of the findings. Although rigorous statistical adjustments were employed to account for confounders, prospective studies are needed to confirm the causal relationship between LDAR and outcomes. Additionally, the study’s focus on baseline LDAR values excludes an exploration of longitudinal trends, which could further enhance its prognostic utility. Another limitation lies in the absence of external validation, particularly across diverse patient populations with varying healthcare practices. Future multicenter studies are needed to confirm its broader applicability, refine dynamic monitoring, and explore integration with other biomarkers for improved predictive models.

Conclusion

This study highlights LDAR as an independent prognostic biomarker for 28-day mortality in septic patients with malignancies. Prospective multicenter studies are

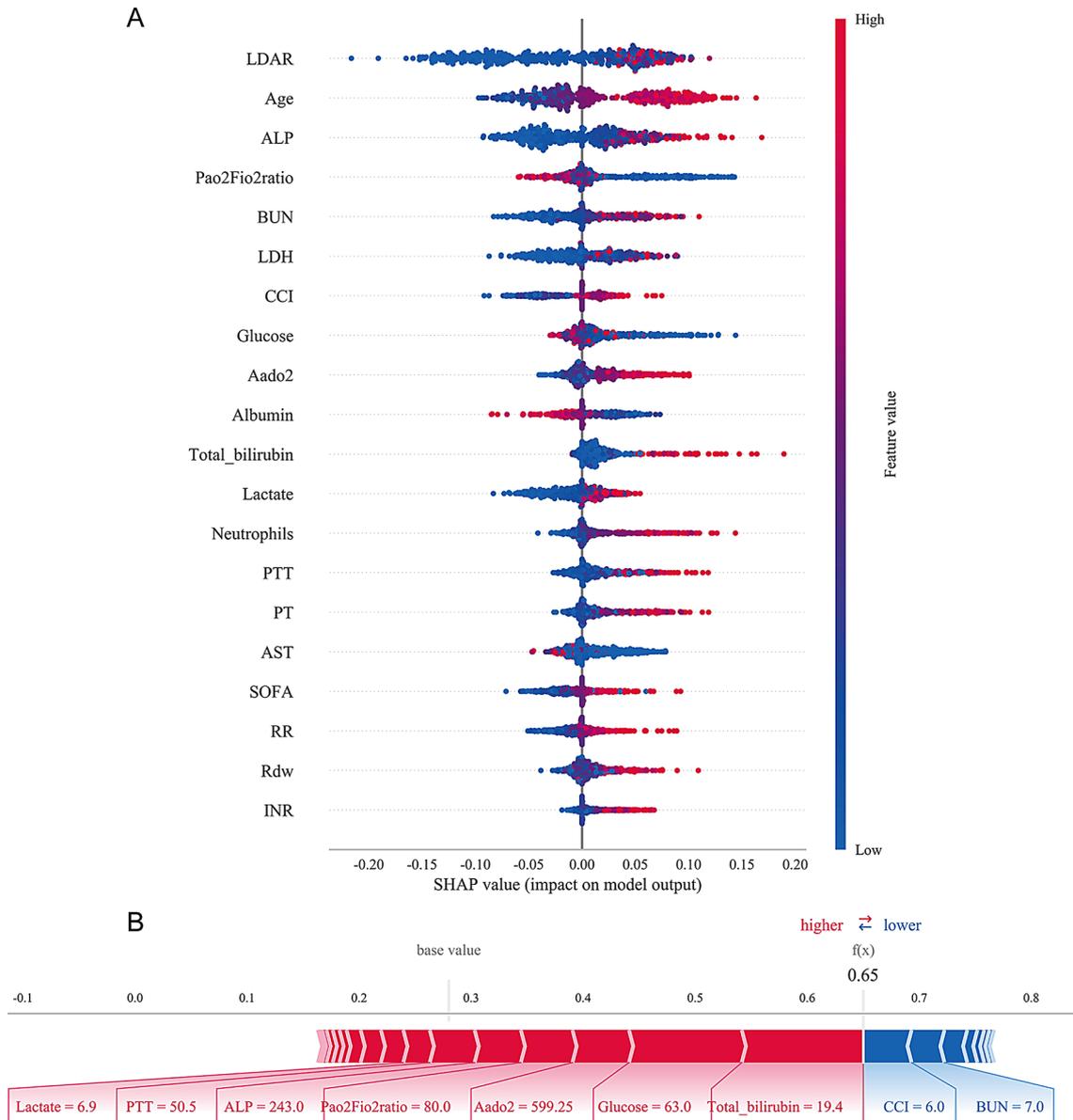


Fig. 7 SHAP Analysis for Model Explainability. **(A)** SHAP (Shapley additive explanations) summary plot demonstrating the impact of each feature on the predictive output of the random forest model for 28-day ICU mortality. The features are ranked by their importance. Each dot represents a single prediction, and the color gradient indicates the feature value (red for high and blue for low). SHAP values on the x-axis quantify the impact of each feature on the model's output. **(B)** SHAP force plot visualizing a specific case in detail. The red bars indicate positive contributions to the prediction of 28-day ICU mortality, whereas the blue bars represent negative contributions. The base value represents the average prediction probability, and the final output probability ($f(x)$) for this case is 0.65, indicating a 65% predicted probability of mortality. The key contributing features for this prediction include LDAR, ALP, BUN, and glucose levels. SHAP: specific additive explanations; LDAR: lactate dehydrogenase to albumin ratio; ALP: alkaline phosphatase; BUN: blood urea nitrogen; LDH: lactate dehydrogenase; CCI: Charlson comorbidity index; PTT: partial thromboplastin time; PT: prothrombin time; AST: aspartate aminotransferase; SOFA: sequential organ failure assessment; RR: respiratory rate; RDW: red cell distribution width; INR: international normalized ratio

needed to confirm these findings and evaluate longitudinal trends.

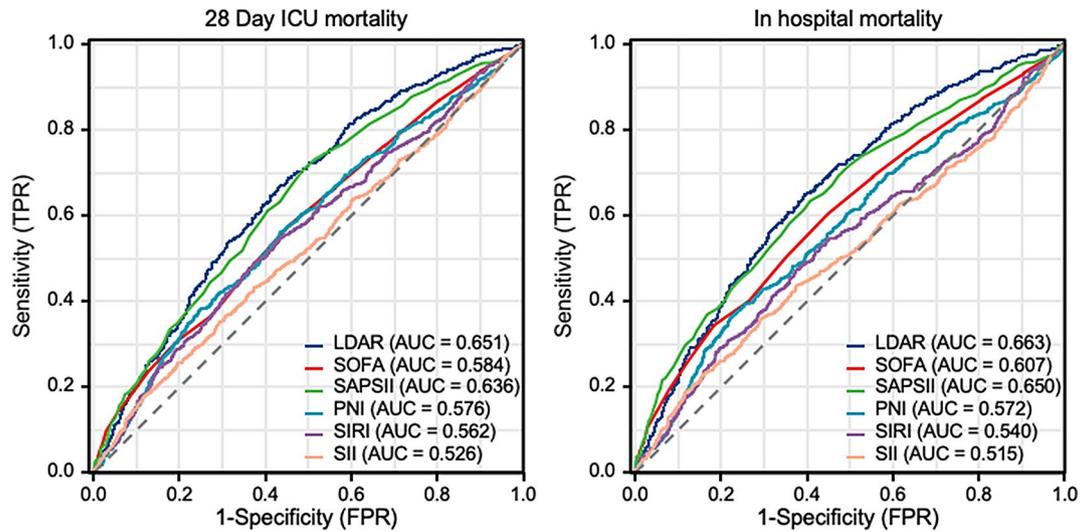


Fig. 8 Receiver operating characteristic curves for mortality prediction. **(A)** The predictive performance of various models for 28-day ICU mortality; **(B)** the predictive performance of the same models for in-hospital mortality. The dashed diagonal lines represent the reference line for random classification (AUC=0.5); ROC: receiver operating characteristic; LDAR: lactate dehydrogenase to albumin ratio; SOFA: sequential organ failure assessment; SAPSII: simplified acute physiology score II; PNI: prognostic nutritional index; SIRI: systemic inflammation response index; and SII: systemic immune-inflammation index

Abbreviations

LDAR	Lactate Dehydrogenase to Albumin Ratio
BMI	Body Mass Index
MAP	Mean Arterial Pressure
SOFA	Sequential Organ Failure Assessment
APSIII	Acute Physiology Score III
SAPSII	Simplified Acute Physiology Score II
OASIS	Oxford Acute Severity of Illness Score
INR	International Normalized Ratio
PT	Prothrombin Time
PTT	Partial Thromboplastin Time
BUN	Blood Urea Nitrogen
ALT	Alanine Aminotransferase
ALP	Alkaline Phosphatase
AST	Aspartate Aminotransferase
LDH	Lactate Dehydrogenase
WBC	White Blood Cell Count
RDW	Red Cell Distribution Width
AADO2	Alveolar-Arterial Oxygen Difference
CRRT	Continuous Renal Replacement Therapy
AKI	Acute Kidney Injury

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-025-14013-2>.

Supplementary Material 1: Additional file 1: Table S1. The variance inflation factor values of the variables. Table S2. Proportion of missing data. Table S3. Baseline characteristics of participants categorized into RCS-derived LDAR groups. Table S4. Hyperparameters of the nine machine learning models. Table S5. The performance of the nine machine learning models.

Supplementary Material 2: Additional file 2. Figure S1. Schoenfeld Residual Plot for $\log_2(\text{LDAR})$ Demonstrating Proportional Hazards. Figure S2. K-M survival curves for 28-day ICU mortality stratified by LDAR level. Figure S3. Calibration curve for model validation. Figure S4. Performance of the random forest model on the test set.

Acknowledgements

We extend our gratitude to all the administrators involved in the collection, organization, and maintenance of the MIMIC database.

Author contributions

Y. S., K. L. and L. Y. contributed equally to this article. Y. Y. contributed to the research design. Y. S. and L. Y. contributed to data collection, data processing and graphing. Y. S. and K. L. contributed to data proofreading and formal analysis. Y. S. and J. W. contributed to writing the manuscript. W. Z. and P. Z. contributed to the review and editing of the manuscript. All the authors have read and approved the final manuscript.

Funding

This work was supported by the Joint Funds for the Innovation of Science and Technology, Fujian Province (Grant numbers: 2023Y9418 and 2024Y9613).

Data availability

The data used in this study were obtained from the MIMIC database, which requires researchers to submit a formal application for access. As a result, the data cannot be made publicly available. However, the corresponding author of this paper, Yongshi Shen, may provide access to the data upon reasonable request, provided that the requester has obtained the necessary permissions through the MIMIC database application process.

Declarations

Ethics approval and consent to participate

The collection of patient information and creation of the research resource was reviewed by the Institutional Review Board at the Beth Israel Deaconess Medical Center, who granted a waiver of informed consent and approved the data sharing initiative.

Consent for publication

Not applicable.

Informed consent

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Intensive Care Unit, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China

²Department of Service Center, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China

Received: 7 February 2025 / Accepted: 25 March 2025

Published online: 08 April 2025

References

- Torres VB, Azevedo LC, Silva UV, et al. Sepsis-Associated outcomes in critically ill patients with malignancies. *Ann Am Thorac Soc*. 2015;12(8):1185–92.
- Wiersinga WJ, van der Poll T. Immunopathophysiology of human sepsis. *EBioMedicine*. 2022;86:104363.
- Schmoch T, Möhnle P, Weigand MA, et al. The prevalence of sepsis-induced coagulopathy in patients with sepsis - a secondary analysis of two German multicenter randomized controlled trials. *Ann Intensive Care*. 2023;13(1):3.
- Dellinger RP, Rhodes A, Evans L, et al. Surviving sepsis campaign. *Crit Care Med*. 2023;51(4):431–44.
- Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intens Care Med*. 2021;47(11):1181–247.
- AbuSara AK, Nazer LH, Hawari FI. ICU readmission of patients with cancer: incidence, risk factors and mortality. *J Crit Care*. 2019;51:84–7.
- Danai PA, Moss M, Mannino DM, Martin GS. The epidemiology of sepsis in patients with malignancy. *Chest*. 2006;129(6):1432–40.
- Nates JL, Pène F, Darmon M, et al. Septic shock in the immunocompromised cancer patient: a narrative review. *Crit Care*. 2024;28(1):285.
- Venet F, Monneret G. Advances in the Understanding and treatment of sepsis-induced immunosuppression. *Nat Rev Nephrol*. 2018;14(2):121–37.
- Shu XP, Xiang YC, Liu F, Cheng Y, Zhang W, Peng D. Effect of serum lactate dehydrogenase-to-albumin ratio (LAR) on the short-term outcomes and long-term prognosis of colorectal cancer after radical surgery. *BMC Cancer*. 2023;23(1):915.
- Luo X, Liu D, Li C, et al. The predictive value of the serum creatinine-to-albumin ratio (sCAR) and lactate dehydrogenase-to-albumin ratio (LAR) in sepsis-related persistent severe acute kidney injury. *Eur J Med Res*. 2025;30(1):25.
- Liang M, Ren X, Huang D, Ruan Z, Chen X, Qiu Z. The association between lactate dehydrogenase to serum albumin ratio and the 28-day mortality in patients with sepsis-associated acute kidney injury in intensive care: a retrospective cohort study. *Ren Fail*. 2023;45(1):2212080.
- Jiménez-Zarazúa O, Vélez-Ramírez LN, Mondragón JD. Biomarkers and sepsis severity as predictors of mechanical ventilation and mortality in COVID-19. *Heliyon*. 2024;10(7):e28521.
- Hu J, Zhou Y. The association between lactate dehydrogenase to serum albumin ratio and in-hospital mortality in patients with pulmonary embolism: a retrospective analysis of the MIMIC-IV database. *Front Cardiovasc Med*. 2024;11 null:1398614.
- Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1.
- Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform*. 2019;20(2):492–503.
- Xiao X, Wu J, Liu Y, Suo Z, Zhang H, Xu H. Increased lactate dehydrogenase to albumin ratio is associated with short-term mortality in septic ICU patients: A retrospective cohort study. *Medicine*. 2024;103(52):e40854.
- Guan X, Zhong L, Zhang J, et al. The relationship between lactate dehydrogenase to albumin ratio and all-cause mortality during ICU stays in patients with sepsis: A retrospective cohort study with propensity score matching. *Heliyon*. 2024;10(6):e27560.
- Zhang F, Zhang M, Niu Z, Sun L, Kang X, Qu Y. Prognostic value of lactic dehydrogenase-to-albumin ratio in critically ill patients with acute respiratory distress syndrome: a retrospective cohort study. *J Thorac Dis*. 2024;16(1):81–90.
- Ye L, Lu J, Yuan M, Min J, Zhong L, Xu J. Correlation between lactate dehydrogenase to albumin ratio and the prognosis of patients with cardiac arrest. *Rev Cardiovasc Med*. 2024;25(2):65.
- Sharma D, Singh M, Rani R. Role of LDH in tumor glycolysis: regulation of LDHA by small molecules for cancer therapeutics. *Semin cancer Biol*. 2022;87(null):184–95.
- Schwartz GG, Tretli S, Klug MG, Robsahm TE. Women who develop ovarian cancer show an increase in serum calcium and a decrease in serum albumin. A longitudinal study in the Janus serum bank cohort. *Gynecol Oncol*. 2020;159(1):264–9.
- Königsbrügge O, Posch F, Riedl J, et al. Association between decreased serum albumin with risk of venous thromboembolism and mortality in cancer patients. *Oncologist*. 2016;21(2):252–7.
- Comandatore A, Franczak M, Smolenski RT, Morelli L, Peters GJ, Giovannetti E. Lactate dehydrogenase and its clinical significance in pancreatic and thoracic cancers. *Semin cancer Biol*. 2022;86(Pt 2):93–100.
- Lu J, Wei Z, Jiang H, et al. Lactate dehydrogenase is associated with 28-day mortality in patients with sepsis: a retrospective observational study. *J Surg Res*. 2018;228:314–21.
- Lee SM, Jo YH, Lee JH. ASSOCIATIONS OF THE SERUM ALBUMIN CONCENTRATION AND SEQUENTIAL ORGAN FAILURE ASSESSMENT SCORE AT DISCHARGE WITH, et al. 1-YEAR MORTALITY IN SEPSIS SURVIVORS: A RETROSPECTIVE COHORT STUDY. *Shock*. 2023;59(4):547–52.
- Yin M, Si L, Qin W, et al. Predictive value of serum albumin level for the prognosis of severe sepsis without exogenous human albumin administration: A prospective cohort study. *J Intensive Care Med*. 2018;33(12):687–94.
- Qin Y, Kernan KF, Fan Z, et al. Machine learning derivation of four computable 24-h pediatric sepsis phenotypes to facilitate enrollment in early personalized anti-inflammatory clinical trials. *Crit Care*. 2022;26(1):128.
- Takkavatakarn K, Oh W, Chan L, et al. Machine learning derived serum creatinine trajectories in acute kidney injury in critically ill patients with sepsis. *Crit Care*. 2024;28(1):156.
- Gao J, Lu Y, Ashrafi N, Domingo I, Alaei K, Pishgar M. Prediction of sepsis mortality in ICU patients using machine learning methods. *BMC Med Inf Decis Mak*. 2024;24(1):228.
- Tjokrowidjaja A, Lord SJ, John T, et al. Pre- and on-treatment lactate dehydrogenase as a prognostic and predictive biomarker in advanced non-small cell lung cancer. *Cancer-am cancer Soc*. 2022;128(8):1574–83.
- Mirzai S, Sarnaik KS, Persits I, et al. Combined prognostic impact of low muscle mass and hypoalbuminemia in patients hospitalized for heart failure: A retrospective cohort study. *J Am Heart Assoc*. 2024;13(3):e030991.
- Huang W, Li C, Wang Z, et al. Decreased serum albumin level indicates poor prognosis of COVID-19 patients: hepatic injury analysis from 2,623 hospitalized cases. *Sci China Life Sci*. 2020;63(11):1678–87.
- Cousillas A, Gallardo E, Carou I, et al. Role of nutritional and inflammatory status in the prognosis of advanced gastric cancer. *J Clin Oncol*. 2017;35(suppl 4):35–35.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.