frontiers | Research Topics

Clinical application of artificial intelligence in emergency and critical care medicine, volume V

Edited by
Rahul Kashyap, Zhongheng Zhang, Ping Zhang and Qinghe Meng

Coordinated by
Ankit Virmani

Published in
Frontiers in Medicine

ISSN 1664-8714
ISBN 978-2-8325-6309-0
DOI 10.3389/978-2-8325-6309-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Clinical application of artificial intelligence in emergency and critical care medicine, volume V

Topic editors

Rahul Kashyap — WellSpan Health, United States

Zhongheng Zhang — Department of Emergency Medicine, Sir Run Run Shaw Hospital, China

Ping Zhang — The Ohio State University, United States

Qinghe Meng — Upstate Medical University, United States

Topic coordinator

Ankit Virmani — Google, United States

Citation

Kashyap, R., Zhang, Z., Zhang, P., Meng, Q., Virmani, A., eds. (2025). *Clinical application of artificial intelligence in emergency and critical care medicine, volume V*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6309-0

Table of contents

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 11 April 2025
DOI 10.3389/fmed.2025.1593416

OPEN ACCESS

EDITED AND REVIEWED BY
Ata Murat Kaynar,
University of Pittsburgh, United States

*CORRESPONDENCE
Zhongheng Zhang
✉ zh_zhang1984@zju.edu.cn

RECEIVED 14 March 2025
ACCEPTED 31 March 2025
PUBLISHED 11 April 2025

CITATION
Wang Y, Kashyap R, Zhang P, Meng Q and Zhang Z (2025) Editorial: Clinical application of artificial intelligence in emergency and critical care medicine, volume V.
Front. Med. 12:1593416.
doi: 10.3389/fmed.2025.1593416

Editorial: Clinical application of artificial intelligence in emergency and critical care medicine, volume V

Yanni Wang[1,2], Rahul Kashyap[3,4], Ping Zhang[5,6], Qinghe Meng[7,8] and Zhongheng Zhang[2,9,10,11]*

[1]Department of Emergency Medicine, The Second Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, [2]Department of Emergency Medicine, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3]Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN, United States, [4]Department of Research, WellSpan Health, York, PA, United States, [5]Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, United States, [6]Department of Biomedical Informatics, The Ohio State University, Columbus, OH, United States, [7]Department of Surgery, State University of New York Upstate Medical University, Syracuse, NY, United States, [8]Sepsis Interdisciplinary Research Center, State University of New York Upstate Medical University, Syracuse, NY, United States, [9]School of Medicine, Shaoxing University, Shaoxing, Zhejiang, China, [10]Key Laboratory of Precision Medicine in Diagnosis and Monitoring Research of Zhejiang Province, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [11]Longquan Industrial Innovation Research Institute, Lishui, China

KEYWORDS

artificial intelligence, prediction, machine learning, critical care, treatment decision-making

Editorial on the Research Topic
Clinical application of artificial intelligence in emergency and critical care medicine, volume V

This Research Topic explores the clinical applications of artificial intelligence (AI) in emergency and critical care medicine through 13 research articles. These studies systematically investigate how advanced data mining and AI techniques enhance risk assessment, diagnostic assistance, prognosis prediction, and treatment decision-making. Leveraging large-scale data resources and machine learning (ML) algorithms, these studies provide detailed analyses of complex conditions such as sepsis, acute pancreatitis, gastroparesis, acute kidney injury, disseminated intravascular coagulation (DIC), and heart failure with sepsis. From a data-driven perspective, this research offers a robust theoretical foundation and practical support for precision medicine.

AI is expected to significantly improve the prognosis of critically ill patients by assisting in disease identification, predicting disease progression, and supporting clinical decision-making (1). One key area of focus is the early prediction of sepsis, Yadgarov et al. conducted a systematic review of studies published from database inception to October 2023, searching Medline, PubMed, Google Scholar, and CENTRAL. Out of 3,953 studies, they analyzed 73 articles encompassing 457,932 sepsis patients and 256 models. Their findings demonstrated that ML models significantly outperformed traditional scoring systems in early sepsis prediction, with neural networks and decision tree models achieving the highest performance. Kim et al. developed a Transformer-based deep learning model for predicting ICU length of stay in sepsis patients, achieving a mean absolute error of only 2.05 days, demonstrating high accuracy and reliability. Beyond sepsis detection, Zhang et al. constructed a 28-day mortality prediction model specifically

for sepsis patients with concurrent heart failure. Using the eICU-CRD database for model development and validating it externally on the MIMIC-IV database, they found that a logistic regression-based model achieved an AUC of 0.746 on the validation set, outperforming more complex algorithms such as XGBoost. The final model identified 10 key predictive features and employed the SHAP method to enhance interpretability, aiding clinicians in early identification of high-risk patients and optimizing resource allocation.

Traditional early warning systems primarily rely on vital signs and basic laboratory tests, whereas AI models integrating multi-source data allow for the incorporation of molecular and biological information into clinical decision-making. Hu et al. explored the combination of metagenomics, radiomics, and ML for sepsis diagnosis. By performing metagenomic sequencing on blood samples from sepsis patients and extracting radiomic features, they developed a fusion model that achieved an AUC close to 0.88 in its best-performing version. The integration of multimodal data helps overcome the limitations of single-indicator approaches, providing a novel strategy for the early and precise diagnosis of sepsis.

Furthermore, interpretable AI-based risk prediction models offer refined and individualized decision support in critical care medicine (2). Zhai et al. developed an interpretable XGBoost model for predicting in-hospital mortality risk in patients with severe pulmonary infections, achieving an AUC of 0.956. More importantly, the model incorporated SHAP and LIME methodologies to enhance interpretability. Such personalized risk assessment tools assist in mortality risk stratification and provide valuable clinical decision support. However, for AI models to truly assist in ICU decision-making, they must evolve from purely "predictive AI" to "actionable AI" capable of comparing the effects of different interventions. This transition necessitates the integration of causal inference to guide optimal treatment selection (3).

Beyond sepsis, AI technologies also hold significant clinical potential in other critical conditions. Tan et al. reviewed advancements in ML applications for predicting disease severity and complications in acute pancreatitis. Liu et al. utilized explainable ML to construct a postoperative gastroparesis risk prediction model, providing empirical evidence for early intervention. In another study, Wei et al. developed a nomogram for predicting early acute kidney injury (AKI) in patients with acute non-variceal upper gastrointestinal bleeding (NVUGIB) using data from the MIMIC-IV database, subsequently implementing it as a web-based clinical calculator.

AI also demonstrates immense potential in personalized treatment strategies. Traditional clinical guidelines often struggle to account for the unique physiological characteristics of individual patients, whereas AI-driven approaches offer the ability to support clinicians in devising more precise and tailored treatment plans (4). Nevertheless, while AI has demonstrated outstanding performance, it is crucial to recognize its limitations and reinforce validation and regulatory oversight in real-world applications to ensure tangible patient benefits (5).

Overall, *Clinical application of artificial intelligence in emergency and critical care medicine, Volume V* presents a comprehensive workflow from data acquisition and feature extraction to model construction and result interpretation. With advancements in multimodal data integration, causal inference modeling, and improvements in AI interpretability, AI is poised to become a powerful tool in sepsis management. We anticipate the deep integration of AI technologies into clinical medicine, fostering more precise and efficient workflows from early disease identification to treatment decision-making, ultimately contributing to lower mortality rates worldwide.

Author contributions

YW: Writing – original draft. RK: Writing – review & editing. PZ: Writing – review & editing. QM: Writing – review & editing. ZZ: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Zhongheng Zhang received funding from the China National Key Research and Development Program (2022YFC2504500), the Huadong Medicine Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (LHDMD24H150001), the National Natural Science Foundation of China (82272180, 82472243), the China National Key Research and Development Program (2023YFC3603104), a collaborative scientific project co-established by the Science and Technology Department of the National Administration of Traditional Chinese Medicine and the Zhejiang Provincial Administration of Traditional Chinese Medicine (GZY-ZJ-KJ-24082), the General Health Science and Technology Program of Zhejiang Province (2024KY1099), and Project of Zhejiang University Longquan Innovation Center (ZJDXLQCXZCJBGS2024016).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Saqib M, Iftikhar M, Neha F, Karishma F, Mumtaz H. Artificial intelligence in critical illness and its impact on patient care: a comprehensive review. *Front Med.* (2023) 10:1176192. doi: 10.3389/fmed.2023.1176192

2. Yang M, Chen H, Hu W, Mischi M, Shan C, Li J et al. Development and validation of an interpretable conformal predictor to predict sepsis mortality risk: retrospective cohort study. *J Med Internet Res*. (2024) 26:e50369. doi: 10.2196/50369

3. Smit JM, Krijthe JH, van Bommel J. The future of artificial intelligence in intensive care: moving from predictive to actionable AI. *Intensive Care Med.* (2023) 49:1114–6. doi: 10.1007/s00134-023-07102-y

4. Li F, Wang S, Gao Z, Qing M, Pan S, Liu Y, et al. Harnessing artificial intelligence in sepsis care: advances in early detection, personalized treatment, and real-time monitoring. *Front Med.* (2025) 11:1510792. doi: 10.3389/fmed.2024.1510792

5. Schinkel M, van der Poll T, Wiersinga WJ. Artificial intelligence for early sepsis detection: a word of caution. *Am J Respir Crit Care Med*. 207:853–4. doi: 10.1164/rccm.202212-2284VP

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 June 2024
DOI 10.3389/fmed.2024.1416396

OPEN ACCESS

EDITED BY
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

REVIEWED BY
Yanfei Shen,
Zhejiang Hospital, China
Guo-wei Tu,
Fudan University, China

*CORRESPONDENCE
Dawei Liu
✉ dwliu98@163.com
Hua Zhao
✉ 13811633584@163.com

RECEIVED 12 April 2024
ACCEPTED 27 May 2024
PUBLISHED 06 June 2024

CITATION
Gao B, Chen R, Zhao H, Zhang H, Wang X and Liu D (2024) Estimated plasma volume status as a simple and accessible predictor of 28-day mortality in septic shock: insights from a retrospective study of the MIMIC-IV database.
Front. Med. 11:1416396.
doi: 10.3389/fmed.2024.1416396

Estimated plasma volume status as a simple and accessible predictor of 28-day mortality in septic shock: insights from a retrospective study of the MIMIC-IV database

Beijun Gao[1], Rongping Chen[1], Hua Zhao[1]*, Hongmin Zhang[2], Xiaoting Wang[1] and Dawei Liu[1]*

[1]Department of Critical Care Medicine, Peking Union Medical College, Chinese Academy of Medical Sciences, Peking Union Medical College Hospital, Beijing, China, [2]Department of Health Care, Peking Union Medical College, Chinese Academy of Medical Sciences, Peking Union Medical College Hospital, Beijing, China

Background: Assessing volume status in septic shock patients is crucial for tailored fluid resuscitation. Estimated plasma volume status (ePVS) has emerged as a simple and effective tool for evaluating patient volume status. However, the prognostic value of ePVS in septic shock patients remains underexplored.

Methods: The study cohort consisted of septic shock patients admitted to the ICU, sourced from the MIMIC-IV database. Patients were categorized into two groups based on 28-day survival outcomes, and their baseline characteristics were compared. According to the ePVS (6.52 dL/g) with a hazard ratio of 1 in the restricted cubic spline (RCS) analysis, patients were further divided into high and low ePVS groups. A multivariable Cox regression model was utilized to evaluate the association between ePVS and 28-day mortality rate. The Kaplan–Meier survival curve was plotted, and all-cause mortality was compared between the high and low groups using the log-rank test.

Results: A total of 7,607 septic shock patients were included in the study, among whom 2,144 (28.2%) died within 28 days. A J-shaped relationship was observed between ePVS at ICU admission and 28-day mortality, with an increase in mortality risk noted when ePVS exceeded 6.52 dL/g. The high ePVS group exhibited notably higher mortality rates compared to the low ePVS group (28-day mortality: 26.2% vs. 30.2%; 90-day mortality: 35% vs. 42.3%). After adjustment for confounding factors, ePVS greater than 6.52 dL/g independently correlated with an increased risk of 28-day mortality (HR: 1.20, 95% CI: 1.10–1.31, $p < 0.001$) and 90-day mortality (HR: 1.25, 95% CI: 1.15–1.35, $p < 0.001$). Kaplan–Meier curves demonstrated a heightened risk of mortality associated with ePVS values exceeding 6.52 dL/g.

Conclusion: A J-shaped association was observed between ePVS and 28-day mortality in septic shock patients, with higher ePVS levels associated with increased risk of mortality.

KEYWORDS

estimated plasma volume status, septic shock, fluid therapy, 28-day mortality, MIMIC-IV database

1 Introduction

Septic shock is a form of circulatory failure characterized by a combination of mechanisms including hypovolemia, vascular tone depression, cardiac dysfunction, and disturbances in microcirculation (1). Fluid infusion is the most used treatment method in clinical practice (2). The purpose of fluid infusion is to increase cardiac output and improve tissue perfusion. However, in practice, fluid infusion can yield different outcomes (3, 4). Sepsis reduces vascular tone through various mechanisms, resulting not only in arterial hypotension due to vasodilation but also in venous dilation, altered blood flow distribution, and microcirculatory dysfunction (5). However, reduced venous tone increases unstressed volume, leading to venous return impairment, causing an increase in total body volume but failing to enhance venous return (6). Therefore, the effect of intravenous fluid infusion is difficult to maintain and can cause subsequent damage, leading to poor prognosis (7). We hypothesize that the total vascular volume in septic patients can reflect the extent of venous dilation in sepsis.

The traditional Strauss et al. formula, developed in 1951, utilizes an equation based on hematocrit and hemoglobin to provide estimations of plasma volume status (ePVS) (8). In 2015, Duarte et al. introduced a single time-point 'instantaneous'-derived measurement of plasma volume for estimating PV derived from the Strauss formula (9). They found that, in cases of myocardial infarction complicated by heart failure (HF), ePVS, as an indicator of total vascular volume, holds significant prognostic value for early cardiovascular events beyond routine clinical evaluations.

ePVS offers a straightforward method to estimate plasma volume. As a surrogate marker for total vascular volume, it has been validated for its reliability, with multiple studies demonstrating its independent association with outcomes in various heart failure phenotypes (10–12). Moreover, ePVS has shown prognostic relevance in patients with acute respiratory distress syndrome and fever (13, 14). We hypothesize that in patients with septic shock, the total vascular volume will increase following fluid resuscitation due to the systemic vasodilation caused by the inflammatory response, and this increase in total vascular volume is related to the prognosis.

Despite the clinical utility and simplicity of ePVS, alongside its cost-effectiveness and efficiency, its adoption in clinical practice remains limited. This study aims to explore the impact of ePVS on the mortality of patients with septic shock, thereby contributing to the optimization of septic shock management.

2 Methods

2.1 The database

The Medical Information Mart for Intensive Care IV (MIMIC-IV version 2.2) database was utilized to gather the data for this investigation (15). The MIMIC-IV database collected clinical data on patients who visited Beth Israel Deaconess Medical Center (BIDMC) between 2008 and 2019. Access to the database is available for download upon completion of an authorized course on their official website. The author, Beijun Gao, has completed the accredited course, had database access, and oversaw data extraction (Record ID: 12338471). All patient information is hidden to protect their privacy.

2.2 Cohort information

2.2.1 Selection of participants

Selection of patients diagnosed with septic shock in version 2.2 of the (MIMIC)-IV database. Septic shock is defined as patients who received appropriate fluid resuscitation but still require vasopressors to maintain mean arterial pressure (MAP) >65 mmHg, and serum lactate levels above 2.0 mmol/L (16), ICD9 and ICD10 codes are used to identify patients with septic shock. Inclusion criteria are as follows: first admission to the intensive care unit (ICU), age over 18 years, and ICU stay of at least 1 day. Exclusion criteria are: (i) multiple admissions; (ii) age < 18 years; (iii) during pregnancy and postpartum period; (iv) hospital stay <24 h; (v) Lack of data on hemoglobin levels and hematocrit values, or substantial baseline data absence.

2.2.2 Variable extraction

We selected the first data point upon the target patient's admission to the ICU. Baseline characteristics of patients include age, gender, weight, history of diabetes, history of hypertension, and history of malignant tumors. Vital signs data extracted from ICU patients include heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), and respiratory rate (RR). Blood gas analysis indices include potassium (K^+), sodium (Na^+), anion gap, and lactate. Laboratory parameters include white blood cell count (WBC), platelet count (PLT), hematocrit (HCT), hemoglobin (HB), potassium (K^+), sodium (Na^+), anion gap, lactate, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, prothrombin time (PT), international normalized ratio (INR), blood urea nitrogen (BUN) and creatinine (Cr). Intervention measures include ventilation, continuous renal replacement therapy (CRRT), and vasopressors. Vasopressor use is defined as the administration of norepinephrine, epinephrine, dopamine, dobutamine, or vasopressin during the first day of ICU hospitalization. Additionally, Charlson comorbidity index, Sequential Organ Failure Assessment (SOFA) score, Acute Physiology And Chronic Health Evaluation (APACHE) score, and worst renal function stage during hospitalization based on the Kidney Disease: Improving Global Outcomes (KDIGO) AKI Guideline Work Group criteria were calculated for each patient (17).

2.3 Evaluation of ePVS

The Duarte formula incorporating hematocrit and hemoglobin was utilized as follows (7):

$$\mathrm{ePVS}(\mathrm{dL/g}) = (100 - \mathrm{hematocrit}(\%)) / \mathrm{hemoglobin}(\mathrm{g/dL})$$

2.4 Grouping and study endpoints

Based on the 28-day follow-up outcomes, enrolled patients were categorized into the survival group ($n = 5{,}463$) and the death group ($n = 2{,}144$). Additionally, patients were further stratified into a high ePVS group ($n = 3{,}789$) and a low ePVS group ($n = 3{,}818$) based on the ePVS value (6.52 dL/g), which was determined through restricted

cubic spline (RCS) analysis, as described later in this study. The primary outcome assessed in this study was 28-day mortality. The secondary outcomes included 90-day mortality, duration of ICU stay, and occurrence of acute kidney injury.

2.5 Statistical analysis

Data management procedures were implemented to address missing data issues. Cases with severely missing data, exceeding 20% of the dataset, were excluded from the analysis (Supplementary Table S1 for details on missing data). Acceptable missing data were imputed using the multiple imputation method with random forests, implemented in the R software package (18). Continuous variables underwent an initial assessment for normal distribution. Those adhering to a normal distribution were summarized as mean (standard deviation) and analyzed using the *t*-test method. Alternatively, continuous variables not conforming to a normal distribution were presented as median (interquartile range) and analyzed using nonparametric methods (Mann–Whitney *U* test). Categorical data were summarized as frequencies and percentages and analyzed using the chi-square method.

To investigate the relationship between ePVS and 28-day all-cause mortality risk in patients with septic shock, RCS analysis was performed. Subsequently, a cut-off value of 6.52 dL/g for ePVS was determined based on the RCS analysis, stratifying patients into high and low ePVS groups. Univariate and multivariate Cox regression analyses were conducted to assess the independent association between increased ePVS and higher 28-day and 90-day all-cause mortality in patients with septic shock. Results were expressed as hazard ratios (HR) with 95% confidence intervals (CI). Model I analysis involved no adjustments for covariates. In Model II, adjustments were made for Age, Gender, and Weight. Model III further adjusted for SOFA score, Charlson Comorbidity Index, APACHE II score, SAPS II score, HR, SBP, DBP, MAP, RR, lactate, K^+, Na^+, anion gap, WBC, PLT, ALT, AST, total bilirubin, PT, INR, BUN, Cr, mechanical ventilation, CRRT, vasopressor use, metastatic cancer, DM, and HT. Kaplan–Meier curves were generated to visualize the survival probability between high and low ePVS groups, with comparison done using the log-rank test. Stratified analyses were conducted based on the variables Gender, Age (>65 vs. <=65), Ventilation, CRRT, and Vasopressor use. Calibration curve was also included to better substantiate the predictive value of ePVS. Data analysis was performed using R programming language version 4.2.0, with statistical significance set at a two-tailed *p*-value of <0.05.

3 Results

3.1 Demographics and baseline characteristics

The study included 7,607 patients, with 5,463 (71.8%) surviving and 2,144 (28.2%) deceased (Figure 1). Deceased patients had a higher mean age (70.62 vs. 66.24 years) and lower body weight (79.60 vs. 82.63 kg) than survivors. Comorbidities like diabetes mellitus (DM: 3.7% vs. 2.8%) and hypertension (HT: 79.9% vs. 77.6%) were more prevalent in the deceased. Severity scores, including SOFA (9.60 vs. 6.99), Charlson (6.80 vs. 5.66), APACHE (73.59 vs. 55.57), and SAPS II (53.06 vs. 41.44), were higher in the deceased. Vital signs such as heart rate, systolic blood pressure, and respiratory rate were slightly elevated in the deceased. Laboratory findings showed deviations in lactate, potassium, WBC count, PLT count, liver function (ALT, AST, total bilirubin), and renal function (BUN, Cr) in the deceased, with significant differences compared to survivors. Mechanical ventilation, CRRT, and vasopressor usage were more frequent in the deceased cohort (Table 1).

3.2 Restricted cubic spline (RCS) analysis between ePVS level and 28-day mortality

The relationship between admission ePVS level and 28-day mortality demonstrates a nonlinear pattern. Figure 2A presents the restricted cubic spline (RCS) curve illustrating this relationship in patients with septic shock, without adjusting for potential confounders. Upon adjustment for potential confounders, as shown in Figure 2B, both figures indicate that the cutoff value of ePVS was identified as 6.52 dL/g, corresponding to a hazard ratio of 1. Significant escalation in mortality risk was observed when ePVS exceeded 6.52 dL/g.

3.3 Outcomes by ePVS level in patients with septic shock

Table 2 presents outcomes based on ePVS levels in patients with septic shock. The 28-day and 90-day all-cause mortality rates were 28.2 and 38.6%, respectively. Notably, the 28-day mortality rate in the high ePVS group (30.2%) was significantly elevated compared to the low ePVS group (26.2%, $p < 0.001$). Similarly, the 90-day mortality rate in the high ePVS group (42.3%) was also higher than that in the low ePVS group (35%, $p < 0.001$). There was no significant difference in length of stay in the ICU between the two groups. Additionally, 39.5% of patients experienced stage 3 acute kidney injury (AKI), with a higher proportion observed in the high ePVS group (40.9%) compared to the low ePVS group (38.1%).

3.4 Survival analysis

The Kaplan–Meier curve showed that the 28-day and 90-day cumulative survival rates were lower in the high ePVS group than that in the low ePVS group (log-rank test, $\chi^2 = 12.6$, $p < 0.001$; $\chi^2 = 35.9$, $p < 0.001$) (Figures 3A,B).

3.5 Correlation between ePVS and all-cause mortality

The Cox regression models presented in Table 3 demonstrate the relationship between ePVS levels and the risk of 28-day and 90-day mortality. In the unadjusted model (Model I), ePVS as a continuous variable shows a significant positive correlation with 28-day and 90-day all-cause mortality (HR 1.05, 95% CI 1.03–1.08, $p < 0.001$ and HR 1.08, 95% CI 1.05–1.10, $p < 0.001$). When ePVS is categorized, higher levels (>6.52) are associated with increased 28-day and 90-day all-cause

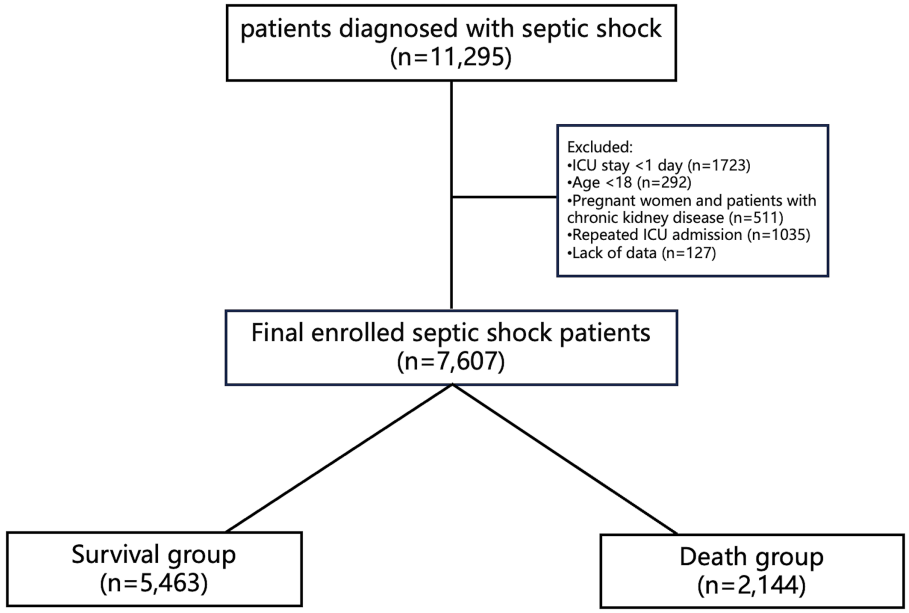

FIGURE 1
Flow chart of the study.

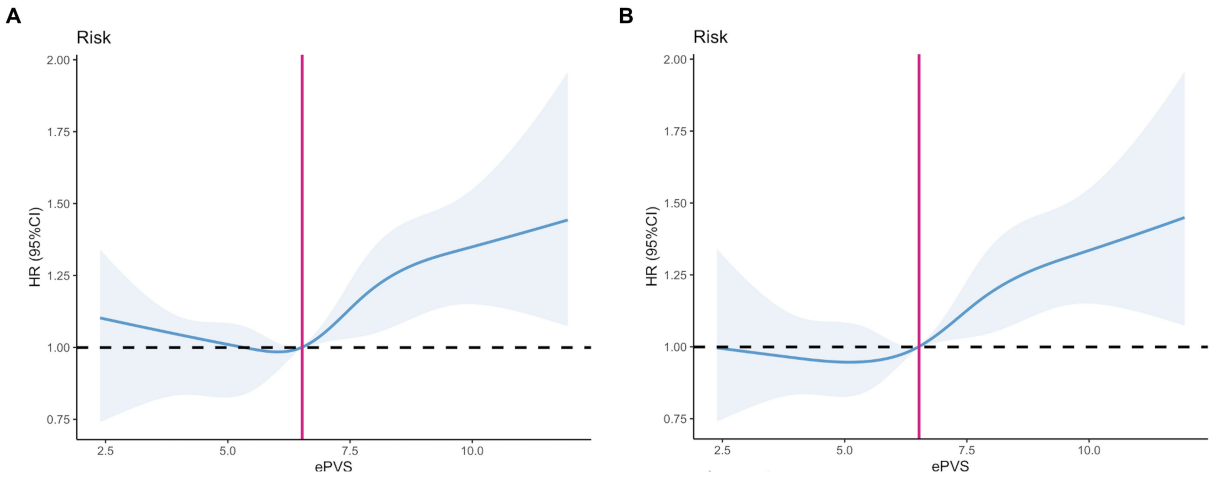

FIGURE 2
Restricted cubic spline (RCS). **(A)** Unadjusted model. **(B)** Adjusted model: Adjusted for Age, Gender, Weight, SOFA score, Charlson Comorbidity Index, APACHE II score, SAPS II score, Heart Rate (HR), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Mean Arterial Pressure (MAP), Respiratory Rate (RR), Lactate, Potassium, Sodium, Anion Gap, White Blood Cell (WBC) count, Platelet (PLT) count, Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Total Bilirubin, Prothrombin Time (PT), International Normalized Ratio (INR), Blood Urea Nitrogen (BUN), Creatinine (Cr), Mechanical Ventilation, Continuous Renal Replacement Therapy (CRRT), Vasopressor Use, Metastatic Cancer, Diabetes Mellitus (DM), and Hypertension (HT).

mortality (HR 1.17, 95% CI 1.07–1.27, $p < 0.001$ and HR 1.25, 95% CI 1.16–1.34, $p < 0.001$). In Model II, ePVS as a continuous variable remains positively correlated with 28-day and 90-day mortality. After categorization, high ePVS (>6.52) is still linked to increased 28-day mortality (HR 1.18, 95% CI 1.08–1.29, $p < 0.001$) and 90-day mortality (HR 1.26, 95% CI 1.17–1.36, $p < 0.001$). Model III results indicate that higher ePVS is an independent risk factor for adverse outcomes in patients with septic shock, with significantly higher 28-day (HR 1.20, 95% CI 1.10–1.31, $p < 0.001$) and 90-day (HR 1.25, 95% CI 1.15–1.35, $p < 0.001$) all-cause mortality rates in the high ePVS group. The calibration curve (Figure 4) demonstrates that the ePVS model reliably predicts 28-day mortality in septic shock patients, with predicted probabilities closely matching observed outcomes. This supports the potential utility of ePVS as a prognostic tool in clinical settings.

3.6 Subgroup analyses

Subgroup analyses revealed significant associations within specific strata (Figure 5). Female gender exhibited a heightened risk for the

TABLE 1 The characteristic of included subjects between different groups.

Variable	Level	Overall $n = 7,607$	Survival $n = 5,463$	Death $n = 2,144$	*p*-value
General information					
Age [mean (SD)]		67.47 (15.08)	66.24 (15.11)	70.62 (14.55)	<0.001
Gender (%)	Female	3,329 (43.8)	2,376 (43.5)	953 (44.4)	0.465
	Male	4,278 (56.2)	3,087 (56.5)	1,191 (55.6)	
Weight [mean (SD)]		81.78 (26.11)	82.63 (26.91)	79.60 (23.84)	<0.001
Comorbidities					
Metastatic cancer (%)	No	7,600 (99.9)	5,459 (99.9)	2,141 (99.9)	0.658
	Yes	7 (0.1)	4 (0.1)	3 (0.1)	
DM (%)	No	7,345 (96.6)	5,260 (96.3)	2,085 (97.2)	0.045
	Yes	262 (3.4)	203 (3.7)	59 (2.8)	
HT (%)	No	5,956 (78.3)	4,242 (77.6)	1,714 (79.9)	0.031
Score system					
SOFA [mean (SD)]		7.73 (3.79)	6.99 (3.41)	9.60 (4.04)	<0.001
Charlson [mean (SD)]		5.98 (2.89)	5.66 (2.83)	6.80 (2.87)	<0.001
APACHE II [mean (SD)]		60.65 (22.49)	55.57 (19.45)	73.59 (24.44)	<0.001
SAPSII [mean (SD)]		44.72 (14.73)	41.44 (13.20)	53.06 (15.11)	<0.001
Vital signs					
HR [mean (SD)]		111.40 (22.44)	110.14 (22.11)	114.61 (22.97)	<0.001
SBP [mean (SD)]		141.26 (22.75)	141.93 (22.51)	139.55 (23.28)	<0.001
DBP [mean (SD)]		86.92 (21.07)	87.14 (20.67)	86.38 (22.04)	0.159
MAP [mean (SD)]		103.52 (28.64)	103.40 (27.40)	103.83 (31.58)	0.557
RR [mean (SD)]		30.14 (6.96)	29.80 (6.93)	30.99 (6.98)	<0.001
Laboratory results					
ePVS [mean (SD)]		6.64 (1.80)	6.58 (1.77)	6.79 (1.85)	<0.001
Lac [mean (SD)]		3.21 (2.77)	2.78 (2.14)	4.31 (3.74)	<0.001
Potassium [mean (SD)]		4.73 (0.96)	4.68 (0.96)	4.83 (0.95)	<0.001
Sodium [mean (SD)]		139.71 (5.89)	139.68 (5.57)	139.81 (6.65)	0.392
Anion gap [mean (SD)]		18.41 (5.77)	17.77 (5.34)	20.06 (6.45)	<0.001
WBC [mean (SD)]		17.13 (12.56)	16.54 (10.68)	18.65 (16.30)	<0.001
PLT [mean (SD)]		230.71 (143.75)	236.69 (144.20)	215.46 (141.49)	<0.001
Hemoglobin [mean (SD)]		10.63 (2.07)	10.69 (2.06)	10.45 (2.07)	<0.001
ALT (median [IQR])		43.00 [20.00, 99.00]	43.00 [20.00, 93.97]	43.00 [20.00, 112.35]	0.042
AST (median [IQR])		71.00 [31.00, 152.00]	67.00 [29.00, 139.16]	82.00 [35.00, 194.25]	<0.001
Total bilirubin (median [IQR])		1.20 [0.50, 2.68]	1.20 [0.50, 2.50]	1.40 [0.60, 3.50]	<0.001
PT [mean (SD)]		21.51 (15.45)	20.26 (13.86)	24.72 (18.52)	<0.001
INR [mean (SD)]		1.99 (1.46)	1.87 (1.32)	2.29 (1.75)	<0.001
Bun [mean (SD)]		40.36 (28.36)	37.60 (26.84)	47.38 (30.83)	<0.001
Cr [mean (SD)]		2.21 (1.98)	2.13 (2.04)	2.41 (1.80)	<0.001
Treatment					
Ventilation (%)	No	4,472 (58.8)	3,403 (62.3)	1,069 (49.9)	<0.001
	Yes	3,135 (41.2)	2,060 (37.7)	1,075 (50.1)	
CRRT (%)	No	6,575 (86.4)	4,930 (90.2)	1,645 (76.7)	<0.001
	Yes	1,032 (13.6)	533 (9.8)	499 (23.3)	
Vasopressor (%)	No	5,668 (74.5)	4,532 (83.0)	1,136 (53.0)	<0.001

(Continued)

TABLE 1 (Continued)

Variable	Level	Overall *n* = 7,607	Survival *n* = 5,463	Death *n* = 2,144	*p*-value
	Yes	1,939 (25.5)	931 (17.0)	1,008 (47.0)	
Outcomes					
AKI stage (%)	0	1,475 (19.4)	1,331 (24.4)	144 (6.7)	<0.001
	1	949 (12.5)	779 (14.3)	170 (7.9)	
	2	2,177 (28.6)	1,728 (31.6)	449 (20.9)	
	3	3,006 (39.5)	1,625 (29.7)	1,381 (64.4)	
Losicu [mean (SD)]		6.89 (8.08)	7.00 (8.89)	6.61 (5.47)	0.059

Results are expressed as mean (SD), median [IQR] or *n* (%). DM, Diabetes Mellitus; HT, Hypertension; SOFA, Sequential Organ Failure Assessment score; Charlson, comorbidity index; APACHE II, Acute Physiology And Chronic Health Evaluation II; SASPII, Severe Acute Pancreatitis Score II; HR, Heart Rate; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; MAP, Mean Arterial Pressure; RR, Respiratory Rate; ePVS, Estimated plasma volume status; Lac, Lactate; WBC, White Blood Cell count; PLT, Platelet count; ALT, Alanine Aminotransferase; AST, Aspartate Aminotransferase; PT, Prothrombin Time; INR, International Normalized Ratio; BUN, Blood Urea Nitrogen; Cr, Creatinine; CRRT, Continuous Renal Replacement Therapy; AKI, Acute Kidney Injury; Losicu, Length of ICU Stay.

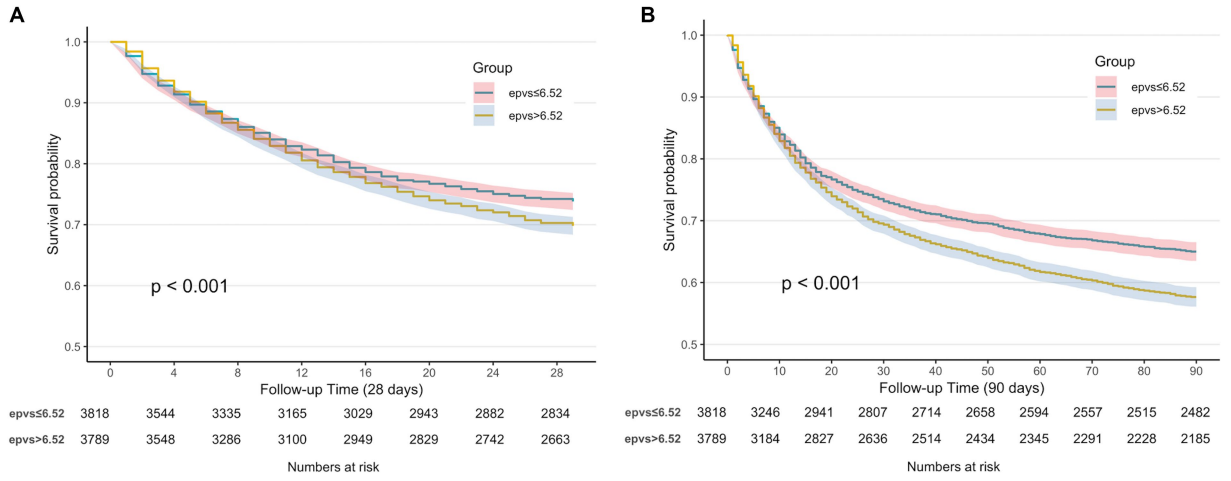

FIGURE 3
(A) Kaplan–Meier survival curve of 28-day cumulative survival rate for low and high ePVS groups. **(B)** Kaplan–Meier survival curve of 90-day cumulative survival rate for low and high ePVS groups.

TABLE 2 The comparison of outcomes between the low ePVS group and high ePVS group.

Outcome	Level	Overall	Low ePVS ≤6.52 dL/g	High ePVS >6.52 dL/g	*p*-value
AKI stage (%)	0	1,475 (19.4)	731 (19.1)	744 (19.6)	0.014
	1	949 (12.5)	478 (12.5)	471 (12.4)	
	2	2,177 (28.6)	1,153 (30.2)	1,024 (27.0)	
	3	3,006 (39.5)	1,456 (38.1)	1,550 (40.9)	
Losicu [mean (SD)]		6.89 (8.08)	7.06 (8.53)	6.72 (7.59)	0.067
28-day mortality (%)		2,144 (28.2)	1,000 (26.2)	1,144 (30.2)	<0.001
90-day mortality (%)		2,940 (38.6)	1,336 (35.0)	1,604 (42.3)	<0.001

Results are expressed as mean (SD), *n* (%). ePVS, Estimated plasma volume status; AKI, Acute Kidney Injury; Losicu, Length of ICU Stay.

primary outcome, with a hazard ratio (HR) of 1.19 (95% CI: 1.06–1.33), compared to male patients (HR: 1.13, 95% CI: 1.00–1.29). Patients aged 65 years or younger demonstrated an HR of 1.38 (95% CI: 1.19–1.59). Non-ventilated patients were at a significantly increased risk for the primary outcome (HR: 1.32, 95% CI: 1.18–1.49) compared to ventilated patients (HR: 1.19, 95% CI: 1.06–1.35). The HR for patients not receiving Continuous Renal Replacement Therapy (CRRT) was 1.21 (95% CI: 1.10–1.34), whereas for those receiving CRRT, the HR was 0.90 (95% CI, 0.76 to 1.08). Non-vasopressor use was associated with an elevated risk of the primary outcome (HR:

TABLE 3 ePVS levels and all-cause in-hospital mortality of septic shock.

	Model I, HR 95%CI, *p* value	Model II, HR, 95%CI, *p* value	Model III, HR, 95%CI, *p* value
28-day mortality			
ePVS (continuous variable)	1.05 (1.03–1.08, $p < 0.001$)	1.06 (1.03–1.08, $p < 0.001$)	1.07 (1.04–1.09, $p < 0.001$)
ePVS [Categorical variables (quartile)]			
≤6.52	Ref	Ref	Ref
>6.52	1.17 (1.07–1.27, $p < 0.001$)	1.18 (1.08–1.29, $p < 0.001$)	1.20 (1.10–1.31, $p < 0.001$)
P for trend	$p < 0.001$	$p < 0.001$	$p < 0.001$
90-day mortality			
ePVS (continuous variable)	1.08 (1.05–1.10, $p < 0.001$)	1.08 (1.06–1.10, $p < 0.001$)	1.08 (1.06–1.10, $p < 0.001$)
ePVS [Categorical variables (quartile)]			
≤6.52	Ref	Ref	Ref
>6.52	1.25 (1.16–1.34, $p < 0.001$)	1.26 (1.17–1.36, $p < 0.001$)	1.25 (1.15–1.35, $p < 0.001$)
P for trend	$p < 0.001$	$p < 0.001$	$p < 0.001$

HR, hazard ratio; CI, confidence interval; Ref, reference.
Model I: No adjustments.
Model II: Adjusted for Age, Gender, and Weight.
Model III: In addition to Model II adjustments, further adjustments were made for SOFA score, Charlson Comorbidity Index, APACHE II score, SAPS II score, Heart Rate (HR), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Mean Arterial Pressure (MAP), Respiratory Rate (RR), Lactate, Potassium, Sodium, Anion Gap, White Blood Cell (WBC) count, Platelet (PLT) count, Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Total Bilirubin, Prothrombin Time (PT), International Normalized Ratio (INR), Blood Urea Nitrogen (BUN), Creatinine (Cr), Mechanical Ventilation, Continuous Renal Replacement Therapy (CRRT), Vasopressor Use, Metastatic Cancer, Diabetes Mellitus (DM), and Hypertension (HT).

1.32, 95% CI: 1.18–1.49) compared to vasopressor use (HR: 1.19, 95% CI: 1.06–1.35).

4 Discussion

In this retrospective cohort study, we analyzed 7,607 ICU patients with septic shock, finding that 2,144 (28.2%) succumbed within 28 days. We identified a J-shaped relationship between estimated plasma volume status (ePVS) at ICU admission and 28-day mortality, with a significant increase in mortality risk when ePVS exceeded 6.52 dL/g. Multivariable Cox regression analysis showed a positive correlation between baseline ePVS above 6.52 dL/g and the risk of death at both 28 and 90 days. Kaplan–Meier curves demonstrated an increased risk of in-hospital mortality for ePVS values over 6.52 dL/g. Calibration curves further confirmed the predictive value of ePVS. Our findings suggest that ePVS, being readily accessible, holds promise as a prognostic tool for patients with septic shock.

ePVS was initially utilized in heart failure patients, with Duarte et al. leading its application by demonstrating its predictive value for early cardiovascular events in heart failure complicating myocardial infarction (9). Subsequent studies have consistently revealed its association with early clinical outcomes of decompensated heart failure and its potential to enhance risk stratification for heart failure (19). Another investigation conducted among US adults unveiled a robust correlation between increasing ePVS and elevated rates of all-cause mortality, cardiovascular mortality, and cancer-related mortality (4). In the emergency department, ePVS has been confirmed to be associated with the diagnosis and prognosis of dyspneic patients (20), as well as in ARDS (14), and also with febrile emergency department patients (21). Furthermore, ePVS has demonstrated a correlation with the severity of lower limb arterial disease and clinical outcomes (22).

In our study involving patients with septic shock, we found that ePVS levels were higher compared to those with cardiovascular diseases. Our cutoff value was set at 6.52 dL/g. This is an intriguing result, as in patients with acute myocardial infarction, an ePVS ≥5.28 mL/g emerged as a risk factor for in-hospital mortality and was associated with an elevated risk of 30-day mortality (23). However, in the report by Kim et al. (24), they investigated ICU patients with sepsis or septic shock. Their findings revealed an ePVS of 7.7 ± 2.1 dL/g, which stood out prominently, emphasizing a significant correlation between ePVS and the amount of intravenous fluid resuscitation in deceased patients. Additionally, they evaluated the utility of ePVS in predicting in-hospital mortality and identified a cutoff point of 7.09 dL/g. They observed a significant association between higher ePVS and increased in-hospital mortality (OR, 1.39; 95% CI, 1.04–1.85, $p = 0.028$). Gianni Turcato et al. conducted three studies on the application of ePVS in emergency department (ED) patients. Among 1,502 febrile patients in the emergency department, the median ePVS value in patients diagnosed with sepsis was 5.54 (4.43–6.51) dL/g, compared to a median ePVS value of 4.51 (3.89–5.24) dL/g in non-septic patients ($p < 0.001$). In multivariate analysis, an ePVS higher than 4.52 dL/g was associated with a odds ratio of 1.824 (95% CI 1.055–3.154, $p = 0.030$) for 30-day mortality (21). For emergency department patients diagnosed with sepsis, it was observed that the average ePVS among those surviving to 30 days was 5.19, while the average ePVS among those who died within 30 days was 5.74 ($p = 0.004$). ePVS emerged as an independent risk factor for 30-day mortality, with an adjusted odds ratio of 1.211 (95% CI 1.004–1.460, $p = 0.045$) (25). A recently published prospective study measured ePVS in 949 infected patients included in the study. The median ePVS value for patients who died within 30 days was higher than that of survivors (5.83 vs. 4.61, $p < 0.001$). Multivariate analysis demonstrated that ePVS, both in continuous and categorical forms

around the median, was an independent risk factor for 30-day mortality even after adjusting for severity, comorbidity, and urgency (13).

We observed that ePVS levels were elevated in patients diagnosed with sepsis compared to those with fever, and notably higher in patients with septic shock. Our study revealed a significant increase in mortality risk when ePVS exceeded 6.52 dL/g. These differing cutoff values may be linked to the pathophysiology of septic shock. Primarily, a hallmark of sepsis is vascular paralysis, characterized by a decrease in arterial pressure and extensive venous dilation in both visceral and cutaneous vascular beds (26). Throughout the progression of sepsis, plasma volume does not decrease but rather increases the unstressed volume, thereby reducing venous return and cardiac output (1, 27). Secondly, the microcirculatory disturbances caused by systemic inflammatory response alter vascular permeability, irreversibly affecting the body's volume regulation and the balance between interstitial and intravascular spaces (27). Additionally, early treatment for septic patients involves intravenous fluid resuscitation to restore tissue perfusion (28). Thus, as mentioned earlier, ePVS correlates with the volume of fluid administered before admission to the intensive care unit (24). The gradual accumulation of resuscitative fluids ultimately leads to intravascular congestion.

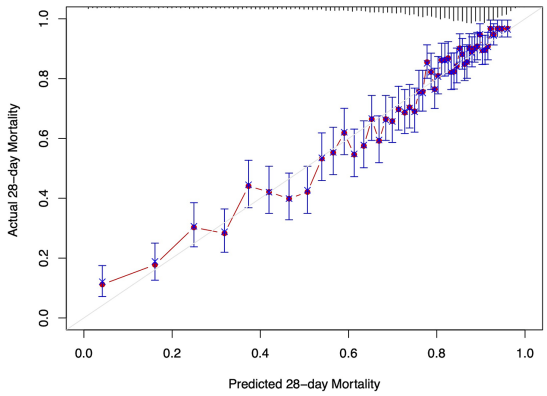

FIGURE 4
Calibration curve for EPVS in predicting 28-day mortality. The *x*-axis represents the predicted 28-day mortality based on ePVS measurements, while the *y*-axis shows the actual 28-day mortality. The blue crosses compare the predicted values with the actual outcomes, and the error bars indicate the confidence intervals of the predictions. The red line shows the trend between predicted and actual values.

The association between ePVS and mortality in septic shock has been unequivocally established in our study. After multiple adjustments for variables, we consistently confirmed that an elevated ePVS serves as an independent risk factor for 28-day mortality in septic shock. One of the reasons for this correlation is the association between ePVS and venous congestion, which can serve as a marker of hemodynamic congestion. Research correlating ePVS with hemodynamic indices has shown a notable correlation between higher ePVS derived from the Duarte formula and higher E/e' ratios. Interestingly, only in females, left ventricular end-diastolic pressure

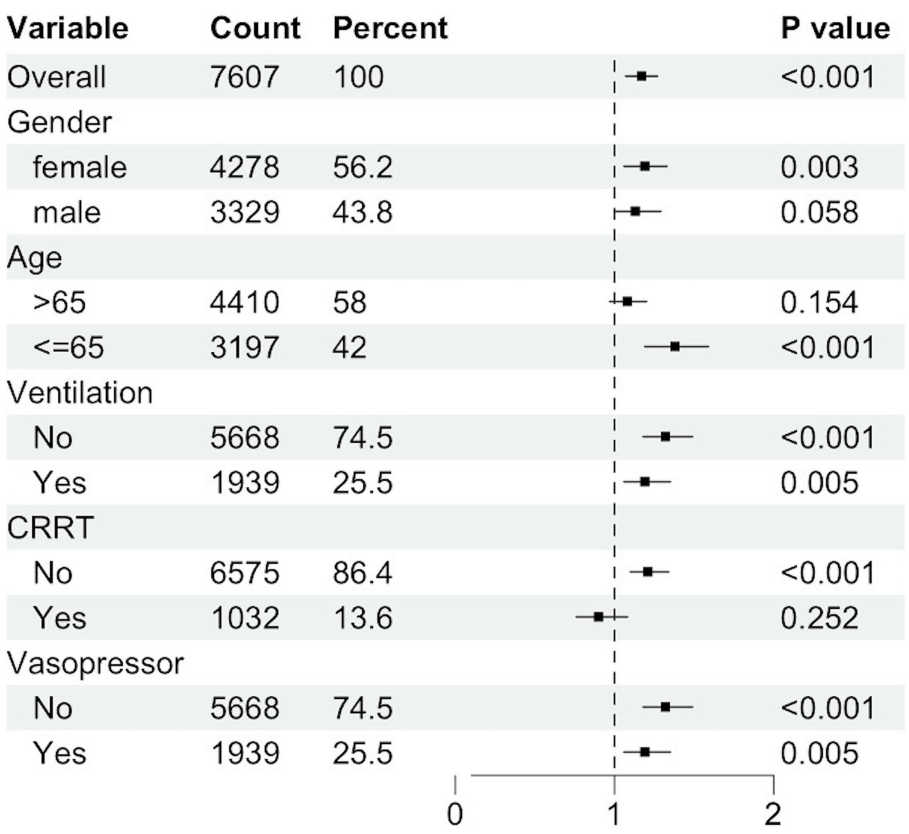

FIGURE 5
Subgroup analysis of the association between ePVS and 28-day mortality.

(LVEDP) is associated with ePVS (29). Moreover, in our subgroup analysis, we observed differences in ePVS between genders, indicating that the influence of gender on plasma volume regulation requires further investigation. While studies have suggested that ePVS is not correlated with pulmonary artery wedge pressure (PAWP) and intracardiac filling pressures (30), this may indirectly suggest that ePVS is more closely associated with the regulation of systemic venous beds and venous return rather than with cardiac function. It is essential to prioritize venous return function during fluid therapy. In summary, our study identified ePVS as an independent predictor of 28-day mortality in septic shock. Elevated ePVS levels may indicate the need for clinicians to prioritize venous return function during fluid therapy and to be vigilant about fluid redistribution in patients with septic shock, thereby assessing their volume status accordingly.

Our study has several limitations. Firstly, retrospective cohort studies inevitably entail biases. However, we attempted to adjust for potential confounders in our data analysis to minimize bias. Secondly, in our inclusion of patients with septic shock in the ICU, we did not extract information on fluid resuscitation in the emergency department, making it difficult to explore the relationship between fluid infusion and ePVS. Thirdly, we only selected ePVS at ICU admission. While we believe ePVS can serve as a continuous target variable to assess its impact on the prognosis of patients with septic shock, we plan to conduct further prospective studies to explore the continuous changes in ePVS and the prognosis of sepsis. Fourthly, although there is a correlation between the ePVS formula and actual plasma volume, this relationship needs to be validated against a gold standard (31). Fifthly, patients known to have chronic anemia were not excluded. Certain ePVS values may be altered due to these conditions. Lastly, due to limitations in the database, there were no echocardiographic data or hemodynamic data related to patient blood volume status and ePVS values in this dataset.

5 Conclusion

In patients with septic shock admitted to the ICU, there exists a J-shaped relationship between the first obtained ePVS values during routine blood tests and the 28-day mortality rate. When blood ePVS exceeds 6.52 dL/g, it is associated with an increased risk of mortality at both 28 and 90 days. However, prospective evidence is needed to confirm these clinical observations and to study the pathophysiological reasons behind elevated ePVS values. Nonetheless, high ePVS levels can serve as important indicators of the severity of illness in patients with septic shock.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The establishment of this database received approval from both the Massachusetts Institute of Technology (Cambridge, MA) and Beth Israel Deaconess Medical Center (Boston, MA), with consent obtained for the original data collection. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

BG: Writing – original draft, Writing – review & editing. RC: Data curation, Writing – review & editing. HuZ: Data curation, Methodology, Writing – review & editing. HoZ: Conceptualization, Supervision, Writing – review & editing. XW: Supervision, Writing – review & editing. DL: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The protocol received financial support from the Capital Clinic Research and Demonstration Application of Diagnosis and Treatment Project (No. Z201100005520038).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1416396/full#supplementary-material

References

1. Bakker J, Kattan E, Annane D, Castro R, Cecconi M, De Backer D, et al. Current practice and evolving concepts in septic shock resuscitation. *Intensive Care Med*. (2022) 48:148–63. doi: 10.1007/s00134-021-06595-9

2. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*. (2021) 47:1181–247. doi: 10.1007/s00134-021-06506-y

3. Vincent J-L, Sakr Y, Sprung CL, Ranieri VM, Reinhart K, Gerlach H, et al. Sepsis in European intensive care units: results of the SOAP study*. *Crit Care Med*. (2006) 34:344–53. doi: 10.1097/01.CCM.0000194725.48928.3A

4. Marawan A, Qayyum R. Estimated plasma volume and mortality: analysis from NHANES 1999–2014. *Clin Res Cardiol*. (2020) 109:1148–54. doi: 10.1007/s00392-020-01606-z

5. Cohen J. The immunopathogenesis of sepsis. *Nature*. (2002) 420:885–91. doi: 10.1038/nature01326

6. Magder S. Volume and its relationship to cardiac output and venous return. *Crit Care*. (2016) 20:1–11. doi: 10.1186/s13054-016-1438-7

7. De Backer D, Cecconi M, Chew MS, Hajjar L, Monnet X, Ospina-Tascón GA, et al. A plea for personalization of the hemodynamic management of septic shock. *Crit Care*. (2022) 26:372. doi: 10.1186/s13054-022-04255-y

8. Strauss MB, Davis RK, Rosenbaum JD, Rossmeisl EC. Water diuresis produced during recumbency by the intravenous infusion of isotonic saline solution. *J Clin Invest*. (1951) 30:862–8. doi: 10.1172/JCI102501

9. Duarte K, Monnez J-M, Albuisson E, Pitt B, Zannad F, Rossignol P. Prognostic value of estimated plasma volume in heart failure. JACC. *Heart Failure*. (2015) 3:886–93. doi: 10.1016/j.jchf.2015.06.014

10. Norrington KD, Turner HK, Barakat MF, Konstantinou K, Kelshikir M, O'Driscoll S, et al. Abstract 17469: prognostic utility of the hemoglobin/hematocrit equation for estimating plasma volume changes during hospitalization for acute decompensated heart failure. *Circulation*. (2012) 126:A17469-A. doi: 10.1161/circ.126.suppl_21.A17469

11. Kobayashi M, Girerd N, Duarte K, Chouihed T, Chikamori T, Pitt B, et al. Estimated plasma volume status in heart failure: clinical implications and future directions. *Clin Res Cardiol*. (2021) 110:1159–72. doi: 10.1007/s00392-020-01794-8

12. Kobayashi M, Rossignol P, Ferreira JP, Aragão I, Paku Y, Iwasaki Y, et al. Prognostic value of estimated plasma volume in acute heart failure in three cohort studies. *Clin Res Cardiol*. (2019) 108:549–61. doi: 10.1007/s00392-018-1385-1

13. Turcato G, Zaboli A, Sibilio S, Brigo F. Estimated plasma volume status is a simple and quick tool that could help define the severity of patients with infection on arrival at the emergency department. *Am J Med Sci*. (2024) 367:343–51. doi: 10.1016/j.amjms.2024.02.003

14. Niedermeyer SE, Stephens RS, Kim BS, Metkus TS. Calculated plasma volume status is associated with mortality in acute respiratory distress syndrome. *Crit Care Explor*. (2021) 3:e0534. doi: 10.1097/CCE.0000000000000534

15. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. (2023) 10:1. doi: 10.1038/s41597-022-01899-x

16. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for Sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287

17. Kellum JA, Lameire NKDIGO AKI Guideline Work Group. Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (part 1). *Crit Care*. (2013) 17:204. doi: 10.1186/cc11454

18. Beesley LJ, Bondarenko I, Elliot MR, Kurian AW, Katz SJ, Taylor JMG. Multiple imputation with missing data indicators. *Stat Methods Med Res*. (2021) 30:2685–700. doi: 10.1177/09622802211047346

19. Fudim M, Lerman JB, Page C, Alhanti B, Califf RM, Ezekowitz JA, et al. Plasma volume status and its association with in-hospital and postdischarge outcomes in decompensated heart failure. *J Card Fail*. (2021) 27:297–308. doi: 10.1016/j.cardfail.2020.09.478

20. Chouihed T, Rossignol P, Bassand A, Duarte K, Kobayashi M, Jaeger D, et al. Diagnostic and prognostic value of plasma volume status at emergency department admission in dyspneic patients: results from the PARADISE cohort. *Clin Res Cardiol*. (2019) 108:563–73. doi: 10.1007/s00392-018-1388-y

21. Turcato G, Zaboli A, Ciccariello L, Pfeifer N. Estimated plasma volume status (ePVS) could be an easy-to-use clinical tool to determine the risk of sepsis or death in patients with fever. *J Crit Care*. (2020) 58:106–12. doi: 10.1016/j.jcrc.2020.05.001

22. Kurokawa T, Otaki Y, Takahashi H, Watanabe T, Shikama T, Tamura H, et al. Impact of estimated plasma volume status on clinical outcomes in patients with lower extremity artery disease who underwent endovascular therapy. *Hypertens Res*. (2023) 46:2005–15. doi: 10.1038/s41440-023-01315-w

23. Chen J, Shen J, Cai D, Wei T, Qian R, Zeng C, et al. Estimated plasma volume status (ePVS) is a predictor for acute myocardial infarction in-hospital mortality: analysis based on MIMIC-III database. *BMC Cardiovasc Disord*. (2021) 21:530. doi: 10.1186/s12872-021-02338-2

24. Kim KH, Cho HJ, Kim SC, Lee J. Prognostic value of estimated plasma volume status in patients with Sepsis. *J Korean Med Sci*. (2022) 37:e145. doi: 10.3346/jkms.2022.37.e145

25. Turcato G, Zaboli A, Sibilio S, Mian M, Brigo F. Estimated plasma volume status can help identify patients with sepsis at risk of death within 30 days in the emergency department. *Emergency Care J*. (2023) 19. doi: 10.4081/ecj.2023.11655

26. Giamarellos-Bourboulis EJ, Aschenbrenner AC, Bauer M, Bock C, Calandra T, Gat-Viks I, et al. The pathophysiology of sepsis and precision-medicine-based immunotherapy. *Nat Immunol*. (2024) 25:19–28. doi: 10.1038/s41590-023-01660-5

27. Messina A, Bakker J, Chew M, De Backer D, Hamzaoui O, Hernandez G, et al. Pathophysiology of fluid administration in critically ill patients. *Intensive Care Med Exp*. (2022) 10:46. doi: 10.1186/s40635-022-00473-4

28. Zampieri FG, Bagshaw SM, Semler MW. Fluid therapy for critically ill adults with sepsis: a review. *JAMA*. (2023) 329:1967–80. doi: 10.1001/jama.2023.7560

29. Kobayashi M, Huttin O, Donal E, Duarte K, Hubert A, Le Breton H, et al. Association of estimated plasma volume status with hemodynamic and echocardiographic parameters. *Clin Res Cardiol*. (2020) 109:1060–9. doi: 10.1007/s00392-020-01599-9

30. Wu Y, Tian P, Liang L, Chen Y, Feng J, Huang B, et al. Estimated plasma volume status adds prognostic value to hemodynamic parameters in advanced heart failure. *Intern Emerg Med*. (2023) 18:2281–91. doi: 10.1007/s11739-023-03422-5

31. Fudim M, Miller WL. Calculated estimates of plasma volume in patients with chronic heart failure—comparison with measured volumes. *J Card Fail*. (2018) 24:553–60. doi: 10.1016/j.cardfail.2018.07.462

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 12 June 2024
DOI 10.3389/fmed.2024.1399527

OPEN ACCESS

EDITED BY
Qinghe Meng,
Upstate Medical University, United States

REVIEWED BY
Elif Keles,
Northwestern Medicine, United States
Chang Hu,
Zhongnan Hospital of Wuhan University, China

*CORRESPONDENCE
Liqin Mo
✉ moliqin001@163.com

RECEIVED 12 March 2024
ACCEPTED 13 May 2024
PUBLISHED 12 June 2024

CITATION
Zhai Y, Lan D, Lv S and Mo L (2024) Interpretability-based machine learning for predicting the risk of death from pulmonary inflammation in Chinese intensive care unit patients.
Front. Med. 11:1399527.
doi: 10.3389/fmed.2024.1399527

Interpretability-based machine learning for predicting the risk of death from pulmonary inflammation in Chinese intensive care unit patients

Yihai Zhai, Danxiu Lan, Siying Lv and Liqin Mo*

Cardiothoracic Surgery Intensive Care Unit, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China

Objective: The objective of this research was to create a machine learning predictive model that could be easily interpreted in order to precisely determine the risk of premature death in patients receiving intensive care after pulmonary inflammation.

Methods: In this study, information from the China intensive care units (ICU) Open Source database was used to examine data from 2790 patients who had infections between January 2019 and December 2020. A 7:3 ratio was used to randomly assign the whole patient population to training and validation groups. This study used six machine learning techniques: logistic regression, random forest, gradient boosting tree, extreme gradient boosting tree (XGBoost), multilayer perceptron, and K-nearest neighbor. A cross-validation grid search method was used to search the parameters in each model. Eight metrics were used to assess the models' performance: accuracy, precision, recall, F1 score, area under the curve (AUC) value, Brier score, Jordon's index, and calibration slope. The machine methods were ranked based on how well they performed in each of these metrics. The best-performing models were selected for interpretation using both the Shapley Additive exPlanations (SHAP) and Local interpretable model-agnostic explanations (LIME) interpretable techniques.

Results: A subset of the study cohort's patients (120/1668, or 7.19%) died in the hospital following screening for inclusion and exclusion criteria. Using a cross-validated grid search to evaluate the six machine learning techniques, XGBoost showed good discriminative ability, achieving an accuracy score of 0.889 (0.874–0.904), precision score of 0.871 (0.849–0.893), recall score of 0.913 (0.890–0.936), F1 score of 0.891 (0.876–0.906), and AUC of 0.956 (0.939–0.973). Additionally, XGBoost exhibited excellent performance with a Brier score of 0.050, Jordon index of 0.947, and calibration slope of 1.074. It was also possible to create an interactive internet page using the XGBoost model.

Conclusion: By identifying patients at higher risk of early mortality, machine learning-based mortality risk prediction models have the potential to

significantly improve patient care by directing clinical decision making and enabling early detection of survival and mortality issues in patients with pulmonary inflammation disease.

KEYWORDS

intensive care unit, infection, mortality, machine learning, precision therapy

1 Introduction

Worldwide, the incidence of infections in intensive care units (ICUs) surpasses that in general wards by approximately 5 to 10 times (1). Particularly prevalent among ICU patients are lower respiratory tract infections, which can constitute 40 to 50% of all infections (2, 3). Among these, lung inflammation is the most common respiratory disease ailment in the lower respiratory tract, contributing significantly to global mortality rates (4).

As the core organ of the respiratory system, impaired lung function can disrupt the balance of oxygen and carbon dioxide in the blood and cause a buildup of metabolic products. This can worsen the body's physiological stress response and lead to serious complications such as acute respiratory failure and sepsis, significantly increasing the risk of death (5, 6). Notably, approximately 20 to 30% of patients with pneumonia admitted to the ICU die within 1 week (7). Thus, early detection of patients with inflammatory lung disease who are at high risk of death is crucial.

Current studies aiming to predict the probability of death in ICU patients encompass various factors, including cerebral infarction (8), acute heart failure (9), sepsis (10), healthcare-associated infections (HAIs) (11), and other domains. However, there have been few investigations on the risk of death from lung inflammation. Sepsis emerges as the most extensively studied area in ICU mortality risk research. Typically triggered by an underlying condition such as a lung infection, its presence indicates that the disease has progressed to a severe level. As a result, early detection of the onset and progression of pulmonary inflammation has major implications for optimizing therapy and improving patient outcomes.

Existing mortality risk models primarily use demographic data from patients outside of China, and Chinese patients are not adequately represented. This limits the ability of existing models to accurately forecast the probability of death in Chinese ICU patients. Hence, patients in China may differ significantly from those in other countries in terms of demographics, disease spectrum, medical procedures, and lifestyle.

Today, determining a patient's risk of death is a challenging clinical task. Machine learning emerges as a potential approach for identifying this risk (12), capable of capturing complex non-linear relationships to accurately identify patterns and features associated with the risk of death by learning from a large amount of clinical data and biochemical indicator data, allowing physicians to make more accurate diagnostic and therapeutic decisions (13).

The goal of this study was to create and verify an interpretable machine learning-based mortality risk prediction model for Chinese ICU patients with pulmonary inflammatory illness. It provides guidance to healthcare practitioners by exploring in-depth the risk factors associated with death. By identifying unfavorable patient outcomes in the early stages of the disease, timely intervention can be implemented, leading to improved patient survival and ultimately enhancing clinical decision making and patient outcomes.

2 Materials and methods

2.1 Study population and outcome

The data used in this study to estimate the probability of death in patients with pulmonary inflammation were obtained from the Critical Care Database version 1.1. This database is an open-source database for intensive care units in Zigong City, Sichuan Province, China, and specifically contains patients with infection (14). The Ethics Committee of the Fourth People's Hospital in Zigong approved the use of this data (Ethics Approval No. 2020-065). The database includes information from 2790 infected individuals (excluding those with COVID-19 pneumonia), such as laboratory test results, baseline characteristics, medication use records, International Classification of Diseases (ICD) codes, nursing records, and follow-up information.

The inclusion criteria for this study were as follows: (1) age $\geq$18 years old and (2) infection site identified as "lung" according to ICD codes. The exclusion criteria were: (1) missing data values $>$25% and (2) missing key variables. A total of 1668 cases were included in the analysis. The patients were divided into two groups: Survivors and Non-survivors, based on their deceased or alive status. The study's results were reported following the criteria for developing and publishing machine learning predictive models in biomedical research (15). **Figure 1** illustrates the flowchart for the patients included in this study and the study design.

2.2 Variable selection and pre-processing

This study selected variables that reflect the disease and treatment effects based on clinical experience and database characteristics, including:

(1) General information: gender, age, history of chronic pulmonary disease, and history of diabetes mellitus;
(2) vital signs: diastolic blood pressure, systolic blood pressure, body temperature, respiration, heart rate, and type of respiratory support;

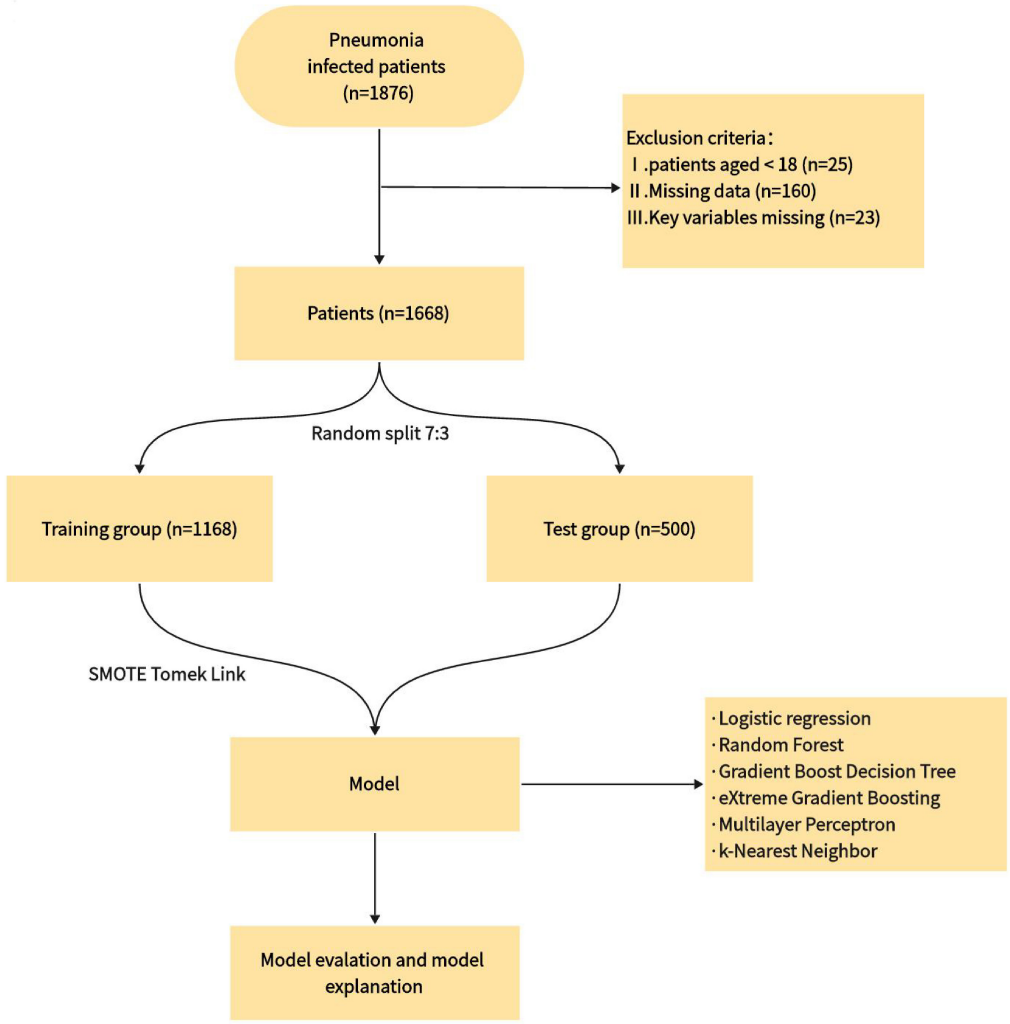

FIGURE 1
Patient selection flowchart and study design routes.

(3) laboratory tests: oxygen saturation in arterial blood (SaO2), white blood cell, albumin, blood creatinine, sodium ions, calcium ions, potassium ions, platelets, Alanine amioTransferase (ALT), Aspartate Aminotransferase (AST), hemoglobin (Hg), activated partial thromboplastin time (APTT), serum total bilirubin, high-sensitivity troponin-i (Tn-i), and international normalized ratio (INR). In total, 25 variables were included.

All variables were checked for outliers and missing values. Missing values greater than 25% were removed, while those less than 25% were addressed using multiple interpolations with the "mice" package in R. Additionally, all variables were mean standardized. Gender, history of chronic obstructive pulmonary disease, history of diabetes, and type of respiratory support were considered discrete variables, while the rest were considered continuous variables. Positive events are represented by a variable value of 1, while negative events are denoted by 0. Vital signs were also selected as the first recorded data upon ICU admission. Supplementary File 1 provides further details.

2.3 Sample equalization processing

The overall mortality rate at discharge in this trial was 7.19%, with a positive-to-negative ratio of approximately 1 to 13. In supervised learning, classification algorithms whose learning goal is overall classification accuracy tend to focus too much on the majority class and fail to learn characteristics from the minority class. To ensure the efficiency of machine learning, this study utilized the SMOTE Tomek Link algorithm, which combines oversampling and undersampling (16). This approach removes noise from samples and balances the sample size.

2.4 Model construction

2.4.1 Machine learning model

In this study, Python software (version 3.10) was used to process the data. Logistic regression (LR), random forest (RF), gradient boost decision tree (GBDT), extreme gradient boosting tree (XGBoost), multilayer perceptron (MLP), and k-nearest

neighbor (KNN) algorithms were used to predict the risk of death in patients with pneumonia.

2.4.2 Model training

The dataset was divided into training and test sets in a 7:3 ratio. To improve the model's generalization ability, 10-fold cross-validation was applied to the test set, and the model's hyperparameters were adjusted using the GridSearchCV method. The model's accuracy was estimated by averaging the data in the test set along with its 95% confidence interval. Eight metrics were used to evaluate the model outcomes: accuracy, precision, sensitivity, F1 score, area under the curve, Brier score, Jordan's index, and calibration slope. Due to minimal variations in the performance metrics among most of the machine learning models, selecting the final model posed a challenge (17). In this study, each measure (such as accuracy, precision) was evaluated from highest to lowest and given a score ranging from 6 to 1, all the points are added together to make the total score. Therefore, The model with the highest score was chosen for further model interpretation.

2.4.3 Model interpretability and variable importance

Variable importance was assessed using the Shapley Additive exPlanations (SHAP) method. For each predicted sample, the model generates a predicted value, and the SHAP value is the value assigned to each feature in that sample (18). SHAP allows for a global evaluation of the model by determining the marginal contribution of features to the model output. Complementing the SHAP method, the Local Interpretable Model-Agnostic Explanations (LIME) method improves the interpretability of the best model and its transparency in clinical practice (19). LIME calculates the risk of premature death and assigns individual weights to each variable, helping to understand changes in estimated probabilities under different observation settings and making the model more distinct.

2.5 Dataset description

The count data in the baseline data are expressed as frequencies and percentages, while the measurement data are expressed as mean ± standard deviation or median (interquartile range), depending on the numerical distribution. The appropriate statistical tests (*t*-test/Chi-square test/non-parametric tests) were used according to the data distribution shape, with $\alpha = 0.05$.

3 Results

The average age of the patients was 67.55 ± 16.37 years, 17.03% had diabetes, and 14.09% had chronic lung disease. There were statistical differences in gender, temperature, systolic blood pressure, diastolic blood pressure, SaO_2, type of respiratory support, APTT, albumin, AST, calcium ions, Tn-i, INR, and white blood cells between surviving and deceased patients. The other characteristics did not show statistical significance. The baseline characteristics of the dataset are summarized in Table 1. The "Total" category represents the information of the entire study population, including survivors and non-survivors groups.

The study adjusted some of the hyperparameters of the models using the GridSearchCV method, and the adjustment space and determined values of the hyperparameters can be found in Supplementary File 2. Table 2 displays the final 10-fold cross-validated model efficacy along with its 95% confidence intervals. In terms of individual model performance, the GBDT model has the highest accuracy, precision, F1-score, AUC value, Brier score, and Youden index, the KNN model had the highest recall, and the MLP model had the highest calibration slope. By calculating the distribution F1 value and AUC value (score = 0.6 F1 + 0.4 AUC), the optimal cutoff value for XGBoost was determined to be 0.510, achieving the highest score of 0.957.

Figure 2 displays an AUC visualization for ten-fold cross-validation. The AUC of the GBDT model was 0.971 (0.957–0.986), followed by the XGBoost model at 0.956 (0.939–0.973) and the RF at 0.955 (0.936–0.974). The probability curves for each model are displayed in Figure 3. The GBDT, XGBoost, and MLP models exhibited the least overlap and demonstrated a large separation between positive and negative events. These models revealed significant differences between patients who died and those who survived, indicating a higher capacity for discrimination. Figure 4 displays the calibration curves for each model, providing further quantification of this discrimination. The calculation of their calibration slopes in Table 2 confirms the improved effectiveness of the GBDT, XGBoost, and MLP models in differentiating patients with various outcomes. The analysis above demonstrates the usefulness of these three models in clinical decision making. Furthermore, Box plots of the six models are in Supplementary Figure 1 in Supplementary Material 2. Among all models, RF and XGBoost perform better in distinguishing positive and negative samples.

After assigning scores to each performance in turn, GBDT exhibited the highest prediction performance score (45 points), followed by XGBoost (36 points, Table 2). Given that the GBDT calibration curve oscillates between rising and falling values around the ideal curve and it performs mediocrely in distinguishing positive and negative samples Supplementary Figure 1 in Supplementary Material 2, XGBoost was chosen for additional model interpretation in this investigation.

Model interpretability, based on the XGBoost model, rates the variables and visually represents their contribution to the probability of death. Figure 5 presents four cases using the LIME validation set, including two death cases (Figures 5A, B) and two survival cases (Figures 5C, D). These charts showcase the top ten factors that have the greatest impact on patient survival or death and explain how these characteristics influence patient outcomes. Specifically, Figure 5A illustrates that male gender, absence of diabetes, absence of chronic pulmonary disease, use of non-invasive ventilation, and presence of low albumin levels (≤29.48 g/L) increase the risk of death. On the other hand, low potassium levels (≤3.24 mmol/L), normal white blood cell counts, normal systolic blood pressure values (128∼159 mmHg), and normal APTT (30.1∼35.74 s) reduce the risk of death. The comprehensive evaluation of this model predicted a probability of death of 0.95 for the patient in question and correctly classified them as deceased.

TABLE 1 Comparison of various characteristics in the two groups of patients (n = 1668).

Characteristics	Total (n = 1668)	Survivors (n = 1548)	Non-survivors (n = 120)	P-value
Age (year)	67.55 ± 16.37	67.63 ± 16.20	66.45 ± 18.53	0.445
Gender				0.028
Male	1025 (61.45%)	940 (56.35%)	85 (5.10%)	
Female	643 (38.55%)	608 (36.45%)	35 (2.10%)	
Temperature (°F)	97.70 (97.16, 97.88)	97.70 (97.16, 98.06)	97.16 (96.80, 97.70)	0.001
Heart rate	98.00 (80.00, 118.00)	98.00 (80.00, 118.00)	98.50 (76.00, 121.00)	0.586
Systolic blood pressure (mmHg)	135.00 (110.00, 162.00)	136.00 (112.00, 163.00)	113.00 (150.75, 89.25)	<0.001
Diastolic blood pressure (mmHg)	80.00 (65.00, 95.00)	80.00 (66.00, 95.00)	72.00 (51.50, 91.50)	<0.001
SaO_2 (%)	98.10 (96.80, 99.00)	98.20 (97.00, 99.00)	98.00 (95.10, 99.18)	<0.001
Respiratory rate	20.00 (16.00, 26.00)	20.00 (16.00, 26.00)	18.00 (15.00, 25.00)	0.195
Diabetes				0.913
Yes	284 (17.03%)	264 (15.83%)	20 (1.20%)	
No	1384 (82.97%)	1284 (76.98%)	100 (6.00%)	
Chronic pulmonary disease				0.766
Yes	235 (14.09%)	217 (13.01%)	18 (1.08%)	
No	1433 (85.91%)	1331 (79.80%)	102 (6.12%)	
Type of respiratory support				0.002
Invasive	499 (29.90%)	448 (26.85%)	51 (3.05%)	
Non-invasive	1169 (70.10%)	1100 (65.96%)	69 (4.14%)	
Activated partial thromboplastin time	29.00 (25.50, 33.10)	28.80 (25.40, 32.70)	31.80 (26.43, 44.80)	<0.001
Alanine aminotransferase	24.30 (15.30, 43.55)	24.00 (15.00, 42.00)	35.30 (19.78, 86.95)	0.053
Albumin	35.40 (30.13, 39.88)	35.50 (30.40, 39.98)	32.20 (27.15, 38.75)	0.003
Aspartate aminotransferase	32.55 (22.40, 61.08)	32.00 (21.93, 58.18)	51.80 (29.03, 96.15)	0.048
Calcium	2.19 (2.07, 2.31)	2.19 (2.08,2.31)	2.16 (2.01, 2.30)	0.046
Creatinine	71.55 (53.03, 104.53)	70.75 (52.73, 103.13)	78.75 (56.98, 126.65)	0.582
Hemoglobin	118.00 (99.00, 137.00)	118.00 (98.25, 137.00)	120.00 (100.50, 139.75)	0.535
High sensitivity troponin I	0.04 (0.01, 0.17)	0.03 (0.01, 0.16)	0.08 (0.02, 0.57)	<0.001
International normalized ratio	1.30 ± 0.53	1.27 ± 0.43	1.65 ± 1.23	<0.001
Platelet	147.00 (106.00, 204.00)	146.00 (106.00, 202.75)	153.50 (111.25, 121.75)	0.344
Potassium	3.58 (3.21, 4.08)	3.58 (3.21, 4.07)	3.54 (3.20, 4.20)	0.227
Sodium	138.65 (135.40,141.20)	138.70 (135.50,141.20)	138.05 (134.70,141.18)	0.740
Total bilirubin	13.70 (9.30, 20.30)	13.70 (9.30, 13.70)	13.90 (8.93, 21.25)	0.448
White blood cell	11.76 (7.99, 16.18)	11.60 (7.98, 16.05)	12.82 (8.30, 18.27)	0.014

In Figure 5D, factors such as female, INR value ≤1.13, and use of invasive ventilation were identified as reduce the risk of death in a patient. Conversely, the absence of diabetes, absence of chronic pulmonary disease and normal body temperatures, systolic blood pressure values, PLT, and sodium levels helped increase the risk of death. The combined evaluation of this model predicted a probability of death of 0.07 for the patient in question and correctly classified them as surviving. Meanwhile, Figures 6A, B demonstrate that gender, SaO_2, Tn-i, INR, and PLT are the top five variables associated with death. The figures use a color scale, ranging from blue to red, to represent values from low to high. The axis at 0 serves as a critical divider: variables positioned to the left are considered protective factors, reducing the risk of death, while those on the right are risk factors, increasing the likelihood of death. For instance, an increase in Tn-i implies a higher risk of death.

In Figure 7, the SHAP dependence plot reveals that within the age group of 50 to 70 years, when systolic blood pressure exceeds 140 mmHg, SHAP values increase significantly and mainly fall within the positive value range. This suggests that hypertension patients in this age group face a higher risk of death from lung inflammation. However, after age 70, high systolic blood pressure seems to act as a protective factor against the risk of death from lung inflammation.

Using the XGBoost model, we explored the interactions among key variables and presented an interaction diagram for the first six

TABLE 2 **Predictive performance of six machine learning models.**

Measure	LR	RF	GBDT	XGBoost	MLP	KNN
Accuracy	0.752 (0.727–0.777)	0.883 (0.852–0.914)	0.919 (0.901–0.937)	0.889 (0.874–0.904)	0.853 (0.821–0.886)	0.860 (0.836–0.883)
Precision	0.751 (0.723–0.780)	0.882 (0.848–0.917)	0.917 (0.895–0.939)	0.871 (0.849–0.893)	0.839 (0.806–0.871)	0.816 (0.793–0.840)
Recall	0.755 (0.712–0.799)	0.885 (0.856–0.915)	0.922 (0.894–0.951)	0.913 (0.890–0.936)	0.875 (0.831–0.918)	0.929 (0.901–0.957)
F1-score	0.752 (0.726–0.777)	0.883 (0.853–0.914)	0.919 (0.901–0.937)	0.891 (0.876–0.906)	0.855 (0.822–0.889)	0.869 (0.847–0.890)
AUC	0.822 (0.789–0.856)	0.955 (0.936–0.974)	0.971 (0.957–0.986)	0.956 (0.939–0.973)	0.920 (0.895–0.944)	0.873 (0.854–0.892)
Brier Score	0.170	0.064	0.032	0.050	0.061	0.112
Youden index	0.763	0.940	0.967	0.947	0.922	0.915
Calibration slope	0.961	1.302	0.957	1.074	0.985	0.886
Total score	13	28	45	36	25	22

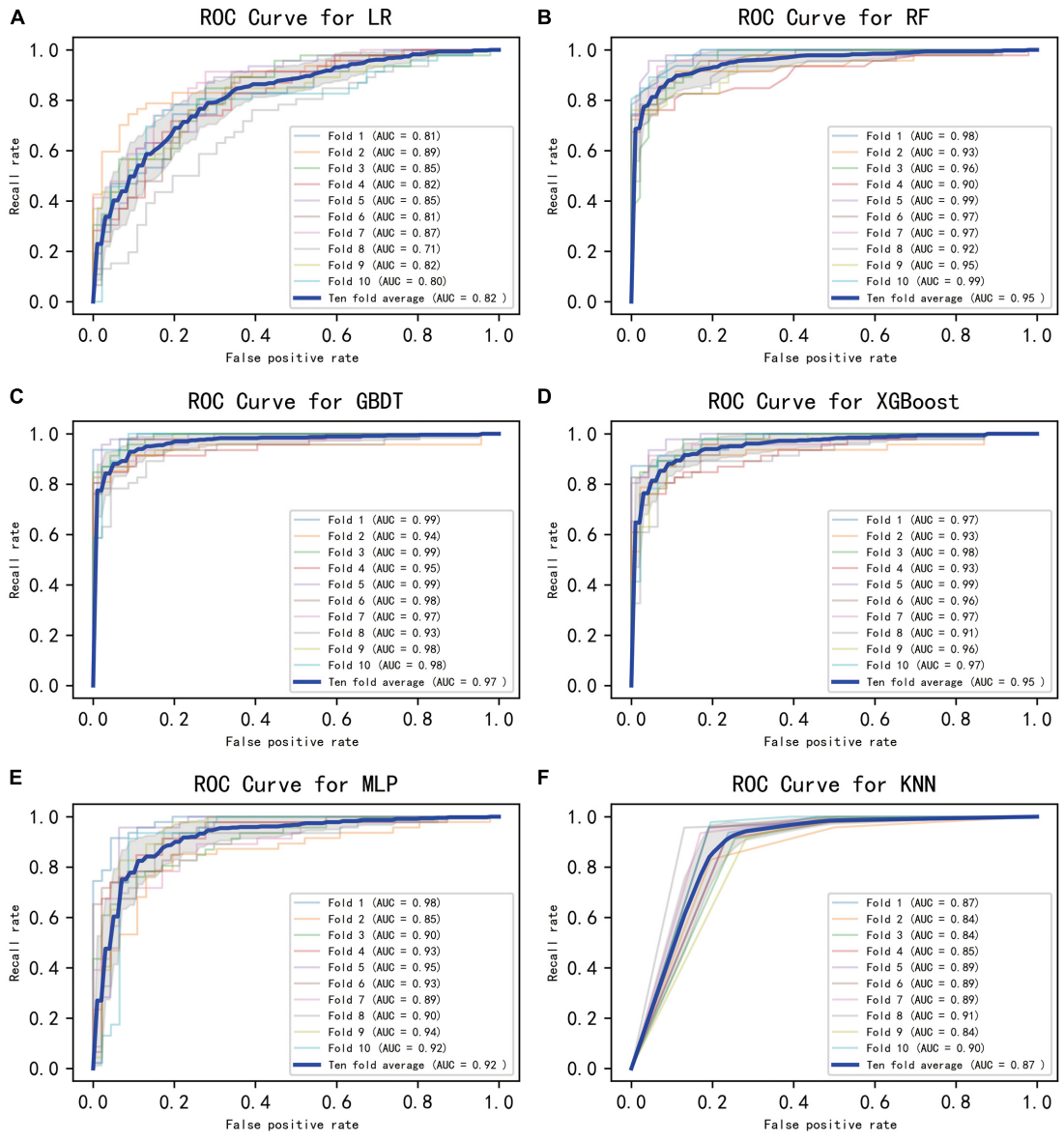

FIGURE 2
Ten-fold cross-validated ROC graphs for six learning models. **(A)** Logistic regression, **(B)** random forest, **(C)** gradient boosting decision tree, **(D)** XGBoost, **(E)** multilayer perceptron, **(F)** K-nearest neighbor. The shaded area in the figure represents the 95% confidence interval.

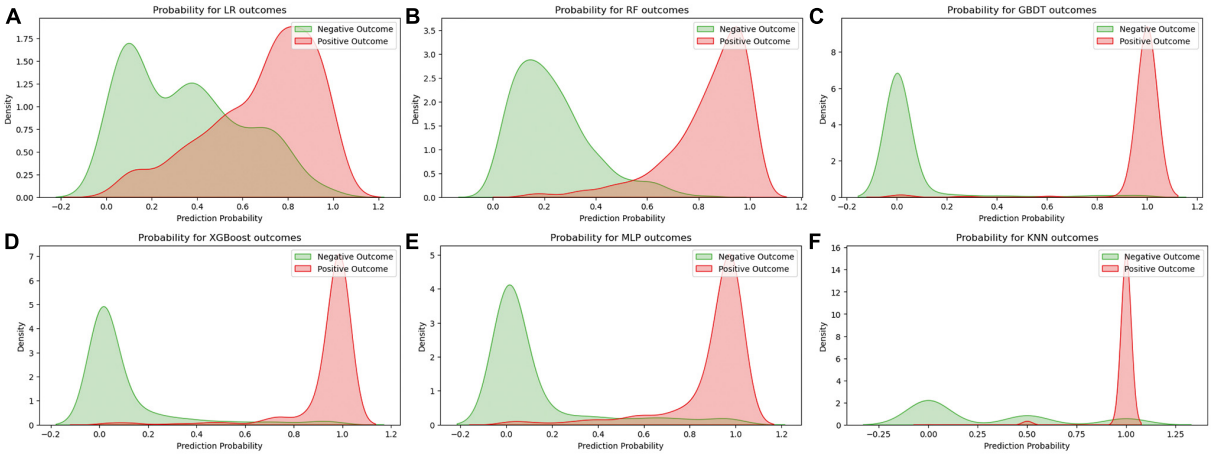

FIGURE 3
Predicted probability curves for the six learning models. **(A)** Logistic regression, **(B)** random forest, **(C)** gradient boosting decision tree, **(D)** XGBoost, **(E)** multilayer perceptron, **(F)** K-nearest neighbor. The green curve indicates patient survival, and the red curve indicates patient death.

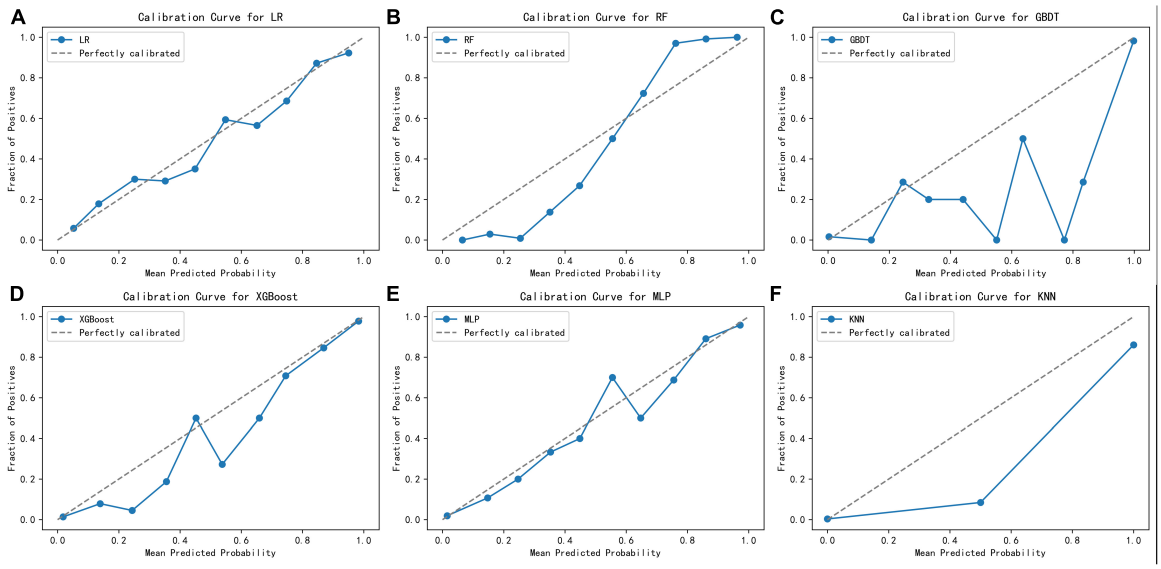

FIGURE 4
Calibration curves for the six learning models. **(A)** Logistic regression, **(B)** random forest, **(C)** gradient boosting decision tree, **(D)** XGBoost, **(E)** multilayer perceptron, **(F)** K-nearest neighbor.

variables (Figure 8). These charts display the interaction between different variables using the distribution of SHAP values. When the interaction between two variables is significant, their corresponding SHAP values are distributed at both ends of the graph. On the other hand, variables with minimal interactions tend to have SHAP values concentrated near zero.

Taking gender and high-sensitivity troponin as an example, as shown in Figure 8, the interaction between these two variables is evident. Areas with a SHAP value of 0 contain mostly blue values, indicating that these variables contribute relatively little to the model output, without significant interactions. In contrast, red values are mainly distributed at both ends of the SHAP value, suggesting that under a specific combination of gender and Tn-i levels, these two variables have a substantial impact on the model prediction. This analysis provides a deeper understanding of the model's behavior.

In addition to the above analyses, we also developed a web-based calculator that can potentially be integrated with hospital information management systems for automated entry and recognition. The website is as follows: https://xgboost-project-app.streamlit.app/. On this website, users can simply input the actual measured values corresponding to the 25 variables mentioned above into the designated content boxes to trigger the model's calculation and prediction process. Figure 9 provides an example diagram of the model home page. The XGBoost model can perform complex calculations and analyses based on these data.

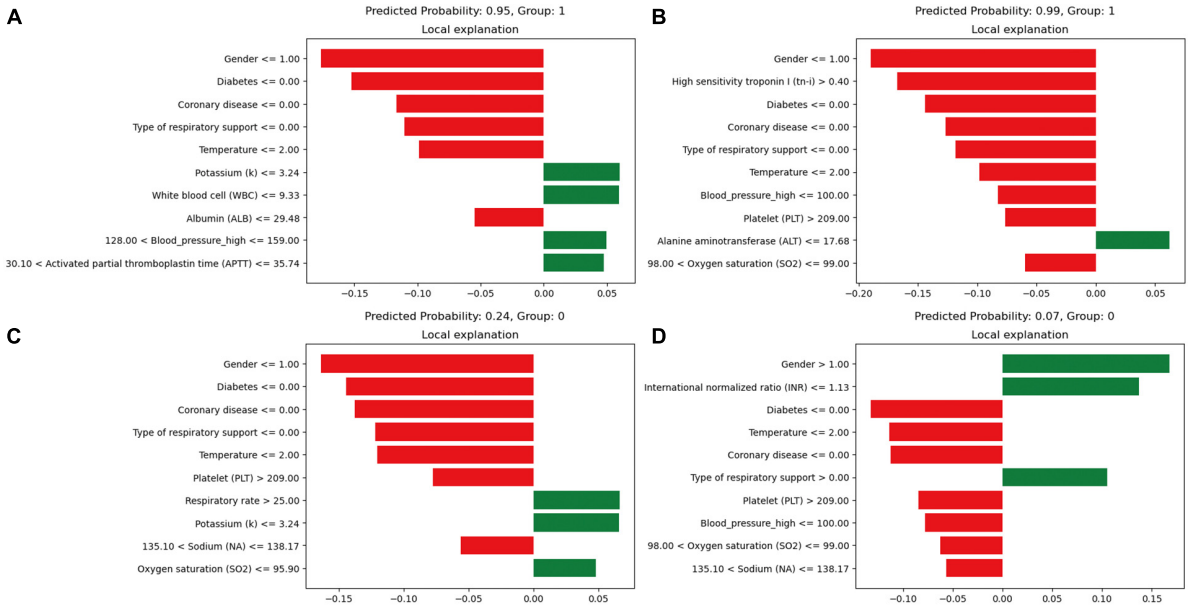

FIGURE 5
Local interpretable model-agnostic explanations (LIME) locally interpretable model agnostic interpretation map. Positive event: 1; Negative event: 0; Male: 1, Female: 2; **(A)** Deceased patients, true-positive cases; **(B)** deceased patients, true-positive cases; **(C)** surviving patients, true-negative cases; **(D)** surviving patients, true-negative cases. The picture presents the top 8 variables that had the greatest impact on survival or death from top to bottom. The length of the bar for each feature indicates the importance (weight) of that feature in making the prediction. A longer bar indicates a feature that contributes more to survival or death. Green bars indicate protective factors and red bars indicate risk factors. *x*-axis indicates the extent to which each predictor variable affects the final probability of a particular patient. The predicted probability of a patient's death, as well as the actual outcome, is shown in each graphic caption.

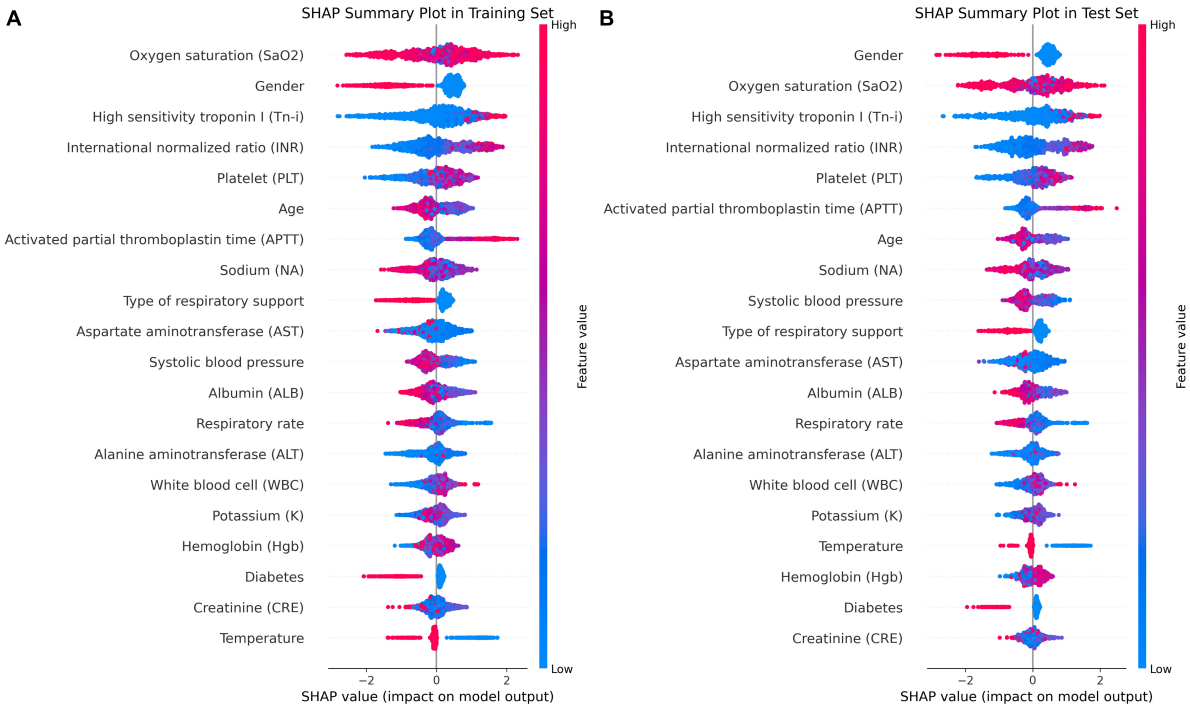

FIGURE 6
Importance ranking of SHAP variables based on the XGBoost model. **(A)** XGBoost SHAP graph on the training set. **(B)** XGBoost SHAP graph on the test set; Each line represents a feature, and the abscissa is the SHAP value. Red dots represent higher feature values, and blue dots represent lower feature values. In terms of Gender, red dots represents female.

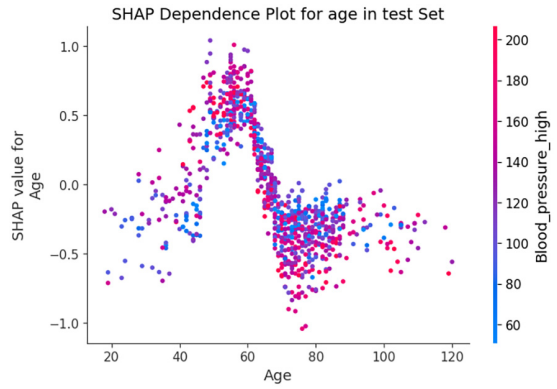

FIGURE 7
Shapley Additive exPlanations (SHAP) dependence plot based on the XGBoost model. SHAP values for specific features exceed zero, representing an increased risk of death.

4 Discussion

Research has indicated that patients admitted to ICU have a comparatively high death rate, which can range from roughly 15 to 40% (20–22). Previous studies have mainly focused on ICU-acquired infections (23) and the forecasting of COVID-19 pneumonia cases and fatalities (24, 25). Moreover, studies indicate that the Acute Physiology and Chronic Health Evaluation II (APACHE II) and the quick Sequential Organ Failure Assessment (qSOFA) scoring systems have a moderate predictive value for mortality among pneumonia patients admitted to the ICU (26, 27).

In practical clinical settings, doctors must undertake strenuous and complex efforts to consider a patient's medical history, physical examination, and trends in vital signs. Accurate, reliable, quick, convenient, and rapid health assessments are crucial for doctors to make decisions that allow them to take appropriate emergency actions in a timely manner, especially for ICU patients. However, predicting the risk of death from pulmonary inflammation with machine learning techniques remains challenging. Therefore, we were able to effectively develop an interpretable machine learning model in this study to predict the in-hospital mortality probability of ICU patients with pulmonary inflammation. Our model excels in rapidly analyzing complex medical data to identify high-risk patients, thereby enabling timely intervention, optimizing resource allocation, and improving outcomes. It also supports personalized medical decision making, helping physicians develop optimal treatment plans for each patient and enhancing the overall efficiency of the healthcare system through precise risk assessments. In short, our model contributes to improved treatment effectiveness and medical resource utilization efficiency.

Prior to this study, previous research has predominantly focused on ICU-acquired infections and the progression of COVID-19 pneumonia, with an emphasis on mortality prediction. However, there has been a scarcity of interpretable machine learning methods tailored for lung inflammation mortality risk prediction. Our model addresses this gap by enabling clinicians to swiftly analyze complex medical data, thereby identifying patients at high risk. This facilitates timely interventions, optimizes the allocation of medical resources, and supports personalized treatment planning, enhancing both patient outcomes and the efficiency of medical care systems. The model's interpretability ensures that clinicians can make informed decisions, thereby

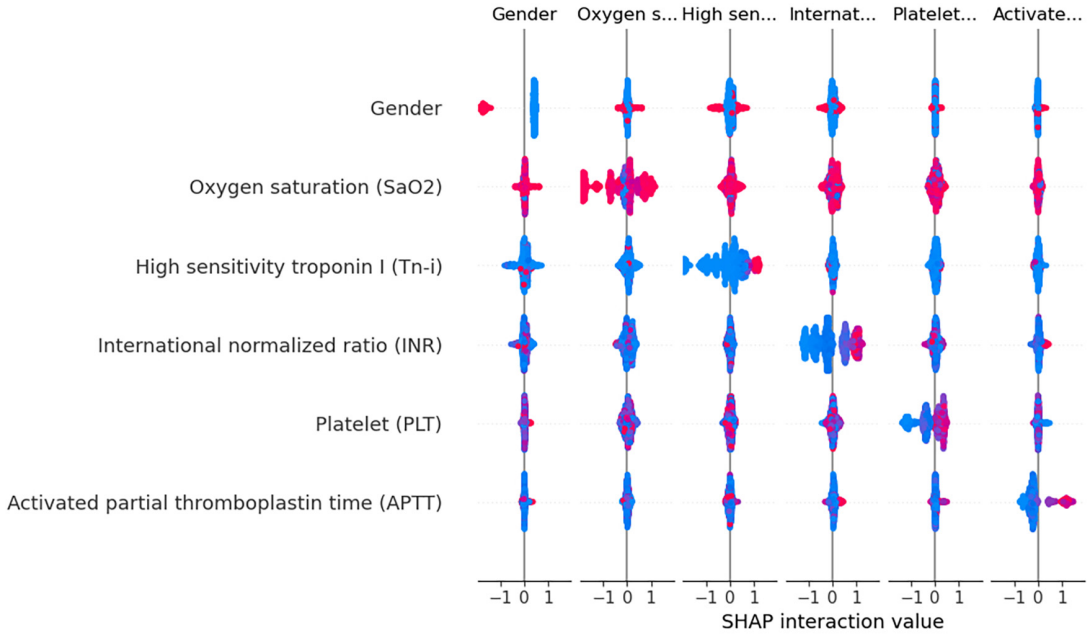

FIGURE 8
Variable interaction graph based on the XGBoost model. Red dots indicate higher values of a feature, while blue dots represent lower values. The red area on the plot signifies that both variables under consideration are registering high values simultaneously. when this interaction is observed on the right side of the SHAP plot, it correlates with an increased risk of death. Conversely, positions on the left side indicate a reduced risk. Specifically regarding the feature "Gender," red dots denote female patients.

improving the overall effectiveness of patient management in critical care settings. Specifically, in our entire cohort, 7.19% (120/1668) of patients experienced early death. Notably, lung infections can progress to sepsis, the leading cause of infection-induced death. Table 3 compares our study to several others, where our study showed excellent performance in specific indicators. Based on the significance of SHAP variables, it was determined that gender, SaO_2, Tn-i, INR, and PLT were the top five important variables associated with early death.

The INR value is a key indicator for measuring the activity of the coagulation system. It reflects the status of blood coagulation function and is an important indicator for evaluating liver dysfunction. In our study, we discovered that patients with pulmonary inflammation exhibited abnormal liver function indicators, such as altered levels of INR, albumin, and ALT. These

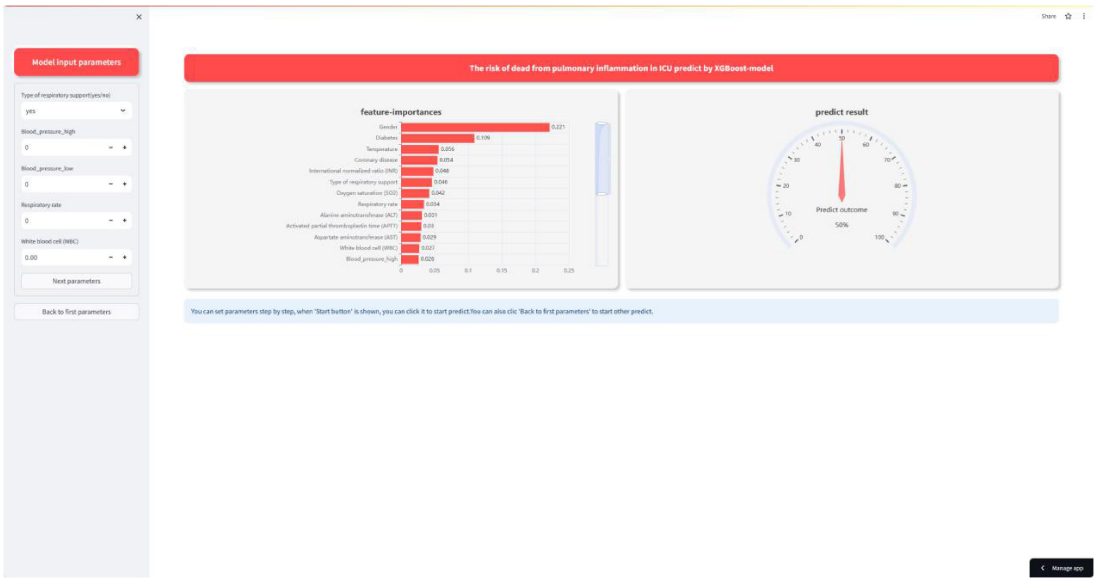

FIGURE 9
Interactive website.

TABLE 3 Comparison with previous studies.

References	Region	Deaths (number of patients)	Disease	Performance
Jeon et al. (41)	Republic of Korea	27.3% (223/816)	Severe pneumonia	ACC: 0.822
				PRE: 0.860
				REC: 0.440
				AUC: 0.856
				Brier score: 0.120
Hu et al. (42)	United States	12.56% (1107/8817)	Sepsis	ACC: 0.895
				AUC: 0.884
Pan et al. (43)	China	47.2% (58/123)	COVID-19	ACC: 0.760
				Sensitivity: 0.667
				AUC: 0.913
				Youden index: 0.733
Wen et al. (44)	China	18.4% (41/223)	Hospital-acquired pneumonia	AUC: 0.863
This work	China	7.19% (120/1668)	Pneumonia	ACC: 0.889
				PRE: 0.871
				REC: 0.913
				AUC: 0.881
				Brier Score: 0.050
				Youden index: 0.947
				Calibration slope: 0.957

abnormalities suggest a potential impairment in liver function. The liver serves as the primary organ responsible for the metabolism and detoxification processes in the human body. Consequently, even a modest decline in liver function can result in metabolic alterations, leading to the accumulation of toxins and worsening the disease's systemic inflammatory response.

Studies have shown that lung inflammation leads to the release of numerous inflammatory mediators, which can, in turn, trigger immune-mediated liver damage, creating a harmful cycle (28). This suggests that when liver dysfunction causes an elevated INR, it impairs coagulation and indicates a weakened ability to respond to inflammation. For instance, studies have found that among patients with liver dysfunction, pulmonary inflammation is one of the most common infectious diseases. In viral pneumonia, the disease can cause cytopathic effects and damage to the endothelial cells, activating platelet and subendothelial aggregation, resulting in hypercoagulability (29, 30). At the same time, the pathogen recognition ability of the immune system and acquired immune system is strengthened, triggering the release of many inflammatory mediators, activating macrophages and T cells to clear viruses and kill infected cells. This not only causes a hypercoagulable state but also severe liver damage. This interaction is directly reflected in the observed increase in INR and exacerbation of liver function in patients with lung inflammation, making these indicators important in predicting a patient's risk of death.

Elevated Tn-i is generally considered a biochemical marker of cardiomyocyte damage, reflecting the degree of damage to the heart muscle cells. In our study, the death group had significantly higher levels of Tn-i compared to the survival group. In cases of lung inflammation, especially severe ones, the heart may be indirectly affected. For example, severe lung infection can trigger a systemic inflammatory response, leading to an increase in inflammatory mediators in the blood, such as interleukin-2 (IL-2), IL-4, IL-6, IL-7, IL-18, and interferon- γ (31), and these mediators lead to cardiac dysfunction and structural damage (32). Specifically, these mediators can cause cardiac dysfunction and structural damage. Infections and inflammatory reactions may also increase the metabolic demand of the heart, and insufficient oxygen supply can further disrupt the metabolism of cardiomyocytes, increasing the risk of cell damage. Additionally, cell infiltration caused by the inflammatory response can lead to inflammatory damage to myocardial tissue and accelerate the release of troponin (33). Therefore, in the context of lung inflammation, elevated troponin is strongly associated with the risk of death.

In conclusion, the key features of the SHAP chart provide crucial insights into the progression and poor prognosis of pneumonia. Most of the indicators support our knowledge from clinical experience. By monitoring these indicators, medical personnel can gain valuable clues that may aid in the early detection of potential risks. This early recognition enables healthcare providers to swiftly implement appropriate interventions. As a result, this proactive approach can significantly enhance the clinical management of patients with pulmonary inflammation disease, ultimately improving their overall care and outcomes.

In the process of building a machine learning model, we utilized various methods for training and optimizing the model, such as the LR, RF, GBDT, XGBoost, MLP, and KNN algorithms. We initially focused on prediction probability plots, visually illustrating how well the model performs under different prediction probabilities. The areas where the curves overlap for positive and negative outcomes are particularly important because they indicate the level of uncertainty the model faces when predicting different outcomes. We noticed a significant overlap in the curves of the LR model, suggesting that the model struggled to distinguish between positive and negative results. This difficulty may be due to the model's linear assumption of the data, which limits its performance.

Another crucial factor in interpreting the model's predictions is the position of the peak on the predicted probability curve. A peak closer to 1 or 0 signifies higher confidence and accuracy in predicting a specific outcome. For example, the GBDT, XGBoost, and MLP models exhibited more concentrated peaks, indicating that these models can provide more accurate predictions when dealing with complex data structures. Furthermore, We conducted a thorough evaluation of multiple machine learning models to determine the most suitable one for deployment. This process can be particularly challenging when the performance metrics of the models are closely matched. To address this, we meticulously assessed each performance indicator, ranking the models from highest to lowest based on their scores. Our analysis revealed that the Gradient Boosting Decision Tree (GBDT) model achieved the highest overall score, closely followed by XGBoost.

Although the GBDT model exhibited strong performance across several metrics, its calibration curve showed significant deviations from the ideal. This was particularly noticeable within the prediction probability range of 30–80%. Moreover, the GBDT model's predicted probabilities were consistently lower than the actual observed probabilities, indicating a potential underestimation issue. These observations necessitate a careful consideration of how the GBDT model's calibration affects its reliability and accuracy in practical applications. On the other hand, although the XGBoost model slightly lags behind GBDT on some performance indicators, its built-in regularization measures and sensitivity to calibration optimization strategies make it more accurate in terms of probabilistic predictions. In the medical field, the requirements for the interpretability and probabilistic accuracy of predictions are particularly stringent. Taking these requirements into consideration, and after thoroughly evaluating the prediction probability curve, calibration curve, box plots, and other performance indicators such as accuracy, we selected the XGBoost model for further application and interpretation. This choice will facilitate its clinical use.

In the field of medical data mining and processing, machine learning has significant advantages over traditional statistical methods. It not only compensates for the limitations of linear models in handling complex data (34), but has also been widely used to develop prediction models for various diseases, such as lung cancer (35), liver cancer (36) and other chronic diseases (37, 38). However, machine learning models are often criticized for their "black box" characteristics in practical applications. This characteristic makes the internal decision making mechanism of the model difficult to intuitively understand, thereby affecting users' trust and acceptance of the model (39). To address this issue, this study incorporates an efficient gradient-boosting machine learning framework: the XGBoost algorithm. It also utilizes a SHAP global variable importance map, LIME personalized explanations, and a web calculator to enhance the interpretability, accuracy, and transparency of the model. This enhancement aims to foster users' trust in the model.

5 Limitations

First, this study was retrospective; therefore, we could not determine the severity of pneumonia. However, the study demonstrates that severe pneumonia comprises approximately 1.3% of all pneumonia patients (40). A more detailed discussion can aid in better disease management by considering the mortality rates of patients with mild/severe pneumonia. Second, this study has limitations regarding the population sample, as its relatively small number of participants may not adequately capture the potential diversity and heterogeneity within the patient population. Third, certain parameters and indicators are absent from the database, which hinders the analysis of factors such as organ failure sequential score and ventilator-specific parameters.

6 Conclusion

This study utilized XGBoost to develop a machine learning model for predicting the risk of death in ICU patients with pulmonary inflammation. The top five important variables were gender, oxygen saturation in arterial blood, high-sensitivity troponin-i, international normalized ratio, and platelets. To gain a deeper understanding of these variables in relation to mortality risk prediction, the LIME method was also used. This model aims to identify patients at a higher risk of early death to guide clinical decision making and improve patient care. However, further research is still needed to expand the sample size and conduct a stratified analysis of patients with mild and severe pneumonia in order to explore more practical treatments for patients with pulmonary inflammation.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://physionet.org/content/icu-infection-zigong-fourth/1.1/.

Ethics statement

The studies involving humans were approved by The Fourth People's Hospital in Zigong. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were obtained from another research group. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

YZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Writing – original draft. DL: Data curation, Formal analysis, Methodology, Project administration, Software, Visualization, Writing – original draft. SL: Writing – original draft, Validation, Project administration, Formal analysis, Data curation, Conceptualization. LM: Conceptualization, Formal analysis, Resources, Validation, Visualization, Writing – review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study received funding support from the 2022 Guangxi Medical and Health Key Discipline Construction Project (Guangxi Health Science and Education Development, 2023-1).

Acknowledgments

We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1399527/full#supplementary-material

References

1. Li Y, Li J, Hu T, Hu J, Song N, Zhang Y, et al. Five-year change of prevalence and risk factors for infection and mortality of carbapenem-resistant *Klebsiella pneumoniae* bloodstream infection in a tertiary hospital in North China. *Antimicrob Resist Infect Control.* (2020) 9:79. doi: 10.1186/s13756-020-00728-3

2. Chang F, Wang X, Huang X, Liu X, Huang L. Analysis on bacterial distribution and change of drug resistance rate in ICUs across southwest China from 2018 to 2022. *Infect Drug Resist.* (2023) 16:5685–96. doi: 10.2147/IDR.S421357

3. Jian ZZ, Zhang N, Ma Y, Man WW. Incidence and trends of nosocomial infection in a tertiary general hospital in China from 2018 to 2021: A retrospective observational study. *Eur Rev Med Pharmacol Sci.* (2023) 27:6760–8.

4. Mattila JT, Fine MJ, Limper AH. Pneumonia-treatment and diagnosis. *Ann Am Thorac Soc.* (2014) 11(Suppl. 4):S189–92. doi: 10.1513/AnnalsATS.201401-027PL

5. Tonelli R, Busani S, Tabbì L, Fantini R, Castaniere I, Biagioni E, et al. Inspiratory effort and lung mechanics in spontaneously breathing patients with acute respiratory failure due to COVID-19: A matched control study. *Am J Respir Crit Care Med.* (2021) 204:725–8. doi: 10.1164/rccm.202104-1029LE

6. Tamminen P, Kerimov D, Viskari H, Aittoniemi J, Syrjänen J, Lehtimäki L. Lung function during and after acute respiratory infection in COVID-19 positive and negative outpatients. *Eur Respir J.* (2022) 59:2102837. doi: 10.1183/13993003.02837-2021

7. Hergens MP, Bell M, Haglund P, Sundström J, Lampa E, Nederby-Öhd J, et al. Risk factors for COVID-19-related death, hospitalization and intensive care: A population-wide study of all inhabitants in Stockholm. *Eur J Epidemiol.* (2022) 37:157–65. doi: 10.1007/s10654-021-00840-7

8. Ouyang Y, Cheng M, He B, Zhang F, Ouyang W, Zhao J, et al. Interpretable machine learning models for predicting in-hospital death in patients in the intensive care unit with cerebral infarction. *Comput Methods Programs Biomed.* (2023) 231:107431. doi: 10.1016/j.cmpb.2023.107431

9. Peng S, Huang J, Liu X, Deng J, Sun C, Tang J, et al. Interpretable machine learning for 28-day all-cause in-hospital mortality prediction in critically ill patients with heart failure combined with hypertension: A retrospective cohort study based on medical information mart for intensive care database-IV and eICU databases. *Front Cardiovasc Med.* (2022) 9:994359. doi: 10.3389/fcvm.2022.994359

10. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost. *J Transl Med.* (2020) 18:462. doi: 10.1186/s12967-020-02620-5

11. Barchitta M, Maugeri A, Favara G, Riela PM, Gallo G, Mura I, et al. Early prediction of seven-day mortality in intensive care unit using a machine learning model: Results from the spin-UTI project. *J Clin Med.* (2021) 10:992. doi: 10.3390/jcm10050992

12. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Med Inform Decis Mak.* (2019) 19:211. doi: 10.1186/s12911-019-0918-5

13. Oh T, Kim D, Lee S, Won C, Kin S, Yang J, et al. Machine learning-based diagnosis and risk factor analysis of cardiocerebrovascular disease based on KNHANES. *Sci Rep.* (2022) 12:2250. doi: 10.1038/s41598-022-06333-1

14. Xu P, Chen L, Zhu Y, Yu S, Chen R, Huang W, et al. Critical care database comprising patients with infection. *Front Public Health.* (2022) 10:852410. doi: 10.3389/fpubh.2022.852410

15. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res.* (2016) 18:e323. doi: 10.2196/jmir.5870

16. Devi D, Biswas S, Kpurkayastha B. Redundancy-driven modified Tomek-link based undersampling: A solution to class imbalance. *Pattern Recogn Lett.* (2017) 93:3–12. doi: 10.1016/j.patrec.2016.10.006

17. Cui Y, Shi X, Wang S, Qin Y, Wang B, Che X, et al. Machine learning approaches for prediction of early death among lung cancer patients with bone metastases using routine clinical characteristics: An analysis of 19,887 patients. *Front Public Health.* (2022) 10:1019168. doi: 10.3389/fpubh.2022.1019168

18. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* (2018) 2:749–60. doi: 10.1038/s41551-018-0304-0

19. Lei M, Wu B, Zhang Z, Qin Y, Cao X, Cao Y, et al. A web-based calculator to predict early death among patients with bone metastasis using machine learning techniques: Development and validation study. *J Med Internet Res.* (2023) 25:e47590. doi: 10.2196/47590

20. Weigl W, Adamski J, Goryński P, Kański A, Hultström M. ICU mortality and variables associated with ICU survival in Poland: A nationwide database study. *Eur J Anaesthesiol.* (2018) 35:949–54. doi: 10.1097/EJA.0000000000000889

21. Donovan J, Glover A, Gregson J, Hitchings AW, Wall EC, Heyderman RS. A retrospective analysis of 20,178 adult neurological infection admissions to United Kingdom critical care units from 2001 to 2020. *BMC Infect Dis.* (2024) 24:132. doi: 10.1186/s12879-024-08976-z

22. Lga ES, De Maio Carrilho CMD, Talizin TB, Cardoso LTQ, Lavado EL, Grion CMC. Risk factors for hospital mortality in intensive care unit survivors: A retrospective cohort study. *Acute Crit Care.* (2023) 38:68–75. doi: 10.4266/acc.2022.01375

23. Vincent JL, Sakr Y, Singer M, Martin-Loeches I, Machado FR, Marshall JC, et al. Prevalence and outcomes of infection among patients in intensive care units in 2017. *JAMA.* (2020) 323:1478–87. doi: 10.1001/jama.2020.2717

24. Grasselli G, Greco M, Zanella A, Albano G, Antonelli M, Bellani G, et al. Risk factors associated with mortality among patients with COVID-19 in Intensive Care Units in Lombardy, Italy. *JAMA Intern Med.* (2020) 180:1345–55.

25. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan. *China. JAMA Intern Med.* (2020) 180:934–43. doi: 10.1001/jamainternmed.2020.0994

26. Carmo TA, Ferreira IB, Menezes RC, Telles GP, Otero ML, Arriaga MB, et al. Derivation and validation of a novel severity scoring system for pneumonia at intensive care unit admission. *Clin Infect Dis.* (2021) 72:942–9. doi: 10.1093/cid/ciaa183

27. Richards G, Levy H, Laterre PF, Feldman C, Woodward B, Bates BM, et al. Curb-65, PSI, and Apache II to assess mortality risk in patients with severe sepsis and community acquired pneumonia in PROWESS. *J Intensive Care Med.* (2011) 26:34–40. doi: 10.1177/0885066610383949

28. Bonnel AR, Bunchorntavakul Creddy KR. Immune dysfunction and infections in patients with cirrhosis. *Clin Gastroenterol Hepatol.* (2011) 9:727–38. doi: 10.1016/j.cgh.2011.02.031

29. Mcconnell MJ, Kondo R, Kawaguchi N, Iwakiri Y. COVID-19 and liver injury: Role of inflammatory endotheliopathy, platelet dysfunction, and thrombosis. *Hepatol Commun.* (2022) 6:255–69. doi: 10.1002/hep4.1843

30. Nardo AD, Schneeweiss-Gleixner M, Bakail M, Dixon, Lax SF, Trauner M. Pathophysiological mechanisms of liver injury in COVID-19. *Liver Int.* (2021) 41:20–32. doi: 10.1111/liv.14730

31. Hasanvand A. COVID-19 and the role of cytokines in this disease. *Inflammopharmacology.* (2022) 30:789–98. doi: 10.1007/s10787-022-00992-2

32. Ren Y, Zhang L, Xu F, Han D, Zheng S, Zhang F, et al. Risk factor analysis and nomogram for predicting in-hospital mortality in Icu patients with sepsis and lung infection. *BMC Pulm Med.* (2022) 22:17. doi: 10.1186/s12890-021-01809-8

33. Burkert FR, Lanser L, Pizzini A, Bellmann-Weiler R, Weiss G. Markers of infection-mediated cardiac damage in influenza and COVID-19. *Pathogens.* (2022) 11:1191. doi: 10.3390/pathogens11101191

34. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: Machine learning and the future of medicine. *J Intern Med.* (2018) 284:603–19. doi: 10.1111/joim.12822

35. Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R. machine learning for early lung cancer identification using routine clinical and laboratory data. *Am J Respir Crit Care Med.* (2021) 204:445–53. doi: 10.1164/rccm.202007-2791OC

36. Turco S, Tiyarattanachai T, Ebrahimkheil K, Eisenbrey J, Kamaya A, Mischi M, et al. Interpretable machine learning for characterization of focal liver lesions by contrast-enhanced ultrasound. *IEEE Trans Ultrason Ferroelectr Freq Control.* (2022) 69:1670–81. doi: 10.1109/TUFFC.2022.3161719

37. Yuan L, Yang L, Zhang S, Xu Z, Qin J, Shi Y, et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: A prospective multicentre clinical cohort study. *EClinicalMedicine.* (2023) 57:101834. doi: 10.1016/j.eclinm.2023.101834

38. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell.* (2023) 186:1772–91. doi: 10.1016/j.cell.2023.01.035

39. Petch J, Di Snelson W. Opening the black box: The promise and limitations of explainable machine learning in cardiology. *Can J Cardiol.* (2022) 38:204–13. doi: 10.1016/j.cjca.2021.09.004

40. Lee HW, Ji E, Ahn S, Yang HJ, Yoon SY, Park TY, et al. A population-based observational study of patients with pulmonary disorders in intensive care unit. *Korean J Intern Med.* (2020) 35:1411–23. doi: 10.3904/kjim.2018.449

41. Jeon ET, Lee HJ, Park TY, Jin KN, Ryu B, Lee HW, et al. Machine learning-based prediction of in-ICU mortality in pneumonia patients. *Sci Rep.* (2023) 13:11527. doi: 10.1038/s41598-023-38765-8

42. Hu C, Li L, Huang W, Wu T, Xu Q, Liu J, et al. Interpretable machine learning for early prediction of prognosis in sepsis: A discovery and validation study. *Infect Dis Ther.* (2022) 11:1117–32. doi: 10.1007/s40121-022-00628-6

43. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic assessment of COVID-19 in the intensive care unit by machine learning methods: Model development and validation. *J Med Internet Res.* (2020) 22:e23128. doi: 10.2196/23128

44. Wen JN, Li N, Guo CX, Shen N, He B. Performance and comparison of assessment models to predict 30-day mortality in patients with hospital-acquired pneumonia. *Chin Med J.* (2020) 133:2947–52. doi: 10.1097/CM9.0000000000001252

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 09 July 2024
DOI 10.3389/fmed.2024.1425799

OPEN ACCESS

EDITED BY
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

REVIEWED BY
Yuichiro Sakamoto,
Saga University, Japan
Rungrote Natesirinilkul,
Chiang Mai University, Thailand

*CORRESPONDENCE
Jingchun Song
✉ songjingchun@126.com

†These authors have contributed equally to this work

RECEIVED 30 April 2024
ACCEPTED 17 June 2024
PUBLISHED 09 July 2024

CITATION
Zeng Q, Lin Q, Zhong L, He L, Zhang N and Song J (2024) Development and validation of a nomogram to predict mortality of patients with DIC in ICU.
Front. Med. 11:1425799.
doi: 10.3389/fmed.2024.1425799

Development and validation of a nomogram to predict mortality of patients with DIC in ICU

Qingbo Zeng[1,2†], Qingwei Lin[1†], Lincui Zhong[1], Longping He[1], Nianqing Zhang[2] and Jingchun Song[1*]

[1]Intensive Care Unit, The 908th Hospital of Chinese PLA Logistic Support Force, Nanchang, China, [2]Intensive Care Unit, Nanchang Hongdu Hospital of Traditional Chinese Medicine, Nanchang, China

Background: Disseminated intravascular coagulation (DIC) is a devastating condition, which always cause poor outcome of critically ill patients in intensive care unit. Studies concerning short-term mortality prediction in DIC patients is scarce. This study aimed to identify risk factors contributing to DIC mortality and construct a predictive nomogram.

Methods: A total of 676 overt DIC patients were included. A Cox proportional hazards regression model was developed based on covariates identified using least absolute shrinkage and selection operator (LASSO) regression. The prediction performance was independently evaluated in the MIMIC-III and MIMIC-IV Clinical Database, as well as the 908th Hospital Database (908thH). Model performance was independently assessed using MIMIC-III, MIMIC-IV, and the 908th Hospital Clinical Database.

Results: The Cox model incorporated variables identified by Lasso regression including heart failure, sepsis, height, SBP, lactate levels, HCT, PLT, INR, AST, and norepinephrine use. The model effectively stratified patients into different mortality risk groups, with a C-index of >0.65 across the MIMIC-III, MIMIC-IV, and 908th Hospital databases. The calibration curves of the model at 7 and 28 days demonstrated that the prediction performance was good. And then, a nomogram was developed to facilitate result visualization. Decision curve analysis indicated superior net benefits of the nomogram.

Conclusion: This study provides a predictive nomogram for short-term overt DIC mortality risk based on a Lasso-Cox regression model, offering individualized and reliable mortality risk predictions.

KEYWORDS

nomogram, Lasso-Cox regression, disseminated intravascular coagulation, prediction, short-term mortality, intensive care unit

Introduction

Disseminated intravascular coagulation (DIC) frequently occurs in critically ill patients admitted to the intensive care unit (ICU) and is associated with high mortality. DIC is a severe syndrome characterized by systemic activation of coagulation, and plenty of thrombin with fibrin deposition in the micro- and macrovascular systems (1, 2), which is provoked by various underlying diseases, such as serious infections, trauma, malignancies, and liver disease. The complex pathophysiology of DIC involves impairment of anticoagulant mechanisms,

uncontrolled activation of the tissue factor pathway, and suppression of fibrinolysis (3, 4). These changes can lead to life-threatening thrombosis and bleeding event, as well as organ dysfunction, which will substantially increase mortality (5, 6). Therefore, the key to therapy for DIC is to control activation of blood coagulation and decrease bleeding and thrombosis risk. Currently, the clinically therapeutic strategies for DIC are limited and prognosis is often poor although treatment of the underlying condition may improve DIC (7). Obviously, it is necessary to identify its potential death risk factors if DIC occurred.

Close monitoring of coagulation parameters is required for patients with DIC given their severely abnormal and dynamic laboratory values and clinical status. However, quantification of mortality risk based on the large volume of data available in electronic health records poses a challenge for ICU physicians. Rapid identification of high-risk patients and timely intervention remain clinically difficult.

Recent advancements in prognostic and diagnostic tools, such as nomogram and artificial intelligence algorithms, have demonstrated efficacy in a range of medical disciplines (8, 9). Previous studies have shown that the models established by these methods can be used in the survival analysis of multiple diseases and show more promising performance than traditional illness severity scoring systems (10–14). Regrettably, prognostic models specifically catering to overt DIC patients within ICU settings remain conspicuously absent. Therefore, this study aims to construct a new prediction model based on the Lasso-Cox method to predict the survival probability of critically ill patients with overt DIC during hospitalization accurately, which would be helpful to provide timely management of overt DIC.

Methods

Source of data

This retrospective observational cohort study utilized data from the MIMIC-III version 1.4 and MIMIC-IV version 2.0 databases. Both MIMIC-III and MIMIC-IV are openly available critical care databases containing ICU patient data from Boston's Beth Israel Deaconess Medical Center. Permission to extract data from these databases was granted following a rigorous deidentification process overseen by the Harvard Medical School's Ethics Review Board and the Massachusetts Institute of Technology (Record ID: 11763035). We also included the patients from the 908th Hospital of Chinese PLA Logistic Support Force as the external validation set. The institutional review board of the 908th Hospital of Chinese PLA Logistic Support Force approved our study and waived the requirement for informed consent because of the retrospective nature of the present study.

Study population and data extraction

The following data were obtained from the MIMIC-III, MIMIC-IV, and the 908th Hospital databases: (1) demographic data; (2) comorbidities, including Atrial fibrillation, sepsis, congestive heart failure, hypertension, chronic obstructive pulmonary disease, and diabetes; (3) outcomes, including ICU stay time, 7-day mortality, and 28-day mortality; (4) severity score, including simplified Acute Physiology Score II (SAPSII), sequential organ failure assessment (SOFA) score; (5) mean value of vital signs and the laboratory test value within 24 h of admission to the ICU; and (6) treatment measures (renal replacement therapy). Data were extracted using PostgreSQL program (version 12). Adult DIC patients (≥18 years) were included as defined according to the ISTH criterion (15): platelet count<$100*10^9$/L = 1 point, <$50*10^9$/L = 2 points; PT prolongation of >3 s = 1 point, >6 s = 2 points; FIB<1.0 g/L = 1 point; a moderate increase in FDP or D-dimer levels = 2 points, a strong increase = 3 points. Overt DIC was diagnosed if a total score was not less than 5. Exclusion criteria were: (1) age < 18 years; (2) pregnant women; (3) patients with congenital coagulopathy; (4) the coagulation function was frequently affected by the pathologic states of tumors and the chemotherapy agent used; and (5) dying or leaving within 24 h after ICU admission.

Statistical analysis

Categorical variables were compared using the chi-squared test, while continuous variables were analyzed using the Student's *t*-test or the Wilcoxon–Mann–Whitney test. Lasso regression was employed to identify significant risk factors. The prediction model including variables screened by Lasso regression were constructed based on Cox regression. The performance of predictive nomogram was evaluated by the C-index and calibration curve. DCA analysis was applied to assess the clinical usefulness. Clinical applicability was assessed using the Kaplan–Meier method. All of statistical analyses were conducted with R software v4.1.1 and SPSS 27.0 software. Variables missing more than 5% of values were handled using the random forest (RF) method, based on the"randomForest"package of R. For variables missing less than 5% of values, median imputation was used. Variables with more than 30% missing values were excluded. A p value of <0.05 was considered statistically significant.

Results

Characteristics of the study participants

The study integrated 39 routinely available variables from ICU patients, comprising of 32 continuous and seven categorical variables (see Table 1). The correlative data of patients in the MIMIC-III were extracted. After data preprocessing and removal of samples with missing values, the MIMIC-III database contributed 148 patients. Employing the same inclusion and exclusion criteria, 386 patients were drawn from MIMIC-IV and 142 additional patients were randomly selected from the 908th Hospital (Figure 1).

A total of 676 overt DIC patients were included. Of these, 32.5% experienced short-time mortality. The patient baseline characteristics are detailed in Table 1. The median patient age was 60, with 382 (56.5%) being male. 179 (26.5%) individuals were diagnosed with hypertension, 215 (31.8%) with atrial fibrillation, 170 (25.1%) with heart failure, 294 (43.5%) with sepsis, 212 (31.4%) with diabetes, and 29 (4.3%) with trauma. There were 140 (20.7%) patients who received continuous renal replacement therapy (CRRT).

10.3389/fmed.2024.1425799

TABLE 1 Comparison of clinical data.

Variables	Training set (*N* = 148)	Testing set (*N* = 386)	External testing set (*N* = 142)
Male, *n* (%)	83 (56.1)	213 (55.2)	86 (60.6)
Age, (years)	63 (51–75)	58 (50–72)	60 (44–75)
Weight (kg)	78 (65–96)	79 (67–98)	75 (59–95)
Height (cm)	169 (163–178)	170 (160–178)	168 (161–176)
Hypertension, *n* (%)	50 (33.8)	82 (21.2)	47 (33.1)
Diabetes, *n* (%)	46 (31.1)	121 (31.3)	45 (31.7)
COPD, *n* (%)	2 (1.3)	7 (1.8)	3 (2.1)
Atrial fibrillation, *n* (%)	42 (28.4)	134 (34.7)	39 (27.5)
Heart failure, *n* (%)	44 (29.7)	90 (23.3)	36 (25.3)
Sepsis, *n* (%)	78 (52.7)	159 (41.2)	57 (40.1)
Trauma, *n* (%)	3 (2.0)	21 (5.4)	5 (3.5)
Temperature (°C)	36.7 (36.2–37.5)	36.7 (36.4–37.2)	36.6 (36.1–37.3)
RR (rate/min)	21 (18–25)	20 (17–23)	20 (18–25)
HR (rate/min)	93 (80–107)	88 (79–104)	91 (76–105)
SBP (mmHg)	110 (102–122)	107 (102–117)	116 (96–135)
DBP (mmHg)	60 (53–67)	58 (52–65)	65 (56–77)
MBP (mmHg)	74 (67–83)	73 (67–79)	77 (65–86)
WBC ($\times 10^9$/L)	10.3 (6.3–14.5)	10.4 (6.4–15.8)	12.1 (8.1–16.5)
RBC ($\times 10^{12}$/L)	3.1 (2.8–3.6)	2.8 (2.5–3.4)	3.0 (2.4–3.5)
HB (g/L)	94 (83–109)	90 (77–104)	90 (72–111)
HCT (%)	28 (25–31)	28 (24–32)	28 (22–33)
PLT ($\times 10^9$/L)	92 (49–199)	93 (58–173)	68 (38–107)
ALT (U/L)	34.5 (20.3–103.8)	34.0 (20.0–74.8)	84.9 (40.6–407.7)
AST (U/L)	55.5 (30.0–159.8)	59.0 (38.0–137.3)	60.1 (28.3–323.0)
Tbil (mg/dL)	1.5 (0.6–4.5)	1.9 (0.9–4.9)	0.8 (0.5–1.6)
BUN (mg/dL)	32 (19–51)	29 (18–46)	33 (21–53)
Cr (mg/dL)	1.4 (0.9–2.4)	1.2 (0.9–2.0)	1.2 (0.8–1.7)
Glu (mg/dL)	117 (101–150)	120 (99–163)	119 (103–161)
PT (s)	16.8 (14.8–19.7)	20.7 (18.7–26.7)	21.9 (18.7–27.5)
INR	1.5 (1.3–2.0)	1.9 (1.7–2.5)	1.8 (1.5–2.1)
FIB (g/L)	3.36 (1.85–5.60)	1.99 (1.39–3.43)	1.36 (0.81–2.57)
D-dimer (mg/L)	6.32 (4.04–9.61)	2.46 (1.12–6.74)	10.66 (4.98–20.49)
PH	7.33 (7.26–7.42)	7.38 (7.28–7.44)	7.34 (7.26–7.42)
PaO2 (mm)	102 (78–130)	137 (76–291)	125 (71–174)
PaCO2 (mm)	39 (33–44)	38 (32–44)	36 (29–42)
Lactate (mmol/L)	2.0 (1.4–3.0)	2.4 (1.6–4.0)	4.3 (2.4–7.4)
CRRT, *n* (%)	39 (26.4)	78 (20.2)	23 (16.2)
Epinephrine, *n* (%)	4 (2.7)	9 (2.3)	6 (4.2)
Norepinephrine, *n* (%)	62 (41.9)	194 (50.2)	79 (55.6)
SAPSS II	50 (36–61)	45 (36–56)	47 (35–58)

Variables selection

We utilized Lasso regression to identify significant predictors and the variation characteristics of these variables were shown in Figure 2. This analysis indicated heart failure, sepsis, height, SBP, lactate, HCT, PLT, INR, AST, and norepinephrine as relevant variables. Finally, Cox regression model was further established based on variables screened by Lasso regression (Table 2).

Development of a multivariate prognostic nomogram

The patients were divided into the training, testing, and external validation sets. The training set was used to develop the predictive model, while the testing and external validation sets were applied to validate. On the basis of the LASSO regression, a prognostic nomogram was constructed, integrating the nine significant factors (Figure 3). This nomogram was subsequently validated using both testing and external validation sets.

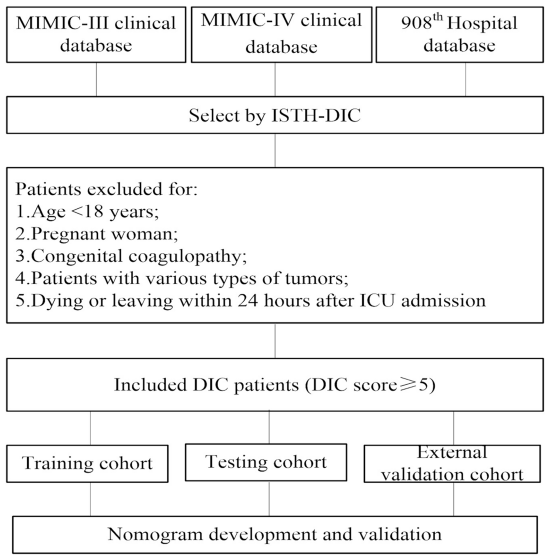

FIGURE 1
Flowchart of patient inclusion in the study.

Validation of the prediction nomogram

The nomogram demonstrated good accuracy for predicting overt DIC patient short-term mortality. The nomogram exhibited the C-index of 0.81 and 0.77 for predicting 7- and 28-day mortality in the training set, and the C-index of 0.63 and 0.67 in the testing set. In the external validation set, C-index was 0.65 and 0.68 in predicting 7- and 28-day mortality. The AUCs for the 7-day mortality probabilities were 0.78, 0.71, and 0.70 in the training, testing, and external validation sets, while the AUCs for the 28-day mortality probabilities 0.82, 0.69, and 0.68, respectively (Figures 4A,B).

Calibration and clinical application of nomogram

The calibration curves of the nomogram established for predicting mortality of 7 and 28 days showed a good consistency between observed and predicted outcomes in the training, testing, and external validation sets (Figures 5A–F). In addition, the decision curve analysis displayed the prediction model provides useful prognostic information to assist clinical decision making (Figures 6A,B).

Risk strata were generated based on tertiles of predicted 7- and 28-day mortality. K-M curves of mortality were plotted based on the risk strata in the training, testing, and external validation sets. The

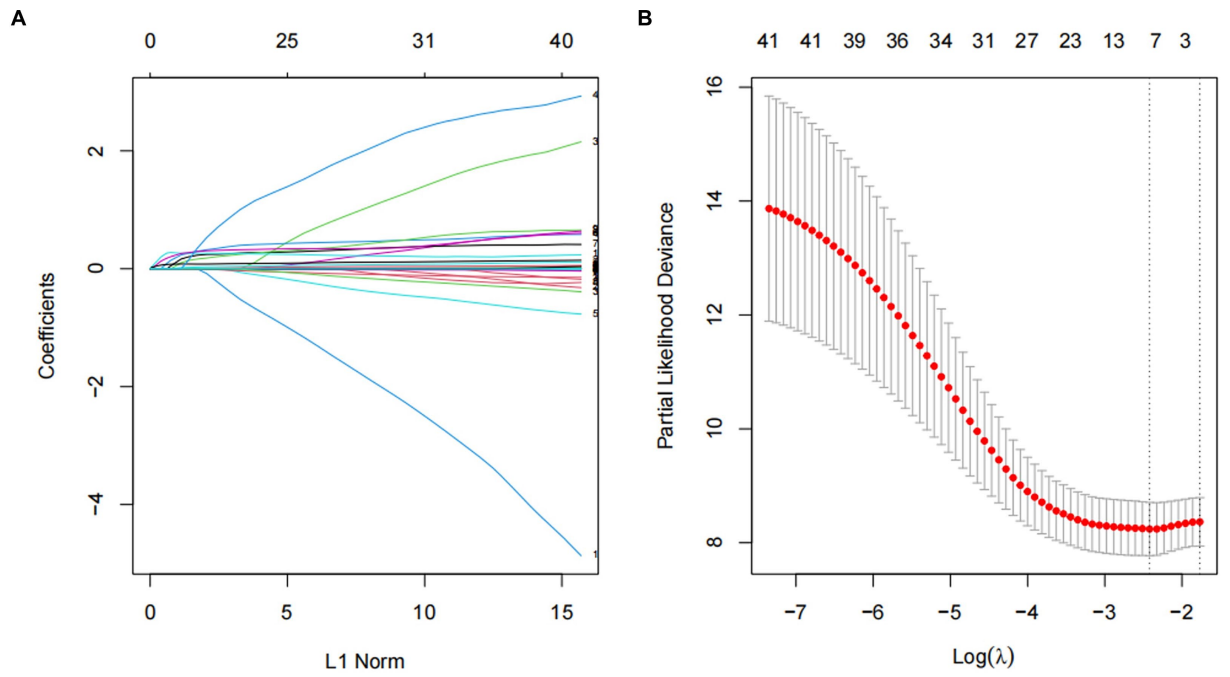

FIGURE 2
Screening of variables based on Lasso regression. **(A)** The variation characteristics of the coefficient of variables; **(B)** the selection process of the optimum value of the parameter λ in the Lasso regression model by cross-validation method.

TABLE 2 Cox regression to predict prognosis based on Lasso regression.

Variables	HR (95%)	Coefficient	*p value*
Heart failure	1.74 (0.97–3.13)	0.550	0.062
Sepsis	1.86 (0.99–3.49)	0.727	0.053
Height	0.99 (0.97–1.00)	−0.015	0.116
SBP	0.99 (0.97–1.02)	−0.005	0.719
Lactate	1.09 (0.98–1.21)	0.101	0.119
HCT	1.07 (1.01–1.14)	0.063	0.019
PLT	0.99 (0.98–1.02)	−0.003	0.031
INR	1.60 (1.17–2.19)	0.499	0.003
AST	1.02 (1.00–1.03)	0.002	0.042
Norepinephrine	1.53 (0.82–2.85)	0.424	0.183

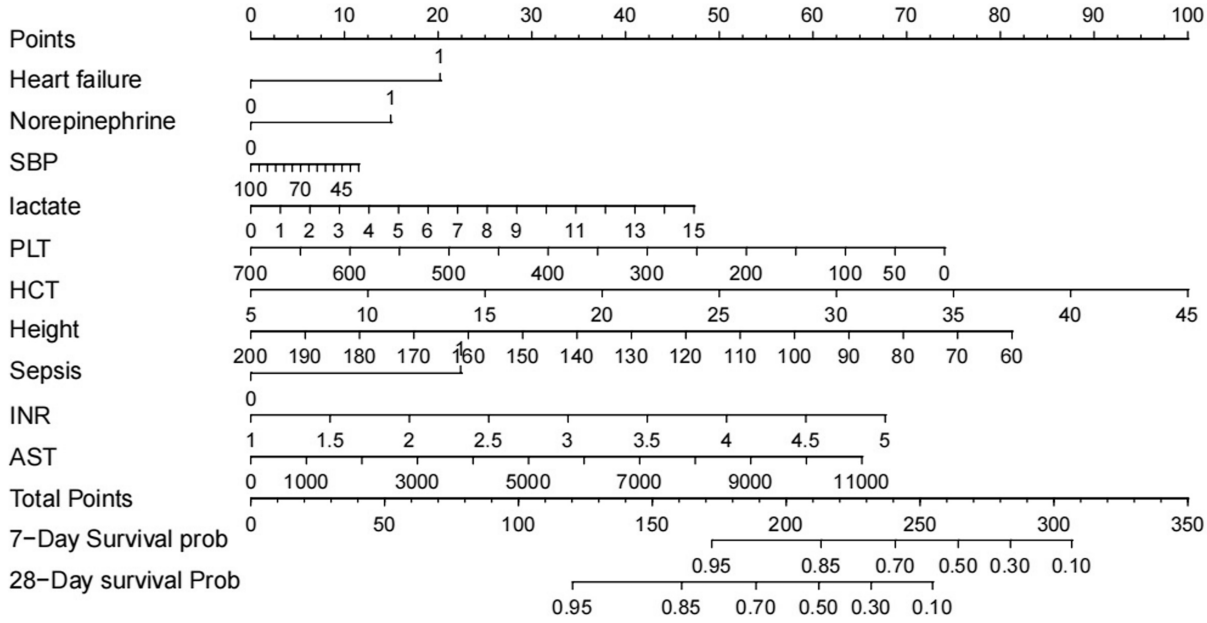

FIGURE 3
Nomogram to predict the risk of short-term mortality of patients with DIC.

results demonstrated that the model can effectively stratify patients and predict 7- and 28-day mortality (Figures 7A–F).

Discussion

Disseminated intravascular coagulation is a devastating condition and associated with high mortality. For instance, 30-day mortality rates in septic DIC have been observed to reach approximately 20%, while the 30-day mortality is up to 45% in a heterogeneous DIC population (16, 17). Therefore, we deemed 7- and 28-day mortality as suitable outcomes for this study. The assessment of the short-term death probability of overt DIC patients is an important reference for physicians to select appropriate intervention times and decide individual treatment strategies. Hence, it is necessary to analyze the risk factors affecting the outcome of overt DIC patients and establish a prognostic model to make the individualized prediction of mortality risk.

This study utilized Lasso regression to screen potential factors and establish a predictive nomogram. Lasso regression provides an advantage over univariate analysis in addressing the problem of multicollinearity among variables (18). And it was confirmed in this study that the Cox regression had a good C-index, which indicated the prediction effect of this nomogram model was better. This model integrated 10 clinically relevant variables encompassing underlying diseases (sepsis and heart failure), vasopressor use, coagulation function status (PLT and INR), hematologic condition and metabolism (HCT and Lactate), liver function status (AST), and demographic data (height and SBP). These indicators could comprehensively assess the specific situation of individuals so as to better predict the mortality risk of overt DIC. All indicators included

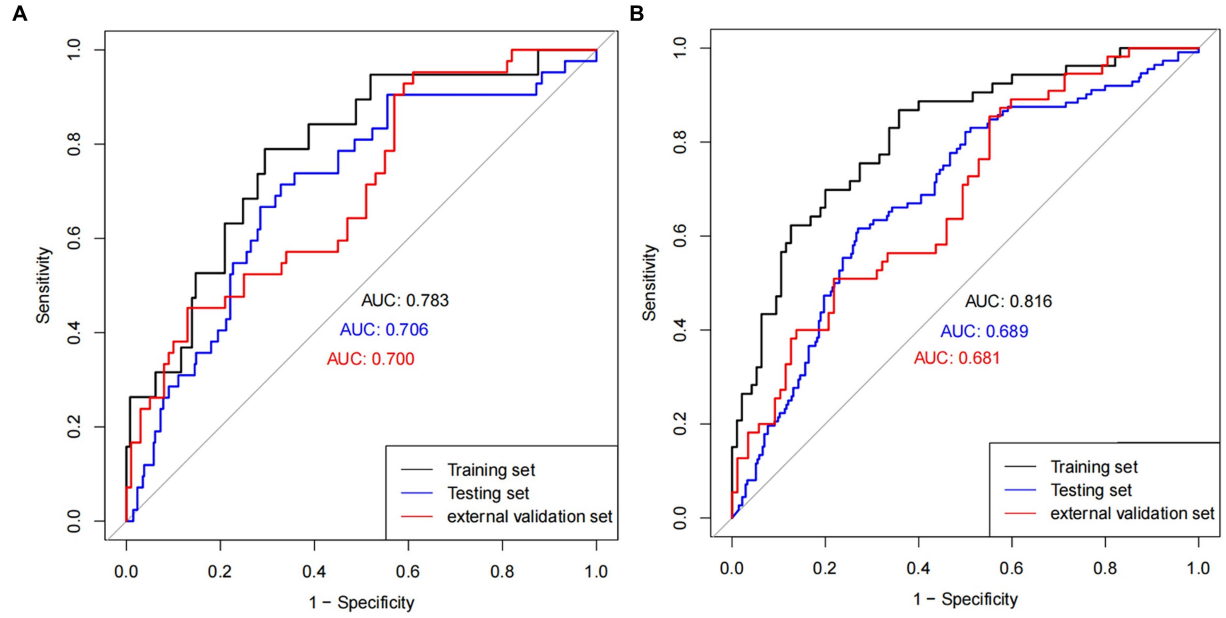

FIGURE 4
ROC plot. ROC curves in **(A)** 7 days and **(B)** 28 days in training, testing, and external validation cohort population.

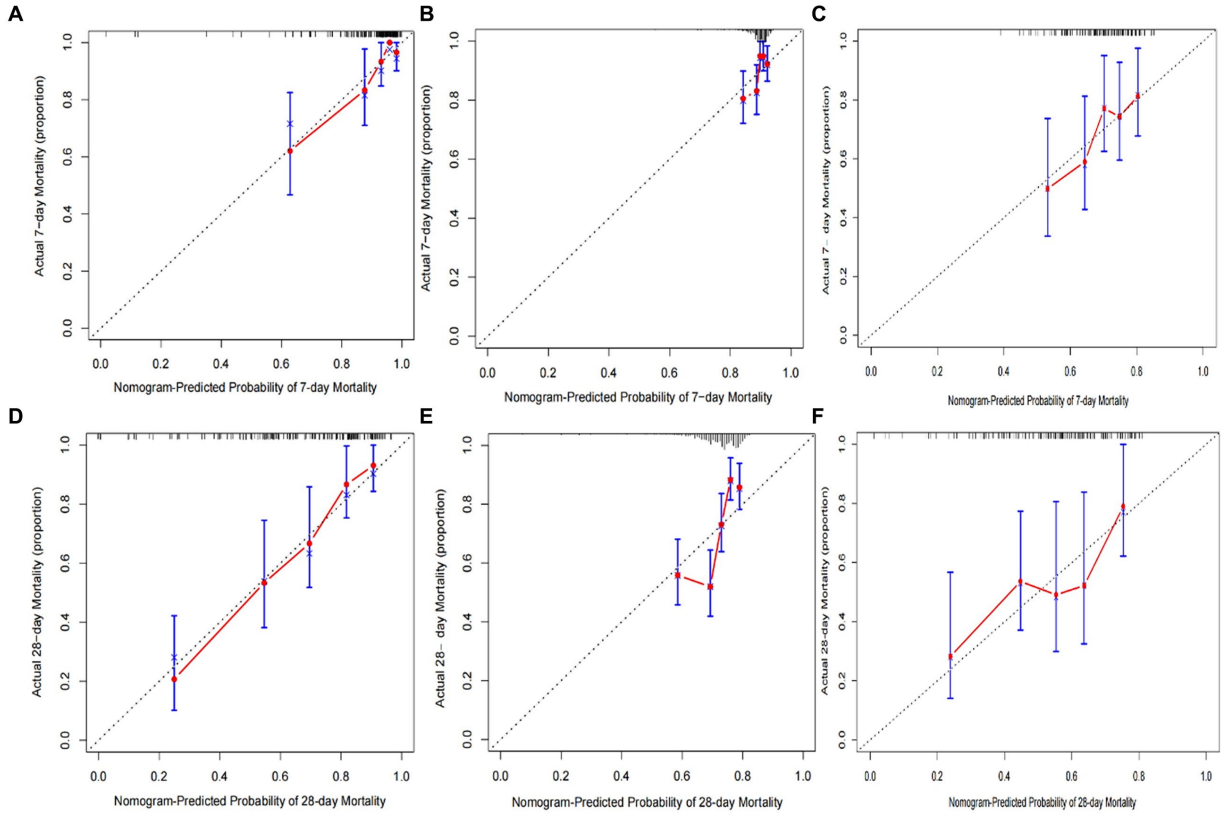

FIGURE 5
Calibration plots of predicted 7- and 28-day mortality based on Cox regression modeling in the training, testing, and external validation sets **(A–F)**.

were scored on the basis of their contribution to the prognosis, and the scores of these variables were finally summed. A vertical line at the position of total scores would be drawn to make it cross with the other two lines representing the predictive risk of death. Additionally, this model does not only incorporate patient death or survival outcomes but also the patient length of stay and survival, thus reflecting the risk

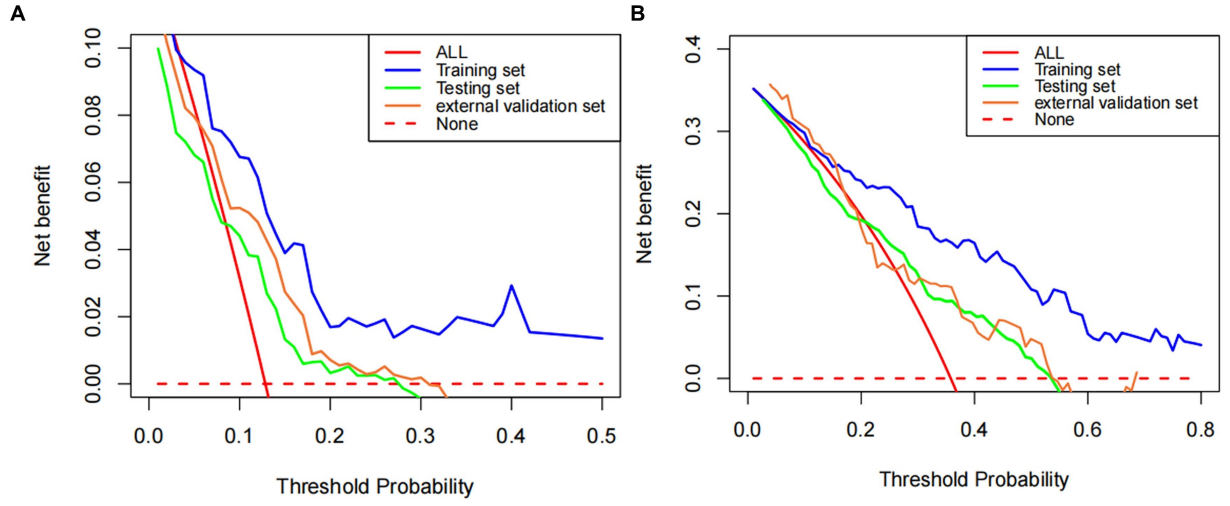

FIGURE 6
DCA of predicted 7- and 28-day mortality based on Cox regression modeling **(A,B)**.

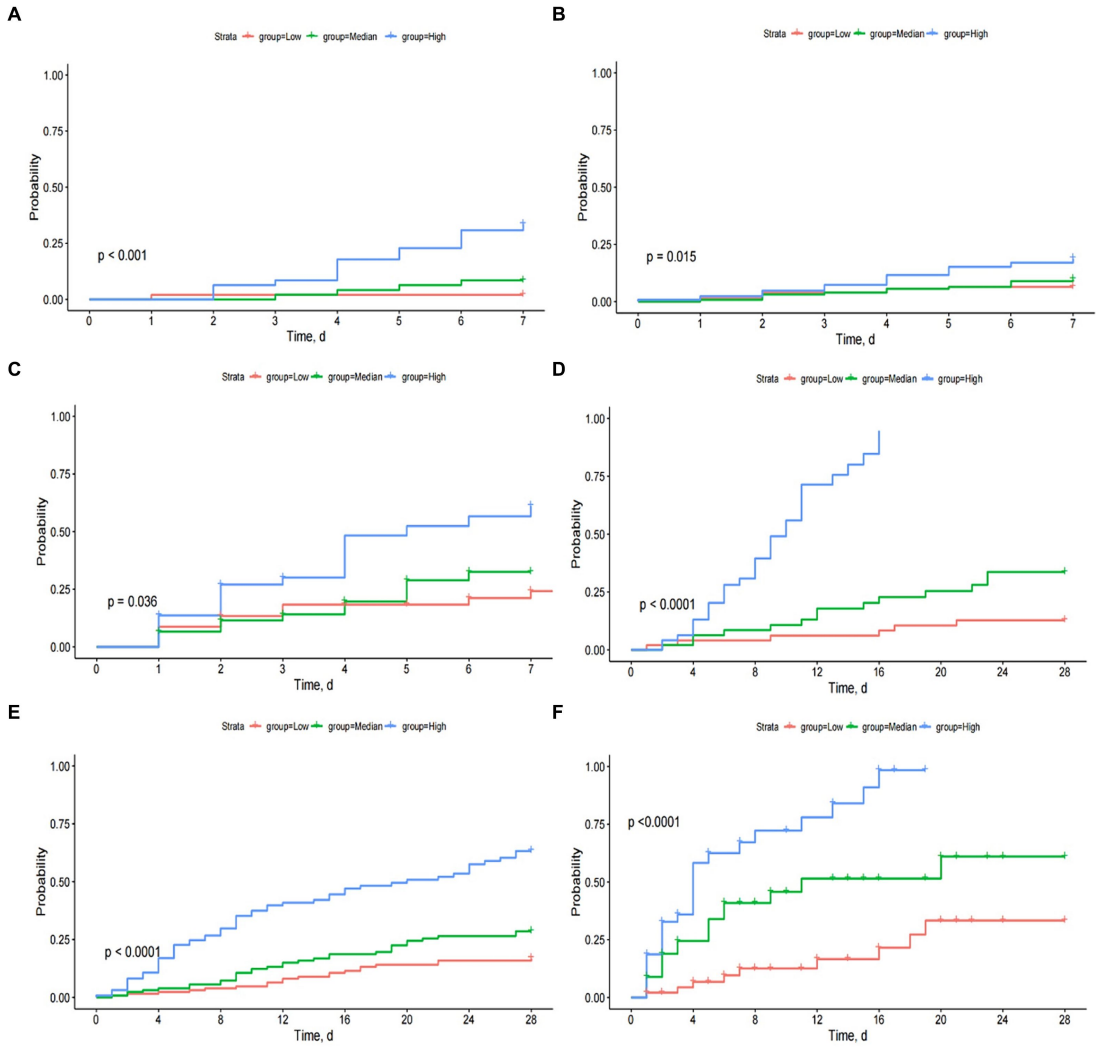

FIGURE 7
The Kaplan–Meier curves by tertiles of predicted 7- and 28-day mortality in training, testing, and external validation sets **(A–F)**.

of events at every stage following ICU admission. In other words, this prediction model could evaluate patients' survival probability in the dimension of survival time.

In terms of model evaluation, we employed DCA, calibration and C-index to estimate the predictive efficacy of our model in the training, testing and external validation cohorts. These methods are widely used to assess a model's performance (19, 20). Despite a decrease in evaluation accuracy in the testing and external validation sets, the developed Lasso-Cox regression-based prediction model remained effective in predicting 7- and 28-day prognoses. The decrease could be attributed to data heterogeneity across the three datasets. From our perspective, the ICU data from different medical centers were biased, as they were collected retrospectively and there are no restrictions on the conditions of laboratory examination equipment and the measurement methods of vital signs.

Our established predictive model facilitated clear stratification of patients into three groups based on individualized death risk. As shown in Figures 6A–F, the model could perform well in predicting short-term mortality and identify patients at high mortality risk, which indicated this nomogram had considerable predictive strength. To validate our result again, survival curves were plotted to explore whether patients with varying predicted mortality risks experienced different outcomes in the testing and external validation sets. As we expected, a higher risk of death did correspond to a poorer prognosis, reaffirming the robustness of our model. Therefore, a prediction model based on the Lasso-Cox regression model possessed a significant reference value for clinicians to identify the individual risk of mortality intuitively. In other words, the higher the mortality probability reflected in the nomogram indicated the shorter the survival time of patients, and these patients need to be treated sooner.

The present study acknowledges certain limitations. Firstly, it is a retrospective study using ICU databases, which may encompass missing data and data collection errors. Secondly, the DIC in this study is primarily attributed to sepsis, which may impact the predictive performance of our prediction model for bleeding-induced DIC. Thirdly, regarding treatment, we only examined CRRT and vasopressor use. The prognosis of overt DIC patients could also be influenced by other therapeutic measures such as plasma transfusion, cryoprecipitate transfusion, and platelet transfusion, but such data were unavailable in the MIMIC database. Finally, to validate our findings, multicenter registry and prospective studies may be necessary.

Conclusion

The present study developed a nomogram for predicting mortality in patients with overt DIC based on the Lasso-Cox regression model. This prediction model may help ICU doctors detect overt DIC patients with a higher mortality risk ahead of time, enabling timely care and allocating appropriately medical resources to increase the overall patient population survival.

Data availability statement

Publicly available datasets were analyzed in this study. The datasets generated and/or analyzed during the current study are available in the MIMIC-III and MIMIC-IV database, https://mimic.physionet.org/iii/ and https://mimic.physionet.org/iv/. The raw data from the 908th hospital supporting the conclusions of this article are available upon reasonable request so long as relevant patient confidentiality regulations can be followed.

Ethics statement

The studies involving humans were approved by the institutional review board of the 908th Hospital of Chinese PLA Logistic Support Force. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

QZ: Conceptualization, Writing – original draft, Writing – review & editing. QL: Data curation, Writing – review & editing. LZ: Data curation, Writing – review & editing. LH: Formal analysis, Writing – review & editing. NZ: Methodology, Writing – review & editing. JS: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study was supported by National Key R&D Program of China (No.2022YFC2304601). The funders were not involved in research design, data collection and manuscript preparation.

Acknowledgments

We thank the patients for participating in our study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Squizzato A, Hunt BJ, Kinasewitz GT, Wada H, Cate H, Thachil J, et al. Supportive management strategies for disseminated intravascular coagulation. *Thromb Haemost*. (2016) 115:896–904. doi: 10.1160/TH15-09-0740

2. Yamakawa K. Special issue on "disseminated intravascular coagulation: current understanding and future perspectives". *J Clin Med*. (2022) 11:3315. doi: 10.3390/jcm11123315

3. Iba T, Levi M, Thachil J, Levy JH. Disseminated intravascular coagulation: the past, present, and future considerations. *Semin Thromb Hemost*. (2022) 48:978–87. doi: 10.1055/s-0042-1756300

4. Smith L. Disseminated intravascular coagulation. *Semin Oncol Nurs*. (2021) 37:151135. doi: 10.1016/j.soncn.2021.151135

5. Iba T, Levy JH, Warkentin TE. Diagnosis and management of sepsis-induced coagulopathy and disseminated intravascular coagulation. *J Thromb Haemost*. (2019) 17:1989–94. doi: 10.1111/jth.14578

6. Costello RA, Leslie SW, Nehring SM. Disseminated intravascular coagulation In: StatPearls. Treasure Island (FL): StatPearls Publishing (2024)

7. Adelborg K, Larsen JB, Hvas AM. Disseminated intravascular coagulation: epidemiology, biomarkers, and management. *Br J Haematol*. (2021) 192:803–18. doi: 10.1111/bjh.17172

8. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine. *N Engl J Med*. (2023) 388:1201–8. doi: 10.1056/NEJMra2302038

9. Park SY. Nomogram: an analogue tool to deliver digital knowledge. *J Thorac Cardiovasc Surg*. (2018) 155:1793. doi: 10.1016/j.jtcvs.2017.12.107

10. Huang H, Liu Y, Wu M, Gao Y, Yu X. Development and validation of a risk stratification model for predicting the mortality of acute kidney injury in critical care patients. *Ann Transl Med*. (2021) 9:323. doi: 10.21037/atm-20-5723

11. Deshmukh F, Merchant SS. Explainable machine learning model for predicting GI bleed mortality in the intensive care unit. *Am J Gastroenterol*. (2020) 115:1657–68. doi: 10.14309/ajg.0000000000000632

12. Kang MW, Kim J, Kim DK, Oh KH, Joo KW, Kim YS, et al. Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. *Crit Care*. (2020) 24:42. doi: 10.1186/s13054-020-2752-7

13. Yang Y, Liang S, Geng J, Wang Q, Wang P, Cao Y, et al. Development of a nomogram to predict 30-day mortality of patients with sepsis-associated encephalopathy: a retrospective cohort study. *J Intensive Care*. (2020) 8:45. doi: 10.1186/s40560-020-00459-y

14. Lu Z, Zhang J, Hong J, Wu J, Liu Y, Xiao W, et al. Development of a nomogram to predict 28-day mortality of patients with Sepsis-induced coagulopathy: an analysis of the MIMIC-III database. *Front Med*. (2021) 8:661710. doi: 10.3389/fmed.2021.661710

15. Taylor FB Jr, Toh CH, Hoots WK, Wada H, Levi M. Towards defifinition, clinical and laboratory criteria, and a scoring system for disseminated intravascular coagulation. *Thromb Haemost*. (2001) 86:1327–30. doi: 10.1055/s-0037-1616068

16. Bakhtiari K, Meijers JC, de Jonge E, Levi M. Prospective validation of the International Society of Thrombosis and Haemostasis scoring system for disseminated intravascular coagulation. *Crit Care Med*. (2004) 32:2416–21. doi: 10.1097/01.CCM.0000147769.07699.E3

17. Gando S, Saitoh D, Ogura H, Mayumi T, Koseki K, Ikeda T, et al. Natural history of disseminated intravascular coagulation diagnosed based on the newly established diagnostic criteria for critically ill patients: results of a multicenter, prospective survey. *Crit Care Med*. (2008) 36:145–50. doi: 10.1097/01.CCM.0000295317.97245.2D

18. Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med*. (1997) 16:385–95. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

19. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. (2008) 26:1364–70. doi: 10.1200/JCO.2007.12.9791

20. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. (2006) 26:565–74. doi: 10.1177/0272989X06295361

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 20 September 2024
DOI 10.3389/fmed.2024.1400166

OPEN ACCESS

EDITED BY
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

REVIEWED BY
Juhua Hu,
University of Washington, United States
Chang Hu,
Zhongnan Hospital of Wuhan University, China

*CORRESPONDENCE
Wei Li
✉ liwei0111@cqu.edu.cn

RECEIVED 13 March 2024
ACCEPTED 09 September 2024
PUBLISHED 20 September 2024

CITATION
Hu X, Zhi S, Wu W, Tao Y, Zhang Y, Li L, Li X, Pan L, Fan H and Li W (2024) The application of metagenomics, radiomics and machine learning for diagnosis of sepsis.
Front. Med. 11:1400166.
doi: 10.3389/fmed.2024.1400166

The application of metagenomics, radiomics and machine learning for diagnosis of sepsis

Xiefei Hu[1,2], Shenshen Zhi[2,3], Wenyan Wu[1,2], Yang Tao[1,4], Yuanyuan Zhang[1,2], Lijuan Li[1,2], Xun Li[1,2], Liyan Pan[1,2], Haiping Fan[1,2] and Wei Li[1,2]*

[1]Clinical Laboratory, Chongqing University Central Hospital, School of Medicine, Chongqing University, Chongqing, China, [2]Chongqing Key Laboratory of Emergency Medicine, Chongqing Emergency Medical Center, Chongqing, China, [3]Department of Blood Transfusion, Chongqing University Central Hospital, School of Medicine, Chongqing University, Chongqing, China, [4]Intensive Care Unit, Chongqing University Central Hospital, School of Medicine, Chongqing University, Chongqing, China

Introduction: Sepsis poses a serious threat to individual life and health. Early and accessible diagnosis and targeted treatment are crucial. This study aims to explore the relationship between microbes, metabolic pathways, and blood test indicators in sepsis patients and develop a machine learning model for clinical diagnosis.

Methods: Blood samples from sepsis patients were sequenced. α-diversity and β-diversity analyses were performed to compare the microbial diversity between the sepsis group and the normal group. Correlation analysis was conducted on microbes, metabolic pathways, and blood test indicators. In addition, a model was developed based on medical records and radiomic features using machine learning algorithms.

Results: The results of α-diversity and β-diversity analyses showed that the microbial diversity of sepsis group was significantly higher than that of normal group ($p < 0.05$). The top 10 microbial abundances in the sepsis and normal groups were Vitis vinifera, *Mycobacterium canettii, Solanum pennellii, Ralstonia insidiosa, Ananas comosus, Moraxella osloensis, Escherichia coli, Staphylococcus hominis, Camelina sativa*, and *Cutibacterium acnes*. The enriched metabolic pathways mainly included Protein families: genetic information processing, Translation, Protein families: signaling and cellular processes, and Unclassified: genetic information processing. The correlation analysis revealed a significant positive correlation ($p < 0.05$) between IL-6 and Membrane transport. Metabolism of other amino acids showed a significant positive correlation ($p < 0.05$) with *Cutibacterium acnes, Ralstonia insidiosa, Moraxella osloensis*, and *Staphylococcus hominis*. Ananas comosus showed a significant positive correlation ($p < 0.05$) with Poorly characterized and Unclassified: metabolism. Blood test-related indicators showed a significant negative correlation ($p < 0.05$) with microorganisms. Logistic regression (LR) was used as the optimal model in six machine learning models based on medical records and radiomic features. The nomogram, calibration curves, and AUC values demonstrated that LR performed best for prediction.

Discussion: This study provides insights into the relationship between microbes, metabolic pathways, and blood test indicators in sepsis. The developed machine learning model shows potential for aiding in clinical diagnosis. However, further research is needed to validate and improve the model.

KEYWORDS

sepsis, metagenomics, blood test indicators, radiomics, machine learning

1 Introduction

Sepsis is a severe organ dysfunction endangering life, resulting from impaired host function triggered by infection (1, 2). Epidemiological survey data show (3) that sepsis is characterized by a high incidence and mortality, and its incidence has been increasing in recent years. Each year, sepsis impacts over 30 million individuals globally and leads to around 6 million deaths (4). According to domestic statistics, around 20.6% of ICU patients experience sepsis, the 90-day overall mortality is 35.5%, and the rate is as high as 51.94% combined with septic shock (5). Sepsis is a serious threat to physical health.

Sepsis has a variety of clinical manifestations, including fever, increased heart rate, shortness of breath, hypotension, changes in consciousness, etc. (6). Additionally, patients may experience systemic multiple organ dysfunction such as pneumonia, acute respiratory distress syndrome (ARDS), renal impairment, and cardiac insufficiency. Severe sepsis can lead to shock and even death. Therefore, an early diagnosis is particularly important for the treatment of sepsis. It is diagnosed based on the evidence of infection and manifestations of systemic inflammation (7). Currently, the diagnosis of sepsis is mainly based on blood culture (8), and white blood cell (WBC) count, classification, C-reactive protein (CRP), and procalcitonin precursor (PCT) are determined for auxiliary diagnosis (9).

With the advancement of big data analysis, genomics research, and biomarker research, the pathogenesis of sepsis will be further clarified (10), which will facilitate molecular biology-oriented diagnosis of sepsis, improve the sensitivity and specificity of diagnosis, and formulate more appropriate diagnostic criteria to reflect the infection and uncontrolled response of the body, thereby further contributing to the early identification and diagnosis of sepsis. It can also reflect the characteristics of dynamic changes in the disease, provide conditions for accurate treatment of sepsis, and improve patient survival. This project aimed to obtain a comprehensive bacterial infectious sepsis-specific pathogenic microorganism through comparative research and whole genome sequencing technology on the metagenomic next-generation sequencing (mNGS) platform, and to establish a prediction model of sepsis integrating radiomics and machine learning algorithms, hoping to provide an implication for its clinical diagnosis.

2 Materials and methods

2.1 Data analysis

Metagenomic sequencing was performed on 25 patients with sepsis, and nine samples from the normal group were used for metagenomic sequencing. After the raw data were exported, low-quality reads were filtered out and the obtained valid data were used for subsequent analyses. The host sequences were first removed, and sequence alignment was used to infer the species composition and relative abundance of the microbial community, followed by plotting of the species abundance profile. The function, consanguinity, and metabolic pathways of each gene were determined by comparing and annotating the genes to a known database. Through integration and statistical analyses of the annotated results, functional modules and metabolic pathways involved in the microbial community were identified, and their roles in the ecosystem were explored.

2.2 Radiomics analysis

The region of interest (ROI)/volume of interest (VOI) usually refers to a lesion that was manually segmented using 3D Slicer v5.1.0. Quantitative features were extracted from the digital images, which were stored in a shared database. The data were mined, and hypotheses were generated or validated. A plugin of the 3D Slicer v5.1.0 software PyRadiomics was used to perform radiomic feature extraction from each ROI. The plugin automatically extracted 851 radioactive features from each ROI. It includes first-order statistical features (energy, entropy, mean, standard deviation, maximum, etc.), shape-and size-based features (maximum 3D diameter, volume, superficial area, etc.), texture features (grayscale co-occurrence matrix GLCM and grayscale run matrix GLRLM), and wavelet-based transform features.

2.3 Construction of a machine learning model

In the training set, the selection was made by 10-fold cross-validation and grid search 10 times, and six classification algorithms (LR: logistic regression; RF: random forest; adaboost: adaptive enhancement; SVM: support vector machine; NB: naive Bayes; GBDT: gradient enhancement decision tree) of the corresponding cohort were obtained. The six classification algorithms completed by the training were called to train the data and build the model, and the prediction results of the different models were obtained. The average value of the multiple accuracies was used as the final model score, and the final model was generated simultaneously. A receiver operating characteristic (ROC) curve was plotted for each training model and test result, and the area under the curve (AUC), accuracy, sensitivity, recall rate, and specificity were calculated.

2.4 Statistical analysis

Statistical analyses were performed using R software (V4.2.2). The measured data were tested for homogeneity and normality of variance. For measurement data following a normal distribution, the mean ± standard deviation was utilized, with t-tests employed for inter-group comparisons. Count data were presented as percentages, and differences between groups were assessed using χ^2 tests, with significance set at $p < 0.05$. Correlation analysis of microorganisms, metabolic pathways, and blood test-related indicators was performed using the Spearman's correlation coefficient. Based on R software (version 4.0.3) and R studio platform, Lasso feature dimensionality reduction, logistic regression model construction, ROC curve plotting, calibration curve plotting, nomographic chart and clinical decision analysis curve were performed using the corresponding software package.

3 Results

3.1 Microbiome composition analysis

The α-diversity analysis showed that ACE, Chao1, Shannon, and Simpson indices of sepsis patients were significantly higher than

those of normal group ($p < 0.05$), indicating that the microbial abundance and diversity of sepsis patients were elevated (Figure 1A). The β-diversity results showed small sample differences between the two groups (Figure 1B). Based on the species annotation results, the top 20 species in terms of abundance were selected at the species level for each sample in the sepsis and normal groups to plot the relative abundance histogram (Figure 2A). The top 10 microbial genera in both groups included *Vitis vinifera*, *Mycobacterium canettii*, *Solanum pennellii*, *Ralstonia insidiosa*, *Ananas comosus*, *Moraxella osloensis*, *Escherichia coli*, *Staphylococcus hominis*, *Camelina sativa*, and *Cutibacterium acnes*. The enriched metabolic pathways were mainly protein families: genetic information processing, translation, protein families: signaling and cellular processes, and unclassified: genetic information processing (Figure 2B).

3.2 Correlation analysis

The analyses revealed (Figures 3A–C) a correlation between metabolic pathways, blood detection indicators, and pathogenic microorganisms, with a significant positive correlation ($p < 0.05$) between IL-6 and membrane transport. Metabolism of other amino acids showed a significant positive correlation ($p < 0.05$) with

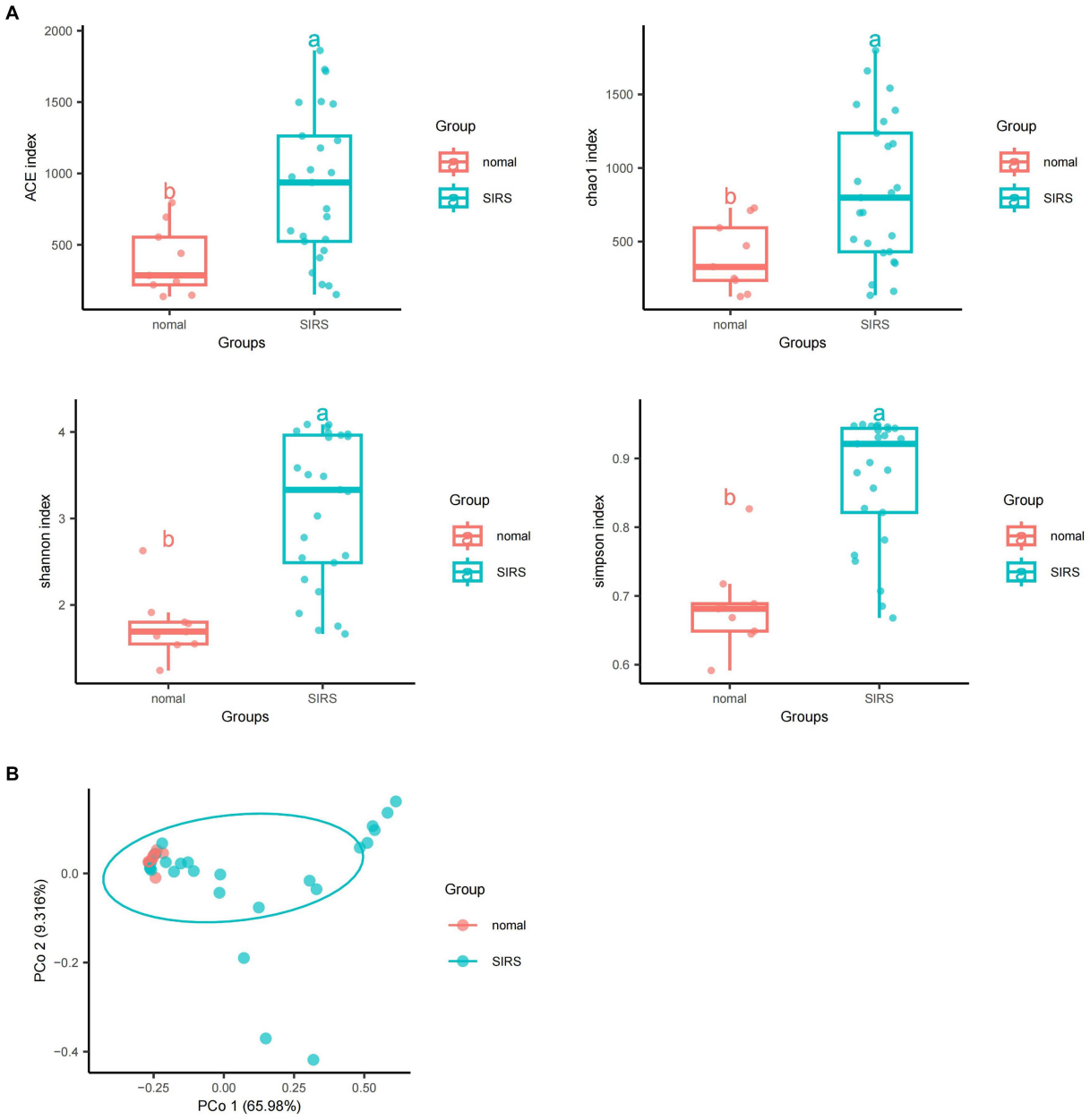

FIGURE 1
Diversity and abundance analyses of pathogenic microorganisms. **(A)** α-diversity analyses showed that ACE, Chao1, Shannon, and Simpson indices of sepsis patients were significantly higher than those of normal group ($p < 0.05$). **(B)** β-diversity analyses showed that the sample difference between the two groups was small.

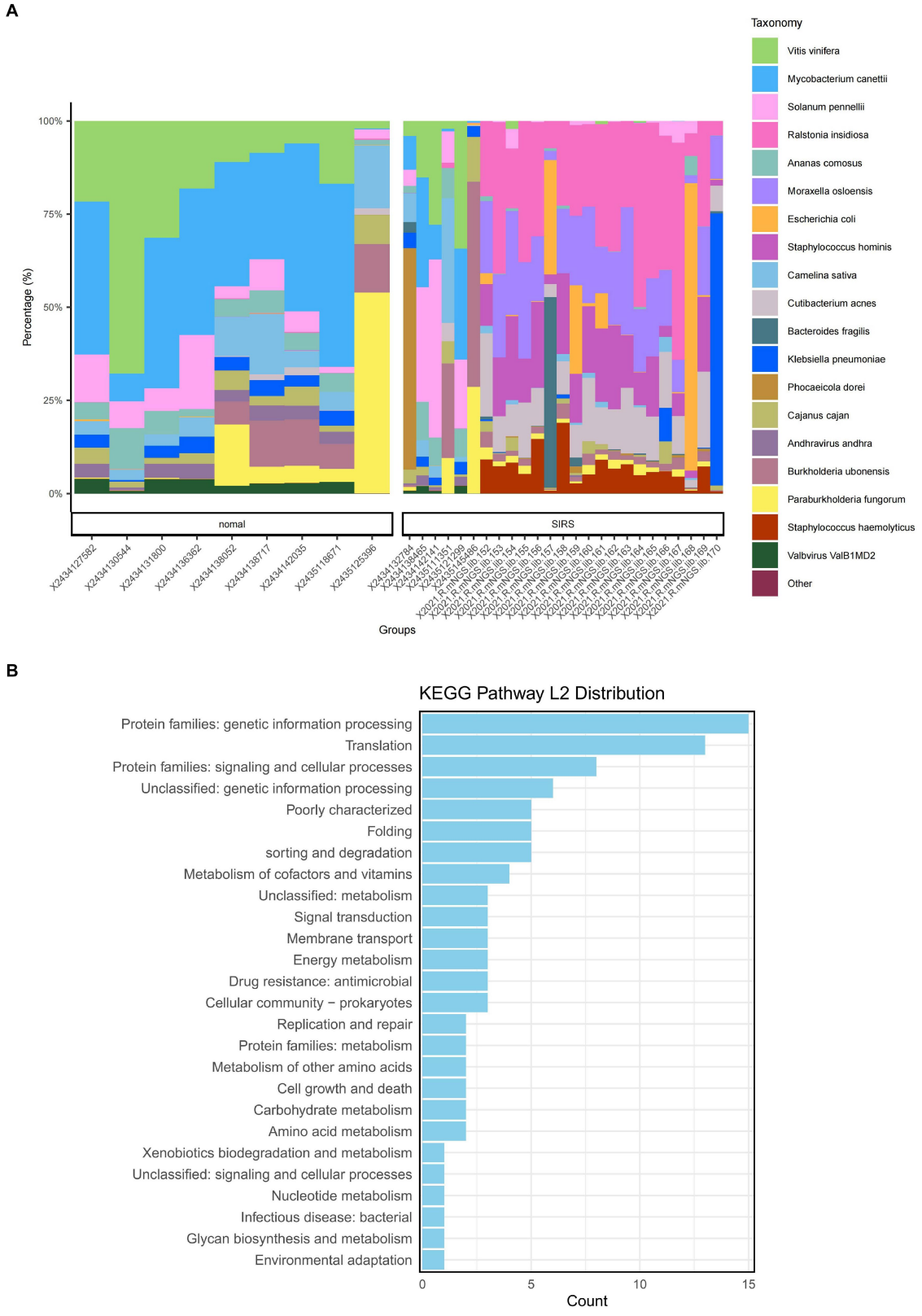

FIGURE 2
Species annotation results and KEGG metabolic pathway enrichment. **(A)** Bar chart of the relevant abundance of the top 20 species in the sepsis and normal groups. **(B)** Enriched metabolic pathways.

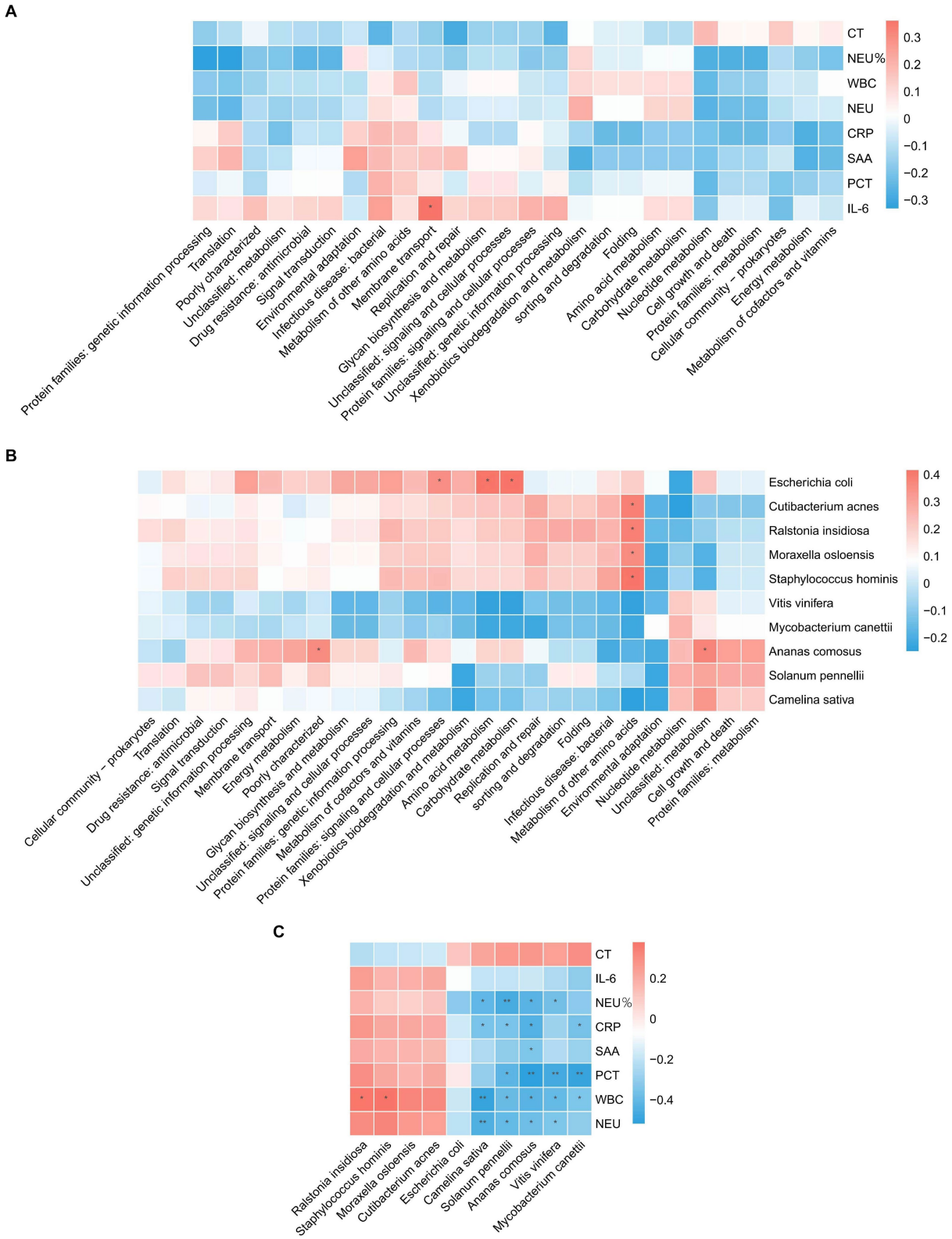

FIGURE 3
Correlation analysis. **(A)** Correlation between blood test indicators and metabolic pathways. **(B)** Correlation of pathogenic microorganisms and metabolic pathways. **(C)** Correlation between blood test indicators and pathogenic microorganisms.

Cutibacterium acnes, *Ralstonia insidiosa*, *Moraxella osloensis*, and *Staphylococcus hominis*. *Ananas comosus* showed a significant positive correlation ($p < 0.05$) with poorly characterized and unclassified: metabolism. *Escherichia coli* showed a significant positive correlation ($p < 0.05$) with protein families: signaling and cellular processes, amino acid metabolism, and carbohydrate metabolism. WBC showed a significant positive correlation ($p < 0.05$) with *Ralstonia insidiosa*, and *Staphylococcus hominis*. *Camelina sativa* showed a significant negative correlation ($p < 0.05$) with NEU%, CRP, WBC, and NEU. *Solanum pennelli* showed a significant negative correlation ($p < 0.05$) with NEU%, CRP, PCT, WBC, and NEU. *Ananas comosus* showed a significant negative correlation ($p < 0.05$) with NEU%, CRP, SAA, PCT, WBC, and NEU. *Vitis vinifera* showed a significant negative correlation ($p < 0.05$) with NEU%, PCT, WBC, and NEU. *Mycobacterium canettii* showed a significant negative correlation ($p < 0.05$) with CRP, PCT, and WBC.

3.3 Construction of a machine learning model

Based on radiomics, five important image features were finally screened using LASSO-Cox regression and 10-fold cross validation including original-shape-sphericity, original-firstorder-10Percentile, wavelet-HHL-glcm-InverseVariance, wavelet-HHH-glszm-ZoneEntropy, and wavelet-LLL-gldm-LargeDependenceLow GrayLevelEmphasis (Figure 4A). Utilizing the previously outlined features, model performance was assessed through the ROC curve, revealing an AUC = 0.791 for the model's ROC curve (Figure 4B). Significant features were extracted from the medical records using LASSO, including occupancy, inflammation, blood lipids, prognosis, advention, and discharge (Figure 5A). Drawing from the characteristics outlined earlier, the model's performance was analyzed via the ROC curve, indicating an AUC = 0.873 for the model's ROC curve (Figure 5B).

The model performance was evaluated and compared using the following seven metrics: AUC, sensitivity, specificity, accuracy, precision, recall (F1), and prAUC. In comparison, among all the machine learning models, the LR model performed the best in classification (AUC value was 0.897 in the training set), and the sensitivity, specificity, accuracy, precision, F1, and prAUC values were also the highest in the LR model (Table 1); therefore, the optimal model was LR. We visualized the LR model and plotted the nomogram for easy clinical application (Figure 6). A calibration curve was utilized to assess the model's performance, and it can be seen that the error between the predicted values and the real values of the prediction model was small, and the result was highly accurate. The AUC = 0.879 for the model's ROC curve, demonstrating that LR had the best predictive power (Figure 7).

4 Discussion

Sepsis is a serious infectious disease, and its pathogenesis involves multiple factors such as the immune system, inflammatory mediators, and vascular endothelium. The clinical manifestations of sepsis are diverse and it is necessary to comprehensively consider infection control, inflammation regulation, and organ support during treatment. Studies have shown that there may be potential benefits in the treatment and prevention of sepsis through the regulation of intestinal flora and the use of microbial preparations. The microbiota in the gastrointestinal tract of the human body is composed of trillions of bacteria that form the mucosal barrier of the intestine and are present in different proportions and numbers in different parts of the intestine, thus playing a defensive and protective role (11). Dysregulation of the microbiome or a reduction in microbial diversity is associated with altered immune responses. Sepsis affects the composition of the intestinal microbiome, which is characterized by loss of diversity, reduced abundance of key

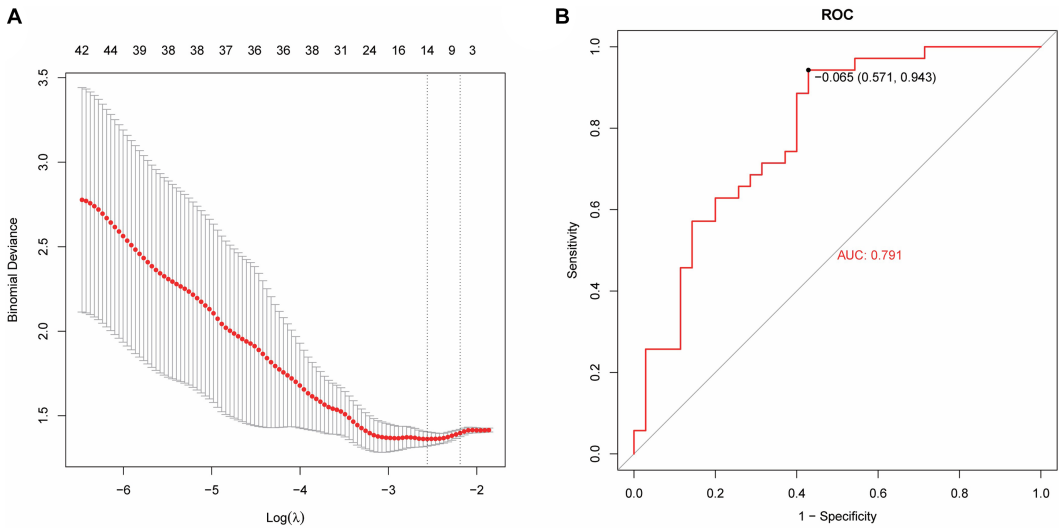

FIGURE 4
Machine learning model establishment based on radiomics. **(A)** Five important image features screened using LASSO-Cox regression and 10-fold cross validation. **(B)** The AUC value of the model ROC curve is 0.791.

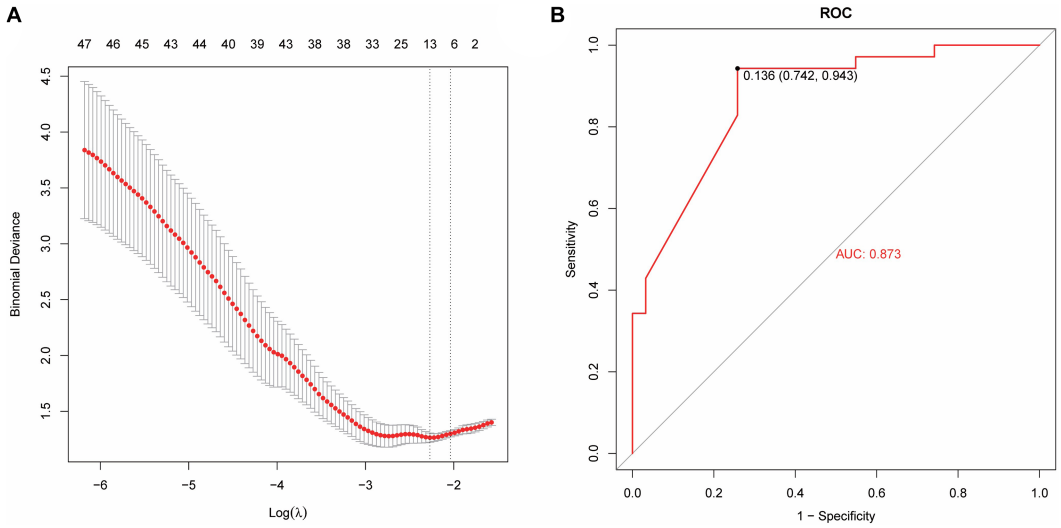

FIGURE 5
A machine learning model based on medical record data. **(A)** Lasso extraction of seven features from the medical records. **(B)** The AUC value of the model ROC curve is 0.873.

TABLE 1 **Model construction and evaluation.**

Model	AUC	Sensitivity	Specificity	Accuracy	Precision	F1	prAUC
adaboost	0.659	0.635	0.630	0.632	0.665	0.695	0.622
GBDT	0.559	0.410	0.685	0.547	0.613	0.609	0.742
LR	0.897	0.800	0.850	0.825	0.842	0.820	0.776
NB	0.775	0.600	0.845	0.723	0.824	0.751	0.615
RF	0.801	0.670	0.670	0.670	0.711	0.716	0.602
SVM	0.785	0.645	0.655	0.650	0.692	0.713	0.620

symbiotic bacteria such as *Faecium* and *Gastrococcus*, weakened colonization capacity of *Proteus* and other conditioning pathogens, and overpropagation, growing as a dominant bacterial group (12). A prospective cohort study of over 200 preterm infants in 2019 found that increased bacterial diversity and anaerobic colonization of the neonatal gut microbiome protected newborns from sepsis (13). The results of metagenomic sequencing technology in this study showed that the microbial abundance and diversity in patients with sepsis were significantly higher than those in the normal group ($p < 0.05$). The top 10 microbial abundances in the sepsis and normal groups were *Vitis vinifera*, *Mycobacterium canettii*, *Solanum pennellii*, *Ralstonia insidiosa*, *Ananas comosus*, *Moraxella osloensis*, *Escherichia coli*, *Staphylococcus hominis*, *Camelina sativa*, and *Cutibacterium acnes*, which mainly include protein families: genetic information processing, translation, protein families: signaling and cellular processes, and unclassified: genetic information processing.

Nowadays, traditional biomarkers such as CRP, PCT and IL-6 are widely used in the diagnosis and evaluation of sepsis (9). Inflammation serves as a defensive reaction aimed at eliminating invading pathogens, mitigating detrimental stimuli, and initiating tissue healing. The inflammatory response is activated when innate immune cells detect antigenic structures via pattern recognition receptors that identify molecular patterns associated with pathogens or damage-related molecular patterns (14). Excessive inflammatory response contributes to tissue damage and organ dysfunction in individuals with sepsis. Neutrophils produce reactive oxygen species through chemotaxis and phagocytosis, leading to widespread inflammation and increased microvascular permeability. Excessive inflammation causes a large number of neutrophil degranulation and proteolytic enzyme release, resulting in systemic and local endothelial damage (15). Therefore, neutrophils reflect the inflammatory response and immune status of the body during sepsis. IL-6 not only activates neutrophils but also delays phagocytosis of senescent and dysfunctional neutrophils, thereby exacerbating the production of post-traumatic inflammatory mediators and promoting the onset of post-traumatic systemic inflammatory response syndrome (16). Normal human serum IL-6 levels are very low, but when the body has an inflammatory response, IL-6 levels are significantly increased, and its level are increased earlier than other acute stage proteins, so it is helpful for the early diagnosis

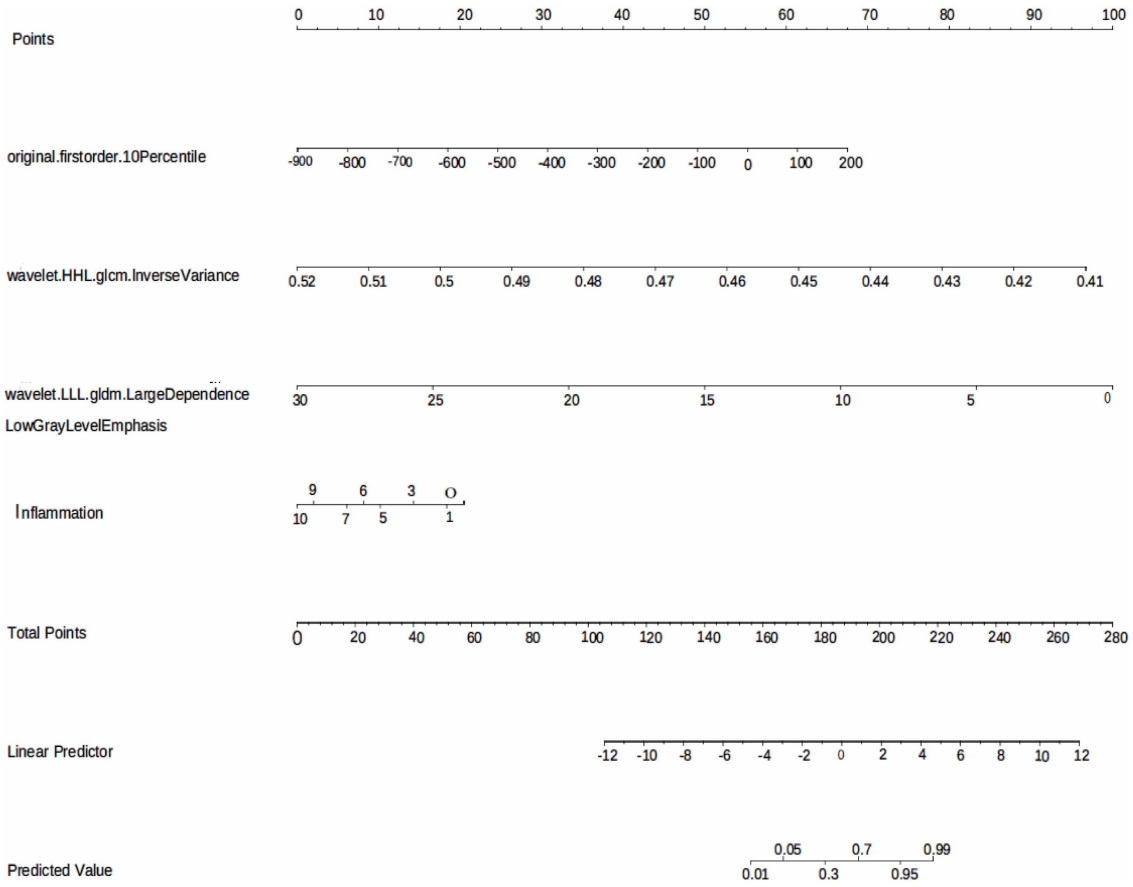

FIGURE 6
The LR model based on the features extracted from medical records and radiomics is visualized using nomogram.

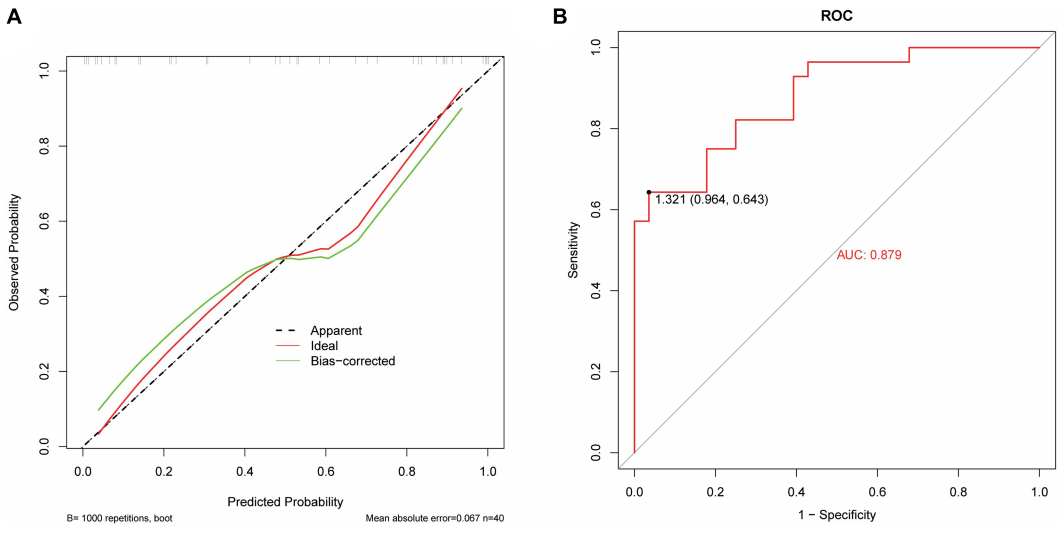

FIGURE 7
Performance evaluation of the LR model. **(A)** The calibrate calibration curves. **(B)** The ROC curve.

of emergency infection patients and can reflect the change in the disease condition (17). CRP is an acute phase protein produced by hepatocytes under the action of IL-6, and the serum CRP level in healthy people is very low; however, it can be significantly increased during bacterial infection, tissue damage, or stress, and it is significantly increased at the early stage of inflammation,

which is a sensitive indicator of bacterial infection (18, 19). PCT is a hormone-free calcitonin peptide. Under normal physiological conditions, serum PCT levels are extremely low (20). However, under the action of inflammatory cytokines, the liver, kidneys, muscles, adipose tissue, and other solid organs of septic shock patients produce a large amount of PCT, resulting in a significant increase in the blood PCT levels of patients (20). Therefore, it can be used to diagnose, evaluate, and predict infectious diseases by measuring serum PCT levels in patients. Serum amyloid A (SAA) can be used as a sensitive indicator to reflect body infection and inflammation management, playing a crucial role in the adjunct diagnosis of infectious diseases (21). WBC is a common indicator of systemic inflammation, and relevant studies have reported that WBC count can diagnose early sepsis and is closely related to its prognosis (22). By analyzing the correlation between sepsis detection indicators and pathogenic microorganisms, it is possible to improve the diagnostic accuracy of sepsis, formulate more effective treatment options, and evaluate the prognosis of patients, which can help improve the recovery and survival rates of patients with sepsis. In this study, the correlation between sepsis-related inflammatory indicators, pathogenic microorganisms, and metabolic pathways was analyzed. The correlation analysis revealed a significant positive correlation ($p < 0.05$) between IL-6 and membrane transport. Metabolism of other amino acids showed a significant positive correlation ($p < 0.05$) with *Cutibacterium acnes*, *Ralstonia insidiosa*, *Moraxella osloensis*, and *Staphylococcus hominis*. *Ananas comosus* showed a significant positive correlation ($p < 0.05$) with poorly characterized and unclassified: metabolism. Blood test-related indicators showed a significant negative correlation ($p < 0.05$) with microorganisms.

Radiomics is the high-throughput extraction of a large amount of information from medical images to achieve lesion segmentation, feature extraction, and model establishment. It assists clinicians in making the most accurate diagnosis by conducting deeper mining, prediction, and analysis of massive amounts of image data information. It can also be intuitively understood as converting visual image information into deep quantitative features (23). In recent years, with the enhancement of computer data processing capabilities, improvement of image recognition technology, and continuous improvement of machine learning algorithms, in-depth data information of massive medical images can be mined and analyzed (24, 25). This capability has been applied to assess the severity of diseases (26), construct disease monitoring systems (automated alerting system) (27), and enable early prediction of diseases (28–31). Zhang et al. (32) found that the establishment of an XGBoost prediction model can predict sepsis-associated delirium earlier and is suitable for patients who are difficult to evaluate using traditional methods. Ge et al. (29) developed a machine learning model to accurately predict the occurrence of sepsis-associated acute brain injury and provide a basis for early intervention and treatment. In this study, the AUC value of the model based on the features extracted by radiomics was 0.791, the AUC of the medical record data features is 0.873; the AUC value of the logistic regression model based on the features extracted from medical records and radiomics was 0.879. It is proven that the model prediction ability is better when the two features are combined.

In this study, the blood samples of patients with sepsis were metagenomically sequenced to explore the complex relationship between microorganisms, metabolic pathways and blood test indicators, which provided a new idea and method for the diagnosis of sepsis. At the same time, a machine learning model based on medical records and radiomic features was developed for clinical diagnosis of sepsis, which filled some gaps in this field. The sample size of this study was small; the results may have been affected by sample selection, and further expansion of the sample size is required to verify the stability of the conclusions. Although the establishment of machine learning models has achieved certain prediction capabilities, they still need to be verified and optimized on larger datasets.

5 Conclusion

Taken together, microbial abundance and diversity were elevated in the sepsis group. Correlation analysis of blood test-related indicators with microbial and metabolic pathways showed significant correlations, which might contribute to further clinical diagnosis and treatment. The LR prediction model based on radiomics and medical record data had good diagnostic efficacy and calibration for identifying patients with sepsis, which is a potential auxiliary tool for clinical decision-making.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Chongqing Emergency Medical Center and Chongqing University Central Hospital (Approval Ethics Review No. 2021-73). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XH: Conceptualization, Formal analysis, Methodology, Project administration, Resources, Writing – original draft, Data curation, Investigation, Writing – review & editing. SZ: Data curation, Investigation, Writing – original draft, Writing – review & editing, Supervision. WW: Investigation, Writing – original draft, Writing – review & editing, Methodology. YT: Investigation, Methodology, Writing – original draft, Writing – review & editing. YZ: Investigation, Methodology, Writing – original draft, Writing – review & editing. LL: Investigation, Methodology, Writing – original draft, Writing – review & editing. XL: Investigation, Methodology, Writing – original draft, Writing – review & editing. LP: Data curation, Writing – review & editing. HF: Data curation, Writing – review & editing. WL: Conceptualization, Formal analysis, Funding

acquisition, Methodology, Project administration, Resources, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant Number: KJZD-M202300101), Emergency Medicine Chongqing Municipal Key Laboratory Talent Innovation Development Joint Fund Project (Grant Number: 2024RCCX06), and Chongqing Advanced Medical Talents Program for Young and Middle Aged People (Grant Number: ZQNYXGDRCGZS2019008).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287

2. Yang J, Hao S, Huang J, Chen T, Liu R, Zhang P, et al. The application of artificial intelligence in the management of sepsis. *Med Rev*. (2021) 3:369–80. doi: 10.1515/mr-2023-0039

3. Huang M, Cai S, Su J. The pathogenesis of sepsis and potential therapeutic targets. *Int J Mol Sci*. (2019) 20:5376. doi: 10.3390/ijms20215376

4. Chicco D, Jurman G. Survival prediction of patients with sepsis from age, sex, and septic episode number alone. *Sci Rep*. (2020) 10:17156. doi: 10.1038/s41598-020-73558-3

5. Xie J, Wang H, Kang Y, Zhou L, Liu Z, Qin B, et al. The epidemiology of sepsis in Chinese ICUs: a national cross-sectional survey. *Crit Care Med*. (2020) 48:e209–18. doi: 10.1097/CCM.0000000000004155

6. Guan W, Xu J, Shi Y, Wang X, Gu S, Xie L. VNN1 as a potential biomarker for sepsis diagnosis and its implications in immune infiltration and tumor prognosis. *Front Med*. (2023) 10:1236484. doi: 10.3389/fmed.2023.1236484

7. Zhang Y, Han J. Rethinking sepsis after a two-year battle with COVID-19. *Cell Mol Immunol*. (2022) 19:1317–8. doi: 10.1038/s41423-022-00909-7

8. Shi J, Lu Z-Q, Lin Q-M, Zeng W, Gu PJ, Yu Q, et al. The role of albumin in the diagnosis of neonatal sepsis over the last 11 years: a retrospective study. *J Inflamm Res*. (2023) 16:2855–63. doi: 10.2147/JIR.S414611

9. Westerdijk K, Simons KS, Zegers M, Wever PC, Pickkers P, de Jager CPC. The value of the neutrophil-lymphocyte count ratio in the diagnosis of sepsis in patients admitted to the intensive care unit: a retrospective cohort study. *PLoS One*. (2019) 14:e0212861. doi: 10.1371/journal.pone.0212861

10. Kojic D, Siegler BH, Uhle F, Lichtenstern C, Nawroth PP, Weigand MA, et al. Are there new approaches for diagnosis, therapy guidance and outcome prediction of sepsis? *World J Exp Med*. (2015) 5:50–63. doi: 10.5493/wjem.v5.i2.50

11. Xie T, Lv J, Wei J, Ye L, Zhang D. Research on the mechanisms and treatment of dysbiosis of intestinal microbiota in sepsis. *J Med Res*. (2018) 47:4. doi: 10.11969/j.issn.1673-548X.2018.08.043

12. McDonald D, Ackermann G, Khailova L, Baird C, Heyland D, Kozar R, et al. Extreme dysbiosis of the microbiome in critical illness. *mSphere*. (2016) 1:e00199. doi: 10.1128/mSphere.00199-16

13. Graspeuntner S, Waschina S, Künzel S, Twisselmann N, Rausch TK, Cloppenborg-Schmidt K, et al. Gut dysbiosis with bacilli dominance and accumulation of fermentation products precedes late-onset sepsis in preterm infants. *Clin Infect Dis*. (2019) 69:268–77. doi: 10.1093/cid/ciy882

14. Slaats J, ten Oever J, van de Veerdonk FL, Netea MG. IL-1β/IL-6/CRP and IL-18/ferritin: distinct inflammatory programs in infections. *PLoS Pathog*. (2016) 12:e1005973. doi: 10.1371/journal.ppat.1005973

15. Sheats MK. A comparative review of equine SIRS, Sepsis, and neutrophils. *Front Vet Sci*. (2019) 6:69. doi: 10.3389/fvets.2019.00069

16. Tang S, Wan M, Huang W, Stanton RC, Xu Y. Maresins: specialized proresolving lipid mediators and their potential role in inflammatory-related diseases. *Mediat Inflamm*. (2018) 2018:1–8. doi: 10.1155/2018/2380319

17. Wu Y, Wang G, Huang Z, Yang B, Yang T, Liu J, et al. Diagnostic and therapeutic value of biomarkers in urosepsis. *Ther Adv Urol*. (2023) 15:17562872231151852. doi: 10.1177/17562872231151852

18. Kudlinski B, Zgoła D, Stolińska M, Murkos M, Kania J, Nowak P, et al. Systemic inflammatory predictors of in-hospital mortality in COVID-19 patients: a retrospective study. *Diagnostics*. (2022) 12:859. doi: 10.3390/diagnostics12040859

19. Ruan G-T, Xie H-L, Gong Y-Z, Ge YZ, Zhang Q, Wang ZW, et al. Prognostic importance of systemic inflammation and insulin resistance in patients with cancer: a prospective multicenter study. *BMC Cancer*. (2022) 22:700. doi: 10.1186/s12885-022-09752-5

20. Hegazy MA, Omar AS, Samir N, Moharram A, Weber S, Radwan WA. Amalgamation of procalcitonin, C-reactive protein, and sequential organ failure scoring system in predicting sepsis survival. *Anesth Essays Res*. (2014) 8:296–301. doi: 10.4103/0259-1162.143115

21. Lv Y-J, Hu Q-L, Huang R, Zhang L, Wu L-F, Fu S. The diagnostic and therapeutic value of the detection of serum amyloid a and C-reactive protein in infants with rotavirus diarrhea. *Int J Gen Med*. (2021) 14:3611–7. doi: 10.2147/IJGM.S319915

22. Rimmer E, Garland A, Kumar A, Doucette S, Houston BL, Menard CE, et al. White blood cell count trajectory and mortality in septic shock: a historical cohort study. *Can J Anaesth*. (2022) 69:1230–9. doi: 10.1007/s12630-022-02282-5

23. Chen C, Geng Q, Song G, Zhang Q, Wang Y, Sun D, et al. A comprehensive nomogram combining CT-based radiomics with clinical features for differentiation of benign and malignant lung subcentimeter solid nodules. *Front Oncol*. (2023) 13:1066360. doi: 10.3389/fonc.2023.1066360

24. Ochiai K, Ozawa T, Shibata J, Ishihara S, Tada T. Current status of artificial intelligence-based computer-assisted diagnosis systems for gastric cancer in endoscopy. *Diagnostics*. (2022) 12:3153. doi: 10.3390/diagnostics12123153

25. Zhang Z, Kashyap R, Su L, Meng Q. Editorial: clinical application of artificial intelligence in emergency and critical care medicine, volume III. *Front Med*. (2022) 9:1075023. doi: 10.3389/fmed.2022.1075023

26. Liu X, Hu P, Yeung W, Zhang Z, Ho V, Liu C, et al. Illness severity assessment of older adults in critical illness using machine learning (ELDER-ICU): an international multicentre study with subgroup bias evaluation. *Lancet Digit health*. (2023) 5:e657–67. doi: 10.1016/s2589-7500(23)00128-0

27. Zhang Z, Chen L, Xu P, Wang Q, Zhang J, Chen K, et al. Effectiveness of automated alerting system compared to usual care for the management of sepsis. *npj Digit Med*. (2022) 5:101. doi: 10.1038/s41746-022-00650-5

28. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. *Front Med*. (2021) 8:607952. doi: 10.3389/fmed.2021.607952

29. Ge C, Deng F, Chen W, Ye Z, Zhang L, Ai Y, et al. Machine learning for early prediction of sepsis-associated acute brain injury. *Front Med*. (2022) 9:962027. doi: 10.3389/fmed.2022.962027

30. Le S, Hoffman J, Barton C, Fitzgerald JC, Allen A, Pellegrini E, et al. Pediatric severe sepsis prediction using machine learning. *Front Pediatr*. (2019) 7:413. doi: 10.3389/fped.2019.00413

31. Su L, Xu Z, Chang F, Ma Y, Liu S, Jiang H, et al. Early prediction of mortality, severity, and length of stay in the intensive care unit of sepsis patients based on sepsis 3.0 by machine learning models. *Front Med*. (2021) 8:664966. doi: 10.3389/fmed.2021.664966

32. Zhang Y, Hu J, Hua T, Zhang J, Zhang Z, Yang M. Development of a machine learning-based prediction model for sepsis-associated delirium in the intensive care unit. *Sci Rep*. (2023) 13:12697. doi: 10.1038/s41598-023-38650-4

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 03 October 2024
DOI 10.3389/fmed.2024.1474311

OPEN ACCESS

EDITED BY
Ping Zhang,
The Ohio State University, United States

REVIEWED BY
Gowthami Sai Kogilathota Jagirdhar,
Saint Michael's Medical Center, United States
Pranjal Sharma,
Northeast Ohio Medical University,
United States

*CORRESPONDENCE
Lu-Huai Feng
✉ 823900187@qq.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 01 August 2024
ACCEPTED 23 September 2024
PUBLISHED 03 October 2024

CITATION
Wei C, Cao H, Huang L and Feng L-H (2024) Development and validation of a web-based nomogram for acute kidney injury in acute non-variceal upper gastrointestinal bleeding patients.
Front. Med. 11:1474311.
doi: 10.3389/fmed.2024.1474311

Development and validation of a web-based nomogram for acute kidney injury in acute non-variceal upper gastrointestinal bleeding patients

Chaolian Wei[1†], Honghua Cao[2†], Lina Huang[3] and Lu-Huai Feng[3*]

[1]Department of Gastric and Abdominal Tumor Surgery, Guangxi Medical University Cancer Hospital, Nanning, China, [2]Department of Gastrointestinal Surgery, Guigang People's Hospital, Guigang, China, [3]Department of Endocrinology and Metabolism Nephrology, Guangxi Medical University Cancer Hospital, Nanning, China

Background: Acute kidney injury (AKI) is a common and serious complication in patients with acute non-variceal upper gastrointestinal bleeding (NVUGIB). Early prediction and intervention are crucial for improving patient outcomes.

Methods: Data for patients presenting with acute NVUGIB in this retrospective study were sourced from the MIMC-IV database. Patients were randomly allocated into training and validation cohorts for further analysis. Independent predictors for AKI were identified using least absolute shrinkage and selection operator regression and multivariable logistic regression analyses in the training cohort. Based on the logistic regression results, a nomogram was developed to predict early AKI onset in acute NVUGIB patients, and implemented as a web-based calculator for clinical application. The nomogram's performance was evaluated through discrimination, using the C-index, calibration curves, and decision curve analysis (DCA) to assess its clinical value.

Results: The study involved 1082 acute NVUGIB patients, with 406 developing AKI. A multivariable logistic regression identified five key AKI predictors: CKD, use of human albumin, chronic liver disease, glucose, and blood urea nitrogen. The nomogram was constructed based on independent predictors. The nomogram exhibited robust accuracy, evidenced by a C-index of 0.73 in the training cohort and 0.72 in the validation cohort. Calibration curves demonstrated satisfactory concordance between predicted and observed AKI occurrences. DCA revealed that the nomogram offered considerable clinical benefit within a threshold probability range of 7% to 54%.

Conclusion: Our nomogram is a valuable tool for predicting AKI risk in patients with acute NVUGIB, offering potential for early intervention and improved clinical outcomes.

KEYWORDS

acute kidney injury, prediction, nomogram, intensive care unit, upper gastrointestinal bleeding

Introduction

Acute non-variceal upper gastrointestinal bleeding (NVUGIB) is a common and serious condition frequently observed in intensive care units (ICU), resulting in substantial morbidity and mortality rates. The in-hospital mortality rate following acute NVUGIB typically ranges from 5% to 15%, but may escalate to 35% among elderly patients with acute kidney injury (AKI) (1). Approximately 5% of hospitalized patients suffer an AKI, with an incidence of 30 - 57% in intensive care units (2), with the average pooled mortality rate of 23% but reached 49.4% in those requiring g kidney replacement therapy (3, 4). Existing literature indicates that AKI occurs in 1–11.4% of patients with acute NVUGIB, and those with acute NVUGIB complicated by AKI have longer hospital stays and higher mortality rates (1, 5). Therefore, early identification of high-risk patients is critical for the prevention of AKI, and early diagnosis and treatment can improve the long-term prognosis of patients (6).

The current diagnostic criteria for AKI as outlined by the Kidney Disease Improving Global Outcomes (KDIGO), an increase in serum creatinine or a decline in urine output remains its key diagnostic criteria (7). However, current clinical detection methods, which rely on creatinine levels and urine output, are inadequate for early AKI diagnosis, AKI is rarely diagnosed and mild cases are often missed (8, 9). In recent years, significant progress has been made in the early diagnosis of AKI due to advancements in information technology, nanotechnology, and biomedicine. Although certain studies have proposed alternative biomarkers—such as cystatin C, neutrophil gelatinase-associated lipocalin, kidney injury molecule-1, and liver-type fatty acid binding protein—for the early detection of kidney damage preceding serum creatinine elevation (4), the diagnostic accuracy of these biomarkers remains limited (4, 10, 11). Consequently, further research is imperative to develop tools capable of predicting AKI at an early stage.

Machine learning, a subset of artificial intelligence, has demonstrated efficacy in predicting AKI through the development of predictive models that analyze extensive datasets pertaining to medical treatments and outcomes. The nomogram, a widely utilized visualization technique in machine learning, serves as a dependable instrument for predicting and quantifying the risk of clinical events (12, 13). While risk prediction models for AKI in cirrhotic patients with gastric variceal bleeding are relatively well-established (14, 15), there remains a notable deficiency in risk prediction models for AKI in patients experiencing upper gastrointestinal bleeding due to acute non-variceal causes. Our study identified a combination of routinely available clinical variables that could be used for the highly precise prediction of acute NVUGIB with AKI in critically ill patients.

Materials and methods

The methodologies described in this article are consistent with the guidelines established in the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement (16).

Ethics approval and consent to participate

The establishment of MIMIC-IV (version 2.2) was approved by the institutional review boards of the Beth Israel Deaconess Medical Center (Boston, MA) and Massachusetts Institute of Technology (Cambridge, MA), thus, this study was granted a waiver of informed consent.

Database

The study utilized data from the publicly available MIMIC-IV database (version 2.2), a robust critical care database situated in the United States. This database encompasses clinical information from a vast cohort of over 190,000 patients and 450,000 hospitalizations spanning the years 2008 to 2019. The data captured within the database includes a comprehensive array of patient demographics, laboratory tests, medications, vital signs, disease diagnoses, drug management, and follow-up survival outcomes.

Participants

The study's criteria for inclusion consisted of adult patients aged 18 years and older who were admitted to the ICU with acute NVUGIB. Exclusion criteria encompassed individuals with a baseline creatinine level suggestive of stage 5 chronic kidney disease (CKD) or those undergoing frequent renal replacement therapy.

Patients were assigned to groups utilizing a pre-seeded random number generator (123) in R software version 4.3.3, and subsequently divided into training and validation sets at a ratio of 7:3.

Data extraction

Data was extracted from the MIMIC-IV database using PostgreSQL tools (V.1.13.1). Variables relevant to the risk of AKI were evaluated a priori, taking into consideration scientific literature, clinical significance, and predictors identified in prior studies (5, 17, 18).

For included patients, we collated data relating to clinical features as follows:

Demographic characteristics: sex, age, race.

Treatment modalities: use of diuretic use, use of aminoglycoside, use of human albumin therapy.

Comorbidities: chronic obstructive pulmonary disease (COPD), hypertension, hypotension, diabetes, heart failure, chronic liver disease, CKD, coronary and acute pancreatitis.

Laboratory test: hemoglobin, blood urea nitrogen (BUN), albumin (Alb), Serum creatinine (Scr), and glucose.

Outcome: AKI occurred during hospitalization.

For all laboratory test result parameters, we use the values at the time AKI occurred.

Missing data handling

In the MIMIC-IV database, a noteworthy prevalence of missing data is observed. However, the exclusion of patients with incomplete data may introduce significant bias into the study. To mitigate the impact of missing data, all variables used in the analyses were thoroughly evaluated. Less than 10% of missing values were identified across all variables. Consequently, imputation was conducted by replacing missing values with means for continuous variables with normal distributions and with medians for continuous variables with skewed distributions (19). Additionally, no dichotomous variables were missing from our study.

Definitions and outcomes

The primary outcome of interest during the ICU stay was AKI, which was defined according to the Kidney Disease Improving Global Outcomes (KDIGO) criteria (7). The use of diuretics, human albumin and aminoglycosides was categorized as any administration of these medications prior to the occurrence of AKI during the ICU stay for any indication. Hypotension was defined as any occurrence of systolic blood pressure less than 90 mmHg or diastolic blood pressure less than 60 mmHg before the onset of AKI.

Statistical analysis

Statistical analyses were conducted using SPSS version 26.0 (IBM, Armonk, NY, USA) and R version 4.2.1. Two-sided *P*-values were employed, with statistical significance defined as $P < 0.05$. Categorical variables were expressed as percentages, while continuous variables were reported as means ± SD, medians, or ranges, depending on their normality of distribution. The chi-square test was utilized for categorical variables, while *t*-tests or Wilcoxon rank sum tests were employed for continuous variables, depending on their distributions.

To enhance the accuracy of forecasts and the comprehensibility of findings, the research employed least absolute shrinkage and selection operator (LASSO) regression analysis for variable selection and regularization (20). The variables identified in the LASSO regression model during the training phase were further examined using univariate logistic regression to determine their predictive significance for AKI (21). Variables demonstrating a *p*-value of less than 0.05 in the initial univariate logistic analyses were subsequently subjected to multivariable logistic regression analysis using a backward stepwise selection method. Additionally, the variance inflation factor (VIF) was calculated among the covariate variables in the multivariable logistic regression analysis, and VIF > 4.0 was interpreted as indicating multicollinearity. Variables with VIF > 4.0 weren't included in the final analysis. After constructing a predictive model through multivariable logistic regression analysis, a clinical prediction nomogram and an interactive web-based application were developed utilizing Shiny apps to estimate the likelihood of AKI.

The performance of the nomogram was evaluated in both the training and validation cohorts through assessments of

TABLE 1 Characteristics of patients in the training and validation cohorts.

Characteristic	Training cohort (*n* = 766)	Validation cohort (*n* = 316)	*p*-value
AKI, *n* (%)			
No	471 (61.5)	205 (64.9)	0.329
Yes	295 (38.5)	111 (35.1)	
Race, *n* (%)			
White	500 (65.3)	208 (65.8)	0.829
Black	77 (10.1)	28 (8.9)	
Other	189 (24.7)	80 (25.3)	
Sex, *n* (%)			
Female	291 (38.0)	108 (34.2)	0.266
Male	475 (62.0)	208 (65.8)	
Age, years	61 (52, 73)	61 (52, 72)	0.864
CKD, *n* (%)			
No	716 (93.5)	286 (90.5)	0.117
Yes	50 (6.5)	30 (9.5)	
COPD, *n* (%)			
No	744 (97.1)	309 (97.8)	0.688
Yes	22 (2.9)	7 (2.2)	
Coronary, *n* (%)			
No	709 (92.6)	298 (94.3)	0.370
Yes	57 (7.4)	18 (5.7)	
Diabetes, *n* (%)			
No	681 (88.9)	277 (87.7)	0.631
Yes	85 (11.1)	39 (12.3)	
Hypotension, *n* (%)			
No	734 (95.8)	304 (96.2)	0.906
Yes	32 (4.2)	12 (3.8)	
Hypertension, *n* (%)			
No	586 (76.5)	247 (78.2)	0.609
Yes	180 (23.5)	69 (21.8)	
Heart faire, *n* (%)			
No	698 (91.1)	289 (91.5)	0.954
Yes	68 (8.9)	27 (8.5)	
Chronic liver disease, *n* (%)			
No	692 (90.3)	279 (88.3)	0.368
Yes	74 (9.7)	37 (11.7)	
Use of human albumin, *n* (%)			
No	507 (66.2)	206 (65.2)	0.807
Yes	259 (33.8)	110 (34.8)	
Acute pancreatitis, *n* (%)			
No	755 (98.6)	305 (96.5)	0.054
Yes	11 (1.4)	11 (3.5)	

(Continued)

TABLE 1 (Continued)

Characteristic	Training cohort (*n* = 766)	Validation cohort (*n* = 316)	*p*-value
Use of diuretic, *n* (%)			
No	612 (79.9)	242 (76.6)	0.257
Yes	154 (20.1)	74 (23.4)	
Use of aminoglycosides, *n* (%)			
No	719 (93.9)	298 (94.3)	0.892
Yes	47 (6.1)	18 (5.7)	
Hemoglobin, g/L	94 (84, 107)	94 (82, 105)	0.207
Scr (umol/L)	80 (62, 133)	80 (62, 139)	0.814
Blood urea nitrogen (mmol/L)	3.4 (2.2, 6.0)	3.7 (2.2, 6.0)	0.441
Albumin (g/L)	30 (26, 35)	30 (26, 35)	0.979
Glucose (mmol/L)	6.4 (5.3, 8.1)	6.2 (5.3, 7.9)	0.447

Scr, Serum creatinine; COPD, chronic obstructive pulmonary disease; CKD, chronic kidney disease; AKI, actue kindey injure.

discrimination and calibration (22). Discrimination was measured using the C-index, which ranges from 0.5 (indicating no discrimination) to 1.0 (indicating perfect prediction), indicating the extent of predictive accuracy. Calibration was assessed by comparing predicted and actual probabilities of AKI occurrence through a visual calibration plot. Internal validation was conducted using 1000 bootstrap resamples to further evaluate the nomogram's predictive accuracy. Additionally, a decision curve analysis (DCA), which determines the net benefit of models and predictors, was performed to assess the clinical value of the nomogram (23).

Results

Characteristics of patients

A total of 1,082 patients presenting with acute NVUGIB were included in the study, with 406 patients (37.5%) testing positive for AKI. The average age of the patients was 62 years, with a majority (63.1%) being male. Patients were randomly divided into training (766 patients) and validation (316 patients) cohorts. Table 1 displays the demographic and clinical characteristics of patients in each cohort. Baseline clinical features were found to be comparable between the two cohorts, with AKI rates of 38.5% and 35.1% in the training and validation cohorts, respectively.

Nomogram variable screening

In this investigation, LASSO regression analysis was employed to identify 13 predictors with significant correlation to AKI from a pool of 21 potential predictors in the training cohort, as depicted in Figures 1A, B. The predictors associated with AKI as identified by the LASSO regression technique are detailed in Table 2 (lambda = 0.01038261). Following this, a multivariable logistic regression analysis was carried out to delve deeper into the variables that successfully passed through both univariate logistic regression and LASSO analyses.

Based on the findings of the stepwise logistic regression analysis, the model incorporating five independent predictors of AKI, including CKD, use of human albumin, chronic liver disease, glucose, and blood urea nitrogen, demonstrated the lowest AIC value within the training cohort. Additionally, the VIF values for all variables were less than 4, suggesting the absence of collinearity among the screened predictors (Table 2).

Nomogram construction and performance in the training cohort

Utilizing the outcomes of multivariate Logistic regression analysis, a nomogram (Figure 2A) was developed to visually represent a model incorporating independent predictors. For instance, a patient presenting with a blood urea nitrogen level of 14.3 mmol/L, glucose level of 12 mmol/L, absence of chronic liver disease and AKD, and using human albumin, possesses a current AKI risk score of 150, which equates to a 78% probability of developing AKI. This nomogram can be accessed online at https://risk-prediction-model-web-calculator20240327.shinyapps.io/AKI_probability_of_UBG/, as depicted in Figure 2B. Users are required to interact with the in-line graph by selecting either "Yes" or "No" from the provided options, inputting pertinent laboratory test results, and subsequently choosing "Predict" to ascertain the probability of AKI occurrence during the patient's ICU admission.

The C-index of the nomogram was calculated to be 0.73 (95% CI: 0.70–0.77) for the training cohort. The calibration curve illustrated in Figure 3A exhibits a satisfactory concordance between the anticipated and actual occurrences for the likelihood of AKI in the training cohort. The lack of statistical significance ($P = 0.580$) in the Hosmer–Lemeshow test implies that the model did not demonstrate overfitting.

External validation of the nomogram 2 in the validation cohort

In the validation cohort, the nomogram exhibited a C-index of 0.72 (95% CI 0.66–0.79) for the assessment of AKI risk. Additionally, a well-calibrated risk estimation was demonstrated through the calibration curve (Figure 3B).

Clinical value of the nomogram

Figure 4 illustrates the outcomes of decision curve analysis for the nomogram, highlighting the high-risk threshold probability at which a clinician can assess a patient's risk of AKI and the potential advantages of intervention. The decision curve indicates that employing the nomogram for AKI prediction can yield substantial benefits when a clinician's threshold probability falls between 7% and 54%, with the nomogram exhibiting greater predictive accuracy compared to a single predictor within this specified range.

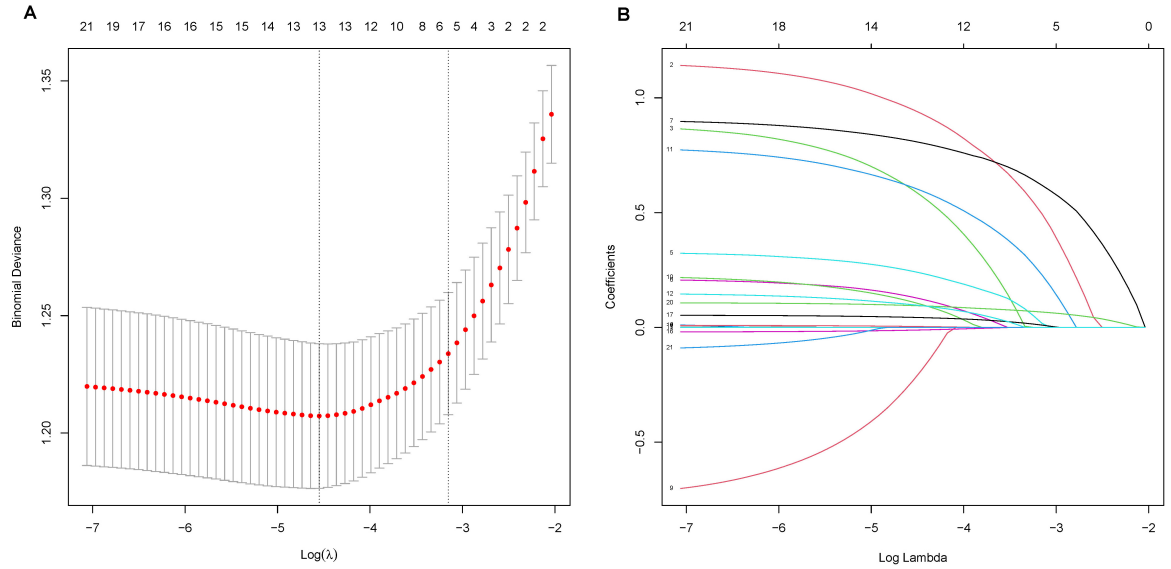

FIGURE 1
LASSO Regression Analysis for Predictor Selection. **(A)** Tuning parameter (lambda) selection in the LASSO model using 10-fold cross-validation. The x-axis represents the log(lambda), and the y-axis represents the binomial deviance. The red dots indicate the average deviance values, and the vertical bars represent the standard errors. The vertical dashed lines represent the optimal values chosen by the minimum criteria and 1 standard error of the minimum criteria (the 1-SE rule). **(B)** LASSO coefficient profiles of the 21 potential predictors. Each curve represents a predictor, with the y-axis displaying the coefficient values and the x-axis representing the log(lambda). The numbers at the top of the plot indicate the number of predictors included in the model at each log(lambda) value.

Discussion

AKI is a serious complication in patients with acute NVUGIB significantly impacting morbidity and mortality. In this study, we developed a nomogram to quantitatively predict the risk of AKI in patients with acute non-variceal UGIB based on patient-specific factors. This predictive tool can be utilized to assess individual AKI risk, thereby facilitating personalized treatment and surveillance strategies. The significance of the present study lies in the development of the nomogram utilizing a substantial cohort of ICU patients diagnosed with acute non-variceal UGIB. Furthermore, the nomogram's performance underwent rigorous assessment and internal validation.

The current AKI diagnostic criteria, established by KDIGO in 2012, rely on changes in serum creatinine and urine output, which can miss early or subclinical AKI, depend on baseline creatinine levels that may not always be known, and are influenced by non-renal factors like hydration and muscle mass. Consequently, research efforts have been directed towards identifying susceptibility and exposure factors associated with AKI to facilitate preemptive preventive measures. These measures include optimizing fluid management, avoiding nephrotoxic medications, closely monitoring renal function in high-risk patients, and using alternative imaging methods to minimize contrast exposure. Nonetheless, the onset and progression of AKI involve intricate pathophysiological mechanisms, rendering the accurate assessment of AKI risk based on individual susceptibility and exposure challenging. This complexity often leads to the potential for over-treatment in preventive strategies. Another particular strength of this study is the consideration of a range of previously reported clinical features and laboratory findings related to AKI (5, 7, 17, 18, 24). In the present study, we similarly noted that CKD, glucose, chronic liver disease were closely related to AKI in patients with acute NVUGIB, which is consistent with most studies on variceal upper gastrointestinal bleeding (20–23). In contrast to variceal upper gastrointestinal bleeding, where albumin administration is generally advantageous for the prevention of AKI, our study identified a positive correlation between albumin use and the risk of AKI in patients with acute NVUGIB. This association may be attributable to enhanced albumin filtration and modified tubular albumin uptake in these individuals. Nonetheless, the relationship between human blood albumin administration and the incidence of AKI remains a subject of ongoing debate, (25–27), and the precise underlying mechanism remains uncertain. It is worth mentioning that we found that blood urea nitrogen was more predictive of AKI than Scr in patients with acute UGIB. This disparity may be attributable to fasting-induced alterations in muscle metabolism and renal blood flow, which influence Scr production and excretion. Despite these physiological changes, Scr levels tend to remain relatively stable and are less affected by short-term dietary modifications. Conversely, blood urea nitrogen levels are more susceptible to variations in protein catabolism and renal perfusion, rendering blood urea nitrogen a more sensitive marker for detecting changes in renal function and assessing the risk of AKI in this patient population.

The capacity to precisely forecast the incidence of AKI in patients experiencing acute NVUGIB holds substantial importance, given that AKI represents a heterogeneous syndrome necessitating individualized care and management approaches (28). In contrast to the population-based or large cohort data utilized by KDIGO clinical practice guidelines, nomograms offer a more individualized approach to delivering prognostic information to patients. To the

TABLE 2 Univariate and multivariate logistic regression analyses of variables relating to AKI in the training cohort.

Variable	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Race				
White	Reference		Reference	
Black	1.03 (0.63, 1.70)	0.896	0.89 (0.51, 1.55)	0.680
Other	1.61 (1.15, 2.26)	0.006	1.40 (0.96, 2.03)	0.080
CKD				
No	Reference	<0.001	Reference	<0.001
Yes	4.54 (2.40, 8.57)		3.62 (1.83, 7.15)	
Chronic liver disease				
No	Reference	<0.001	Reference	0.003
Yes	2.92 (1.78, 4.79)		2.27 (1.33, 3.88)	
Use of human albumin				
No	Reference	<0.001	Reference	<0.001
Yes	3.18 (2.33, 4.35)		2.45 (1.76, 3.42)	
Acute pancreatitis				
No	Reference	0.445	NA	
Yes	0.59 (0.16, 2.26)			
COPD				
No	Reference	0.019	Reference	0.080
Yes	2.88 (1.19, 6.96)		2.39 (0.90, 6.31)	
Hypertension				
No	Reference	0.017	Reference	0.315
Yes	1.51 (1.08, 2.12)		1.22 (0.83, 1.81)	
Diabetes				
No	Reference	0.002	Reference	0.221
Yes	2.05 (1.30, 3.22)		1.40 (0.82, 2.41)	
Use of diuretic				
No	Reference	0.007	Reference	0.297
Yes	1.64 (1.15, 2.34)		1.24 (0.83, 1.86)	
Hemoglobin	0.91 (0.76, 1.10)	0.348	NA	
Blood urea nitrogen	1.81 (1.51, 2.15)	<0.001	1.48 (1.23, 1.78)	<0.001
Albumin	0.76 (0.63, 0.93)	0.006	0.86 (0.70, 1.07)	0.177
Glucose	1.23 (1.08, 1.40)	0.002	1.21 (1.06, 1.39)	0.005

best of our knowledge, this study represents the first attempt to develop an AKI risk prediction model that independently evaluates previously proposed risk variables for their inclusion in a formal nomogram specifically for patients with acute NVUGIB. Most studies on acute NVUGIB primarily concentrate on assessing the severity of bleeding, patient prognosis, and the risk of rebleeding. For instance, the Rockall score is employed to determine the necessity for further endoscopic intervention, while the Forrest score is utilized to evaluate the risk of endoscopic rebleeding (29). However, AKI, a complication associated with more severe short-term and long-term prognoses, is frequently overlooked. The lack of timely and effective interventions for AKI significantly exacerbates the complexity and cost of treatment. Therefore, we utilized clinically accessible laboratory results and assessed patients' susceptibility to AKI to develop a nomogram aimed at providing individualized AKI risk predictions for patients with acute NVUGIB in the intensive care unit. This approach aligns with the contemporary emphasis on personalized medicine.

The primary and ultimate justification for employing the nomogram lies in its capacity to assess the necessity for individualized supplementary treatment or care. Nevertheless, the metrics of risk-prediction performance, including discrimination and calibration, fail to encapsulate the clinical implications associated with specific levels of discrimination or degrees of miscalibration (30–32). Therefore, to substantiate the clinical utility

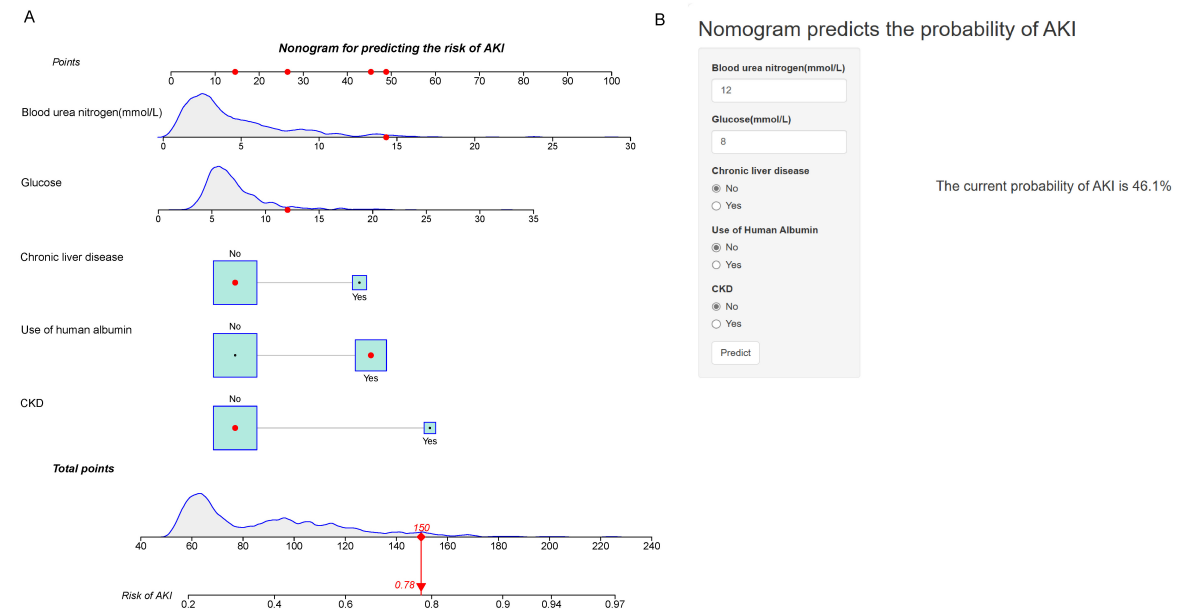

FIGURE 2
Nomogram for Predicting the Risk of AKI. **(A)** Nomogram to predict the risk of AKI in patients with acute NVUGIB. The nomogram integrates multiple predictors, including blood urea nitrogen (mmol/L), glucose (mmol/L), chronic liver disease, use of human albumin, and CKD. Each predictor has a corresponding point scale, which is used to calculate the total points. The total points are then used to determine the risk of AKI, displayed on the bottom scale. **(B)** Web-based calculator interface for predicting AKI risk. Users input values for blood urea nitrogen, glucose, chronic liver disease, use of human albumin, and CKD status to obtain the predicted probability of AKI. The example provided shows a current probability of AKI at 46.1%.

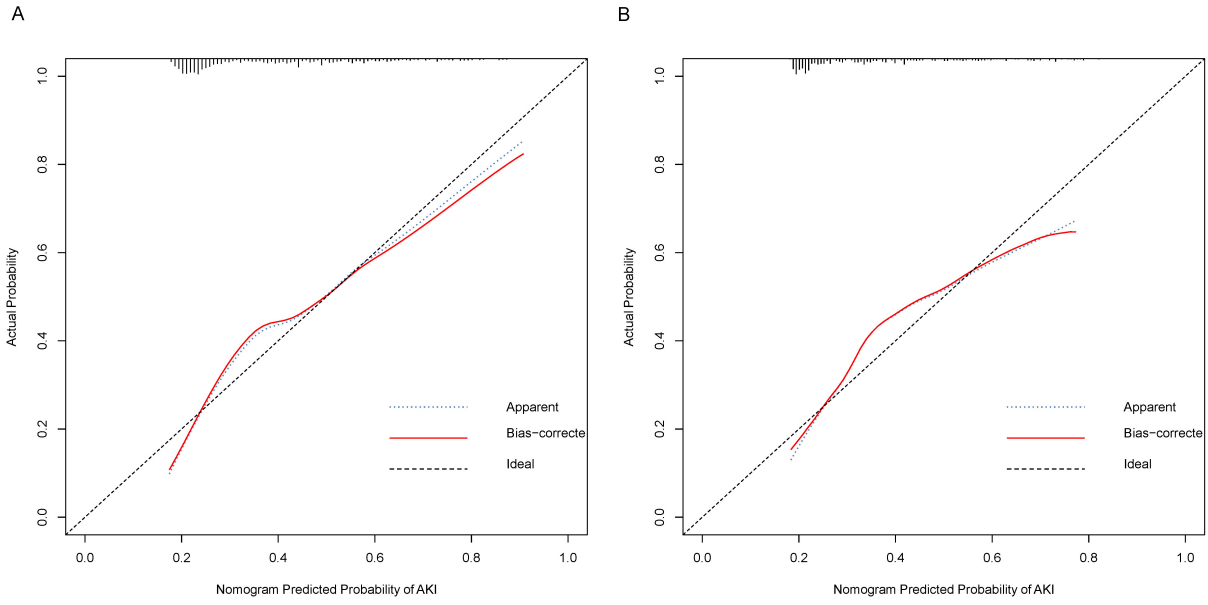

FIGURE 3
Calibration Curves for Nomogram Predicted Probability of AKI. These calibration plots compare the predicted probabilities of AKI against the actual observed probabilities, demonstrating the accuracy of the nomogram model. **(A)** Calibration curve for the training cohort. **(B)** Calibration curve for the validation cohort. The x-axis represents the nomogram predicted probability of AKI, while the y-axis shows the actual probability. The plot includes the apparent calibration (blue dotted line), the bias-corrected calibration (red solid line), and the ideal calibration (black dashed line). The closer the red line is to the black line, the more accurate the model.

of our nomogram, we undertook an evaluation to ascertain whether decisions informed by the nomogram would result in improved patient outcomes. Considering the inherent difficulties of executing a multi-institutional prospective validation, particularly due to the complexities involved in aggregating clinical data from multiple institutions, we opted to utilize decision curve analysis as an alternative methodological approach in this study. This study introduces an innovative methodology for evaluating the clinical

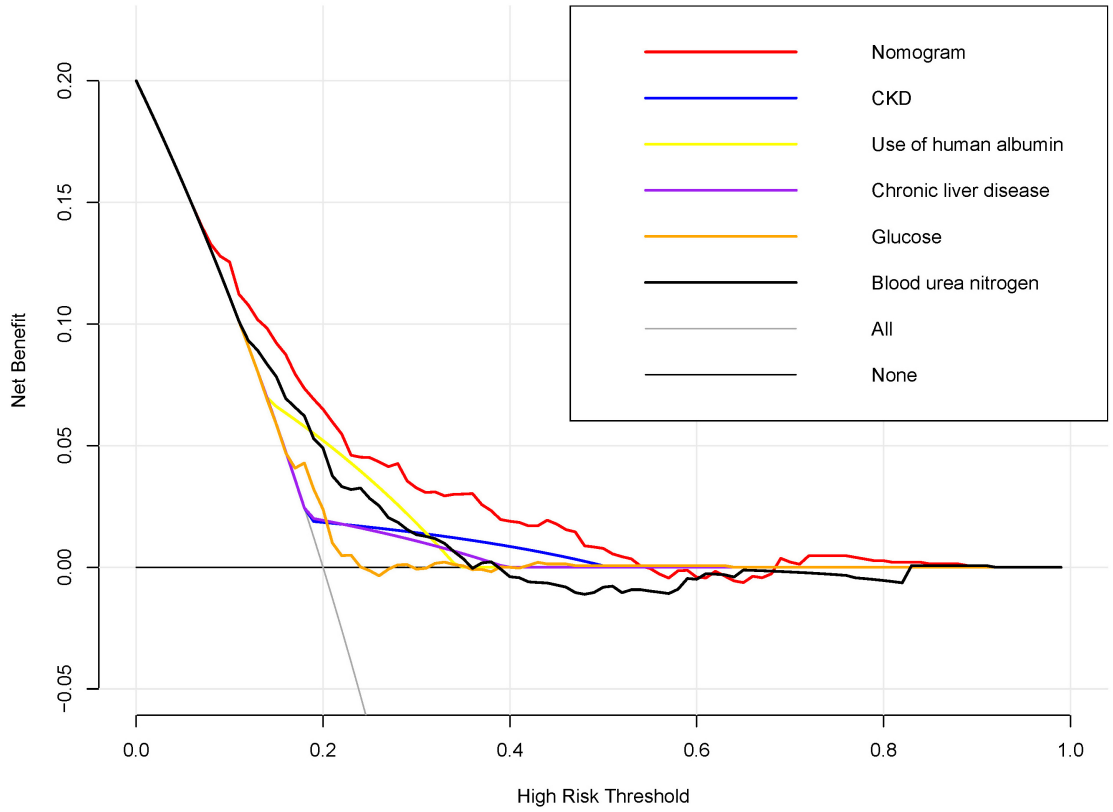

FIGURE 4
Decision Curve Analysis for the Nomogram Predicting AKI in acute NVUGIB Patients. Decision curve analysis was conducted to compare the net benefits of the nomogram against individual predictors across a spectrum of high-risk thresholds. The x-axis denotes the high-risk threshold for predicting acute kidney injury (AKI), while the y-axis indicates the net benefit. The red line, representing the nomogram, demonstrates superior net benefit across most thresholds when compared to individual predictors, including chronic kidney disease (CKD), use of human albumin, chronic liver disease, glucose levels, and blood urea nitrogen.

implications of decisions grounded in threshold probability, thereby facilitating the calculation of net benefit (16, 33). The decision curve analysis conducted herein reveals that employing a nomogram for the prediction of AKI yields greater benefits when the threshold probability for physicians between 7% and 54%, as opposed to the strategies of treating all patients or treating none.

This study has limitations. First, its monocentric design within a single ICU, which restricts the generalizability of the findings and the dynamic online nomogram to other centers or countries. Further research is needed to validate the model in varied settings. Second, the MIMIC database's lack of novel biomarkers, such as cystatin C, neutral gelatin-associated lipocalcin, NT-proBNP, uNGAL, and uAGT, precluded us from enhancing the model's predictive capacity. Third, the retrospective design of the study inherently limited our ability to eliminate bias. However, rigorous inclusion criteria were applied to ensure that both the control and case groups accurately reflected real-world conditions.

Conclusion

Our study presents an innovative online nomogram that incorporates clinical risk factors to facilitate personalized prediction of AKI in patients with acute NVUGIB upon ICU admission. This predictive tool holds significant potential in identifying acute NVUGIB patients who are most likely to benefit from targeted interventions for the prevention and management of AKI.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the institutional review boards of the Beth Israel Deaconess Medical Center (Boston, MA) and Massachusetts Institute of Technology (Cambridge, MA). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the study did not disclose any patient privacy.

Author contributions

CW: Writing – original draft, Software, Methodology, Investigation, Formal analysis. HC: Writing – original draft, Validation, Software, Resources, Methodology, Formal analysis, Data curation. LH: Writing – original draft, Validation, Software, Resources, Methodology, Formal analysis. L-HF: Writing – review and editing, Writing – original draft, Software, Methodology, Formal analysis.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of the article. This work was supported by grants from the Natural Science Foundation of Guangxi (NO. 2021GXNSFAA220036).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Benedeto-Stojanov D, Bjelaković M, Stojanov D, Aleksovski B. Prediction of in-hospital mortality after acute upper gastrointestinal bleeding: Cross-validation of several risk scoring systems. *J Int Med Res.* (2022) 50:3000605221086442. doi: 10.1177/03000605221086442

2. Hoste E, Bagshaw S, Bellomo R, Cely C, Colman R, Cruz D, et al. Epidemiology of acute kidney injury in critically ill patients: The multinational AKI-EPI study. *Intensive Care Med.* (2015) 41:1411–23. doi: 10.1007/s00134-015-3934-7

3. Kellum J, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders H. Acute kidney injury. *Nat Rev Dis Prim.* (2021) 7:52.

4. Ronco C, Bellomo R, Kellum J. Acute kidney injury. *Lancet.* (2019) 394:1949–64.

5. Cakmak U, Merhametsiz O, Gok Oguz E, Ercan Z, Haspulat A, Ozkan S, et al. Effects of acute kidney injury on clinical outcomes in patients with upper gastrointestinal bleeding. *Ren Fail.* (2016) 38:176–84.

6. Joannidis M, Druml W, Forni L, Groeneveld A, Honore P, Hoste E, et al. Prevention of acute kidney injury and protection of renal function in the intensive care unit: Update 2017: Expert opinion of the working group on prevention, AKI section, European society of intensive care medicine. *Intensive Care Med.* (2017) 43:730–49. doi: 10.1007/s00134-017-4832-y

7. Kellum J, Lameire N. Diagnosis, evaluation, and management of acute kidney injury: A KDIGO summary (Part 1). *Crit Care.* (2013) 17:204. doi: 10.1186/cc11454

8. Grams M, Waikar S, MacMahon B, Whelton S, Ballew S, Coresh J. Performance and limitations of administrative data in the identification of AKI. *Clin J Am Soc Nephrol.* (2014) 9:682–9.

9. Bhatraju P, Wurfel M, Himmelfarb J. Trajectory of kidney function: The canary in sepsis. *Am J Respir Crit Care Med.* (2020) 202:1211–2. doi: 10.1164/rccm.202007-2627ED

10. Bonavia A, Singbartl K. Kidney injury and electrolyte abnormalities in liver failure. *Semin Respir Crit Care Med.* (2018) 39:556–65.

11. Angeli P, Gines P, Wong F, Bernardi M, Boyer T, Gerbes A, et al. Diagnosis and management of acute kidney injury in patients with cirrhosis: Revised consensus recommendations of the International Club of Ascites. *Gut.* (2015) 64:531–7.

12. Feng L, Bu K, Ren S, Yang Z, Li B, Deng C. Nomogram for predicting risk of digestive carcinoma among patients with type 2 diabetes. *Diabetes Metab Syndrome Obes Targets Ther.* (2020) 13:1763–70. doi: 10.2147/DMSO.S251063

13. Iasonos A, Schrag D, Raj G, Panageas K. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* (2008) 26:1364–70.

14. Hsieh Y, Lee K, Chen P, Su C, Hou M, Lin H. Acute kidney injury predicts mortality in cirrhotic patients with gastric variceal bleeding. *J Gastroenterol Hepatol.* (2017) 32:1859–66.

15. Feng L, Lu Y, Ren S, Liang H, Wei L, Jiang J. Development and validation of a dynamic online nomogram for predicting acute kidney injury in cirrhotic patients upon ICU admission. *Front Med (Lausanne).* (2023) 10:1055137. doi: 10.3389/fmed.2023.1055137

16. Collins G, Reitsma J, Altman D, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ.* (2015) 350: g7594.

17. Wu P, Wu C, Lin C, Wu V. Long-term risk of upper gastrointestinal hemorrhage after advanced AKI. *Clin J Am Soc Nephrol.* (2015) 10:353–62.

18. Alkhatib A, Lam A, Shihab F, Adler D. RIFLE criteria accurately identifies renal dysfunction and renal failure in elderly patients with upper gastrointestinal hemorrhage: A pilot study. *South Med J.* (2009) 102:580–4. doi: 10.1097/SMJ.0b013e3181a5cec9

19. Zhang Z. Missing data imputation: Focusing on single imputation. *Ann Transl Med.* (2016) 4:9.

20. Kaufman J. Acute kidney injury in CKD: Role of metabolic acidosis. *Kidney Int Rep.* (2022) 7:2555–7.

21. Yang L. How acute kidney injury contributes to renal fibrosis. *Adv Exp Med Biol.* (2019) 1165:117–42.

22. Lee H, Seo Y. Current knowledge about biomarkers of acute kidney injury in liver cirrhosis. *Clin Mol Hepatol.* (2022) 28:31–46.

23. Cullaro G, Verna E, Duarte-Rojo A, Kappus M, Ganger D, Rahimi R, et al. Frailty and the risk of acute kidney injury among patients with cirrhosis. *Hepatol Commun.* (2022) 6:910–9.

24. Gameiro J, Fonseca J, Outerelo C, Lopes J. Acute kidney injury: From diagnosis to prevention and treatment strategies. *J Clin Med.* (2020) 9:1704.

25. Frenette A, Bouchard J, Bernier P, Charbonneau A, Nguyen L, Rioux J, et al. Albumin administration is associated with acute kidney injury in cardiac surgery: A propensity score analysis. *Crit Care.* (2014) 18:602.

26. Lee E, Kim W, Kim J, Chin J, Choi D, Sim J, et al. Effect of exogenous albumin on the incidence of postoperative acute kidney injury in patients undergoing off-pump coronary artery bypass surgery with a preoperative albumin level of less than 4.0 g/dl. *Anesthesiology.* (2016) 124:1001–11. doi: 10.1097/ALN.0000000000001051

27. James M, Grams M, Woodward M, Elley C, Green J, Wheeler D, et al. A meta-analysis of the association of estimated GFR, albuminuria, diabetes mellitus, and hypertension with acute kidney injury. *Am J Kidney Dis.* (2015) 66:602–12. doi: 10.1053/j.ajkd.2015.02.338

28. Arora T, Martin M, Grimshaw A, Mansour S, Wilson F. Prediction of outcomes after acute kidney injury in hospitalised patients: Protocol for a systematic review. *BMJ Open.* (2020) 10:e042035.

29. Yin L, Yu W. Retrospective analysis of risk factors for non-variceal upper gastrointestinal bleeding and construction of a nomogram prediction model. *Am J Transl Res.* (2023) 15:3385–93.

30. Sauerbrei W, Boulesteix A, Binder H. Stability investigations of multivariable regression models derived from low- and high-dimensional data. *J Biopharm Stat.* (2011) 21:1206–31. doi: 10.1080/10543406.2011.629890

31. Localio A, Goodman S. Beyond the usual prediction accuracy metrics: Reporting results for clinical decision making. *Ann Intern Med.* (2012) 157:294–5. doi: 10.7326/0003-4819-157-4-201208210-00014

32. Van Calster B, Vickers A. Calibration of risk prediction models: Impact on decision-analytic performance. *Med Decis Mak.* (2015) 35:162–9.

33. Balachandran V, Gonen M, Smith J, DeMatteo R. Nomograms in oncology: More than meets the eye. *Lancet Oncol.* (2015) 16:e173–80. doi: 10.1016/S1470-2045(14)71116-7

frontiers | Frontiers in Medicine

TYPE Correction
PUBLISHED 26 March 2025
DOI 10.3389/fmed.2025.1593738

OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE
Lu-Huai Feng
✉ 823900187@qq.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 14 March 2025
ACCEPTED 17 March 2025
PUBLISHED 26 March 2025

CITATION
Wei C, Cao H, Huang L and Feng L-H (2025) Corrigendum: Development and validation of a web-based nomogram for acute kidney injury in acute non-variceal upper gastrointestinal bleeding patients. *Front. Med.* 12:1593738. doi: 10.3389/fmed.2025.1593738

Corrigendum: Development and validation of a web-based nomogram for acute kidney injury in acute non-variceal upper gastrointestinal bleeding patients

Chaolian Wei[1†], Honghua Cao[2†], Lina Huang[3] and Lu-Huai Feng[3*]

[1]Department of Gastric and Abdominal Tumor Surgery, Guangxi Medical University Cancer Hospital, Nanning, China, [2]Department of Gastrointestinal Surgery, Guigang People's Hospital, Guigang, China, [3]Department of Endocrinology and Metabolism Nephrology, Guangxi Medical University Cancer Hospital, Nanning, China

KEYWORDS

acute kidney injury, prediction, nomogram, intensive care unit, upper gastrointestinal bleeding

A Corrigendum on
Development and validation of a web-based nomogram for acute kidney injury in acute non-variceal upper gastrointestinal bleeding patients

by Wei, C., Cao, H., Huang, L., and Feng, L. H. (2024). *Front Med.* 11:1474311. doi: 10.3389/fmed.2024.1474311

In the published article, there was an error in affiliation(s) [1 and 3]. Instead of "[The Affiliated Tumor Hospital of Guangxi Medical University]", it should be "[Guangxi Medical University Cancer Hospital]".

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 03 October 2024
DOI 10.3389/fmed.2024.1410702

OPEN ACCESS

EDITED BY
Qinghe Meng,
Upstate Medical University, United States

REVIEWED BY
Markus Graf,
Heilbronn University, Germany
Ahmet Çifci,
Mehmet Akif Ersoy University, Türkiye
Nehal Hassan,
Newcastle University, United Kingdom

*CORRESPONDENCE
Xiaohong Huang
✉ 2430874657@qq.com

†These authors share first authorship

RECEIVED 01 April 2024
ACCEPTED 17 September 2024
PUBLISHED 03 October 2024

CITATION
Zhang Q, Xu L, He W, Lai X and Huang X (2024) Survival prediction for heart failure complicated by sepsis: based on machine learning methods. *Front. Med.* 11:1410702. doi: 10.3389/fmed.2024.1410702

Survival prediction for heart failure complicated by sepsis: based on machine learning methods

Qitian Zhang[1†], Lizhen Xu[2†], Weibin He[1], Xinqi Lai[1] and Xiaohong Huang[1*]

[1]Department of Cardiology, Zhangzhou Affiliated Hospital of Fujian Medical University, Zhangzhou, Fujian, China, [2]Department of Endocrinology, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fuzhou, China

Background: Heart failure is a cardiovascular disorder, while sepsis is a common non-cardiac cause of mortality. Patients with combined heart failure and sepsis have a significantly higher mortality rate and poor prognosis, making early identification of high-risk patients and appropriate allocation of medical resources critically important.

Methods: We constructed a survival prediction model for patients with heart failure and sepsis using the eICU-CRD database and externally validated it using the MIMIC-IV database. Our primary outcome is the 28-day all-cause mortality rate. The Boruta method is used for initial feature selection, followed by feature ranking using the XGBoost algorithm. Four machine learning models were compared, including Logistic Regression (LR), eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Gaussian Naive Bayes (GNB). Model performance was assessed using metrics such as area under the curve (AUC), accuracy, sensitivity, and specificity, and the SHAP method was utilized to visualize feature importance and interpret model results. Additionally, we conducted external validation using the MIMIC-IV database.

Results: We developed a survival prediction model for heart failure complicated by sepsis using data from 3891 patients in the eICU-CRD and validated it externally with 2928 patients from the MIMIC-IV database. The LR model outperformed all other machine learning algorithms with a validation set AUC of 0.746 (XGBoost: 0.726, AdaBoost: 0.744, GNB: 0.722), alongside accuracy (0.685), sensitivity (0.666), and specificity (0.712). The final model incorporates 10 features: age, ventilation, norepinephrine, white blood cell count, total bilirubin, temperature, phenylephrine, respiratory rate, neutrophil count, and systolic blood pressure. We employed the SHAP method to enhance the interpretability of the model based on the LR algorithm. Additionally, external validation was conducted using the MIMIC-IV database, with an external validation AUC of 0.699.

Conclusion: Based on the LR algorithm, a model was constructed to effectively predict the 28-day all-cause mortality rate in patients with heart failure complicated by sepsis. Utilizing our model predictions, clinicians can promptly identify high-risk patients and receive guidance for clinical practice.

KEYWORDS

heart failure, sepsis, eICU-CRD, MIMIC-IV, machine learning, hospital mortality

1 Introduction

The 2018 medical insurance data reveals that sepsis and heart failure, respectively, ranked first and second in 30-day readmission rates among patients (1). Sepsis is defined as a dysregulated host response to infection, leading to organ failure (2). In 2017, an estimated 48.9 million cases of sepsis were recorded globally, resulting in 11 million sepsis-related deaths, which accounted for 19.7% of all global deaths (3). The mortality rates of sepsis in intensive care units and hospitals are reported to be 25.8 and 35.3%, respectively (4), with annual losses exceeding $24 billion (5, 6). Heart failure is a cardiovascular disorder characterized by high incidence and mortality rates, representing an escalating global epidemic (7). Over 64 million individuals worldwide are afflicted with heart failure, severely compromising their quality of life (8). Chronic heart failure is the leading complication in septic patients, with two-thirds of critically ill cases having prior heart failure (9, 10). Heart failure patients may exhibit underlying circulatory dysfunction and impaired cardiac reserve, placing them at increased risk if they develop sepsis. Alon et al. discovered that heart failure patients admitted for sepsis had a higher mortality rate compared to those without heart failure (51% vs. 41%; $p = 0.015$) (11). Walker et al. studied the effect of sepsis on heart failure patient mortality and found it caused one-fourth of deaths (12). The high incidence and mortality rates stress the importance of early identification, assessment, and management of heart failure patients with sepsis.

Currently, there are no identified predictive models for survival in patients with heart failure complicated by sepsis. The Sequential Organ Failure Assessment (SOFA), Simplified Acute Physiology Score II (SAPS II), and Acute Physiology Score III (APS III) are frequently utilized assessment tools for predicting disease prognosis (13, 14). Despite their extensive utilization, they exhibit limitations such as the complexity of assessment, insufficient specificity, and potential suitability restricted to specific disease types or clinical contexts. The current research trend is to integrate novel biomarkers (15, 16) into established scoring systems or to revamp these systems (17) to improve their predictive accuracy for disease prognosis. In clinical practice, machine learning is widely applied for result prediction, diagnosis, medical image interpretation, disease risk assessment, and treatment planning (18, 19). Compared to traditional statistical methods, machine learning excels in handling complex data, exhibiting higher accuracy and efficiency (20). In the past, the application of machine learning was constrained by limited interpretability. However, with the emergence of techniques like SHAP, users can now professionally understand model predictions with greater clarity (21).

Our research aims to build survival prediction models using various machine learning algorithms to assess the overall in-hospital mortality rate among patients with heart failure complicated by sepsis. We utilize the eICU-CRD database to build machine learning models, selecting the one with optimal predictive performance. Subsequently, we conduct external validation using the MIMIC-IV database. Additionally, the SHAP method is used to explain model predictions and assess the importance of features. The objective of this study is to identify critically ill patients and offer guidance for clinical practice.

2 Materials and methods

2.1 Data sources and study population

This study draws data from two primary sources: the eICU Collaborative Research Database (eICU-CRD) and The Medical Information Mart for Intensive Care IV database (MIMIC-IV). The eICU-CRD database encompasses various ICU units across the United States, offering a comprehensive array of clinical data, physiological parameters, and medical events. Spanning from 2014 to 2015, it meticulously documents information for over 200,000 patients, facilitating medical research endeavors and data-informed clinical decision-making (22). On the other hand, MIMIC-IV (version 2.2) represents an extensive repository of intensive care data, featuring detailed records of more than 190,000 ICU patients from 2008 to 2019 (23). This database is characterized by its exhaustive collection of clinical details, including demographic profiles, laboratory findings, and medication histories, serving as invaluable resources for rigorous clinical investigations.

We identified patients with heart failure complicated by sepsis from both the eICU-CRD and MIMIC-IV databases using ICD-9 and ICD-10 codes. The exclusion criteria for the study population are: (1) age under 18 years, (2) ICU stay duration less than 24 h, and (3) clinical information missing rate exceeding 30% at data collection. For patients with multiple hospital admissions or ICU visits, only the first ICU experience during the initial hospital admission is considered. Heart failure was defined as a syndrome resulting from structural or functional cardiac abnormalities that lead to inadequate cardiac output and congestion in the systemic or pulmonary circulation, encompassing all types of heart failure with different ejection fractions. Sepsis was diagnosed based on the Sepsis-3.0 guidelines, which define it as life-threatening organ dysfunction caused by a dysregulated host response to infection. A SOFA score ≥ 2 (or a qSOFA score ≥ 2 for suspected infection in non-ICU settings) was used to diagnose sepsis.

2.2 Data extraction and preprocessing

In this study, we included ICU patients diagnosed with heart failure and sepsis, and extracted the following data: (1) Demographics: age, gender, height, and weight; (2) Vital Signs: temperature (T), heart rate (HR), respiratory rate (R), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), and peripheral oxygen saturation (SpO2); (3) Laboratory parameters: complete blood count, liver and kidney function tests, electrolytes, lipid profile, blood gas analysis, coagulation function, cardiac enzymes, and BNP; (4) Comorbidities: hypertension, diabetes mellitus, hyperlipidemia, chronic obstructive pulmonary disease (COPD), pneumonia, chronic kidney disease (CKD), and atrial fibrillation (AF); (5) Medication data: angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers (ACEI/ARB), beta Blockers, furosemide, spironolactone, dobutamine, dopamine, epinephrine, milrinone, norepinephrine, and phenylephrine; and (6) Other Indicators: Ventilation and 24-h fluid balance. The primary outcome is the 28-day all-cause mortality rate.

Initially, we transformed certain indicators, such as computing BMI from height and weight and determining 28-day in-hospital

mortality using hospitalization duration and survival status. Variables with over 30% missing data were removed, and missing values in the remaining features were imputed using KNN. Outliers were identified using the 1.5 times interquartile range method, particularly focusing on BMI, mechanical ventilation time, and 24-h fluid balance, and were subsequently removed. Additionally, Spearman correlation coefficients were calculated to evaluate variable relationships, while VIF values assessed multicollinearity. Variables with high correlation or VIF exceeding 5 underwent pre-screening. Continuous variables were standardized for model stability, and categorical variables were transformed into dummy variables via one-hot encoding. Despite minor sample imbalances in the outcome variable, we chose not to employ sample balancing techniques.

2.3 Model construction and evaluation

The Boruta method is used for initial feature screening, determining feature importance by comparing them with randomly generated "shadow features" (24). The XGBoost method is employed for importance ranking of the preliminarily selected features. Model construction and validation are conducted using the EICU dataset, with 10-fold cross-validation to generate training and validation sets, and Logistic Regression (LR), eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Gaussian Naive Bayes (GNB) models are established and validated. Model performance is evaluated on the validation set using metrics such as the area under the curve (AUC) for discrimination, calibration curve for accuracy, and DCA curve for clinical utility, as well as metrics including accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score. The final predictive model is optimized using hyperparameter tuning and grid search. Additionally, the MIMIC dataset is utilized as external validation data, following the same data processing methods as the EICU dataset, with evaluation based on metrics including AUC, accuracy, sensitivity, and specificity to assess model generalization performance.

2.4 Model interpretation

SHAP (SHapley Additive exPlanations) is a technique based on game theory's Shapley values (25). It's used to interpret machine learning predictions by dissecting the contribution of each feature. This enhances model transparency and ensures fair decision-making. We employ SHAP to analyze the outcomes of our top-performing machine learning model. This method not only identifies crucial features for optimizing model performance but also provides detailed insights through feature contribution charts, summary plots, and explanations for individual predictions. These tools help us understand the extent of each feature's influence, whether it's positive or negative, and how they collectively impact model outcomes.

2.5 Statistical analysis

For continuous variables, we display using mean and standard deviation, and comparison is done using *t*-tests (or Wilcoxon rank-sum tests); for categorical variables, presentation is in percentages, and comparison is conducted using chi-square tests (or Fisher's exact tests). A p-value <0.05 is deemed statistically significant. All statistical analyses were performed using R version 4.2.3 and Python version 3.11.4.

3 Results

3.1 Baseline characteristics

According to the inclusion and exclusion criteria, our study cohort comprised a total of 6819 patients with heart failure and sepsis. Among these, 3891 cases from the eICU-CRD were used for model construction, while 2928 cases from the MIMIC-IV database were used for external validation. As shown in Figure 1, the screening process is illustrated. During the selection process, patients with ICU stays less than 1 day or under 18 years old were excluded. Subsequently, data processing involved removing outliers and handling missing values. In the eICU-CRD database, 560 patients (14.4%) died within 28 days, compared to 660 patients (22.5%) in the MIMIC-IV database. Differences in baseline characteristics are summarized in Table 1. In the eICU-CRD database, compared to the survival group, patients in the death group exhibited higher age, white blood cell count, neutrophil count, TBIL (total bilirubin), ALT (alanine aminotransferase), BUN (blood urea nitrogen), respiratory rate, fluid balance, and mechanical ventilation time, and lower BMI (body mass index), calcium, blood pressure, and peripheral oxygen pressure. Differences in comorbidities, such as atrial fibrillation, hypertension, and pneumonia, were also observed between the two groups. Additionally, there were differences in medication usage between the two groups, including the use of ACEI/ARB (ACE inhibitors/angiotensin receptor blockers), beta-blockers, furosemide, spironolactone, dobutamine, dopamine, epinephrine, norepinephrine, and phenylephrine.

3.2 Feature selection

We eliminated features with a missing rate exceeding 30%, as demonstrated in Appendix Figure 1. Features with notably high missing rates are primarily found in laboratory tests such as cardiac enzymes, blood gas analysis, lipid profile, and coagulation function. Additionally, guided by the correlation heatmap showing features with correlation coefficients greater than 0.5 and features with VIF exceeding 5, as illustrated in Appendix Figure 2, we conducted further screening. Prior to model construction, we excluded features with high correlation and VIF, including hemoglobin, lymphocyte count, chloride, aspartate aminotransferase, blood urea nitrogen, mean blood pressure, and diastolic blood pressure.

The Boruta method, based on random forests, assesses feature importance by comparing original features with randomly generated "shadow features." We applied Boruta for initial feature selection, as depicted in Figure 2. Green denotes important features included in the model to enhance predictive capability; red represents unimportant features excluded from consideration; yellow indicates features with uncertain importance requiring further investigation. Blue represents shadow features for comparison but not used in model training. Boruta identified 22 initial features, including age, WBC (white blood

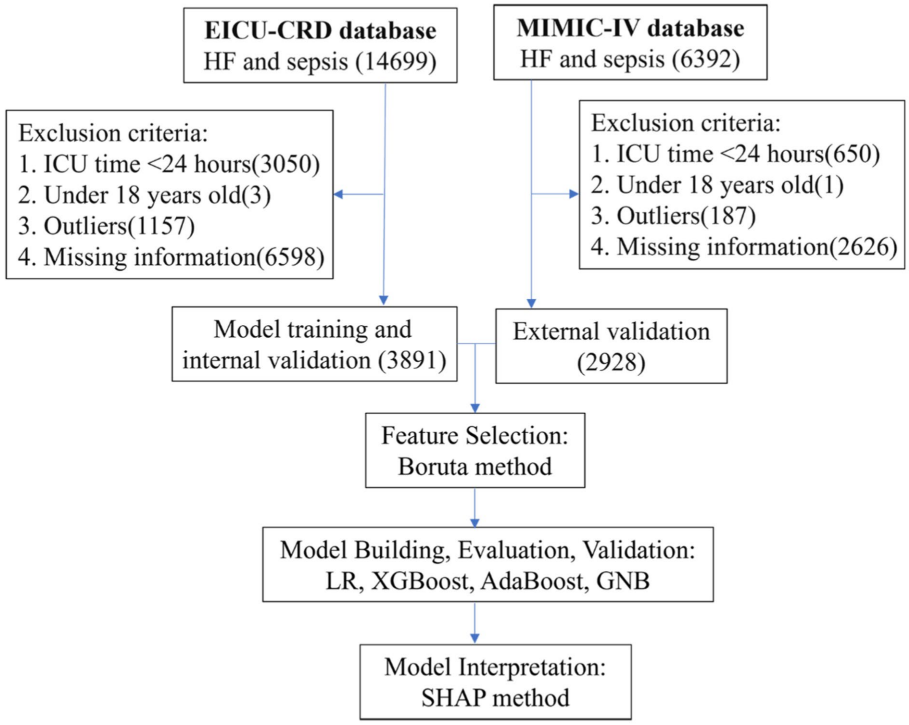

FIGURE 1
Flowchart of patient selection and research methodology.

cell count), NE (neutrophil count), MONO (monocyte count), PLT (platelet count), sodium, calcium, TBIL (total bilirubin), alanine aminotransferase, creatinine, T (temperature), R (respiratory rate), SBP (systolic blood pressure), oxygen saturation, ventilation time, BMI, atrial fibrillation, dopamine, epinephrine, norepinephrine, phenylephrine, and ventilation.

The XGBoost algorithm ranks feature importance based on split frequency and gain in decision trees. Appendix Figure 3 shows our feature importance ranking using XGBoost. The top 10 variables, in descending order of importance, are: age, ventilation, norepinephrine, WBC, TBIL, T, phenylephrine, R, NE, SBP.

3.3 Model construction

This study utilized four binary classification machine learning algorithms, Logistic Regression (LR), eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Gaussian Naive Bayes (GNB), to construct predictive models. Employing the eICU database, we implemented a 10-fold cross-validation technique to establish training and validation sets, followed by evaluation on a separate test cohort. Figure 3 and Table 2 illustrate the performance of these models. The ROC curve (Figure 4A) highlights LR's superior performance, achieving an AUC of 0.746 in the test cohort, compared to XGBoost (0.726), AdaBoost (0.744), and GNB (0.722). Furthermore, LR's calibration curve (Figure 4B) closely aligns with the ideal line, indicating excellent calibration. Decision curve analysis (DCA) (Figure 4C) indicates LR's highest net benefit within the 0–80% threshold range. The precision-recall (PR) curve (Figure 4D) illustrates LR's higher recall at sustained high precision. Additionally, LR demonstrates robust performance across various metrics, including accuracy (0.685), sensitivity (0.666), specificity (0.712), positive predictive value (0.285), negative predictive value (0.914), and F1 score (0.397). Consequently, we selected the LR algorithm for model construction, incorporating 10 variables: age, ventilation, norepinephrine, WBC, TBIL, T, phenylephrine, R, NE, and SBP. Through hyperparameter tuning and grid search optimization, we established the model parameters as follows: tol (convergence measure): 1e-06, penalty (regularization type): l2, max_iter (number of iterations): 100, C (regularization factor): 1.0.

3.4 Model interpretation

This study employs the SHAP method to interpret model results, presenting both SHAP summary plots and SHAP force plots. In the SHAP summary plot, the Y-axis represents features, while the X-axis indicates the impact of features on outcomes. Each point represents a sample, with red indicating high-risk values and blue indicating low-risk values. As shown in Figure 3A, the LR model's feature importance from top to bottom is: age, ventilation, norepinephrine, T, R, TBIL, SBP, WBC, NE, phenylephrine. Older age (red points) correlates with higher SHAP estimated values, predicting an increased risk of mortality. Additionally, higher white blood cell count, total bilirubin, and respiratory rate are associated with increased mortality risk. Patients using ventilation, norepinephrine, and phenylephrine also show increased mortality risk. Furthermore, lower temperature and lower systolic blood pressure are associated with increased mortality risk. In the SHAP force plot, each Shapley value is represented by an arrow, indicating whether it positively (increases) or negatively (decreases) affects the prediction. As illustrated

10.3389/fmed.2024.1410702

TABLE 1 Baseline characteristics of the eICU-CRD and MIMIC-IV databases, categorized by survival and death groups.

	eICU-CRD			MIMIC-IV	
	Survival	Death	*p*	Survival	Death
	(*N* = 3331)	(*N* = 560)		(*N* = 2268)	(*N* = 660)
Age	69.5 (13.9)	74.5 (12.2)	<0.001	71.3 (13.4)	76.7 (11.2)
Gender			0.737		
0	1516 (45.5%)	250 (44.6%)		929 (41.0%)	283 (42.9%)
1	1815 (54.5%)	310 (55.4%)		1339 (59.0%)	377 (57.1%)
BMI	29.9 (7.63)	28.5 (7.12)	<0.001	29.2 (6.32)	28.0 (6.16)
WBC	11.5 (6.79)	15.5 (20.7)	<0.001	13.6 (8.91)	15.2 (10.3)
RBC	3.73 (0.77)	3.69 (0.76)	0.263	3.44 (0.77)	3.54 (0.79)
NE	77.7 (12.5)	81.1 (12.3)	<0.001	79.2 (10.6)	81.6 (12.0)
LYM	11.6 (8.84)	8.65 (8.24)	<0.001	12.1 (8.76)	8.44 (7.73)
MONO	7.36 (3.94)	6.70 (4.20)	0.001	5.03 (3.27)	5.35 (3.24)
PLT	202 (88.5)	200 (104)	0.649	196 (99.9)	212 (121)
Hb	10.9 (2.28)	10.9 (2.21)	0.941	10.2 (2.26)	10.4 (2.26)
Na	138 (5.27)	138 (6.19)	0.047	139 (4.94)	138 (6.21)
K	4.23 (0.78)	4.29 (0.81)	0.077	4.29 (0.73)	4.47 (0.88)
Cl	102 (6.68)	102 (7.60)	0.122	104 (6.66)	102 (7.61)
Ca	8.49 (0.74)	8.33 (0.89)	<0.001	8.29 (0.79)	8.25 (0.94)
GLU	151 (76.1)	155 (88.4)	0.341	151 (74.3)	169 (86.4)
TBIL	0.93 (0.91)	1.21 (1.35)	<0.001	1.01 (1.55)	1.57 (3.70)
ALT	90.1 (403)	179 (480)	<0.001	100 (466)	160 (509)
AST	124 (644)	266 (783)	<0.001	155 (698)	279 (1258)
BUN	36.0 (23.6)	42.3 (26.0)	<0.001	32.9 (25.1)	43.3 (28.6)
Cr	2.11 (2.09)	2.07 (1.49)	0.58	1.69 (1.63)	2.09 (1.75)
T	36.7 (0.79)	36.5 (1.02)	<0.001	36.7 (2.04)	36.5 (2.42)
HR	89.3 (21.0)	90.6 (21.9)	0.169	88.2 (19.6)	91.2 (21.2)
R	21.7 (6.71)	22.8 (7.09)	0.001	19.0 (6.42)	20.9 (6.55)
SBP	125 (28.4)	115 (26.2)	<0.001	118 (24.0)	116 (25.1)
DBP	69.4 (18.5)	64.9 (17.7)	<0.001	67.7 (144)	65.6 (19.2)
MBP	87.7 (19.7)	81.5 (18.3)	<0.001	77.5 (18.0)	79.4 (38.5)
SPO2	96.6 (4.80)	95.5 (7.82)	0.001	97.3 (18.7)	95.8 (5.93)
AF			<0.001		
0	2585 (77.6%)	382 (68.2%)		550 (24.3%)	162 (24.5%)
1	746 (22.4%)	178 (31.8%)		1718 (75.7%)	498 (75.5%)
CKD			0.137		
0	2749 (82.5%)	447 (79.8%)		1355 (59.7%)	385 (58.3%)
1	582 (17.5%)	113 (20.2%)		913 (40.3%)	275 (41.7%)
COPD			0.468		
0	2726 (81.8%)	466 (83.2%)		1121 (49.4%)	368 (55.8%)
1	605 (18.2%)	94 (16.8%)		1147 (50.6%)	292 (44.2%)
Diabetes			0.689		
0	3253 (97.7%)	549 (98.0%)		2061 (90.9%)	577 (87.4%)
1	78 (2.34%)	11 (1.96%)		207 (9.13%)	83 (12.6%)
Hyperlipidemia			0.687		

(Continued)

TABLE 1 (Continued)

	eICU-CRD			MIMIC-IV	
	Survival	Death	*p*	Survival	Death
	(*N* = 3331)	(*N* = 560)		(*N* = 2268)	(*N* = 660)
0	3104 (93.2%)	525 (93.8%)		1670 (73.6%)	415 (62.9%)
1	227 (6.81%)	35 (6.25%)		598 (26.4%)	245 (37.1%)
Hypertension			0.002		
0	2435 (73.1%)	445 (79.5%)		1464 (64.6%)	367 (55.6%)
1	896 (26.9%)	115 (20.5%)		804 (35.4%)	293 (44.4%)
Pneumonia			<0.001		
0	2676 (80.3%)	396 (70.7%)		1055 (46.5%)	287 (43.5%)
1	655 (19.7%)	164 (29.3%)		1213 (53.5%)	373 (56.5%)
ACEI/ARB			<0.001		
0	2839 (85.2%)	537 (95.9%)		1393 (61.4%)	574 (87.0%)
1	492 (14.8%)	23 (4.11%)		875 (38.6%)	86 (13.0%)
Betablockers			<0.001		
0	2132 (64.0%)	409 (73.0%)		341 (15.0%)	236 (35.8%)
1	1199 (36.0%)	151 (27.0%)		1927 (85.0%)	424 (64.2%)
Furosemide			<0.001		
0	1782 (53.5%)	346 (61.8%)		252 (11.1%)	142 (21.5%)
1	1549 (46.5%)	214 (38.2%)		2016 (88.9%)	518 (78.5%)
Spironolactone			0.001		
0	3177 (95.4%)	552 (98.6%)		2082 (91.8%)	624 (94.5%)
1	154 (4.62%)	8 (1.43%)		186 (8.20%)	36 (5.45%)
Dobutamine			<0.001		
0	3180 (95.5%)	505 (90.2%)		2100 (92.6%)	559 (84.7%)
1	151 (4.53%)	55 (9.82%)		168 (7.41%)	101 (15.3%)
Dopamine			0.001		
0	3186 (95.6%)	517 (92.3%)		2128 (93.8%)	554 (83.9%)
1	145 (4.35%)	43 (7.68%)		140 (6.17%)	106 (16.1%)
Epinephrine			<0.001		
0	3240 (97.3%)	520 (92.9%)		1814 (80.0%)	559 (84.7%)
1	91 (2.73%)	40 (7.14%)		454 (20.0%)	101 (15.3%)
Milrinone			0.86		
0	3232 (97.0%)	542 (96.8%)		1351 (59.6%)	214 (32.4%)
1	99 (2.97%)	18 (3.21%)		917 (40.4%)	446 (67.6%)
Norepinephrine			<0.001		
0	2871 (86.2%)	362 (64.6%)		1344 (59.3%)	370 (56.1%)
1	460 (13.8%)	198 (35.4%)		924 (40.7%)	290 (43.9%)
Phenylephrine			<0.001		
0	3215 (96.5%)	495 (88.4%)		2024 (89.2%)	612 (92.7%)
1	116 (3.48%)	65 (11.6%)		244 (10.8%)	48 (7.27%)
Ventilation			<0.001	111 (145)	119 (110)
0	2338 (70.2%)	272 (48.6%)			
1	993 (29.8%)	288 (51.4%)		88 (3.88%)	50 (7.58%)

(Continued)

TABLE 1 (Continued)

	eICU-CRD			MIMIC-IV	
	Survival (*N* = 3331)	Death (*N* = 560)	*p*	Survival (*N* = 2268)	Death (*N* = 660)
Ventilation hour	357 (1664)	819 (2425)	<0.001	2180 (96.1%)	610 (92.4%)
Balance	−520.14 (4968)	1272 (7270)	<0.001	1991 (4988)	2771 (3971)

BMI, Body mass index; WBC, White blood cell; RBC, Red blood cell; NE, Neutrophil; LYM, Lymphocyte; MONO, Monocyte; PLT, Platelet; Hb, Hemoglobin; Na, Sodium; K, Potassium; Cl, Chloride; Ca, Calcium; GLU, Glucose; TBIL, Total bilirubin; ALT, Alanine aminotransferase; AST, Aspartate aminotransferase; BUN, Blood urea nitrogen; Cr, Creatinine; T, Temperature; HR, Heart rate; R, Respiratory rate; SBP, Systolic blood pressure; DBP, Diastolic blood pressure; MBP, Mean blood pressure; SPO2, Oxygen saturation; AF, Atrial fibrillation; CKD, Chronic kidney disease; COPD, Chronic obstructive pulmonary disease; ACEI/ARB, Angiotensin-converting enzyme inhibitor/angiotensin receptor blocker; betablockers, Beta-blockers; balance: fluid balance in the previous 24 h.

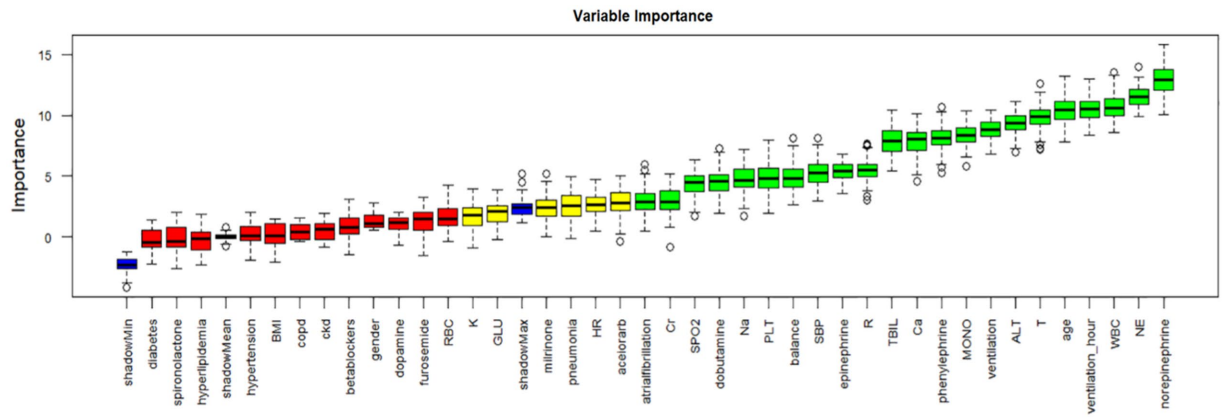

FIGURE 2
Feature selection analyzed by Boruta algorithm.

in Figure 3B, increases in white blood cell count, decreases in temperature, and increases in total bilirubin push the predicted mortality risk higher, while younger age and lower neutrophil count push the predicted mortality risk lower. It's important to note that due to standardization of numerical variables to a mean of 0 and a variance of 1, the data in the plots are not in their original scale.

3.5 External validation

We selected 2928 cases of patients with heart failure and sepsis from the MIMIC-IV database for external validation. In the MIMIC database, these patients had a 28-day in-hospital mortality rate of 22.5%, slightly lower than that of the eICU-CRD database (14.4%). Prior to external validation, we applied the same data processing methods to the MIMIC data as we did to the eICU-CRD data. The validation results revealed an AUC of 0.699 and a Brier score of 0.169. Additionally, the accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score were 0.699, 0.156, 0.403, 0.673, 0.648, and 0.261, respectively. With the AUC difference between the external validation and validation/test sets being less than 0.1, we conclude that the LR model demonstrates favorable stability.

4 Discussion

This study represents the pioneering application of machine learning algorithms to forecast in-hospital mortality among patients with heart failure and sepsis. Our model can be applied to heart failure patients with sepsis upon ICU admission. Our model exhibits exceptional performance in distinguishing between survival and mortality outcomes, coupled with robust calibration and clinical relevance. The utilization of external validation bolsters the model's reliability and generalizability, validating its efficacy across diverse datasets and fortifying the study's scientific robustness and credibility. Leveraging SHAP for visual interpretation of model outcomes enhances the interpretability of predictive results. Furthermore, the model's reliance on a concise and readily accessible set of predictive variables underscores its suitability for clinical deployment. Our model could be integrated into clinical decision support systems within hospitals, especially in ICU. The model would automatically calculate the mortality risk for patients with heart failure complicated by sepsis using routinely collected clinical data, with outputs presented to clinicians via the electronic health record system. This would provide real-time risk assessments to help prioritize care and optimize resource allocation.

Our research findings suggest that the Logistic Regression (LR) model exhibits superior performance in predicting the survival rate of patients with heart failure complicated by sepsis. Moreover, studies indicate that the LR algorithm performs effectively in forecasting various clinical binary outcomes (26, 27). LR offers several advantages, including its simplicity, broad applicability, and straightforward result interpretation, establishing it as a pivotal and dependable modeling technique for binary classification problems (28). However, LR has its limitations; it is sensitive to the quality of feature engineering, vulnerable to outliers, and unable to handle

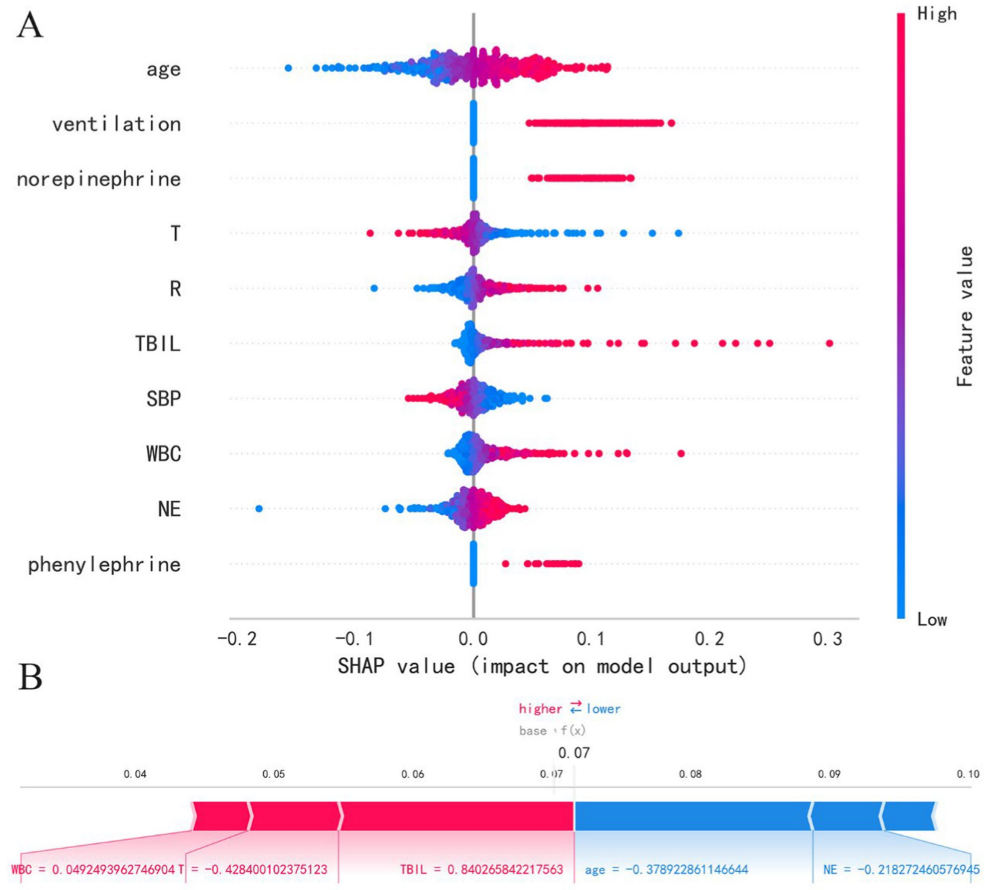

FIGURE 3
(A) SHAP summary plot and **(B)** SHAP force plot.

TABLE 2 Model performance comparison: AUC, accuracy, sensitivity, specificity, PPV, NPV, F1 score, and Brier score.

Models	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	Brierscore
Validation set								
LR	0.746	0.685	0.666	0.712	0.285	0.914	0.397	0.119
XGBoost	0.726	0.718	0.710	0.637	0.268	0.910	0.388	0.116
AdaBoost	0.744	0.672	0.724	0.645	0.265	0.922	0.387	0.221
GNB	0.722	0.683	0.644	0.702	0.269	0.914	0.377	0.160
External validation								
LR	0.699	0.403	0.673	0.648	0.261	0.896	0.376	0.169

LR, Logistic regression; XGBoost, eXtreme gradient boosting; AdaBoost, Adaptive boosting; GNB, Gaussian Naive Bayes; AUC, Area under the curve; PPV, Positive predictive value; NPV, Negative predictive value.

complex non-linear relationships. Moreover, in situations with large feature spaces and predominantly sparse features, LR's performance may be limited, potentially resulting in overfitting (29). Before model construction, we conducted comprehensive data preprocessing, encompassing correlation and multicollinearity assessments, outlier and missing value handling, and data standardization, aiming to enhance the LR model's efficacy. Additionally, external validation has confirmed that the LR model we constructed avoids overfitting and demonstrates reliable generalization ability.

Sepsis and heart failure are common complications in critically ill patients, characterized by complex pathological conditions. Cardiac dysfunction in sepsis, indicated by reduced EF, may accelerate the progression to septic shock by lowering cardiac output and metabolic demand (30). Treatment strategies for sepsis and heart failure often conflict, influenced by varying severity and patient conditions (31). Fluid resuscitation, recommended in sepsis management guidelines, addresses tissue perfusion deficits but may exacerbate congestive symptoms and worsen prognosis in heart failure (32, 33). Our study indicates that higher fluid balance predicts increased mortality in

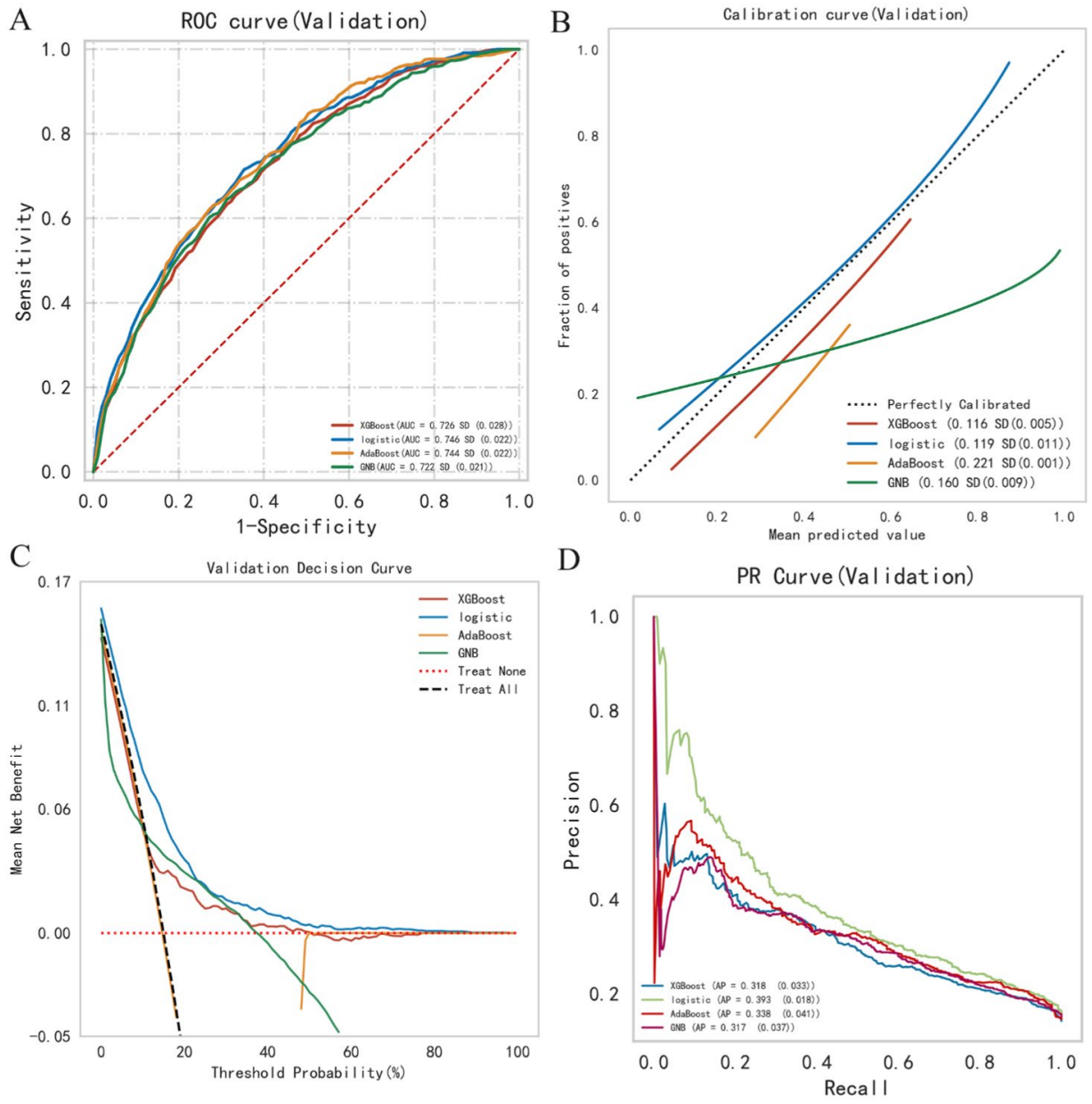

FIGURE 4
Summary plot of machine learning performance evaluation. **(A)** ROC curve, **(B)** calibration plot, **(C)** DCA curve, **(D)** PR curve.

heart failure and sepsis. Singh et al. found that septic patients receiving >3L of fluid experienced reduced EF and higher in-hospital mortality (34). Additionally, other studies have shown that higher fluid balance during hospitalization is associated with increased mortality in patients with heart failure combined with sepsis (35–37). Zhang et al. discovered that higher fluid balance within 24 h of admission is strongly associated with in-hospital mortality in patients with heart failure and sepsis (OR 2.53, 95% CI 1.60–3.99, $p < 0.001$) (31). Due to myocardial edema and oxidative stress, excessive fluid intake is a factor contributing to myocardial injury. For patients with high fluid balance, increased atrial and venous pressures can lead to fluid shift into the interstitium, exacerbating tissue edema, causing tissue distortion and microcirculatory disturbances, thereby resulting in cellular metabolic dysregulation (38, 39). There remains controversy surrounding fluid resuscitation. Duttuluri et al. retrospectively evaluated heart failure patients with severe sepsis, finding increased in-hospital mortality and intubation rates in the hypotensive subgroup receiving inadequate fluid (<30 mL/kg) (40).

Our research reveals that patients with elevated respiratory rates, hypotension, and those necessitating interventions such as norepinephrine, phenylephrine, or ventilation, exhibit a heightened risk of mortality prediction. Norepinephrine and phenylephrine are typically employed to augment cardiac contractility and blood pressure for organ perfusion maintenance, while ventilation is essential for respiratory support. This elevated predictive risk likely reflects the severity of patients' conditions and the associated potential hazards they face. Moreover, it underscores the necessity for prompt and assertive therapeutic interventions tailored to these patients,

alongside vigilant monitoring and comprehensive support measures. Sepsis guidelines recommend norepinephrine as the first-line vasopressor for sepsis and septic shock (33). De Backer et al. found in their study that among 280 cases of cardiogenic shock patients, norepinephrine was more effective than dopamine, significantly reducing the 28-day mortality rate ($p = 0.03$) (41). Additionally, a meta-analysis from 2015 also indicated that in the treatment of septic shock, norepinephrine, compared to dopamine, could lower the mortality rate (RR: 0.89; 95% CI: 0.81–0.98) (42). Additionally, there are studies indicating that compared to adrenaline, norepinephrine carries a lower risk of tachycardia [29] and is associated with reduced mortality risk (43, 44).

With the exacerbation of an aging society, the incidence of sepsis among the elderly is gradually increasing, making it one of the leading causes of mortality in this demographic (45). Age has been demonstrated as an independent risk factor for mortality in sepsis patients, with mortality rates showing a linear increase with advancing age (46). Our research findings indicate that advanced age is associated with a higher predictive risk of mortality in patients with sepsis complicated by heart failure. Elderly patients commonly exhibit compromised immune function, diminished organ reserve, and a higher prevalence of comorbidities such as diabetes and coronary artery disease compared to younger counterparts (47). Sepsis in this demographic frequently presents and swiftly evolves into multi-organ failure. De Matteis et al. studied 6930 elderly patients with heart failure and found that in-hospital mortality increased with advancing age, with infection correlating with an elevated risk of in-hospital death (48). We also found that elevated levels of white blood cells and norepinephrine were associated with poor outcomes. In bacterial and fungal infections, elevated blood neutrophil levels serve as early and sensitive indicators of inflammation (49). Elevated white blood cell or neutrophil counts in sepsis patients suggest immune system activation and intensified inflammatory response, potentially indicating an excessively activated inflammatory state associated with increased mortality risk. Heightened vigilance and proactive therapeutic interventions are warranted to mitigate inflammation and prevent further deterioration in such cases. Additionally, low body temperature and elevated total bilirubin increase the risk of mortality assessment. Observing low body temperature or elevated total bilirubin (TBIL) in sepsis patients may suggest a severe condition and poor prognosis. Low body temperature could indicate suppressed inflammatory response or impaired metabolic function, compromising the body's resistance to infection. Elevated TBIL may signify impaired liver function, possibly due to infection or inflammation.

This study has several limitations. Firstly, we acknowledge that the quality and completeness of the data in the MIMIC-IV and eICU-CRD databases may have certain limitations, especially with the potential absence of key clinical variables (such as ejection fraction or NT-proBNP), which could affect the accuracy of the model's predictions. Future studies will need to incorporate more comprehensive data to improve the model and conduct further validation to enhance its accuracy. Secondly, as this is a retrospective study, the data primarily comes from ICU patients, which may introduce selection bias, limiting the model's broader applicability to other clinical settings. Therefore, we suggest that future research validate the model using multi-center data to reduce selection bias and improve its generalizability. Additionally, the imbalance between survival and death in the dataset may affect the model's performance in predicting mortality. Lastly, the data were collected at different time points, leading to potential temporal discrepancies, which may cause data drift and result in inconsistent model performance across different periods. Thus, future research should validate and adjust the model using data from various timeframes to address these challenges and ensure the model's robustness in different temporal and clinical settings.

5 Conclusion

In this study, we constructed a machine learning model to predict 28-day all-cause mortality in ICU patients with heart failure complicated by sepsis. The final Logistic Regression model incorporates commonly used clinical indicators such as age, mechanical ventilation, respiratory rate, blood pressure, white blood cell count, and vasopressor use. This combination of variables enables the model to predict short-term mortality risk early, upon ICU admission, providing a clinical alert for high-risk patients and assisting clinicians in more effectively allocating medical and nursing resources. Furthermore, the model's generalizability and potential clinical utility were validated across two large ICU databases (eICU-CRD and MIMIC-IV). Despite its strong predictive performance, further updates and validations with larger, multicenter patient cohorts are required to enhance the model's generalizability and practical application in broader clinical settings.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

QZ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LX: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. WH: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft,

Writing – review & editing. XL: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. XH: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to acknowledge the contributions of specific colleagues and institutions that aided the efforts of the authors in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1410702/full#supplementary-material

References

1. Hajj J, Blaine N, Salavaci J, Jacoby D. The "Centrality of Sepsis": A review on incidence, mortality, and cost of care. *Healthcare*. (2018) 6. doi: 10.3390/healthcare6030090

2. Napolitano LM. Sepsis 2018: definitions and guideline changes. *Surg Infect*. (2018) 19:117–25. doi: 10.1089/sur.2017.278

3. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study. *Lancet*. (2020) 395:200–11. doi: 10.1016/S0140-6736(19)32989-7

4. Vincent JL, Marshall JC, Namendys-Silva SA, Francois B, Martin-Loeches I, Lipman J, et al. Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. *Lancet Respir Med*. (2014) 2:380–6. doi: 10.1016/S2213-2600(14)70061-X

5. Liu V, Escobar GJ, Greene JD, Soule J, Whippy A, Angus DC, et al. Hospital deaths in patients with sepsis from 2 independent cohorts. *JAMA*. (2014) 312:90–2. doi: 10.1001/jama.2014.5804

6. Liang L, Moore B, Soni A. National Inpatient Hospital Costs: the Most expensive conditions by payer, 2017 In: Healthcare cost and utilization project (HCUP) statistical briefs. Rockville, MD: Agency for Healthcare Research and Quality (US) (2006)

7. Gupta AK, Tomasoni D, Sidhu K, Metra M, Ezekowitz JA. Evidence-based Management of Acute Heart Failure. *Can J Cardiol*. (2021) 37:621–31. doi: 10.1016/j.cjca.2021.01.002

8. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. *Eur J Heart Fail*. (2020) 22:1342–56. doi: 10.1002/ejhf.1858

9. Pandolfi F, Guillemot D, Watier L, Brun-Buisson C. Trends in bacterial sepsis incidence and mortality in France between 2015 and 2019 based on National Health Data System (Systeme national des donnees de Sante (SNDS)): a retrospective observational study. *BMJ Open*. (2022) 12:e058205. doi: 10.1136/bmjopen-2021-058205

10. Jentzer JC, van Diepen S, Hollenberg SM, Lawler PR, Kashani KB. Shock severity assessment in cardiac intensive care unit patients with Sepsis and mixed septic-cardiogenic shock. *Mayo Clin Proc Innov Qual Outcomes*. (2022) 6:37–44. doi: 10.1016/j.mayocpiqo.2021.11.008

11. Alon D, Stein GY, Korenfeld R, Fuchs S. Predictors and outcomes of infection-related hospital admissions of heart failure patients. *PLoS One*. (2013) 8:e72476. doi: 10.1371/journal.pone.0072476

12. Walker AMN, Drozd M, Hall M, Patel PA, Paton M, Lowry J, et al. Prevalence and predictors of Sepsis death in patients with chronic heart failure and reduced left ventricular ejection fraction. *J Am Heart Assoc*. (2018) 7:e009684. doi: 10.1161/JAHA.118.009684

13. Kopczynska M, Sharif B, Cleaver S, Spencer N, Kurani A, Lee C, et al. Red-flag sepsis and SOFA identifies different patient population at risk of sepsis-related deaths on the general ward. *Medicine*. (2018) 97:e13238. doi: 10.1097/MD.0000000000013238

14. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/north American multicenter study. *JAMA*. (1993) 270:2957–63. doi: 10.1001/jama.270.24.2957

15. Yu H, Nie L, Liu A, Wu K, Hsein YC, Yen DW, et al. Combining procalcitonin with the qSOFA and sepsis mortality prediction. *Medicine*. (2019) 98:e15981. doi: 10.1097/MD.0000000000015981

16. Olejarova M, Dobisova A, Suchankova M, Tibenska E, Szaboova K, Koutun J, et al. Vitamin D deficiency - a potential risk factor for sepsis development, correlation with inflammatory markers, SOFA score and higher early mortality risk in sepsis. *Bratislavske Lekarske Listy*. (2019) 120:284–90. doi: 10.4149/BLL_2019_040

17. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med*. (2020) 18:462. doi: 10.1186/s12967-020-02620-5

18. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl*. (2004) 86:334–8. doi: 10.1308/147870804290

19. Deo RC. Machine learning in medicine. *Circulation*. (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593

20. Lee A, Taylor P, Kalpathy-Cramer J, Tufail AJO. Machine learning has arrived! *Ophthalmol Retina*. (2017) 124:1726–8. doi: 10.1016/j.ophtha.2017.08.046

21. Lubo-Robles D, Devegowda D, Jayaram V, Bedle H, Marfurt KJ, Pranter MJ. (2020). Machine learning model interpretability using SHAP values: application to a seismic facies classification task. SEG Int Expos Ann Meet.

22. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data*. (2018) 5:180178. doi: 10.1038/sdata.2018.178

23. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA. Mark RJPAoahpocm. *Mimic-iv*. (2020):49–55.

24. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform*. (2019) 20:492–503. doi: 10.1093/bib/bbx124

25. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *arXiv*. (2017) 30:7874. doi: 10.48550/arXiv.1705.07874

26. Liu L. (2018). Research on logistic regression algorithm of breast cancer diagnose data by machine learning. 2018 international conference on Robots & Intelligent System (ICRIS).

27. Nishadi AST. Predicting heart diseases in logistic regression of machine learning algorithms by Python Jupyterlab. *Comput Sci Med*. (2019) 3:1–6.

28. Maalouf M. Logistic regression in data analysis: an overview. *Int. J. Data Anal Tech Strat*. (2011) 3:281–99. doi: 10.1504/IJDATS.2011.041335

29. Bisong E. Building machine learning and deep learning models on Google cloud platform. Berlin: Springer (2019).

30. Ferro TN, Goslar PW, Romanovsky AA, Petersen SR. Smoking in trauma patients: the effects on the incidence of sepsis, respiratory failure, organ failure, and mortality. *J Trauma*. (2010) 69:308–12. doi: 10.1097/TA.0b013e3181e1761e

31. Zhang B, Guo S, Fu Z, Wu N, Liu Z. Association between fluid balance and mortality for heart failure and sepsis: a propensity score-matching analysis. *Anesthesiology*. (2022) 22:324. doi: 10.1186/s12871-022-01865-5

32. Levy MM, Evans LE, Rhodes A. The surviving Sepsis campaign bundle: 2018 update. *Eur J Intensive Care Med*. (2018) 44:925–8. doi: 10.1007/s00134-018-5085-0

33. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Med*. (2017) 43:304–77. doi: 10.1007/s00134-017-4683-6

34. Singh H, Iskandir M, Sachdev S, Simmons B, Rabines A, Hassen GW. The effect of initial volume resuscitation for Sepsis in patients with congestive heart failure: is it associated with higher mortality. *J Card Fail*. (2016) 22:S54–5. doi: 10.1016/j.cardfail.2016.06.161

35. Shen Y, Huang X, Zhang WJCC. Association between fluid intake and mortality in critically ill patients with negative fluid balance: a retrospective cohort study. *Crit Care*. (2017) 21:1–8. doi: 10.1186/s13054-017-1692-3

36. Boyd JH, Forbes J, Nakada TA, Walley KR, Russell JA. Fluid resuscitation in septic shock: a positive fluid balance and elevated central venous pressure are associated with increased mortality. *Crit Care Med*. (2011) 39:259–65. doi: 10.1097/CCM.0b013e3181feeb15

37. Micek ST, McEvoy C, McKenzie M, Hampton N, Doherty JA, Kollef MH. Fluid balance and cardiac function in septic shock as predictors of hospital mortality. *Crit Care*. (2013) 17:R246–9. doi: 10.1186/cc13072

38. Marik PE, Byrne L, van Haren F. Fluid resuscitation in sepsis: the great 30 mL per kg hoax. *J Thorac Dis*. (2020) 12:S37–47. doi: 10.21037/jtd.2019.12.84

39. Prowle JR, Kirwan CJ, Bellomo R. Fluid management for the prevention and attenuation of acute kidney injury. *Nat Rev Nephrol*. (2014) 10:37–47. doi: 10.1038/nrneph.2013.232

40. Duttuluri M, Rose K, Shapiro J, Mathew J, Jean R, Kurtz S, et al. Fluid resuscitation dilemma in patients with congestive heart failure presenting with severe sepsis/septic shock. D45 critical care: Circulatory hemodymanics, shock, cardiovascular disease, and fluid management. New York, NY: American Thoracic Society (2016). A7048 p.

41. De Backer D, Biston P, Devriendt J, Madl C, Chochrad D, Aldecoa C, et al. Comparison of dopamine and norepinephrine in the treatment of shock. *N Engl J Med*. (2010) 362:779–89. doi: 10.1056/NEJMoa0907118

42. Avni T, Lador A, Lev S, Leibovici L, Paul M, Grossman A. Vasopressors for the treatment of septic shock: systematic review and Meta-analysis. *PLoS One*. (2015) 10:e0129305. doi: 10.1371/journal.pone.0129305

43. Léopold V, Gayat E, Pirracchio R, Spinar J, Parenica J, Tarvasmäki T, et al. Epinephrine and short-term survival in cardiogenic shock: an individual data meta-analysis of 2583 patients. *Eur J Intensive Care Med*. (2018) 44:847–56. doi: 10.1007/s00134-018-5222-9

44. Léopold V, Gayat E, Pirracchio R, Spinar J, Parenica J, Tarvasmäki T, et al. Correction to: epinephrine and short-term survival in cardiogenic shock: an individual data meta-analysis of 2583 patients. *Intensive Care Med*. (2018) 44:2022–3. doi: 10.1007/s00134-018-5372-9

45. Rowe TA, McKoy JM. Sepsis in Older Adults. *Infect Dis Clin N Am*. (2017) 31:731–42. doi: 10.1016/j.idc.2017.07.010

46. Martin GS, Mannino DM, Moss M. The effect of age on the development and outcome of adult sepsis. *Crit Care Med*. (2006) 34:15–21. doi: 10.1097/01.ccm.0000194535.82812.ba

47. Carbajal-Guerrero J, Cayuela-Dominguez A, Fernandez-Garcia E, Aldabo-Pallas T, Marquez-Vacaro JA, Ortiz-Leyba C, et al. Epidemiology and long-term outcome of sepsis in elderly patients. *Med Intensiva*. (2014) 38:21–32. doi: 10.1016/j.medin.2012.12.006

48. de Matteis G, Covino M, Burzo ML, Della Polla DA, Franceschi F, Mebazaa A, et al. Clinical characteristics and predictors of in-hospital mortality among older patients with acute heart failure. *JCM*. (2022) 11:439. doi: 10.3390/jcm11020439

49. Honda T, Uehara T, Matsumoto G, Arai S, Sugano MJCCA. Neutrophil left shift and white blood cell count as markers of bacterial infection. *Clin Chim Acta*. (2016) 457:46–53. doi: 10.1016/j.cca.2016.03.017

frontiers | Frontiers in Medicine

TYPE Systematic Review
PUBLISHED 16 October 2024
DOI 10.3389/fmed.2024.1491358

OPEN ACCESS

EDITED BY
Qinghe Meng,
Upstate Medical University, United States

REVIEWED BY
Mercedes Sanchez Barba,
University of Salamanca, Spain
Khan Md Hasib,
Bangladesh University of Business and Technology, Bangladesh

*CORRESPONDENCE
Valery V. Likhvantsev
lik0704@gmail.com

RECEIVED 04 September 2024
ACCEPTED 08 October 2024
PUBLISHED 16 October 2024

CITATION
Yadgarov MY, Landoni G, Berikashvili LB, Polyakov PA, Kadantseva KK, Smirnova AV, Kuznetsov IV, Shemetova MM, Yakovlev AA and Likhvantsev VV (2024) Early detection of sepsis using machine learning algorithms: a systematic review and network meta-analysis.
Front. Med. 11:1491358.
doi: 10.3389/fmed.2024.1491358

Early detection of sepsis using machine learning algorithms: a systematic review and network meta-analysis

Mikhail Ya Yadgarov[1], Giovanni Landoni[2,3], Levan B. Berikashvili[1], Petr A. Polyakov[1], Kristina K. Kadantseva[1], Anastasia V. Smirnova[1], Ivan V. Kuznetsov[1], Maria M. Shemetova[1], Alexey A. Yakovlev[1] and Valery V. Likhvantsev[1,4]*

[1]Federal Research and Clinical Centre of Intensive Care Medicine and Rehabilitology, Moscow, Russia, [2]Department of Anaesthesia and Intensive Care, IRCCS San Raffaele Scientific Institute, Milan, Italy, [3]Department of Anesthesiology, Vita-Salute San Raffaele University, Milan, Italy, [4]Department of Anesthesiology, I.M. Sechenov First Moscow State Medical University, Moscow, Russia

Background: With machine learning (ML) carving a niche in diverse medical disciplines, its role in sepsis prediction, a condition where the 'golden hour' is critical, is of paramount interest. This study assesses the factors influencing the efficacy of ML models in sepsis prediction, aiming to optimize their use in clinical practice.

Methods: We searched Medline, PubMed, Google Scholar, and CENTRAL for studies published from inception to October 2023. We focused on studies predicting sepsis in real-time settings in adult patients in any hospital settings without language limits. The primary outcome was area under the curve (AUC) of the receiver operating characteristic. This meta-analysis was conducted according to PRISMA-NMA guidelines and Cochrane Handbook recommendations. A Network Meta-Analysis using the CINeMA approach compared ML models against traditional scoring systems, with meta-regression identifying factors affecting model quality.

Results: From 3,953 studies, 73 articles encompassing 457,932 septic patients and 256 models were analyzed. The pooled AUC for ML models was 0.825 and it significantly outperformed traditional scoring systems. Neural Network and Decision Tree models demonstrated the highest AUC metrics. Significant factors influencing AUC included ML model type, dataset type, and prediction window.

Conclusion: This study establishes the superiority of ML models, especially Neural Network and Decision Tree types, in sepsis prediction. It highlights the importance of model type and dataset characteristics for prediction accuracy, emphasizing the necessity for standardized reporting and validation in ML healthcare applications. These findings call for broader clinical implementation to evaluate the effectiveness of these models in diverse patient groups.

Systematic review registration: https://inplasy.com/inplasy-2023-12-0062/, identifier, INPLASY2023120062.

KEYWORDS

sepsis, machine learning, network meta-analysis, decision trees, neural networks

1 Introduction

Sepsis is a critical medical condition characterized by a substantial risk of mortality (1). Prompt identification of sepsis is crucial for the successful treatment of this life-threatening condition. Adhering to the 'golden hour' principle, which suggests that patient outcomes are significantly improved when treatment is initiated within the first hour following diagnosis, is pivotal for enhancing patient survival rates. Concurrently, there is a robust endorsement for employing systematic screening procedures for early sepsis identification (2).

The accuracy of current clinical scales and diagnostic methodologies in detecting and predicting sepsis seems to be significantly suboptimal, leading to delays in therapeutic interventions (3–5). Despite the widespread use of traditional sepsis scoring systems, such as SOFA, NEWS, MEWS, SIRS, and SAPS II, these tools exhibit several limitations, including their reliance on static thresholds and suboptimal predictive performance. As a result, traditional sepsis scoring systems often lack the sensitivity and specificity required for timely, accurate sepsis detection.

This gap underscores the urgent need for more precise and reliable diagnostic and prognostic tools. In this regard, there is a shift in focus towards innovative approaches such as machine learning (6–13). Particularly, right-aligned models are drawing significant attention for their capacity to predict the development of sepsis hours before its clinical confirmation (14). Evidence increasingly suggests that machine learning methodologies offer a distinct advantage over traditional sepsis scoring systems (6).

To date, three meta-analyses have been conducted in this area of study (6, 15, 16), with one demonstrating the superiority of machine learning over traditional clinical scales in sepsis prognosis (6). In second research, the evidence presented lacks robustness (16), whereas in a third investigation, the focus was solely on the comparative assessment of different machine learning methodologies (15). However, the significant clinical heterogeneity, not entirely unambiguous diagnostic criteria, diverse prognostic time frames, and differing approaches to data preprocessing and model development across patient populations preclude definitive conclusions about the prognostic efficacy of these machine learning models.

In response to these challenges, our objective was to conduct a pioneering network meta-analysis to address this heterogeneity and to surpass the confines of previous research. Through meta-regression, we aimed to identify key factors that influence the effectiveness of predictive models, thereby guiding the development of an optimal model for sepsis prognosis, tailored to the complexities of clinical scenarios.

2 Materials and methods

This study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions (PRISMA-NMA) guidelines (17) and the Cochrane Handbook recommendations (18). The study protocol was registered with the International Platform of Registered Systematic Review and Meta-analysis Protocols (INPLASY) under the registration number INPLASY 2023120062 (doi: 10.37766/inplasy2023.12.0062). The completed PRISMA-NMA checklist is presented in Supplementary Table S1.

2.1 Search strategy

We performed a systematic search of the literature across Medline, PubMed, Google Scholar, and the Cochrane Central Register of Controlled Trials (CENTRAL) from inception to October 2023. The search was conducted by two independent investigators. Backward and forward citation tracking was also employed to identify additional studies, leveraging the Litmaps service (19). No language restrictions were applied. Details of the search strategy, including full queries, are provided in Supplementary Appendix A.

2.2 Eligibility criteria and study selection

Following the automatic removal of duplicate records, two independent researchers screened the remaining studies for eligibility. We applied the PICOS (Population, Intervention, Comparator, Outcome, and Study design) framework to guide study selection (Supplementary Appendix B).

Studies were considered eligible if they focused on real-time prediction of sepsis onset (right alignment (14)) in adult patients across any hospital setting. Both prospective and retrospective diagnostic test accuracy studies were included. The target condition was the onset of sepsis, defined by Sepsis-3 criteria (20) or other operational definitions provided by the authors (Supplementary Table S5).

Studies were excluded if they met one of the following criteria: (1) were review articles, case reports or case series without control groups; (2) had no sepsis definition criteria; (3) reported no AUC for sepsis development; (4) reported AUC for other outcome (e.g., reported data on mortality only); (5) focused on pediatric patients; (6) reported no data on patient cohort (sample size, age, sex, etc.); (7) were published as conference papers or preprints only.

Any disagreement was solved by consultation until consensus was reached. Divergences were resolved by consensus with the involvement of the supervisor.

2.3 Outcome measures and data extraction

A standardized data collection form was developed specifically for this review. Three independent authors used this form to systematically evaluate the full text, supplemental materials, and additional files of all included studies. Data extraction was performed independently by three authors, with any discrepancies resolved through discussion to achieve consensus.

Extracted information encompassed: (1) Basic study details such as the first author, publication year, country, journal, study design, data collection period, mean age, sex, hospital mortality, prediction method, and sample size; (2) ML model characteristics: data source, prediction model, sepsis definition criteria, department, prediction window, external validation, imputation, features; (3) Outcome data: area under the curve of the receiver operating characteristic (AUC) as performance metric.

ML prediction models were grouped as detailed in Supplementary Appendix C.

In an attempt to reduce the number of comparisons, when multiple models from the same group were used in a single article (employing different factor selection and optimization methods), the analysis focused on the highest AUC value. The standard deviation (SD) for AUC was either extracted directly from the article, requested from the authors, converted from the 95% confidence interval (CI) according to the Cochrane Handbook recommendations, or imputed using the Iterative Imputer algorithm based on a Bayesian regression model (Python's sklearn library).

Data on the AUC metrics for traditional scoring systems were also extracted if available. These systems included SOFA (Sequential Organ Failure Assessment), qSOFA (quick SOFA), NEWS/NEWS2 (National Early Warning Score), MEWS (Modified Early Warning Score), SAPS II (Simplified Acute Physiology Score), and SIRS (Systemic Inflammatory Response Syndrome). In cases where multiple traditional scoring systems were used, the best metric was considered.

2.4 Data analysis and synthesis

Traditional meta-analysis was conducted to calculate pooled AUCs. Inter-study heterogeneity was evaluated using the I-squared (I2) statistic and the Cochrane Q test; random-effects model (restricted maximum–likelihood, REML) was used. Statistical significance was set at 0.05 for hypothesis testing. We conducted a meta-regression analysis, leveraging the REML random-effects model, to ascertain if the AUC metrics might be affected by covariates such as study design and ML model characteristics (21). All covariates were first tested in a univariate model, significant covariates were then considered for a multivariable model. The results of the meta-regression were graphically represented using bubble-plots.

We also conducted a frequentist, random-effects Network Meta-Analysis (NMA) using CINeMA (Confidence in Network Meta-Analysis) approach (22), CINeMA software (23), ROB-MEN web application (24) and STATA 17.0 (StataCorp, College Station, TX) software. Articles were included in the NMA if they compared any two ML models with different ML models or any ML model with a traditional scoring system. The Mean Difference (MD) with corresponding 95% CI was calculated for AUCs. Results of NMA were presented using network plots, league tables, contribution tables and NMA forest plots. To assess between-study heterogeneity, we utilized Bayesian NMA with $\tau 2$ calculation. A $\tau 2$ value exceeding the clinically important effect size ($MD \geq 0.15$) indicated significant heterogeneity.

2.5 Internal validity and risk of bias assessment

The internal validity and risk of bias were assessed by three independent reviewers (MY, AS, IK) using the 'Quality Assessment of Diagnostic Accuracy Studies' (QUADAS-2) tool (25) combined with an adapted version of the 'Joanna Briggs Institute Critical Appraisal checklist for analytical cross-sectional studies' (26) (Supplementary Table S2). Publication bias and small-study effects were assessed using Bayesian NMA meta-regression and funnel plot analysis (for comparisons with 10 or more studies). The certainty of evidence was assessed with GRADE methodology integrated in CINeMA approach. We conducted a sensitivity analysis using studies with low to moderate risk of bias.

3 Results

3.1 Study characteristics

The initial literature search identified 3,953 studies from various databases, with an additional 24 studies from other sources (Figure 1). After removing duplicates and abstract screening, 97 papers underwent eligibility screening. A total of 256 models from 73 studies (457,932 septic patients) were included (14, 27–98) with major exclusions list presented in Supplementary Table S1. The specialty journal with the largest number of articles was Critical Care Medicine (39, 41, 53, 62, 87).

Most of the studies included in the analysis were conducted in the ICU ($n = 49$; 67.1%), followed by hospital wards ($n = 12$; 16.4%) and emergency departments (ED, $n = 9$; 12.3%) (Supplementary Tables S4–S8). The median sepsis prevalence across the studies was 14.3% (IQR 7.3–32.4%), with the mean patient age ranging from 35 to 70 years. The median (IQR) mortality rate was 2.3% (6.9–14.8%). Sepsis was most frequently defined by the Sepsis-3 criteria (57.5%), with other definitions including Sepsis-2, ICD-9, ICD-10, and SIRS. Prediction windows varied widely, ranging from immediate (0 h) to 7 days. External validation was performed in 16 studies (21.9%), and imputation techniques were employed in 44 studies (60.3%). Notably, 53% of studies utilized public datasets such as MIMIC, eICU, and the Computing in Cardiology Challenge 2019, while the remaining studies relied on proprietary hospital datasets.

3.2 Pooled AUCs

The pooled AUC for machine learning models was 0.825 (95% CI 0.809–0.840, $p < 0.001$) across 73 studies (Supplementary Table S9). In comparison, the AUC for the SOFA score was 0.667 (95% CI 0.586–0.748) across 17 studies, for qSOFA 0.612 (95% CI 0.574–0.650) across 16 studies, for NEWS/NEWS2 0.719 (95% CI 0.674–0.764) across 9 studies, for MEWS 0.651 (95% CI 0.612–0.690) across 12 studies, for SIRS 0.666 (95% CI 0.643–0.688) across 19 studies, and for SAPS II 0.662 (95% CI 0.589–0.736) across 2 studies (all $p < 0.001$,

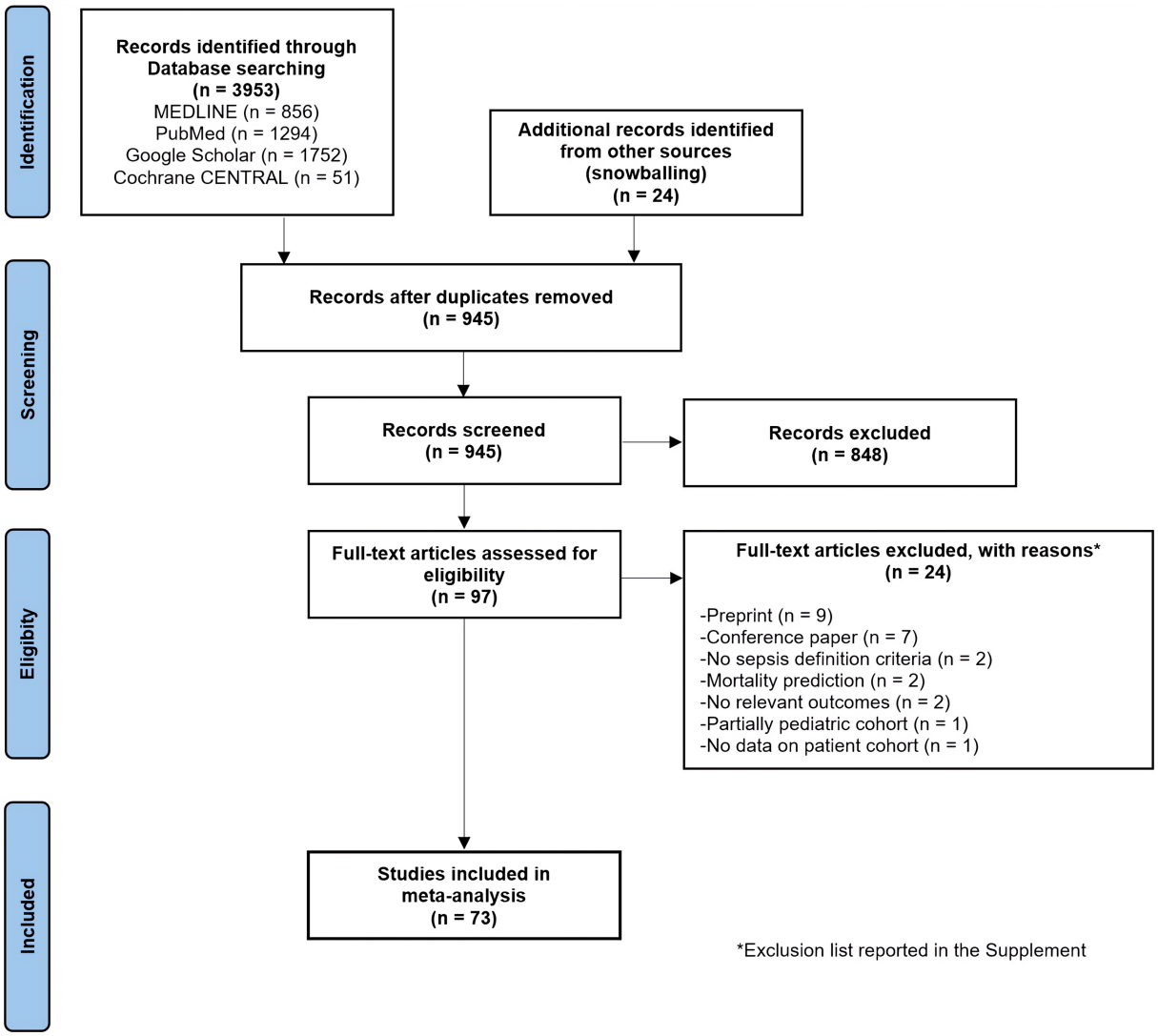

FIGURE 1
PRISMA flow diagram for study selection.

Supplementary Table S9). Heterogeneity across studies was high ($I^2 > 95\%$, $p < 0.001$).

3.3 Network meta-analysis

3.3.1 ML vs. scoring systems

All ML models exhibited a significant performance advantage over traditional scoring systems when performing a NMA (Figures 2, 3, Supplementary Tables S10–S12).

> *A network plot is a visual tool in network meta-analyses, showing interventions (groups) as nodes and their direct comparisons as connecting lines. This visual tool helps in understanding the complex relationships and the extent of evidence available for each comparison in the network meta-analysis.*

> *Network of retrospective diagnostic test accuracy studies comparing the AUCs of various machine learning models and best of scoring system used. The size of nodes and width of the edges are proportional to the number of studies. The colors of edges and nodes refer to the average risk of bias: low (green), moderate (yellow), and high (red).*

3.3.2 ML models

As indicated by the NMA results, Neural Network Models (NNM) and Decision Tree (DT) models exhibited the highest AUC metrics (Figure 3, Supplementary Tables S11, S12).

3.4 Meta-regression

In the multivariable model, only ML model type, dataset type (with non-freely available hospital datasets showing higher AUCs),

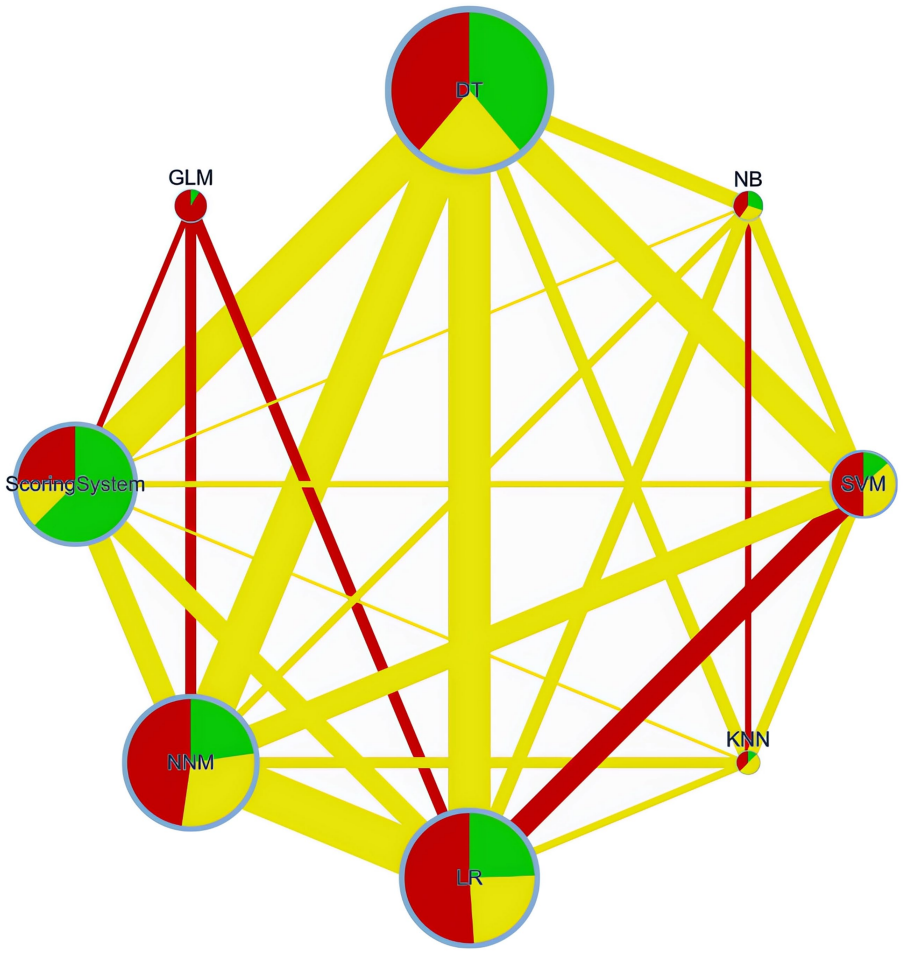

FIGURE 2
Network plot. DT, Decision Tree; NNM, Neural Network Model; SVM, Support Vector Machine; LR, Logistic Regression; NB, Naïve Bayes; GLM, Generalized Linear Model; KNN, K-Nearest Neighbors.

and prediction window (showing a negative association) had significant contributions to the AUC. (Supplementary Table S13, Supplementary Figures S1, S2).

3.5 Risk of bias and GRADE assessment

The overall risk of bias of the 73 enrolled studies was judged as 'low' in 29 studies, with 'some concerns' in 14 studies and 'high' in 30 studies (Supplementary Figure S3). The main sources of bias identified were insufficient description of the study population and data sources, along with the use of different sepsis definitions.

Risk of bias bar chart is presented in Supplementary Figure S4. Publication bias and small-study effects assessment results are summarized in Supplementary Table S14 and Supplementary Figure S5. Between-study variance was not significant ($\tau^2 = 0.095$, with the clinically important effect size stated as 0.1). Contribution matrices are presented in Supplementary Tables S15, S16.

The CINeMA ratings can be found in Supplementary Figure S6. The level of evidence supporting the superiority of ML models over traditional scoring systems was categorized as 'low'.

4 Discussion

4.1 Key findings

This network meta-analysis is the first to comprehensively evaluate the performance of (ML) models in sepsis prediction, demonstrating that ML algorithms, particularly neural network models and decision trees, significantly outperform traditional scoring systems. These findings underscore the enhanced ability of ML models to analyse and interpret complex clinical data, pointing to a potential paradigm shift in sepsis prediction strategies.

A critical aspect of our findings relates to the impact of the type of ML model and the nature of the dataset on model performance. The choice of the ML model itself emerged as a significant determinant of model performance in our study. This finding indicates that the inherent characteristics and algorithms of different ML models substantially influence their ability to predict sepsis effectively. Models utilizing freely available datasets exhibited lower AUCs, which could be a result of overfitting in models based on non-freely available hospital datasets.

Another key finding from our study is the temporal dynamics in prediction accuracy. We observed a negative association

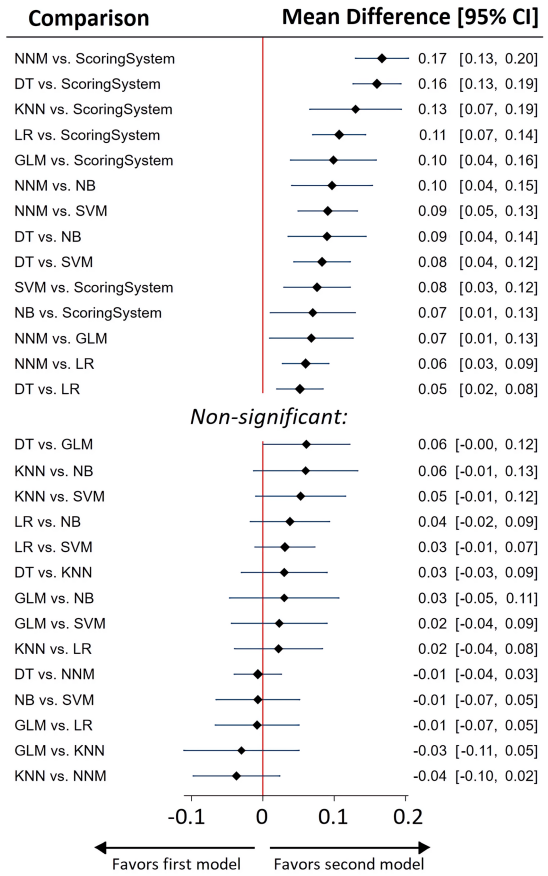

FIGURE 3
Network meta-analysis summary forest plot for predictive efficacy of various ML models and best of scoring system used for sepsis prediction.

between prediction window and AUC, indicating that ML models are more accurate in short-term than in long-term sepsis prediction.

Interestingly, factors such as the size of the training dataset, sepsis prevalence, the department in which the study was conducted, the presence of imputation and external validation, the use of laboratory indicators, and the number of predictors did not significantly influence the quality of the predictive models.

4.2 Relationship with previous studies

The results of our systematic review and meta-analysis can be compared with those of 3 previous meta-analyses. Islam et al. in 2019 were the first demonstrated that ML approach outperforms existing scoring systems in predicting sepsis (6). Fleuren et al. (15) suggested that ML models can accurately predict the onset of sepsis with good discrimination in retrospective cohorts, and this study was the first to indicate that the choice of ML model could impact AUC. The authors also suggested that NNM had advantages over DT, and that the inclusion of body temperature and laboratory indicators enhanced prediction quality. The only other meta-analysis performed so far demonstrated the superiority of XGBoost and random forest models but with high heterogeneity (I2) (16). In other systematic reviews, quantitative meta-analysis was not conducted due to significant heterogeneity among studies (10–13). We were the first to apply a NMA technique, which allowed us to overcome high heterogeneity of previous meta-analyses. This approach enables comparisons between two ML models or an ML model and a traditional scoring system within the same study on a single patient cohort, employing a unified approach and standardized definition of sepsis. In our research, NNM did not demonstrate superiority over DT, and the use of body temperature and laboratory indicators as predictors did not enhance the predictive quality.

4.3 Significance of the study findings

Our network meta-analysis, which evaluated 73 articles encompassing 457,932 septic patients, revealed that ML algorithms significantly outperform traditional sepsis scoring systems. The integration of ML into sepsis prediction marks a significant step forward in improving the early diagnosis and management of this life-threatening condition in emergency and intensive care settings. Unlike traditional scoring systems, ML models can process vast amounts of real-time clinical data, offering early warning systems that may identify sepsis before the appearance of clinical symptoms, thereby facilitating timely and targeted interventions. This has profound implications for clinical practice, as prompt treatments such as early antibiotic administration are known to significantly improve patient outcomes, particularly when initiated within the 'golden hour' (3–5).

Our study makes a key contribution by identifying the factors that impact the effectiveness of sepsis prediction models. Specifically, we found that the number of predictors and sepsis prevalence do not substantially influence model performance, challenging the traditional assumption that larger datasets and a perfectly balanced cohort (with a 50/50% split) are essential for robust predictions. Instead, our findings underscore the importance of data quality and the careful selection of relevant predictors, which has direct implications for how ML models should be developed and deployed in real-world clinical environments.

The heterogeneity in external validation and imputation methods across studies underscores a significant gap in standardizing ML model development and validation for sepsis prediction. While we did not find a notable impact of external validation on AUC, its role in enhancing the robustness and generalizability of prediction models should not be underestimated. Furthermore, it's noteworthy that even in the context of established and stringent diagnostic guidelines for sepsis, there exists a number of studies where researchers have opted to utilize alternative definitions of sepsis in their studies.

Another notable aspect of our research pertains to the 'black box' nature of some ML models, which limits the clinician's ability to understand the logic behind decision-making. Our study demonstrates that the use of DT is not inferior to the more complex NNM. This is a pivotal finding as DT models offer greater transparency in decision-making processes, which is crucial for clinical applications where understanding the rationale behind predictions is as important as the predictions themselves.

While ML models show substantial promise in predicting sepsis onset, their clinical utility remains limited by the challenge of initiating treatment before the appearance of clinical symptoms.

4.4 Strengths and limitations

This research is the first to quantitatively demonstrate the superiority of ML models over traditional scoring systems using NMA. Furthermore, this study is the first to employ NMA to reveal the advantages of NNM and DT over other ML models in the prediction of sepsis. Through meta-regression, we identified several critical factors that influence model performance, providing valuable insights for future model development. The application of the CINeMA approach provided a structured methodology to rate the certainty of our evidence, enhancing the reliability of our findings.

However, limitations of our study must be acknowledged: we found high clinical heterogeneity among the included studies and therefore used random-effects modelling and sensitivity analyses; while the AUC was pragmatically chosen as the summary measure, it may not be as effective in imbalanced datasets, yet it remains the most frequently reported measure in this field.

4.5 Future studies and prospects

The growing body of evidence supporting the advantages of ML models over traditional scoring systems in sepsis prediction underscores the need to integrate these technologies into routine clinical practice. Future research should focus on conducting well-structured prospective trials to evaluate how ML-predicted sepsis outcomes influence the timing and initiation of antibiotic therapy. A critical component of these trials will be assessing the time interval between ML model predictions and clinical recognition by healthcare providers, as delays in treatment initiation can significantly affect patient outcomes.

We propose a randomized, double-blind controlled trial comparing the efficacy of early antibiotic therapy initiated based on ML predictions versus placebo during this pre-recognition window. Such a study could provide definitive evidence regarding the clinical utility of ML-based early warning systems and their potential to reduce mortality and morbidity in sepsis by enabling earlier interventions.

5 Conclusion

Our systematic review and network meta-analysis revealed that machine learning models, specifically neural network models and decision trees, exhibit superior performance in predicting sepsis compared to traditional scoring systems. This study highlights the significant impact of machine learning model type and dataset characteristics on prediction accuracy. Despite the promise of machine learning models in clinical settings, their potential is yet to be fully realized due to study heterogeneity and the variability in sepsis definitions. To bridge this gap, there is an urgent need for standardized reporting and validation frameworks, ensuring that machine learning tools are both reliable and generalizable in diverse clinical settings.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MY: Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. GL: Investigation, Writing – original draft, Writing – review & editing. LB: Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. PP: Formal analysis, Writing – original draft, Writing – review & editing. KK: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. AS: Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. IK: Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. MS: Visualization, Writing – original draft, Writing – review & editing. AY: Supervision, Writing – original draft, Writing – review & editing. VL: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1491358/full#supplementary-material

References

1. Fleischmann C, Scherag A, Adhikari NKJ, Hartog CS, Tsaganos T, Schlattmann P, et al. Assessment of global incidence and mortality of hospital-treated sepsis current estimates and limitations. *Am J Respir Crit Care Med*. (2016) 193:259–72. doi: 10.1164/rccm.201504-0781OC

2. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*. (2021) 47:1181–247. doi: 10.1007/s00134-021-06506-y

3. Rababa M, Hamad DB, Hayajneh AA. Sepsis assessment and management in critically ill adults: a systematic review. *PLoS One*. (2022) 17:e0270711. doi: 10.1371/journal.pone.0270711

4. Il KH, Park S. Sepsis: early recognition and optimized treatment. *Tuberc Respir Dis (Seoul)*. (2019) 82:6–14. doi: 10.4046/trd.2018.0041

5. Im Y, Kang D, Ko RE, Lee YJ, Lim SY, Park S, et al. Time-to-antibiotics and clinical outcomes in patients with sepsis and septic shock: a prospective nationwide multicenter cohort study. *Crit Care*. (2022) 26:19. doi: 10.1186/s13054-021-03883-0

6. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Prog Biomed*. (2019) 170:1–9. doi: 10.1016/j.cmpb.2018.12.027

7. Schinkel M, van der Poll T, Wiersinga WJ. Artificial intelligence for early Sepsis detection a word of caution. *Am J Respir Crit Care Med*. (2023) 207:853–4. doi: 10.1164/rccm.202212-2284VP

8. van der Vegt AH, Scott IA, Dermawan K, Schnetler RJ, Kalke VR, Lane PJ. Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT clinical AI implementation framework. *J Am Med Informatics Assoc*. (2023) 30:1349–61. doi: 10.1093/jamia/ocad075

9. Komorowski M, Green A, Tatham KC, Seymour C, Antcliffe D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine*. (2022) 86:104394. doi: 10.1016/j.ebiom.2022.104394

10. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of Sepsis in the ICU using machine learning: a systematic review. *Front Med*. (2021) 8:607952. doi: 10.3389/fmed.2021.607952

11. Deng HF, Sun MW, Wang Y, Zeng J, Yuan T, Li T, et al. Evaluating machine learning models for sepsis prediction: a systematic review of methodologies. *iScience*. (2022) 25:103651. doi: 10.1016/j.isci.2021.103651

12. Yan MY, Gustad LT, Nytrø Ø. Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. *J Am Med Informatics Assoc*. (2022) 29:559–75. doi: 10.1093/jamia/ocab236

13. Kausch SL, Moorman JR, Lake DE, Keim-Malpass J. Physiological machine learning models for prediction of sepsis in hospitalized adults: an integrative review. *Intensive Crit Care Nurs*. (2021) 65:103035. doi: 10.1016/j.iccn.2021.103035

14. Lauritsen SM, Thiesson B, Jørgensen MJ, Riis AH, Espelund US, Weile JB, et al. The framing of machine learning risk prediction models illustrated by evaluation of sepsis in general wards. *Npj Digit Med*. (2021) 4:1–12. doi: 10.1038/s41746-021-00529-x

15. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. (2020) 46:383–400. doi: 10.1007/s00134-019-05872-y

16. Yang Z, Cui X, Song Z. Predicting sepsis onset in ICU using machine learning models: a systematic review and meta-analysis. *BMC Infect Dis*. (2023) 23:635–21. doi: 10.1186/s12879-023-08614-0

17. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. (2015) 162:777–84. doi: 10.7326/M14-2385

18. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. *Cochrane Handb Syst Rev Interv*. (2019):1–694. doi: 10.1002/9781119536604

19. Litmaps Literature Map Software for Lit Reviews and Research. (2023) Available at: https://www.litmaps.com/ (Accessed November 24, 2023).

20. Singer M, Deutschman CS, Seymour C, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287

21. Harbord RM, Higgins JPT. Meta-regression in Stata. *Stata J*. (2008) 8:493–519. doi: 10.1177/1536867x0800800403

22. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del GC, Egger M, et al. Cinema: an approach for assessing confidence in the results of a network meta-analysis. *PLoS Med*. (2020) 17:e1003082. doi: 10.1371/JOURNAL.PMED.1003082

23. Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: software for semiautomated assessment of the confidence in the results of network meta-analysis. *Campbell Syst Rev*. (2020) 16:e1080. doi: 10.1002/cl2.1080

24. Chiocchia V, Nikolakopoulou A, Higgins JPT, Page MJ, Papakonstantinou T, Cipriani A, et al. ROB-MEN: a tool to assess risk of bias due to missing evidence in network meta-analysis. *BMC Med*. (2021) 19:304. doi: 10.1186/s12916-021-02166-3

25. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Quadas-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. (2011) 155:529–36. doi: 10.7326/0003-4819-155-8-201110180-00009

26. Aromataris E, Munn Z, Moola S, Tufanaru C, Sears K, Sfetcu R, et al. Checklist for analytical cross sectional studies. (2020). Available at: https://synthesismanual.jbi.global/ (Accessed November 24, 2023).

27. Abromavičius V, Plonis D, Tarasevičius D, Serackis A. Two-stage monitoring of patients in intensive care unit for sepsis prediction using non-overfitted machine learning models. *Electron*. (2020) 9:1–14. doi: 10.3390/electronics9071133

28. Camacho-Cogollo JE, Bonet I, Gil B, Iadanza E. Machine learning models for early prediction of Sepsis on large healthcare datasets. *Electron*. (2022) 11:1507. doi: 10.3390/electronics11091507

29. Chen M, Hernández A. Towards an explainable model for Sepsis detection based on sensitivity analysis. *Irbm*. (2022) 43:75–86. doi: 10.1016/j.irbm.2021.05.006

30. Chen Q, Li R, Lin CC, Lai C, Chen D, Qu H, et al. Transferability and interpretability of the sepsis prediction models in the intensive care unit. *BMC Med Inform Decis Mak*. (2022) 22:343–10. doi: 10.1186/s12911-022-02090-3

31. Chen C, Chen B, Yang J, Li X, Peng X, Feng Y, et al. Development and validation of a practical machine learning model to predict sepsis after liver transplantation. *Ann Med*. (2023) 55:624–33. doi: 10.1080/07853890.2023.2179104

32. Choi JS, Trinh TX, Ha J, Yang MS, Lee Y, Kim YE, et al. Implementation of complementary model using optimal combination of hematological parameters for Sepsis screening in patients with fever. *Sci Rep*. (2020) 10:273–10. doi: 10.1038/s41598-019-57107-1

33. Delahanty RJ, Alvarez JA, Flynn LM, Sherwin RL, Jones SS. Development and evaluation of a machine learning model for the early identification of patients at risk for Sepsis. *Ann Emerg Med*. (2019) 73:334–44. doi: 10.1016/j.annemergmed.2018.11.036

34. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. *JMIR Med Informatics*. (2016) 4:e28. doi: 10.2196/medinform.5909

35. Duan Y, Huo J, Chen M, Hou F, Yan G, Li S, et al. Early prediction of sepsis using double fusion of deep features and handcrafted features. *Appl Intell*. (2023) 53:17903–19. doi: 10.1007/s10489-022-04425-z

36. El-Rashidy N, Abuhmed T, Alarabi L, El-Bakry HM, Abdelrazek S, Ali F, et al. Sepsis prediction in intensive care unit based on genetic feature optimization and stacked deep ensemble learning. *Neural Comput & Applic*. (2022) 34:3603–32. doi: 10.1007/s00521-021-06631-1

37. Fagerström J, Bång M, Wilhelms D, Chew MS. LiSep LSTM: a machine learning algorithm for early detection of septic shock. *Sci Rep*. (2019) 9:15132–8. doi: 10.1038/s41598-019-51219-4

38. Amrollahi F, Shashikumar SP, Razmi F, Nemati S. Contextual Embeddings from clinical notes improves prediction of Sepsis. *AMIA Annu Symp Proc*. (2020) 2020:197–202.

39. Faisal M, Scally A, Richardson D, Beatson K, Howes R, Speed K, et al. Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Crit Care Med*. (2018) 46:612–8. doi: 10.1097/CCM.0000000000002967

40. Gholamzadeh M, Abtahi H, Safdari R. Comparison of different machine learning algorithms to classify patients suspected of having sepsis infection in the intensive care unit. *Inform Med Unlock*. (2023) 38:101236. doi: 10.1016/j.imu.2023.101236

41. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe Sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med*. (2019) 47:1485–92. doi: 10.1097/CCM.0000000000003891

42. Goh KH, Wang L, Yeow AYK, Poh H, Li K, Yeow JJL, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun*. (2021) 12:711–0. doi: 10.1038/s41467-021-20910-4

43. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One*. (2017) 12:e0174708. doi: 10.1371/journal.pone.0174708

44. Ibrahim ZM, Wu H, Hamoud A, Stappen L, Dobson RJB, Agarossi A. On classifying sepsis heterogeneity in the ICU: insight using machine learning. *J Am Med Informatics Assoc*. (2020) 27:437–43. doi: 10.1093/jamia/ocz211

45. Kaji DA, Zech JR, Kim JS, Cho SK, Dangayach NS, Costa AB, et al. An attention based deep learning model of clinical events in the intensive care unit. *PLoS One*. (2019) 14:e0211057. doi: 10.1371/journal.pone.0211057

46. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med*. (2017) 89:248–55. doi: 10.1016/j.compbiomed.2017.08.015

47. Khojandi A, Tansakul V, Li X, Koszalinski RS, Paiva W. Prediction of Sepsis and in-hospital mortality using electronic health records. *Methods Inf Med*. (2018) 57:185–93. doi: 10.3414/ME18-01-0014

48. Kijpaisalratana N, Sanglertsinlapachai D, Techaratsami S, Musikatavorn K, Saoraya J. Machine learning algorithms for early sepsis detection in the emergency department: a retrospective study. *Int J Med Inform*. (2022) 160:104689. doi: 10.1016/j.ijmedinf.2022.104689

49. Aşuroğlu T, Oğul H. A deep learning approach for sepsis monitoring via severity score estimation. *Comput Methods Prog Biomed*. (2021) 198:105816. doi: 10.1016/j.cmpb.2020.105816

50. Kuo YY, Huang ST, Chiu HW. Applying artificial neural network for early detection of sepsis with intentionally preserved highly missing real-world data for simulating clinical situation. *BMC Med Inform Decis Mak*. (2021) 21:290–11. doi: 10.1186/s12911-021-01653-0

51. Kwon J, Myoung LYR, Jung MS, Lee YJ, Jo YY, Kang DY, et al. Deep-learning model for screening sepsis using electrocardiography. *Scand J Trauma Resusc Emerg Med*. (2021) 29:1–12. doi: 10.1186/s13049-021-00953-8

52. Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, et al. Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artif Intell Med*. (2020) 104:101820. doi: 10.1016/j.artmed.2020.101820

53. Li X, Xu X, Xie F, Xu X, Sun Y, Liu X, et al. A time-phased machine learning model for real-time prediction of Sepsis in critical care. *Crit Care Med*. (2020) 48:E884–8. doi: 10.1097/CCM.0000000000004494

54. Lin PC, Chen KT, Chen HC, Islam MM, Lin MC. Machine learning model to identify sepsis patients in the emergency department: algorithm development and validation. *J Pers Med*. (2021) 11:1055. doi: 10.3390/jpm11111055

55. Liu Z, Khojandi A, Li X, Mohammed A, Davis RL, Kamaleswaran R. A machine learning–enabled partially observable Markov decision process framework for early Sepsis prediction. *INFORMS J Comput*. (2022) 34:2039–57. doi: 10.1287/ijoc.2022.1176

56. Liu F, Yao J, Liu C, Shou S. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis. *BMC Surg*. (2023) 23:267–13. doi: 10.1186/s12893-023-02151-y

57. Maharjan J, Thapa R, Calvert J, Attwood MM, Shokouhi S, Casie Chetty S, et al. A new standard for Sepsis prediction algorithms: using time-dependent analysis for earlier clinically relevant alerts. *SSRN Electron J*. (2022). doi: 10.2139/ssrn.4130480

58. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*. (2018) 8:e017833. doi: 10.1136/bmjopen-2017-017833

59. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Qual*. (2017) 6:e000158. doi: 10.1136/bmjoq-2017-000158

60. Bao C, Deng F, Zhao S. Machine-learning models for prediction of sepsis patients mortality. *Med Intens*. (2023) 47:315–25. doi: 10.1016/j.medine.2022.06.024

61. Moor M, Bennett N, Plečko D, Horn M, Rieck B, Meinshausen N, et al. Predicting sepsis using deep learning across international sites: a retrospective development and validation study. *eClinicalMedicine*. (2023) 62:102124. doi: 10.1016/j.eclinm.2023.102124

62. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of Sepsis in the ICU. *Crit Care Med*. (2018) 46:547–53. doi: 10.1097/CCM.0000000000002936

63. Nesaragi N, Patidar S. An explainable machine learning model for early prediction of Sepsis using ICU data. *Infect Sepsis Dev*. (2021). doi: 10.5772/intechopen.98957

64. Oei SP, van Sloun RJ, van der Ven M, Korsten HH, Mischi M. Towards early sepsis detection from measurements at the general ward through deep learning. *Intell Med*. (2021) 5:100042. doi: 10.1016/j.ibmed.2021.100042

65. Persson I, Östling A, Arlbrandt M, Söderberg J, Becedas D. A machine learning Sepsis prediction algorithm for intended intensive care unit use (NAVOY Sepsis): proof-of-concept study. *JMIR Form Res*. (2021) 5:e28000. doi: 10.2196/28000

66. Rafiei A, Rezaee A, Hajati F, Gheisari S, Golzan M. SSP: early prediction of sepsis using fully connected LSTM-CNN model. *Comput Biol Med*. (2021) 128:104110. doi: 10.1016/j.compbiomed.2020.104110

67. Rangan ES, Pathinarupothi RK, Anand KJS, Snyder MP. Performance effectiveness of vital parameter combinations for early warning of sepsis - an exhaustive study using machine learning. *JAMIA Open*. (2022) 5:1–11. doi: 10.1093/jamiaopen/ooac080

68. Rosnati M, Fortuin V. MGP-AttTCN: an interpretable machine learning model for the prediction of sepsis. *PLoS One*. (2021) 16:e0251248. doi: 10.1371/journal.pone.0251248

69. Sadasivuni S, Saha M, Bhatia N, Banerjee I, Sanyal A. Fusion of fully integrated analog machine learning classifier with electronic medical records for real-time prediction of sepsis onset. *Sci Rep*. (2022) 12:1–11. doi: 10.1038/s41598-022-09712-w

70. Sharma DK, Lakhotia P, Sain P, Brahmachari S. Early prediction and monitoring of sepsis using sequential long short term memory model. *Expert Syst*. (2022) 39:e12798. doi: 10.1111/exsy.12798

71. Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med*. (2019) 109:79–84. doi: 10.1016/j.compbiomed.2019.04.027

72. Scherpf M, Gräßer F, Malberg H, Zaunseder S. Predicting sepsis with a recurrent neural network using the MIMIC III database. *Comput Biol Med*. (2019) 113:103395. doi: 10.1016/j.compbiomed.2019.103395

73. Schamoni S, Lindner HA, Schneider-Lindner V, Thiel M, Riezler S. Leveraging implicit expert knowledge for non-circular machine learning in sepsis prediction. *Artif Intell Med*. (2019) 100:101725. doi: 10.1016/j.artmed.2019.101725

74. Shashikumar SP, Stanley MD, Sadiq I, Li Q, Holder A, Clifford GD, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *J Electrocardiol*. (2017) 50:739–43. doi: 10.1016/j.jelectrocard.2017.08.013

75. Shashikumar SP, Li Q, Clifford GD, Nemati S. Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol Meas*. (2017) 38:2235–48. doi: 10.1088/1361-6579/aa9772

76. Shashikumar SP, Josef CS, Sharma A, Nemati S. DeepAISE – an interpretable and recurrent neural survival model for early prediction of sepsis. *Artif Intell Med*. (2021) 113:102036. doi: 10.1016/j.artmed.2021.102036

77. Singh YV, Singh P, Khan S, Singh RS. A machine learning model for early prediction and detection of Sepsis in intensive care unit patients. *J Healthc Eng*. (2022) 2022:1–11. doi: 10.1155/2022/9263391

78. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir Res*. (2017) 4:e000234. doi: 10.1136/bmjresp-2017-000234

79. Taneja I, Reddy B, Damhorst G, Dave Zhao S, Hassan U, Price Z, et al. Combining biomarkers with EMR data to identify patients in different phases of Sepsis. *Sci Rep*. (2017) 7:10800–12. doi: 10.1038/s41598-017-09766-1

80. Tang G, Luo Y, Lu F, Li W, Liu X, Nan Y, et al. Prediction of Sepsis in COVID-19 using laboratory indicators. *Front Cell Infect Microbiol*. (2021) 10:586054. doi: 10.3389/fcimb.2020.586054

81. Valik JK, Ward L, Tanushi H, Johansson AF, Färnert A, Mogensen ML, et al. Predicting sepsis onset using a machine learned causal probabilistic network algorithm based on electronic health records data. *Sci Rep*. (2023) 13:11760–12. doi: 10.1038/s41598-023-38858-4

82. Bedoya AD, Futoma J, Clement ME, Corey K, Brajer N, Lin A, et al. Machine learning for early detection of sepsis: an internal and temporal validation study. *JAMIA Open*. (2020) 3:252–60. doi: 10.1093/jamiaopen/ooaa006

83. van Wyk F, Khojandi A, Mohammed A, Begoli E, Davis RL, Kamaleswaran R. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. *Int J Med Inform*. (2019) 122:55–62. doi: 10.1016/j.ijmedinf.2018.12.002

84. Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, et al. A machine learning model for accurate prediction of Sepsis in ICU patients. *Front Public Heal*. (2021) 9:754348. doi: 10.3389/fpubh.2021.754348

85. Wang Z, Yao B. Multi-branching temporal convolutional network for Sepsis prediction. *IEEE J Biomed Heal Inform*. (2022) 26:876–87. doi: 10.1109/JBHI.2021.3092835

86. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med*. (2021) 181:1065–70. doi: 10.1001/jamainternmed.2021.2626

87. Yang M, Liu C, Wang X, Li Y, Gao H, Liu X, et al. An explainable artificial intelligence predictor for early detection of Sepsis. *Crit Care Med*. (2020) 48:E1091–6. doi: 10.1097/CCM.0000000000004550

88. Yang D, Kim J, Yoo J, Cha WC, Paik H. Identifying the risk of Sepsis in patients with Cancer using digital health care records: machine learning-based approach. *JMIR Med Inform*. (2022) 10:e37689. doi: 10.2196/37689

89. Yu SC, Gupta A, Betthauser KD, Lyons PG, Lai AM, Kollef MH, et al. Sepsis prediction for the general Ward setting. *Front Digit Heal*. (2022) 4:848599. doi: 10.3389/fdgth.2022.848599

90. Yuan KC, Tsai LW, Lee KH, Cheng YW, Hsu SC, Lo YS, et al. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int J Med Inform*. (2020) 141:104176. doi: 10.1016/j.ijmedinf.2020.104176

91. Zargoush M, Sameh A, Javadi M, Shabani S, Ghazalbash S, Perri D. The impact of recency and adequacy of historical information on sepsis predictions using machine learning. *Sci Rep*. (2021) 11:20869. doi: 10.1038/s41598-021-00220-x

92. Zhang D, Yin C, Hunold KM, Jiang X, Caterino JM, Zhang P. An interpretable deep-learning model for early prediction of sepsis in the emergency department. *Patterns*. (2021) 2:100196. doi: 10.1016/j.patter.2020.100196

93. Bloch E, Rotem T, Cohen J, Singer P, Aperstein Y. Machine learning models for analysis of vital signs dynamics: a case for Sepsis onset prediction. *J Healthc Eng*. (2019) 2019:1–11. doi: 10.1155/2019/5930379

94. Zhang TY, Zhong M, Cheng YZ, Zhang MW. An interpretable machine learning model for real-time sepsis prediction based on basic physiological indicators. *Eur Rev Med Pharmacol Sci*. (2023) 27:4348–56. doi: 10.26355/eurrev_202305_32439

95. Zhang S, Duan Y, Hou F, Yan G, Li S, Wang H, et al. Early prediction of sepsis using a high-order Markov dynamic Bayesian network (HMDBN) classifier. *Appl Intell*. (2023) 53:26384–99. doi: 10.1007/s10489-023-04920-x

96. Zhao X, Shen W, Wang G. Early prediction of Sepsis based on machine learning algorithm. *Comput Intell Neurosci*. (2021) 2021:6522633. doi: 10.1155/2021/6522633

97. Burdick H, Pino E, Gabel-Comeau D, Gu C, Roberts J, Le S, et al. Validation of a machine learning algorithm for early severe sepsis prediction: a retrospective study predicting severe sepsis up to 48 h in advance using a diverse dataset from 461 US hospitals. *BMC Med Inform Decis Mak*. (2020) 20:1–10. doi: 10.1186/s12911-020-01284-x

98. Calvert JS, Price DA, Chettipally UK, Barton CW, Feldman MD, Hoffman JL, et al. A computational approach to early sepsis detection. *Comput Biol Med*. (2016) 74:69–73. doi: 10.1016/j.compbiomed.2016.05.003

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 01 November 2024
DOI 10.3389/fmed.2024.1501761

OPEN ACCESS

EDITED BY
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

REVIEWED BY
Zhonghua Lu,
The Second Affiliated Hospital of Anhui Medical University, China
Zhengyu He,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Jianfeng Xie
✉ xie820405@126.com
Lili Huang
✉ caowangyou007@126.com

RECEIVED 25 September 2024
ACCEPTED 21 October 2024
PUBLISHED 01 November 2024

CITATION
Zhang S, Chen H, Xie J and Huang L (2024) RIG012 assists in the treatment of pneumonia by inhibiting the RIG-I-like receptor signaling pathway. *Front. Med.* 11:1501761. doi: 10.3389/fmed.2024.1501761

RIG012 assists in the treatment of pneumonia by inhibiting the RIG-I-like receptor signaling pathway

Shi Zhang[1,2], Hanbing Chen[1], Jianfeng Xie[1]* and Lili Huang[1]*

[1]Jiangsu Provincial Key Laboratory of Critical Care Medicine, Department of Critical Care Medicine, Zhongda Hospital, School of Medicine, Southeast University, Nanjing, China, [2]Department of Respiratory and Critical Care Medicine, Central Hospital Affiliated to Shandong First Medical University, Jinan, China

Objective: Pneumonia is a common clinical condition primarily treated with antibiotics and organ support. Exploring the pathogenesis to identify therapeutic targets may aid in the adjunct treatment of pneumonia and improve survival rates.

Methods: Transcriptomic data from peripheral blood of 183 pneumonia patients were analyzed using Gene Set Variation Analysis (GSVA) and univariate Cox regression analysis to identify signaling pathways associated with pneumonia mortality. A pneumonia mouse model was established via airway injection of *Klebsiella pneumoniae*, and pathway-specific blockers were administered via tail vein infusion to assess whether the identified signaling pathways impact the mortality in pneumonia.

Results: The combination of GSVA and Cox analysis revealed 17 signaling pathways significantly associated with 28-day mortality in pneumonia patients ($P < 0.05$). Notably, the RIG-I-like receptor signaling pathway exhibited the highest hazard ratio of 2.501 with a 95% confidence interval of [1.223–5.114]. Infusion of RIG012 via the tail vein effectively inhibited the RIG-I-like receptor signaling pathway, significantly ameliorated lung injury in pneumonia mice, reduced pulmonary inflammatory responses, and showed a trend toward improved survival rates.

Conclusion: RIG012 may represent a novel adjunctive therapeutic agent for pneumonia.

KEYWORDS

pneumonia, RIG012, RIG-I, GSVA, treatment

Introduction

Pneumonia is a common clinical syndrome that severely jeopardizes public health. It accounts for ~10–20% of admissions to intensive care units (ICU). The hospital mortality rate for pneumonia patients varies between 12 and 38%, with mortality rates as high as 40–45% among patients with severe pneumonia in the ICU (1, 2). Thus, pneumonia represents a significant clinical challenge that both clinicians and specialists in critical care medicine must confront.

Currently, the treatment of pneumonia primarily relies on antibiotics and organ support therapies, with a notable lack of adjunctive treatments targeting its underlying pathogenesis (3, 4). Investigating the mechanisms associated with the prognosis of pneumonia patients may provide insights for developing adjunctive therapies, thereby enhancing the efficacy of pneumonia management and improving patient outcomes.

Utilizing algorithms such as machine learning for the secondary analysis of patient transcriptomic data offers new perspectives for understanding diseases and aids in uncovering mechanisms associated with disease progression and prognosis (5, 6). Gene Set Variation Analysis (GSVA) is an algorithm developed by Guinney et al. that transforms transcriptomic data into signaling pathway information through a machine learning framework (7). This algorithm reduces the dimensionality of high-dimensional transcriptomic data into actionable insights about signal transduction pathways, thereby facilitating further exploration of disease-related mechanisms and drug development (7, 8).

This project conducts GSVA on transcriptomic data from 183 pneumonia patients to identify signaling pathways associated with patient prognosis. Additionally, animal experiments are employed to validate the bioinformatics findings.

Materials and methods

The transcriptomic data of pneumonia patients

The transcriptomic data of pneumonia patients are derived from high-throughput sequencing of peripheral blood from ICU patients uploaded to the Gene Expression Omnibus (GEO) public database (accession number GSE65682) by Scicluna et al. The dataset includes information on patients diagnosed with pneumonia, encompassing 28-day survival data and high-throughput sequencing information. This dataset was chosen because it represents the largest publicly available collection of high-throughput sequencing data for pneumonia patients, along with corresponding survival information.

GSVA

The transcriptomic data were transformed into signaling pathway data using the GSVA algorithm developed by Guinney and the publicly available GSVA R package (Bioconductor—GSVA). The principles of the algorithm and detailed procedures are referenced in the documentation provided with Guinney's R package.

Statistical analysis

Univariate Cox analysis was performed to identify signaling pathways associated with the 28-day mortality rate in pneumonia patients, using a significance threshold of $P < 0.05$. Based on the median of the signaling pathway data, patients were categorized into high-risk and low-risk groups. Survival analysis was conducted using the log-rank test to compare the survival outcomes between the two groups. Differential analysis was carried out using one-way analysis of variance, with $P < 0.05$ considered statistically significant.

Mouse model of pneumonia

According to previous studies on the RIG-I pathway, there is no significant association between the RIG-I pathway and sex (10–12). The study subjects consisted of male C57BL/6 mice aged 6–8 weeks. A pneumonia model was established by intratracheal injection of *Klebsiella pneumoniae* (KP) at a dosage of 60×10^8 CFU/kg over a duration of 24 h. The sham-operated group received an equal volume of PBS via intratracheal injection. Each group comprised six mice, with three designated for survival analysis and the other three euthanized after 24 h to collect lung tissue for subsequent experiments.

KP: SHBCCD11105CMCC46117; Shanghai Bioresource Collection Center, Shanghai, China.

RIG012 treatment

Two hours after intratracheal injection of *Klebsiella pneumoniae* (KP), the RIG-I-like signaling pathway-specific inhibitor RIG012 was administered via tail vein infusion at a dosage of 5 mg/kg. Based on the RIG012 reagent specifications and prior research data on RIG012 (21), we selected this dosage, and experimental evidence confirms that 5 mg/kg is indeed effective.

The treatment control group received an equivalent volume of PBS.

RIG012: HY-147124, MedChemExpress, America.

Western blots

Based on previous studies (9–12), the stimulation and variation of the RIG-I-like signaling pathway were evaluated through measuring the pIRF3 levels in mouse lung tissue proteins. Proteins were separated using an SDS-PAGE-PVDF membrane system. The primary antibody (1:1000) was incubated at room temperature for 2 h. Following thorough rinsing of the PVDF membrane, a secondary antibody was applied for 1 h. Protein detection was performed using enhanced chemiluminescence. The antibody information is as follows:

βTubulin Abcam, rabbit mAb, #ab108342.

IRF3 Abcam, rabbit mAb, #ab68481.

pIRF3 Cell Signaling Technology, rabbit mAb, 29047.

Evaluation of lung injury

After paraffin embedding, the right upper lobe of the lung was sectioned into 5-micron-thick slices for histopathological examination. The sections were stained with hematoxylin and eosin (H&E). Histopathological features, including edema, inflammation, hemorrhage, atelectasis, necrosis, and hyaline membrane formation, were scored on a scale from 0 to 4. The total score reflects the severity of lung injury.

qPCR

Based on our previous experimental foundation, the pneumonia model in mice becomes relatively stable 24 h after the intratracheal injection of KP, making it suitable for further research (20). Therefore, quantitative PCR (qPCR) was employed to assess the expression levels of the pro-inflammatory cytokines interleukin 1 beta (IL1B) and tumor necrosis factor (TNF), as well as the anti-inflammatory cytokines IL10 and transforming growth factor-beta (TGFB) in lung tissue homogenates from each group of mice 24 h post-injection of KP. This analysis aimed to evaluate the impact of RIG012 on the pulmonary inflammatory response in pneumonia-induced mice. The primer information is as follows:

IL1B
F: GCCACCTTTTGACAGTGATG
R: CGTCACACACCAGCAGGTTA
TNF
F: AGGCACTCCCCCAAAAGATG
R: CCACTTGGTGGTTTGTGAGTG
IL10
F: GGTTGCCAAGCCTTATCGGA
R: GACACCTTGGTCTTGGAGCTTA
TGFB
F: ACTGGAGTTGTACGGCAGTG
R: GGGGCTGATCCCGTTGATTT

Results

RIG-I like receptor signaling pathway was screened as a significant risk factor for mortality in pneumonia patients

A total of 183 pneumonia patients were enrolled in the study, with a male-to-female ratio of 111:72 and an age range of 18–88 years. The 28-day mortality rate was 21.9% (40/183). Through GSVA, transcriptomic microarray data were converted into 185 signaling pathway datasets.

Univariate Cox analysis revealed that 17 signaling pathways were associated with the 28-day mortality of pneumonia patients, with a significance level of $P < 0.05$ (Figure 1A). A heatmap displayed the distribution of these 17 signaling pathways among the 183 pneumonia patients (Figure 1B). Among these, the RIG-I-like receptor signaling pathway exhibited the highest hazard ratio (HR) of 2.501, with a 95% confidence interval of [1.223–5.114]. Therefore, the RIG-I-like pathway is identified as the focal point of this research project.

Using the median value of the RIG-I Like receptor pathway data as the cutoff, pneumonia patients were categorized into high-risk and low-risk groups. Survival analysis indicated that patients with high expression of the RIG-I Like receptor pathway had a significantly higher mortality rate compared to the low-expression group ($P = 0.039$), shown in Figure 1C.

These results suggest that the RIG-I Like receptor signaling pathway may be a critical risk factor for mortality in pneumonia patients, and inhibiting this pathway may potentially reduce the mortality rate in these patients.

RIG012 effectively inhibits RIG-I like receptor pathway in KP mouse

To investigate whether the inhibition of the RIG-I Like receptor pathway can improve pneumonia prognosis, we first established a pneumonia model in mice by intratracheally injecting KP. Subsequently, the small molecule compound RIG012, a specific inhibitor of the RIG-I signaling pathway, was administered via tail vein injection, while the control group received an equal volume of PBS.

Western blot analysis of lung tissue homogenates indicated that RIG012 effectively inhibited the activity of the RIG-I Like receptor pathway, with a significance level of $P < 0.001$ (Figure 2).

RIG012 effectively improves lung injury in KP mice

To evaluate whether RIG012 can enhance pneumonia prognosis, lung tissues from each group of mice were collected to prepare pathological sections. Lung injury scores were utilized to assess the protective effects of RIG012 on lung tissue. The results indicated that RIG012 significantly improved the lung injury scores in KP mice, with a significance level of $P < 0.001$ (Figure 3). Further survival analysis revealed that all mice in the RIG012 treatment group survived, while two mice in the control group (Ctrl) died within 7 days. These results suggest that RIG012 may potentially improve outcomes in pneumonia.

RIG012 effectively reduces pulmonary inflammatory response

Recent studies have indicated that the RIG-I Like receptor signaling pathway is a significant mechanism contributing to inflammatory lung injury. We hypothesized that the mechanism through which RIG012 improves pneumonia prognosis in mice is by reducing pulmonary inflammatory responses. To validate this hypothesis, lung tissue homogenates from each group of mice were collected, and quantitative qPCR was employed to assess the expression levels of pro-inflammatory factors IL1B and TNF, as well as the anti-inflammatory factors IL10 and TGFB.

The results indicated that RIG012 significantly reduced the expression of pro-inflammatory factors in KP mice while increasing the levels of anti-inflammatory factors, $P < 0.05$ (Figure 4).

Discussion

Pneumonia is a common clinical condition, with mortality rates as high as 40% in severe cases. The treatment of pneumonia primarily relies on antibiotics. However, if pneumonia progresses to acute respiratory distress syndrome and sepsis, leading to multiple organ dysfunction, necessary organ support therapies can be crucial in sustaining the lives of patients with severe pneumonia, providing a critical window for antibiotic efficacy. Nevertheless, there is currently a lack of adjunctive therapies

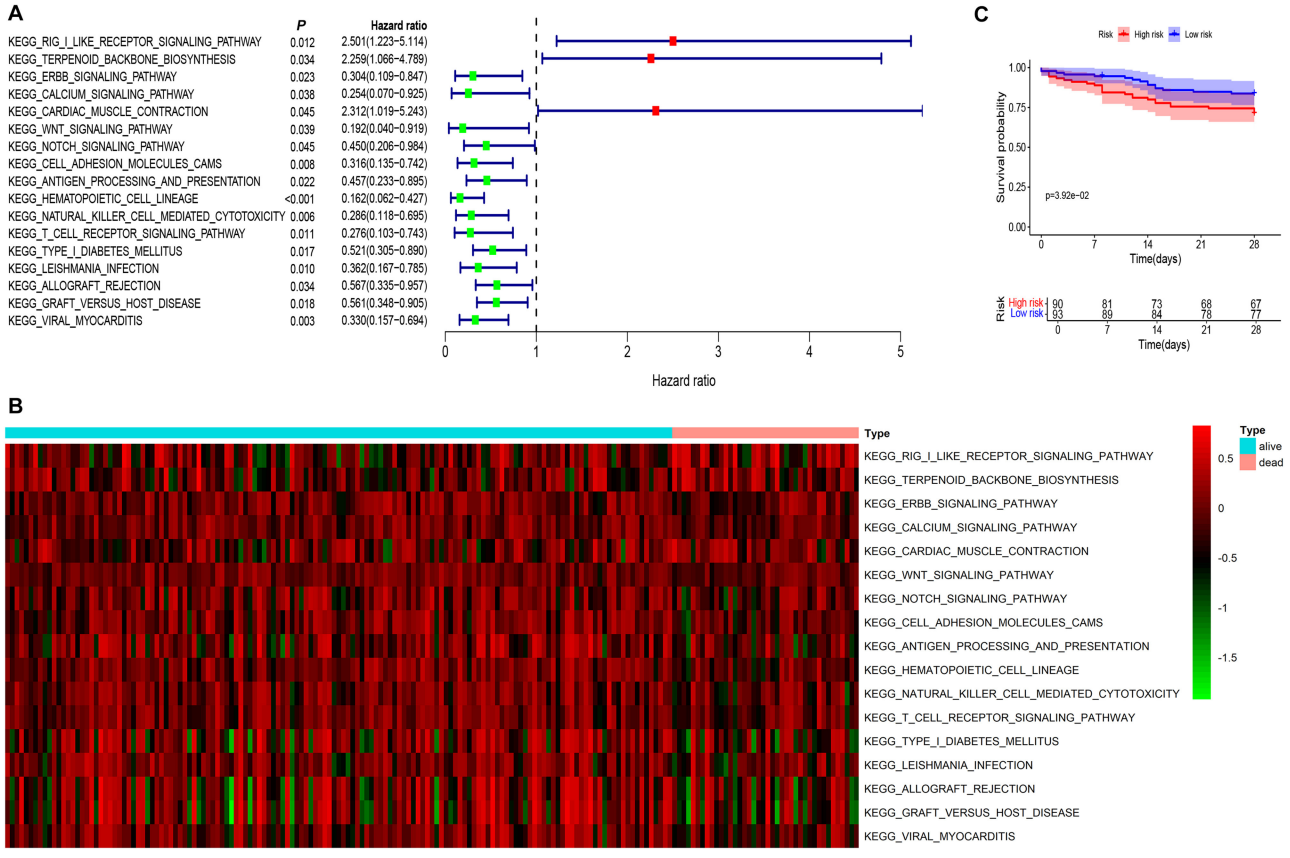

FIGURE 1
RIG-I like receptor pathway was screened as a significant risk factor for poor prognosis in pneumonia patients. **(A)** The combined analysis of GSVA and univariate COX regression identified 17 signaling pathways associated with the 28-day mortality of pneumonia patients, with a significance level of $P < 0.05$. Among these, the RIG-I Like receptor signaling pathway exhibited the highest hazard ratio of 2.501, with a 95% confidence interval of [1.223–5.114]. **(B)** A heatmap displayed the distribution of these 17 signaling pathways among the 183 pneumonia patients. **(C)** Survival analysis indicated that patients with high expression of the RIG-I Like receptor signaling pathway had a significantly higher mortality rate compared to the low-expression group ($P = 0.039$).

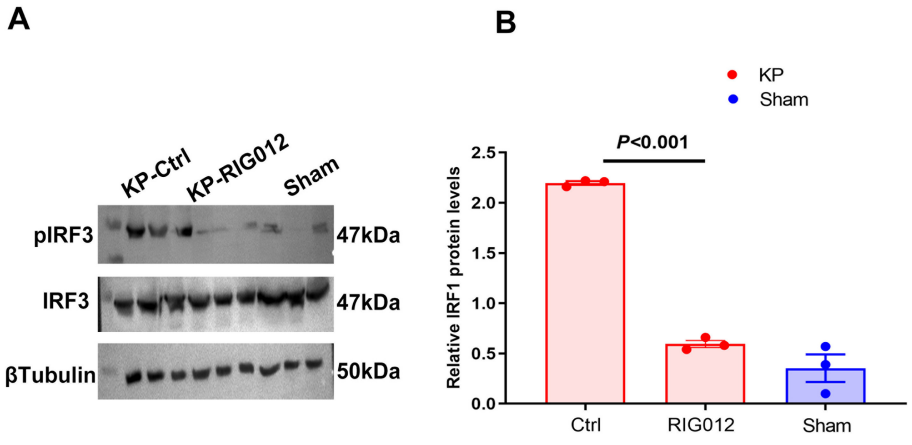

FIGURE 2
RIG012 effectively inhibits RIG-I like receptor signaling pathway in KP mouse. **(A, B)** Western blot analysis of lung tissue homogenates indicated that RIG012 effectively inhibited the activity of the RIG-I like pathway, $P < 0.001$.

targeting the pathophysiological mechanisms of pneumonia (1–4). Pneumonia can cause dysregulation of the immune system and may involve other molecular mechanisms within the body. Therefore, inhibiting harmful molecular pathways could potentially enhance the effectiveness of antibiotic therapy in treating pneumonia (13, 14). Severe pneumonia can lead to sepsis and acute respiratory

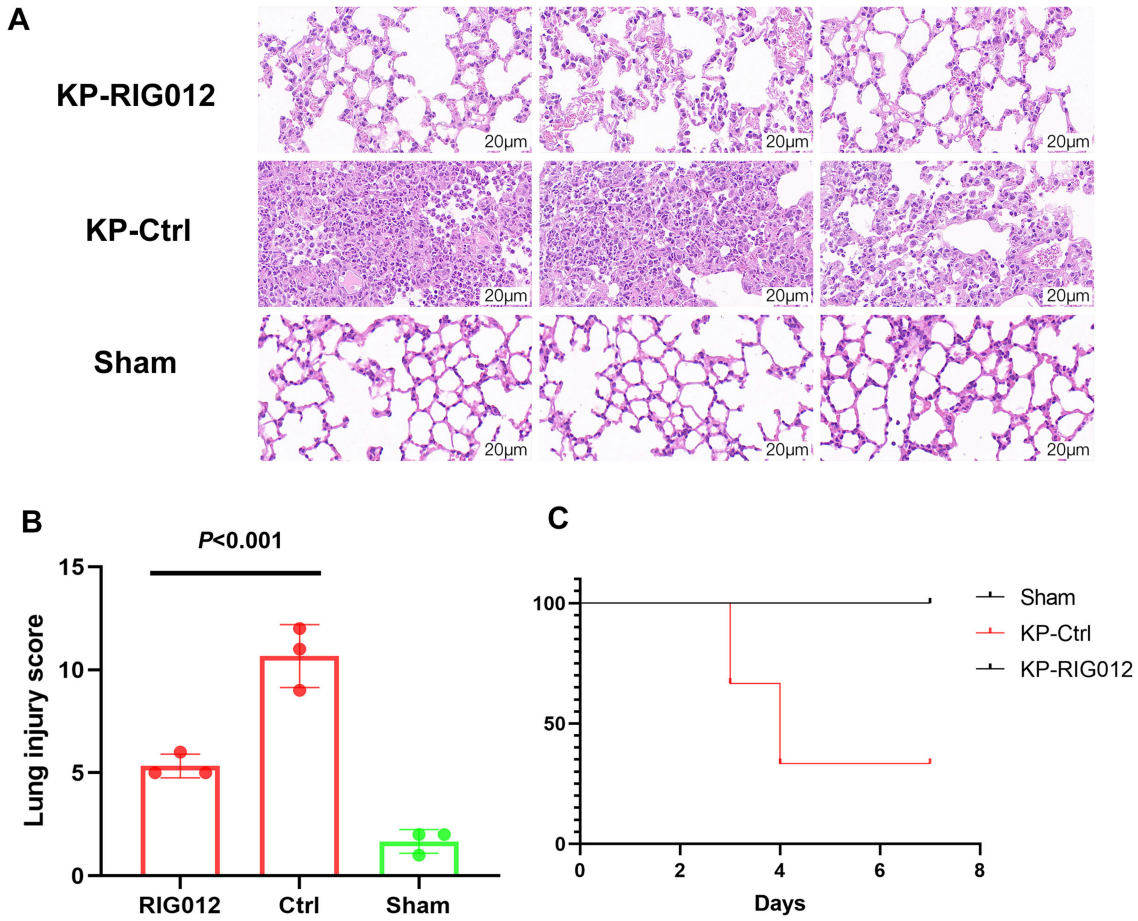

FIGURE 3
RIG012 effectively improves lung injury in KP mice. **(A)** H&E staining of lung tissues from mice in the RIG012 treatment group, treatment control group, and sham surgery group. **(B)** Lung injury scores were utilized to assess the protective effects of RIG012 on lung tissue. **(C)** Further survival analysis revealed that all mice in the RIG012 treatment group survived, while two mice in the control group (Ctrl) died within 7 days. KP, *Klebsiella pneumoniae*.

distress syndrome (ARDS), further increasing mortality rates. This project focuses on patients with pneumonia complicated by sepsis, which constitutes a more severe form of pneumonia. The results indicate that RIG012 treatment may be effective in this context.

GSVA is a non-parametric, unsupervised analytical method primarily utilized for assessing gene set enrichment results in transcriptomic data from microarrays. This technique involves transforming the gene expression matrix across different samples into an expression matrix for gene sets, thereby evaluating whether specific metabolic pathways are enriched among the samples (7, 8). In simpler terms, GSVA converts the gene expression matrix (with gene names as row labels and sample names as column labels) into a pathway matrix (with pathway names as row labels and sample names as column labels). When combined with traditional statistical methods such as survival analysis, the GSVA algorithm can effectively uncover pathogenic mechanisms associated with disease prognosis.

The RIG-I Like receptor pathway is a crucial immune activation pathway in the human body. Initially, research suggested that this pathway was primarily involved in the antiviral defense mechanisms by facilitating the binding of RIG-I protein to viral RNA, resulting in the phosphorylation and nuclear translocation of the transcription factor IRF3, which in turn initiates the transcription of interferons. Recent studies have demonstrated that the RIG-I Like receptor pathway, in addition to mediating interferon production, also interacts with classic inflammatory signaling pathways such as NFκB, contributing to excessive inflammatory responses (15–20). In the context of pneumonia, the pathogenic microbes not only inflict direct damage to lung tissue but also mediate excessive inflammatory responses that can compromise the epithelial-endothelial barrier of the lungs. This may elucidate the mechanism by which GSVA analysis associates the RIG-I Like receptor pathway with poor prognosis in pneumonia patients.

Our study found that the RIG-I-like receptor pathway has the highest hazard ratio (HR) value of 2.501, with a wide confidence interval of [1.223–5.114]. This suggests that for each unit increase in the RIG-I-like pathway, the risk of mortality from pneumonia increases 2.5-fold. This is also the rationale for selecting the RIG-I-like pathway as the primary focus of our

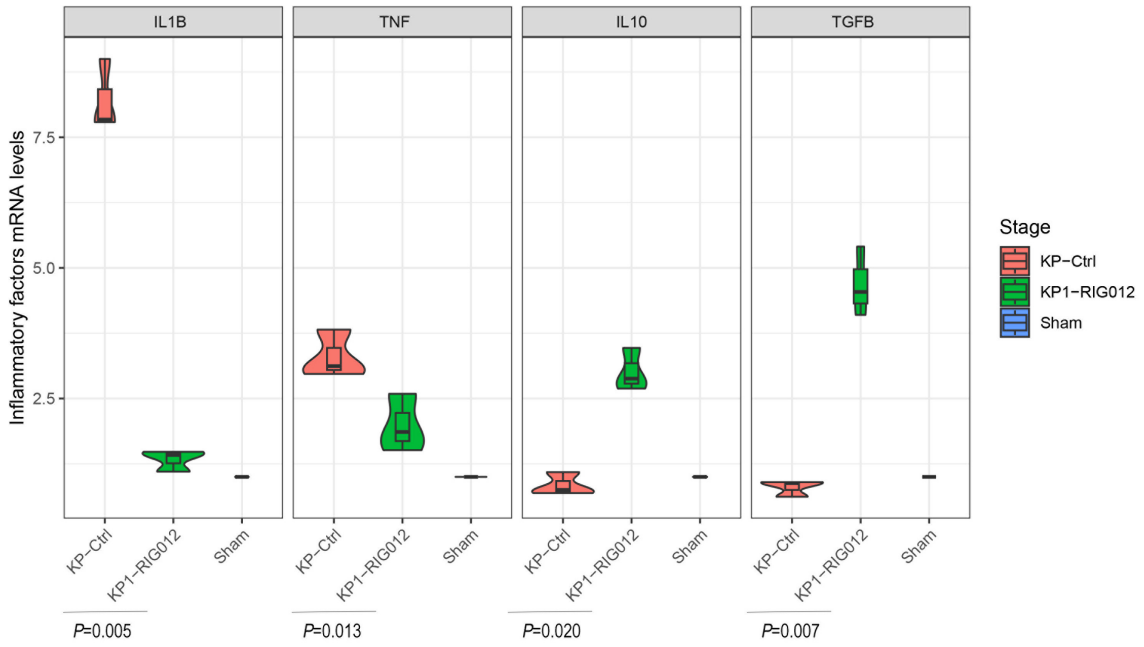

FIGURE 4
RIG012 effectively reduces pulmonary inflammatory response. The qPCR was employed to assess the expression levels of pro-inflammatory factors IL1B and TNF, as well as the anti-inflammatory factors IL10 and TGFB in lung tissue homogenates from each group of mice.

research. Additionally, the wide confidence interval of the HR indicates the presence of patient heterogeneity. Future studies that incorporate a larger sample size and conduct comprehensive subgroup analyses may be beneficial in exploring and elucidating this phenomenon.

RIG012 may serve as an effective adjunctive therapy for pneumonia. RIG012 is a small molecular compound that specifically targets RIG-I Like receptors, effectively inhibiting the RIG-I Like receptor pathway (21). Our animal studies have also demonstrated that RIG012 significantly suppresses the activity of this signaling pathway. Additionally, we observed that RIG012 can effectively mitigate lung damage caused by *Klebsiella pneumoniae* and shows a trend toward improving mortality rates in pneumonia-induced mice. Further experiments suggest that RIG012 may alleviate excessive pulmonary inflammation by inhibiting the RIG-I Like receptor pathway, thereby serving as an adjunctive treatment for lung damage resulting from pneumonia.

This study has certain limitations. First, the included mice were all male, and although previous research suggests that the RIG-I-like signaling pathway is not associated with sex, the issue of sex bias warrants attention. Second, the RIG012 treatment experiments were based on prior research and experimental data regarding dosage and duration; future studies should incorporate concentration and time gradient experiments. Finally, the databases utilized in this research are public and lack clinical information, such as the specific pathogens causing pneumonia. Future investigations should conduct larger-scale clinical studies to clarify the relationship between the RIG-I-like pathway and pneumonia prognosis, as well as to draw conclusions from subgroup analyses considering factors such as sex and pathogens.

Conclusion

Through bioinformatics analysis and validation in animal experiments, we have discovered that RIG012 may represent a novel adjunctive therapeutic agent for pneumonia.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. The animal study was approved by the Animal Ethics Committee of Southeast University Medical School. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

SZ: Writing – review & editing, Writing – original draft. HC: Writing – review & editing, Writing – original draft. JX: Writing –

review & editing, Writing – original draft. LH: Writing – review & editing, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by National Natural Science Foundation of China (Grant Number: 82302470).

Acknowledgments

Thank you for all the authors, reviewers, and editors.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1501761/full#supplementary-material

References

1. Torres A, Cilloniz C, Niederman MS, Menéndez R, Chalmers JD, et al. Pneumonia. *Nat Rev Dis Primers.* (2021) 7:25. doi: 10.1038/s41572-021-00259-0

2. File TM Jr, Ramirez JA. Community-acquired pneumonia. *N Engl J Med.* (2023) 389:632–41. doi: 10.1056/NEJMcp2303286

3. Niederman MS, Torres A. Severe community-acquired pneumonia. *Eur Respir Rev.* (2022) 31:220123. doi: 10.1183/16000617.0123-2022

4. Cillóniz C, Torres A, Niederman MS. Management of pneumonia in critically ill patients. *Br Med J.* (2021) 375:e065871. doi: 10.1136/bmj-2021-065871

5. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol.* (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0

6. Ahsan MA, Liu Y, Feng C, Zhou Y, Ma G, Bai Y, et al. Bioinformatics resources facilitate understanding and harnessing clinical research of SARS-CoV-2. *Brief Bioinform.* (2021) 22:714–25. doi: 10.1093/bib/bbaa416

7. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformat.* (2013) 14:7. doi: 10.1186/1471-2105-14-7

8. Ferreira MR, Santos GA, Biagi CA, Silva Junior WA, Zambuzzi WF. GSVA score reveals molecular signatures from transcriptomes for biomaterials comparison. *J Biomed Mater Res A.* (2021) 109:1004–14. doi: 10.1002/jbm.a.37090

9. He QQ, Huang Y, Nie L, Ren S, Xu G, Deng F, et al. MAVS integrates glucose metabolism and RIG-I-like receptor signaling. *Nat Commun.* (2023) 14:5343. doi: 10.1038/s41467-023-41028-9

10. Jiang Y, Zhang H, Wang J, Chen J, Guo Z, Liu Y, et al. Exploiting RIG-I-like receptor pathway for cancer immunotherapy. *J Hematol Oncol.* (2023) 16:8. doi: 10.1186/s13045-023-01405-9

11. Lefkopoulos S, Polyzou A, Derecka M, Bergo V, Clapes T, Cauchy P, et al. Repetitive elements trigger RIG-I-like receptor signaling that regulates the emergence of hematopoietic stem and progenitor cells. *Immunity.* (2020) 53:934–51.e9. doi: 10.1016/j.immuni.2020.10.007

12. Sprokholt JK, Kaptein TM, van Hamme JL, Overmars RJ, Gringhuis SI, Geijtenbeek TBH. RIG-I-like receptor triggering by dengue virus drives dendritic cell immune activation and TH1 differentiation. *J Immunol.* (2017) 198:4764–71. doi: 10.4049/jimmunol.1602121

13. Long ME, Mallampalli RK, Horowitz JC. Pathogenesis of pneumonia and acute lung injury. *Clin Sci.* (2022) 136:747–69. doi: 10.1042/CS20210879

14. Quinton LJ, Walkey AJ, Mizgerd JP. Integrative physiology of pneumonia. *Physiol Rev.* (2018) 98:1417–64. doi: 10.1152/physrev.00032.2017

15. Jiang M, Zhang S, Yang Z, Lin H, Zhu J, Liu L, et al. Self-recognition of an inducible host lncRNA by RIG-I feedback restricts innate immune response. *Cell.* (2018) 173:906–19.e13. doi: 10.1016/j.cell.2018.03.064

16. Kotla S, Gustin KE. Proteolysis of IFIH1 and IPS-1 is not required for inhibition of the type I IFN response by poliovirus. *Virol J.* (2015) 12:158. doi: 10.1186/s12985-015-0393-2

17. Kawai T, Takahashi K, Sato S, Coban C, Kumar H, Kato H, et al. an adaptor triggering RIG-I- and IFIH1-mediated type I interferon induction. *Nat Immunol.* (2005) 6:981–8. doi: 10.1038/ni1243

18. Zhang S, Chu C, Wu Z, Liu F, Xie J, Yang Y, et al. IFIH1 contributes to M1 macrophage polarization in ARDS. *Front Immunol.* (2021) 11:580838. doi: 10.3389/fimmu.2020.580838

19. Wang A, Kang X, Wang J, Zhang S. IFIH1/IRF1/STAT1 promotes sepsis associated inflammatory lung injury via activating macrophage M1 polarization. *Int Immunopharmacol.* (2023) 114:109478. doi: 10.1016/j.intimp.2022.109478

20. Zhang S, Huang W, Wu X, Chen H, Wang L, Chao J, et al. IBR1, a novel endogenous IFIH1-binding dsRNA, governs IFIH1 activation and M1 macrophage polarisation in ARDS. *Clin Transl Med.* (2024) 14:e70027. doi: 10.1002/ctm2.70027

21. Rawling DC, Jagdmann GE Jr, Potapova O, Pyle AM. Small-molecule antagonists of the RIG-I innate immune receptor. *ACS Chem Biol.* (2020) 15:311–7. doi: 10.1021/acschembio.9b00810

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 January 2025
DOI 10.3389/fmed.2024.1496869

OPEN ACCESS

EDITED BY
Qinghe Meng,
Upstate Medical University, United States

REVIEWED BY
Dung Tran,
Hospices Civils de Lyon, France
Mowafaq Salem Alzboon,
Jadara University, Jordan

*CORRESPONDENCE
Hongsheng Wu
✉ crazywu2007@126.com
Shengmin Zhang
✉ zsmin2008@163.com

RECEIVED 17 September 2024
ACCEPTED 10 December 2024
PUBLISHED 06 January 2025

CITATION
Wu H, Liao B, Ji T, Ma K, Luo Y and Zhang S (2025) Comparison between traditional logistic regression and machine learning for predicting mortality in adult sepsis patients.
Front. Med. 11:1496869.
doi: 10.3389/fmed.2024.1496869

Comparison between traditional logistic regression and machine learning for predicting mortality in adult sepsis patients

Hongsheng Wu*, Biling Liao, Tengfei Ji, Keqiang Ma, Yumei Luo and Shengmin Zhang*

Hepatobiliary Pancreatic Surgery Department, Huadu District People's Hospital of Guangzhou, Guangzhou, China

Background: Sepsis is a life-threatening disease associated with a high mortality rate, emphasizing the need for the exploration of novel models to predict the prognosis of this patient population. This study compared the performance of traditional logistic regression and machine learning models in predicting adult sepsis mortality.

Objective: To develop an optimum model for predicting the mortality of adult sepsis patients based on comparing traditional logistic regression and machine learning methodology.

Methods: Retrospective analysis was conducted on 606 adult sepsis inpatients at our medical center between January 2020 and December 2022, who were randomly divided into training and validation sets in a 7:3 ratio. Traditional logistic regression and machine learning methods were employed to assess the predictive ability of mortality in adult sepsis. Univariate analysis identified independent risk factors for the logistic regression model, while Least Absolute Shrinkage and Selection Operator (LASSO) regression facilitated variable shrinkage and selection for the machine learning model. Among various machine learning models, which included Bagged Tree, *Boost Tree, Decision Tree*, *LightGBM*, *Naïve Bayes*, *Nearest Neighbors*, *Support Vector Machine (SVM)*, and *Random Forest (RF)*, the one with the maximum area under the curve (AUC) was chosen for model construction. Model validation and comparison with the Sequential Organ Failure Assessment (SOFA) and the Acute Physiology and Chronic Health Evaluation (APACHE) scores were performed using receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA) curves in the validation set.

Results: Univariate analysis was employed to assess 17 variables, namely gender, history of coronary heart disease (CHD), systolic pressure, white blood cell (WBC), neutrophil count (NEUT), lymphocyte count (LYMP), lactic acid, neutrophil-to-lymphocyte ratio (NLR), red blood cell distribution width (RDW), interleukin-6 (IL-6), prothrombin time (PT), international normalized ratio (INR), fibrinogen (FBI), D-dimer, aspartate aminotransferase (AST), total bilirubin (Tbil), and lung infection. Significant differences ($p < 0.05$) between the survival and non-survival groups were observed for these variables. Utilizing stepwise regression with the "backward" method, independent risk factors, including systolic pressure, lactic acid, NLR, RDW, IL-6, PT, and Tbil, were identified. These factors were then incorporated into a logistic regression model, chosen based on the minimum Akaike Information Criterion (AIC) value (98.65). Machine learning techniques were also applied, and the RF model, demonstrating the maximum Area Under

the Curve (AUC) of 0.999, was selected. LASSO regression, employing the lambda.1SE criteria, identified systolic pressure, lactic acid, NEUT, RDW, IL6, INR, and Tbil as variables for constructing the RF model, validated through ten-fold cross-validation. For model validation and comparison with traditional logistic models, SOFA, and APACHE scoring.

Conclusion: Based on deep machine learning principles, the RF model demonstrates advantages over traditional logistic regression models in predicting adult sepsis prognosis. The RF model holds significant potential for clinical surveillance and interventions to enhance outcomes for sepsis patients.

KEYWORDS

machine learning, random forest, logistic regression, adult sepsis, mortality

Introduction

Sepsis represents a critical condition marked by organ dysfunction resulting from an imbalanced host response to infection, leading to high mortality rates and substantial healthcare costs (1, 2). Despite the establishment of the initial consensus definitions (Sepsis-1) in 1991, the global incidence of sepsis continues to rise, making the true epidemiology of sepsis a subject of ongoing concern. Further exploration of high-risk factors associated with sepsis-related mortality is essential (3). In clinical settings, the evaluation of sepsis severity and the identification of risk factors for mortality often rely on scoring systems such as Sequential Organ Failure Assessment (SOFA), Quick Sequential Organ Failure Assessment (qSOFA), and Acute Physiology and Chronic Health Evaluation (APACHE) (4–6). However, these scoring systems involve numerous parameters, posing challenges for clinical practitioners. Consequently, there has been a growing interest in exploring the effectiveness of biomarkers and clinical prediction models in predicting the prognosis of sepsis patients (7–9).

Over the past few years, linear regression models have dominated the clinical landscape for predicting sepsis mortality (10, 11). However, their limitations, including the inability to handle non-linearity among variables, sensitivity to outlier values, and the need to meet the linear regression hypothesis, constrain their utility with non-linear and imbalanced datasets (12). Machine learning (ML) is a subfield of artificial intelligence (AI) that focuses on developing systems capable of learning from data or improving performance. Specifically, machine learning is a technique that enables computers to create models by training algorithms using datasets (13). Previous studies had indicated that ML models play a crucial role in predicting the prognosis of sepsis patients. ML models had demonstrated superior predictive accuracy compared to traditional statistical methods. By leveraging complex algorithms, these models can identify non-linear relationships and interactions between variables that may be overlooked by simpler models, leading to more precise predictions of sepsis mortality (14, 15). On the other hand, the key strengths of ML models is their ability to handle high-dimensional data effectively. They can incorporate a vast array of clinical variables, which allows for a more comprehensive understanding of the patient's condition and the factors contributing to mortality risk (16, 17). Consequently, these above benefits position ML as an essential tool in the prediction of sepsis mortality, aiding in the improvement of clinical decision-making and patient outcomes.

Based on the methodological review mentioned above, we employed both traditional generalized linear regression and ML models to assess their predictive capabilities in adult sepsis's mortality during their hospitalization duration. Notably, we conducted internal validation for both models and compared their performance with SOFA and APACHE scores in terms of discrimination, calibration, and clinical practicality. This comprehensive analysis offers profound insights into mortality risk adjustment for observational adult sepsis datasets, contributing valuable information to the understanding of predictive models and their applicability in clinical settings. This study closely complied with TRIPOD guidelines (18) and the PROBAST risk of bias tool (19).

Methods

Source of clinical data

The clinical data for this cross-sectional study were obtained from the electronic medical records of Huadu District People's Hospital of Guangzhou, Southern Medical University. The study focused on adult patients diagnosed with sepsis during hospitalization from January 2020 to December 2022, adhering to the Sepsis-3 definition (20, 21). Exclusion criteria included patients under 18 years old, those with malignant tumors, individuals with immunosuppression, those who died or withdrew treatment within 24 h of admission, and cases where clinical data could not be extracted. Following these criteria, 606 cases of adult sepsis were included in the study. Due to it directly reflects the sepsis patient's survival and is a key performance indicator during the hospitalization duration, we define the mortality as the outcome of this study.

Variables extraction

Variable extraction involved retrieving general information (gender, age, and body mass index), medical history [hypertension, diabetes, and coronary heart disease (CHD)], clinical signs (temperature, heart rate, systolic pressure, and infection site), laboratory examination results [white blood cell count (WBC), platelet count, neutrophil (NEUT) and lymphocyte (LYMP) counts, neutrophil-to-lymphocyte ratio (NLR), red cell distribution width (RDW), C-reactive protein, procalcitonin, lactic acid, prothrombin time (PT), international normalized ratio (INR), fibrinogen (FIB), D-dimer, creatinine, alanine transaminase (ALT), aspartate transaminase (AST), total bilirubin (Tbil), and interleukin-6 (IL-6)], etiologic detection (Gram-positive bacteria, Gram-negative

bacteria, or fungal), and severity scores of sepsis (SOFA and APACHE) from the electronic medical record system. All data were extracted within the first 24 h of patient admission. For missing values, multiple imputation was performed using the "*mice*" package in R software.

Model construction of logistic regression

The study divided the 606 adult sepsis cases randomly into a training set ($n = 435$) and a validation set ($n = 171$) at a ratio of 7:3. Based on whether the patient died or not between 24 h after admission and discharge, participants were categorized into a survival group (421) and a non-survival group (185). For traditional logistic regression model construction, univariate analysis identified significant risk factors ($p < 0.05$), which were then included in binary logistic regression. The stepwise regression with the "backward" method was employed to achieve the optimal model with the least AIC value.

Machine learning model selection and construction

For ML model selection, eight integrated algorithms, including *Bagged Tree*, *Boost Tree*, *Decision Tree*, *LightGBM*, *Naïve Bayes*, *Nearest Neighbors*, *Support SVM*, and *RF*, were considered. In the "*tidymodels*" framework of R software, workflow sets were used to compare these models, perform resampling, and tune parameters. Because of its ability to provide a comprehensive measure of a model's performance across all classification thresholds, we select AUC as an optimum index in order to offer more nuanced view of model performance. The ML model with the highest AUC value was chosen for model construction. In order to perform variable shrinkage and selection, which may avoid the overfitting of the ML model, we utilized LASSO regression with ten-fold cross-validation. The count of variables in the ultimate model was ascertained based on the specific location of lambda.1SE, a coefficient that signifies the ideal equilibrium between model intricacy and forecasting accuracy.

Models validation and comparison

Models were validated and compared using discrimination, calibration, clinical benefit, and generalization. Discrimination was assessed by calculating the AUC of the ROC, while calibration was evaluated using calibration curves and the Hosmer-Lemeshow test. Decision curve analysis (DCA) curves were employed to assess the clinical benefit of the models. To estimate generalization, logistic regression, and ML models were compared with SOFA and APACHE scores using discrimination, calibration, and DCA for both the training and validation sets. The research design flowchart is depicted in Figure 1.

Variables importance

During traditional logistic regression, the importance of variables is determined by assessing the absolute value of each regression coefficient from the covariate. A larger absolute value indicates a more significant and important predictor. Variable importance is a key characteristic of ML models. In ML models, if changing the value of a variable leads to false prediction results, it implies that the variable is sensitive to classification outcomes and holds greater importance. The calculation of variable importance in the ML model such as RF involves determining the importance of each single decision tree, and by considering the number of trees set in the RF, the average of these values yields the overall variable importance of the RF model (22, 23).

Ethics statement

Data extraction and collection for this study were approved by the Ethics Committee of Huadu District People's Hospital of Guangzhou (Registration Number: 2023088). Due to the retrospective nature of the study, the Ethics Committee of Huadu District People's Hospital of Guangzhou waived the need of obtaining informed consent. And we had confirmed that the method of this research was performed in accordance with the regulation of Ethics Committee of Huadu District People's Hospital of Guangzhou.

Statistical analysis

R version 4.1.3 was used for data analysis and the creation of statistical figures. Missing values in this cross-sectional study were addressed through multiple imputations using the "*mice*" package. The study population of adult sepsis patients was divided into training and validation sets using the "*caret*" package. Descriptive statistics, including mean ± standard deviation for continuous data with normal distribution and median (upper and lower quartiles) for non-normally distributed data, were employed to characterize average values. For univariate analysis, the Chi-square test was used to analyze differences in categorical data, while *t*-tests and Mann–Whitney *U* tests were employed for normally and non-normally distributed continuous data, respectively.

For logistic regression model construction, the "*glm*" function was used to conduct univariate analysis and binary logistic regression. The final model was determined by stepwise regression with the least AIC value using the "backward" method. For machine learning models, the Least Absolute Shrinkage and Selection Operator (LASSO) regression was utilized for variable shrinkage and selection, and the "*glmnet*" package was employed with parameter tuning using lambda.1SE under ten-fold cross-validation to remove irrelevant variables. The framework of "*tidymodels*" facilitated model selection, construction, workflow settings, validation, and comparison of predictive capabilities. Random forest was selected as the machine learning model based on the highest AUC and accuracy values, and the "*randomForest*" package was used to fit the model. To optimize the out-of-bag (OOB) error and improve predictive efficacy, the "*tuneRF*" function was employed.

Discrimination of the models was investigated using ROC and AUC with the "*pROC*" package. Calibration was assessed using the "*calibration*" function from the "*rms*" package, and the Hosmer-Lemeshow test was performed. Net benefits, reflecting model clinical

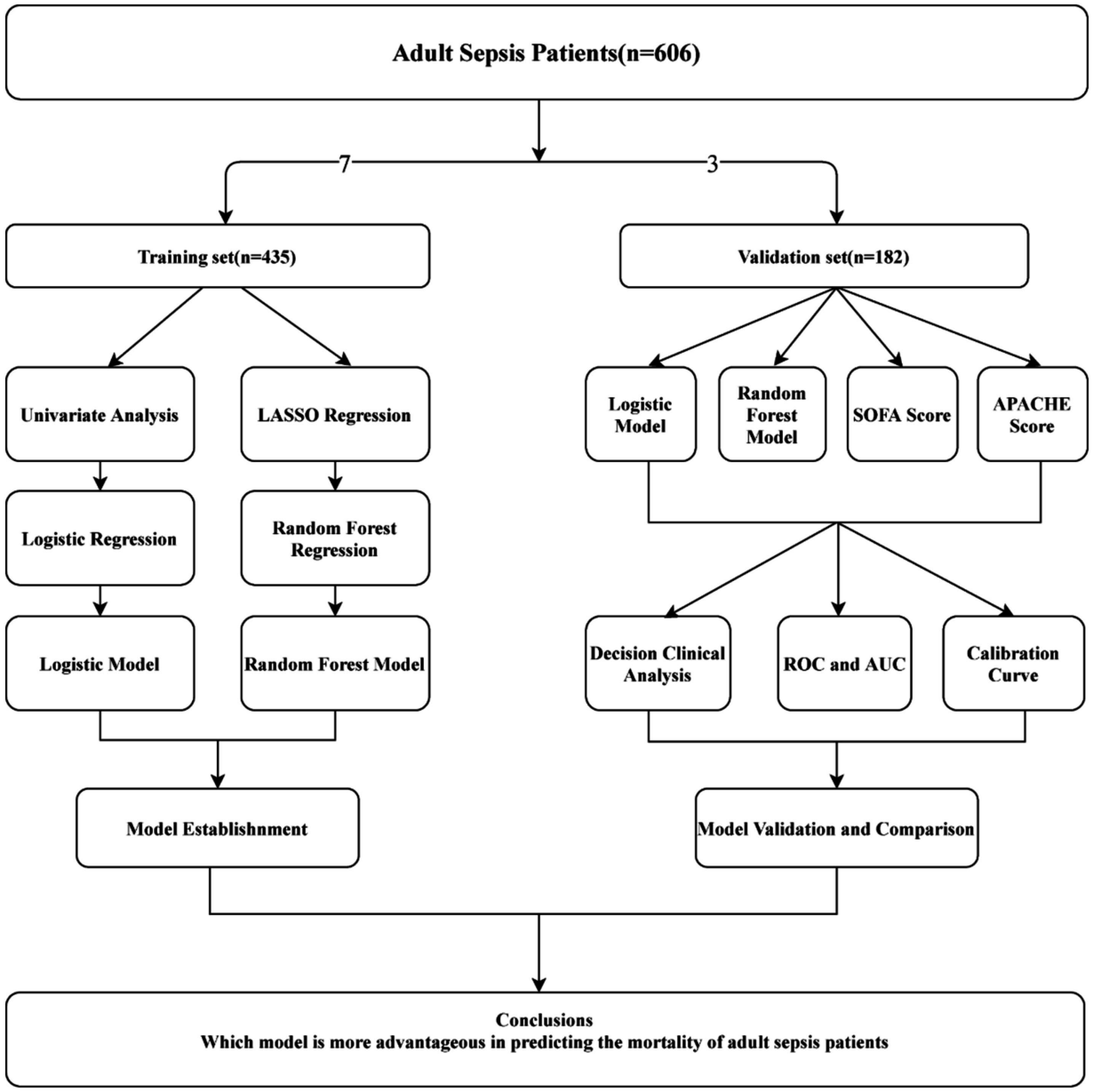

FIGURE 1
Flowchart illustrating the research design.

practicality, were calculated and compared using DCA with the "*ggDCA*" package.

Results

Baseline analysis and data splitting

The study included a total of 606 patients diagnosed with sepsis, categorized into a survival group ($n = 421$) and a non-survival group ($n = 185$) based on hospital stay duration. To facilitate model construction and validation, a random allocation resulted in a training set ($n = 435$) and a validation set ($n = 171$) at a 7:3 ratio. Details of the baseline analysis and data splitting are presented in Table 1.

Logistic regression model construction

We initially conducted univariate analysis for risk factor selection in the training set for the logistic regression model. The results of the univariate analysis revealed that with significant differences ($p < 0.05$) between the survival and non-survival groups for including variables, gender, CHD, systolic pressure, WBC, NEUT, LYMP, lactic acid, NLR, RDW, IL6, PT, INR, FBI, D-dimer, AST, Tbil, and lung infection were brought into multiple variable regression (Table 2). Based on these 17 variables, we utilized logistic step regression (backward step method) to optimize the model according to the Akaike Information Criterion (AIC). The results showed that with the least AIC value (98.65), the logistic model (OR: 1.012, 95% CI: 2.218–3.216) included systolic pressure, lactic acid,

TABLE 1 Baseline analysis and data splitting.

Characteristics		All patients (n = 606)	Survival (n = 421)	Non-survival (n = 185)	Training set (n = 435)	Validation set (n = 171)	p-value
Gender							0.914
Female	n (%)	269 (44.4%)	207(49.2%)	62(33.5%)	192 (44.1%)	77 (45.0%)	
Male	n (%)	337 (55.6%)	214(50.8%)	123(66.5%)	243 (55.9%)	94 (55.0%)	
Age	*$\overline{x}$ ± sd	64.0 ± 17.5	63.6 ± 16.5	64.9 ± 17.6	64.7 ± 18.2	63.2 ± 15.8	0.163
Infection site							0.806
Respiratory system	n (%)	232 (38.3%)	132(31.4%)	100(54.1%)	166 (38.2%)	66 (38.6%)	
Urinary system	n (%)	161 (26.6%)	100(23.6%)	61(33.0%)	113 (26.0%)	48 (28.1%)	
Digestive system	n (%)	213 (35.1%)	189(45.0%)	24(12.9%)	156 (35.9%)	57 (33.3%)	
Pathology							0.332
Gram-positive	n (%)	331 (54.6%)	223(53.0%)	108(58.4%)	230 (52.9%)	101 (59.1%)	
Gram-negative	n (%)	252 (41.6%)	180(42.8%)	72(38.9%)	189 (43.4%)	63 (36.8%)	
Fungal	n (%)	23 (3.80%)	18(4.2%)	5(2.7%)	16 (3.68%)	7 (4.09%)	
Diabetes							0.632
No	n (%)	411 (67.8%)	288(68.4%)	123(66.5%)	298 (68.5%)	113 (66.1%)	
Yes	n (%)	195 (32.2%)	133(31.6%)	62(33.5%)	137 (31.5%)	58 (33.9%)	
Hypertension							0.193
No	n (%)	360 (59.4%)	262(62.2%)	98(53.0%)	266 (61.1%)	94 (55.0%)	
Yes	n (%)	246 (40.6%)	159(37.8%)	87(47.0%)	169 (38.9%)	77 (45.0%)	
CHD							0.304
No	n (%)	390 (64.4%)	298(70.8%)	92(49.7%)	274 (63.0%)	116 (67.8%)	
Yes	n (%)	216 (35.6%)	123(29.2%)	93(50.3%)	161 (37.0%)	55 (32.2%)	
BMI	$\overline{x}$ ± sd	24.3 ± 4.6	24.3 ± 4.6	24.3 ± 4.7	24.2 ± 4.7	23.9 ± 4.5	0.693
Systolic pressure (mmHg)	$\overline{x}$ ± sd	124 ± 23.5	131.1 ± 18.0	93.66 ± 24.7	124.8 ± 19.2	124.6 ± 20.3	0.907
Heart rate (time/min)	$\overline{x}$ ± sd	122 ± 9.6	116.9 ± 11.58	118.1 ± 9.1	122 ± 8.8	121 ± 10.2	0.887
Temperature (°C)	$\overline{x}$ ± sd	38.5 ± 0.7	38.3 ± 0.7	38.3 ± 0.6	38.3 ± 0.6	38.5 ± 0.5	0.387
WBC ($\times 10^9$/L)	*M[P_{25};P_{75}]	12.7 [7.6;18.2]	11.9 [7.5;16.1]	15.0 [8.2;23.4]	12.9 [8.1;18.3]	11.8 [6.9;17.6]	0.107
Platelet ($\times 10^9$/L)	M[P_{25};P_{75}]	172 [111;249]	180.0 [125.0;252.0]	150.0 [63.0;242.5]	172 [114;246]	169 [95.5;263]	0.715
NEUT ($\times 10^9$/L)	M[P_{25};P_{75}]	10.6 [6.2;16.0]	9.8 [5.9;14.2]	14.1 [7.6;20.6]	10.7 [6.54;16.2]	9.85 [5.35;15.2]	0.149
LYMP ($\times 10^9$/L)	M[P_{25};P_{75}]	0.85 [0.5;1.4]	1.0 [0.7;1.5]	0.5 [0.3;0.6]	0.86 [0.5;1.4]	0.85 [0.5;1.4]	0.973

(Continued)

10.3389/fmed.2024.1496869

TABLE 1 (Continued)

Characteristics		All patients (n = 606)	Survival (n = 421)	Non-survival (n = 185)	Training set (n = 435)	Validation set (n = 171)	p-value
NLR	M[P_{25};P_{75}]	12.5 [6.6;22.1]	10.3 [5.5;16.0]	26.2 [13.0;46.5]	12.4 [6.8;22.8]	12.7 [5.9;20.4]	0.431
RDW	$\bar{x}$ ± sd	44.7 ± 8.4	45.2 ± 8.2	50.5 ± 10.5	44.3 ± 8.7	45.2 ± 9.4	0.368
CRP (mg/ml)	M[P_{25};P_{75}]	117 [47.6;179]	112.3 [40.5;166.7]	135.0 [56.9;194.2]	119 [46.6;179]	117 [54.9;176]	0.957
PCT (ng/ml)	M[P_{25};P_{75}]	12.0 [2.42;41.1]	9.1 [1.6;36.0]	18.0 [6.3;53.5]	12.0 [2.49;40.7]	12.0 [2.20;39.5]	0.879
Lactic acid (mmol/L)	M[P_{25};P_{75}]	2.20 [1.5;4.4]	1.9 [1.4;2.6]	4.8 [3.1;8.7]	2.15 [1.6;4.2]	2.50 [1.5;5.3]	0.390
PT (S)	$\bar{x}$ ± sd	15.1 ± 2.4	14.9 ± 2.2	18.9 ± 8.8	15.1 ± 3.1	14.9 ± 3.4	0.811
INR	M[P_{25};P_{75}]	1.19 [1.1;1.4]	1.2 [1.1;1.3]	1.3 [1.2;1.7]	1.19 [1.08;1.35]	1.17 [1.08;1.38]	0.634
FIB	$\bar{x}$ ± sd	4.8 ± 1.6	5.2 ± 1.8	4.4 ± 2.1	4.91 ± 1.7	4.6 ± 1.9	0.181
D dimer (ng/m)	M[P_{25};P_{75}]	2,532 [1,230;4,955]	2,183 [1,190;4,197]	3,365 [1,597;7,461]	2,544 [1,233;5,239]	2,251 [1,232;4,607]	0.569
Creatinine (μmol/L)	M[P_{25};P_{75}]	125 [74.0;246]	106.0 [70.0;214.8]	162.0 [92.5;289.0]	117 [72.6;234]	131 [79.3;272]	0.099
ALT (U/L)	M[P_{25};P_{75}]	28.0 [16.5;52.0]	24.0 [15.7;43.1]	36.0 [20.1;74.0]	29.0 [16.7;52.6]	26.0 [15.7;49.9]	0.350
AST (U/L)	M[P_{25};P_{75}]	31.0 [20.0;62.0]	26.0 [18.9;47.0]	57.1 [29.2;156.5]	31.0 [20.2;61.0]	31.5 [19.6;67.2]	0.859
Tbil (μmol/L)	M[P_{25};P_{75}]	15.6 [10.2;24.6]	14.5 [9.9;21.9]	18.0 [10.9;34.2]	15.7 [10.3;25.0]	14.5 [9.28;23.4]	0.428
IL6 (pg/ml)	M[P_{25};P_{75}]	3.70 [2.30;5.80]	3.0 [1.9;4.2]	6.5 [5.1;7.7]	3.70 [2.30;5.75]	3.70 [2.30;5.90]	0.696
SOFA	M[P_{25};P_{75}]	6.1 [4.2;9.5]	5.0 [4.0;6.1]	12.0 [9.1;14.0]	6.00 [4.0;9.0]	6.0 [4.0;9.0]	0.369
APACHE	$\bar{x}$ ± sd	32.0 ± 8.1	29.2 ± 7.5	40.9 ± 9.1	31.0 ± 78.3	34.0 ± 8.6	0.052

*$\bar{x}$ ± sd: mean ± standard deviation; M[P_{25};P_{75}]: M, median; P_{25}, Lower quartile; P_{75}, Upper quartile.

10.3389/fmed.2024.1496869

TABLE 2 Univariate analysis of risk factors between the survival group and non-survival group.

Characteristics	*B	*SE	*OR	95%CI	Z	P-value
Gender	0.699	0.303	2.01	1.11–3.65	2.308	0.021
Age	0.014	0.008	1.01	1–1.03	1.605	0.108
Diabetes	−0.142	0.336	0.87	0.45–1.68	−0.424	0.671
Hypertension	−0.016	0.308	0.98	0.54–1.8	−0.052	0.959
CHD	0.663	0.302	1.94	1.07–3.51	2.198	0.028
BMI	−0.033	0.033	0.97	0.91–1.03	−0.974	0.33
Systolic pressure	−0.104	0.014	0.9	0.88–0.93	−7.444	<0.001
Heart rate	0.001	0.016	1	0.97–1.03	0.091	0.928
Temperature	0.015	0.239	1.02	0.64–1.62	0.064	0.949
WBC	0.055	0.018	1.06	1.02–1.09	3.051	0.002
Platelet	−0.002	0.001	1	1–1.02	−1.643	0.1
NEUT	0.068	0.02	1.07	1.03–1.11	3.446	0.001
LYMP	−1.605	0.35	0.2	0.1–0.4	−4.586	<0.001
NLR	0.066	0.013	1.07	1.04–1.1	5.167	<0.001
RDW	0.071	0.018	1.07	1.04–1.11	4.016	<0.001
CRP	0.004	0.002	1	1–1.01	1.957	0.05
PCT	0.005	0.004	1.01	1–1.01	1.354	0.176
Lactic acid	0.496	0.084	1.64	1.39–1.94	5.929	<0.001
PT	0.529	0.093	1.7	1.41–2.04	5.686	<0.001
INR	2.751	0.613	15.65	4.71–52.05	4.49	<0.001
FIB	−0.172	0.079	0.84	0.72–0.98	−2.188	0.029
D dimer	0	0	1	1.23–2.45	4.438	<0.001
Creatinine	0.001	0.001	1	1–1.02	1.408	0.159
ALT	0.004	0.002	1	1–1.01	1.261	0.175
AST	0.005	0.001	1	1–1.01	3.198	0.001
Tbil	0.015	0.005	1.01	1–1.02	2.719	0.007
IL-6	0.728	0.102	2.07	1.7–2.53	7.157	<0.001
Infection site						
Respiratory system	1.839	0.325	6.29	3.33–11.89	5.664	<0.001
Urinary system	−0.17	0.349	0.84	0.43–1.67	−0.487	0.626

(Continued)

10.3389/fmed.2024.1496869

10.3389/fmed.2024.1496869

TABLE 2 (Continued)

Characteristics	*B	*SE	*OR	95%CI	Z	*P*-value
Digestive system	−1.941	0.395	0.14	0.75–1.95	−4.916	0.56
Pathology						
Gram-positive	0.235	0.303	1.26	0.7–2.29	0.774	0.439
Gram-negative	−0.14	0.309	0.87	0.47–1.59	−0.453	0.65
Fungal	−0.611	0.805	0.54	0.11–2.63	−0.759	0.448

*B: regression coefficient; SE: Standard Error; OR: Odds Ratio.

NLR, RDW, IL6, PT, and Tbil as the final determinants (Supplementary Table S1).

Machine learning model and variables selection

As depicted in Supplementary Figure S1, a comprehensive comparison of various machine learning models revealed that RF stood out with high accuracy and an AUC value of 0.99. Therefore, RF was chosen as the preferred methodology for model construction. Regarding variable selection, LASSO regression, employing ten-fold cross-validation with lambda.1SE criteria, identified systolic pressure, lactic acid, NEUT, RDW, IL6, INR, and Tbil as the chosen variables for RF model construction (Figure 2). During the RF modeling process, we initially set 500 decision trees for preliminary model calculation in the training set. To determine the optimal parameter for mortality prediction in sepsis patients, we utilized the OOB error as a measure of the model's performance index. The results demonstrated that when the iteration reached 141 decision trees, the error rates of both OOB and model classification showed a noticeable decrease, reaching a stable state. This observation illustrated that the RF model achieved the most stable and optimal situation (Figure 3).

Model validation and multi-models comparison

To assess the predictive efficacy of traditional logistic and RF models, we conducted assessments of discrimination, calibration, and clinical net benefits. Additionally, we compared the performance of logistic regression and RF models with SOFA (4) and APACHE (6) to explore clinical practicality. Discrimination results indicated that the among the predictive models of RF, logistic, SOFA, and APACHE, the AUCs and their corresponding 95% confidence intervals (CIs) were significantly larger ($P_{\text{Delong's test}} < 0.05$) in both training and validation sets compared to other three models (Figures 4A,B). For model calibration, we observed that calibration curves of RF were notably closer to the ideal reference line compared to other models in both training and validation sets, which indicated that comparing to other models, the RF model associated with better fitting goodness and predictive ability (Figures 5A,B). Results of clinical practicality, as indicated by the Area Under Decision Curve (AUDC), showed that in the training set (Figure 6A) and validation set (Figure 6B), comparing with other three models, the AUDCs of RF model were with the highest values. These findings illustrated that the RF model yielded optimal clinical net benefit for predicting mortality in adult sepsis patients.

Variables importance of logistic and RF models

The variable importance calculations from both the logistic regression and RF models are presented in Supplementary Figure S2. In predicting mortality in the adult sepsis cohort, the logistic regression model identified systolic pressure, lactic acid, IL6, and NLR as the most important variables, followed by Tbil, PT, and

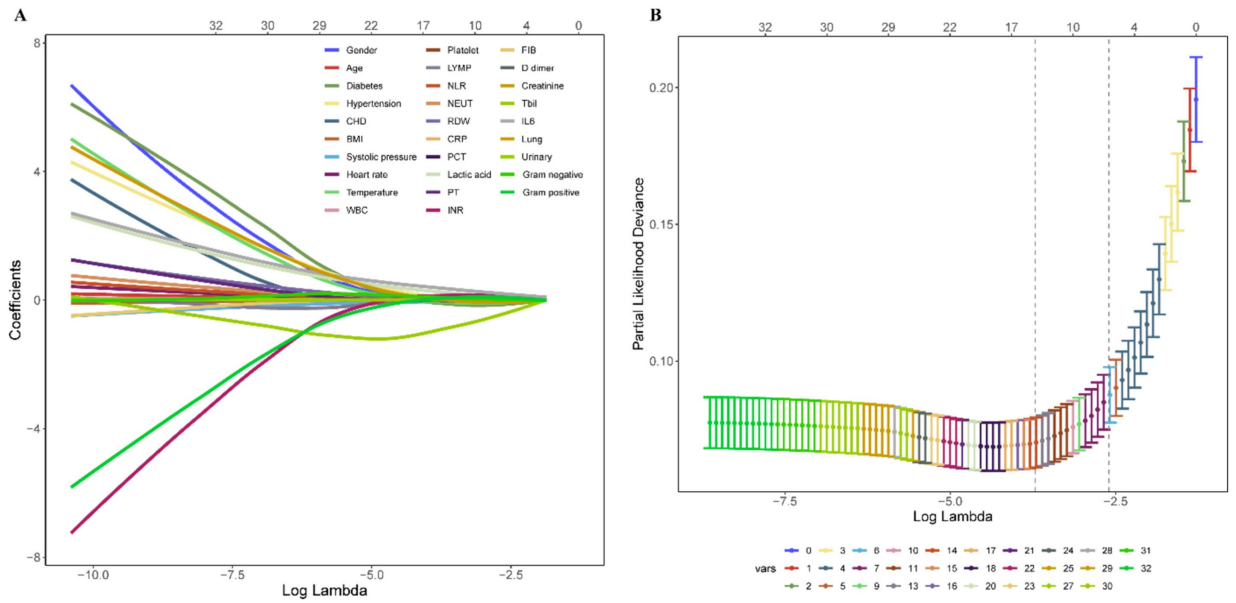

FIGURE 2
Variable shrinkage and selection by LASSO regression. **(A)** Shrinkage pathway of LASSO regression. **(B)** Based on ten-fold cross-validation, seven variables, including systolic pressure, lactic acid, NEUT, RDW, IL6, INR, and Tbil, were chosen using the lambda.1SE criteria.

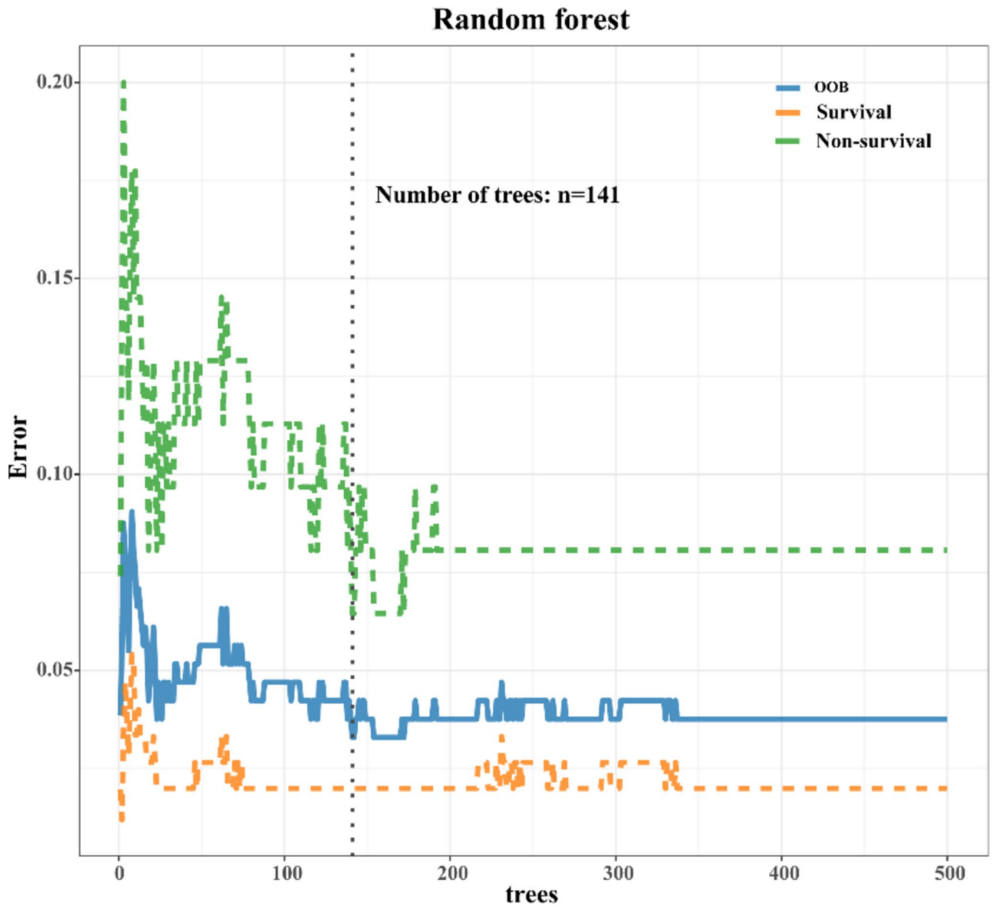

FIGURE 3
Error rate chart of RF model. As the iteration reached 141 decision trees, the error rates of both out-of-bag (OOB) and model classification showed a noticeable decrease, eventually reaching a steady state.

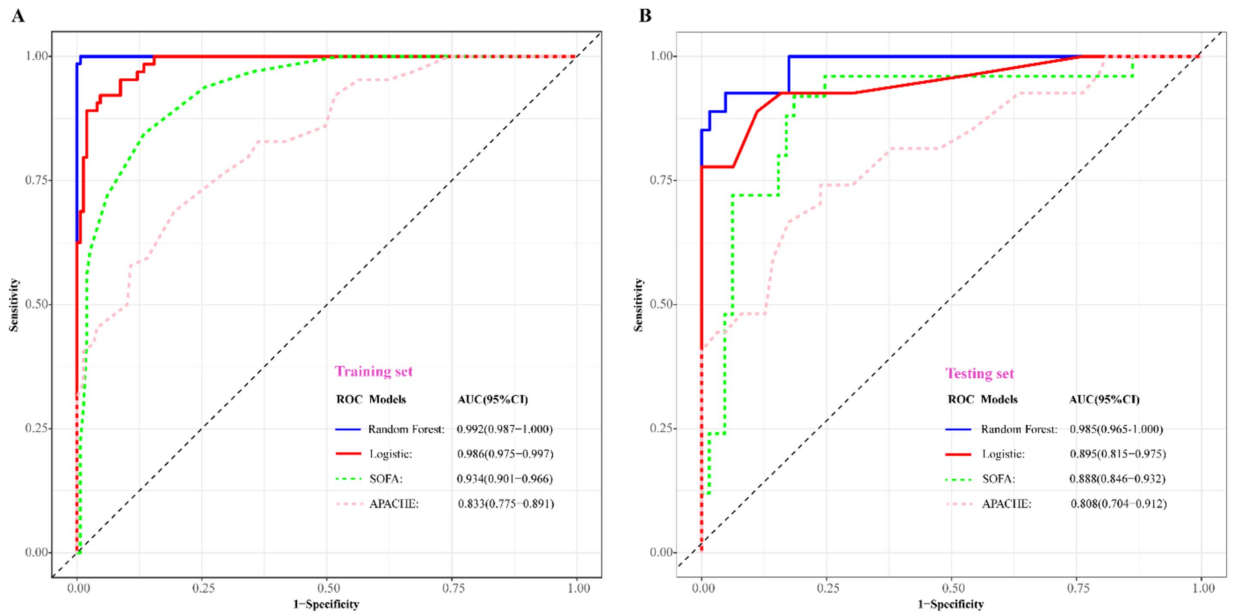

FIGURE 4
Comparison of discriminative ability among RF, logistic regression, SOFA, and APACHE scoring system. **(A)** Training set; **(B)** validation set. The blue solid ROC curves with the largest AUC values both in training set and validation set represented that RF associated with the best discrimination among the four models. AUC, area under curve; SOFA, sequential organ failure assessment scoring; APACHE, acute physiology and chronic health evaluation scoring.

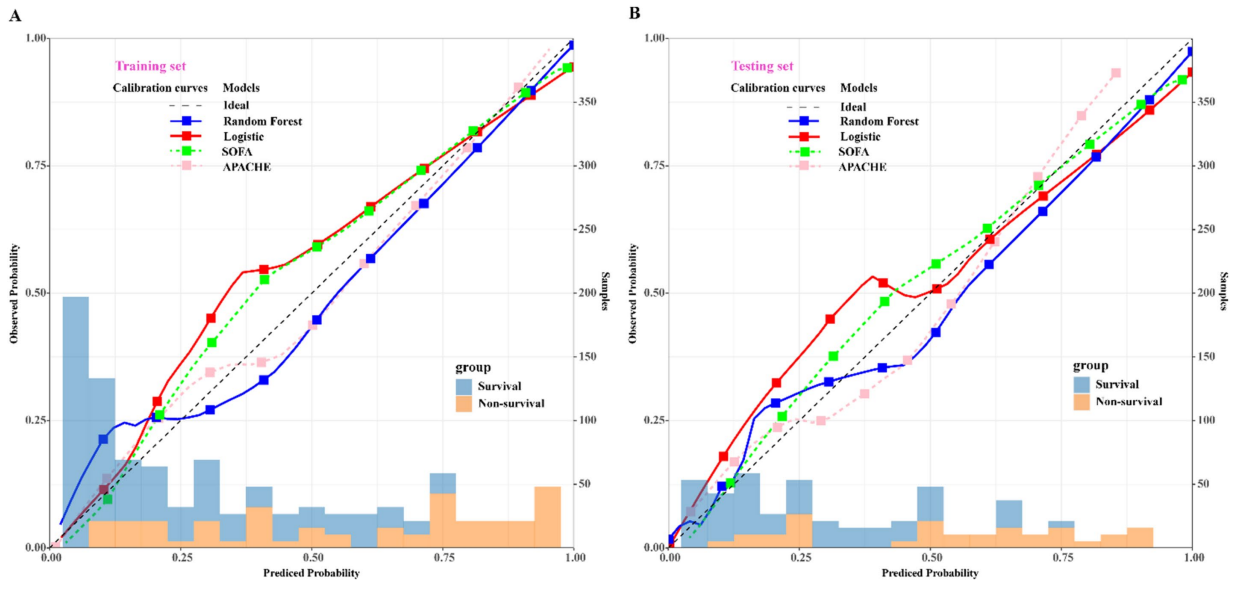

FIGURE 5
Comparison of calibration curves among RF, logistic regression, SOFA, and APACHE scoring system. **(A)** Training set; **(B)** validation set. The blue solid calibration curves which were notably closer to the ideal reference line both in training set and validation set represented that RF associated with the best goodness-of-fit and accuracy of prediction among the four models. SOFA, sequential organ failure assessment scoring; APACHE, acute physiology and chronic health evaluation scoring. The left *x-axis* represents the observed probability; the right *x*-axis represents the sample size, *y*-axis represents the predicted probability.

RDW. Consistently, the RF model also highlighted systolic pressure, lactic acid, IL6, and NLR as the most crucial variables for predicting mortality. However, the variables with relatively less importance in the RF model were RDW, NEUT, and Tbil, in contrast to the logistic regression model.

Discussion

In this study, we investigated the risk factors predicting the mortality of adult patients with sepsis, employing both the traditional logistic regression approach and the RF approach. Overall, both models

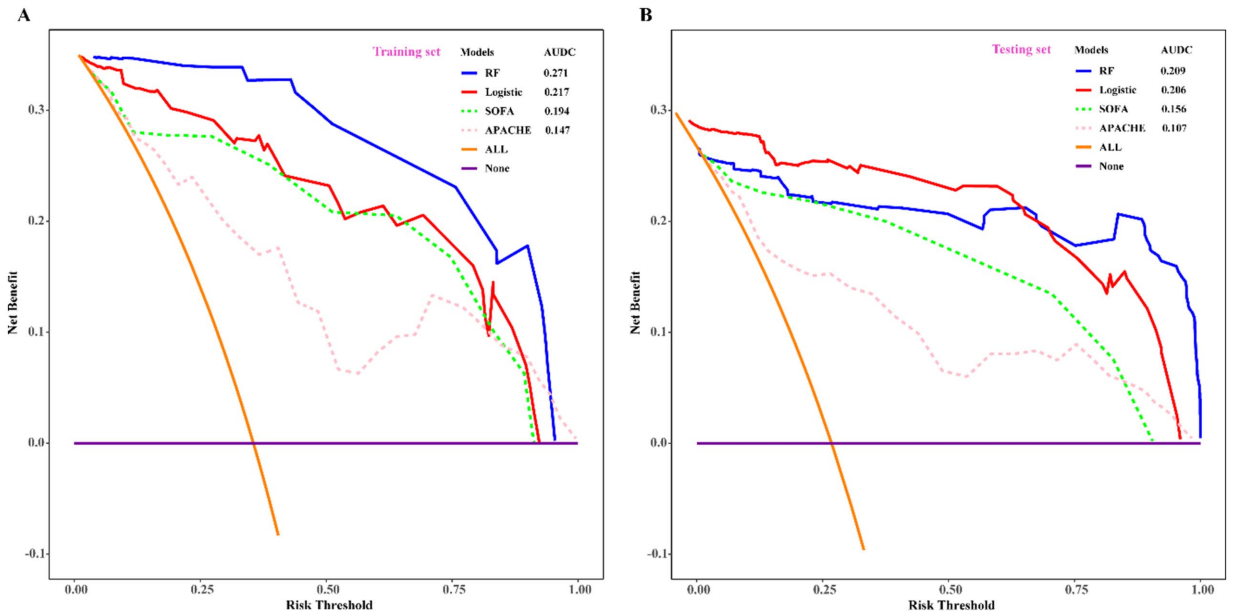

FIGURE 6
Comparison of decision curve analysis among RF, logistic regression, SOFA, and APACHE scoring system. **(A)** Training set; **(B)** validation set. With the highest value of AUDC and net benefit both in training set and validation set, RF was considered as the optimum model which associated with the best clinical practicality. SOFA, sequential organ failure assessment scoring; APACHE, acute physiology and chronic evaluation scoring. AUDC, area under DCA curve.

yielded similar results, with only slight differences in the included variables, with the inclusion of PT as a risk factor in the logistic regression model, while NEUT was included in the RF model. To assess the predictive capabilities of these models for adult sepsis prognosis, we conducted comprehensive validations, considering discrimination, calibration, and clinical benefits. Among the three above criterion of model assessment, calibration is a critical aspect of evaluating the performance of clinical prediction models. It refers to the degree to which the predicted probabilities of an event match the actual observed outcomes. A well-calibrated model is one where the predicted probabilities are reliable indicators of the likelihood of the event occurring in practice. This is particularly important in clinical settings, where accurate predictions can guide treatment decisions and patient management. Additionally, we compared the models with the widely used SOFA and APACHE scoring systems based on these criteria. The results of model validation and comparison demonstrated that the RF model exhibited significant superiority over the logistic regression model, as well as over the SOFA and APACHE scoring systems, in predicting mortality in adult sepsis patients.

Application of biomarkers in adult sepsis prediction

Sepsis represents an aberrant inflammatory response triggered by pathogenic microorganism infection. There is an increasing consensus suggesting that the immune system's activation in the early stages and its subsequent inhibition in the later stages can both contribute to alterations in circulating levels of inflammatory mediators (24–26). While the exact mechanisms of sepsis remain incompletely understood, studies have highlighted the crucial role of biomarkers in sepsis diagnosis and prognosis prediction, significantly impacting the risk of mortality (27–29). Our study exhibited that besides systolic pressure, biomarkers such as lactic acid, RDW, NLR, IL6, NEUT, and Tbil were incorporated into the traditional logistic and RF models we constructed. A closer examination through variable importance analysis revealed that lactic acid, NLR, and IL6 played pivotal roles in determining the significance of variables in both models.

Lactic acid, a metabolic byproduct of anaerobic glucose fermentation, poses a threat to the human body when present at elevated levels. High concentrations of lactic acid not only inhibit the activity of various essential enzymes but also mitigate the sensitivity of endothelial cells to vasoactive drugs (30). Furthermore, in the context of microbial infection or sepsis, lactic acid assumes a critical role in suppressing immune cells, potentially leading to immune suppression and severe consequences for the individual (31). Elevated levels of lactic acid in patients with sepsis are associated with poor outcomes, as they reflect inadequate perfusion and oxygen delivery to tissues. Studies have shown that high lactate levels correlate with increased mortality rates in septic patients, making it a valuable prognostic marker. Over the years, numerous studies have underscored the association between elevated lactic acid levels and increased mortality rates in sepsis (4, 32, 33).

The NLR serves as a biomarker calculated by the ratio of neutrophil to lymphocyte counts, encompassing both the innate immune response, primarily mediated by neutrophils, and adaptive immunity, supported by lymphocytes (34). Neutrophils act as the frontline defenders against pathogen invasion through processes like chemotaxis and phagocytosis. Upon activation by pathogens, various cytokines, granular proteins, and reactive oxygen species (ROS) are produced and released by neutrophils (35). While this activation is crucial for pathogen resistance, excessive activation leading to

increased production of ROS and cytokines may damage vascular endothelial cells through different mechanisms, resulting in tissue hypoperfusion and life-threatening organ failure (36). Consequently, an elevated neutrophil count, or a decreased lymphocyte count, contributes to an increased NLR, serving as a predictor of disease severity and poor prognosis in various conditions such as severe trauma (37), stroke (38), malignant tumor (39, 40) and sepsis (41, 42). Previous studies on NLR in predicting sepsis prognosis have demonstrated its independent association with high in-hospital mortality rates, showcasing significant advantages over conventional scores like SOFA or APACHE (43, 44). In summary, NLR stands out as a valuable biomarker for predicting mortality in sepsis patients.

Pro-inflammatory cytokines play a critical role in sepsis pathogenesis. IL-6, a member of the 4-helical cytokine family, activates signaling pathways by binding to an 80-kDa cytokine receptor (IL-6R). IL-6 plays a pivotal role in the immune response to infection, and it is released by various cells, including macrophages and T cells, in response to inflammatory stimuli. During sepsis, IL-6 is produced in response to pathogenic stimuli, and IL-6R is generated by neutrophils. Consequently, the IL-6/IL-6R complex triggers the phosphorylation and redistribution of VE-cadherin, leading to vascular endothelial damage and leakage (45). Excessive vascular endothelial damage and leakage in sepsis patients can result in blood pressure decline, hemodynamic collapse, irreversible septic shock, and even death. Clinical predictive models have consistently shown that IL-6 holds favorable predictive value for sepsis severity and prognosis. Elevated levels of IL-6 suggest severe illness and poor prognosis (46, 47). Moreover, studies have indicated that immunotherapeutic blockade of IL6 could reduce the mortality rate in sepsis (48).

Application of advanced statistical methods to complement common approaches

The RF algorithm possesses numerous statistical and computational advantages. This algorithm employs integrated learning, wherein its fundamental component is typically a decision tree, placing it within the broader category of integrated learning methods (49, 50). The terminology "random" and "forest" in RF signifies the amalgamation of classifiers, where each tree functions as an individual classifier. Notably, RF operates with hundreds of trees in parallel, collectively forming a forest. RF consolidates the results of all classification votes, designating the category with the highest votes as the final output, aligning with the Bagging concept and reflecting the core idea of RF (51). In contrast to the traditional logistic regression algorithm, RF demonstrates several distinct advantages: (1) RF employs an integrated algorithm with exceptionally high accuracy; (2) The randomness in model construction reduces susceptibility to overfitting; (3) It can handle discrete, continuous, or high-dimensional data without requiring data normalization; (4) The OOB feature allows obtaining unbiased estimates of true errors during model generation without losing training data. In the present study, the RF model demonstrated its superiority in predicting the prognosis of adult sepsis, exhibiting better discrimination, calibration, and clinical decision-making compared to traditional statistical methods (52, 53). Although RF model improves prediction accuracy by integrating multiple decision trees, but this also makes their decision-making process relatively complex and difficult to explain. Each decision tree is trained based on a randomly selected subset of features, which increases the model's diversity but also makes it challenging to interpret. So as to address these limitations, we can solve these problems by conducting feature importance analysis, visualizing individual decision trees, employing local explainability methods, and integrating doctors' experiences and expertise, it is possible to address the limitations of interpretability to a certain extent.

However, the present study has limitations that should be acknowledged. Firstly, being retrospective and cross-sectional, it relies on some laboratory results reflecting the patient's condition at specific time points, which may not be generalizable to the entire population. We expect to validate the current research and strengthen the impact of this study through prospective research. Secondly, we explicitly state that our initial predictor selection was based on univariate analysis, which may not capture the full complexity of the relationships between the predictors and the outcome variable. Thirdly, despite RF's significant advantages in predicting sepsis mortality compared to traditional regression, its interpretability limitation remains noteworthy. Finally, due to this is a single center study and without testing on an independent dataset, the model's accuracy could be artificially inflated, reducing its generalizability. Therefore, integrating RF with traditional regression approaches could enhance the predictive capabilities of healthcare research in the future.

Conclusion

In conclusion, logistic regression and RF models were developed to predict mortality in adult sepsis patients, with both models identifying consistent risk factors. The RF model outperformed traditional regression and the SOFA and APACHE scoring systems, highlighting its superiority in mortality prediction.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Ethics Committee of Huadu District People's Hospital of Guangzhou. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

HW: Conceptualization, Formal analysis, Writing – original draft. BL: Data curation, Writing – original draft. TJ: Data curation, Writing – original draft. KM: Formal analysis, Software, Writing – original draft. YL: Formal analysis, Writing – review & editing. SZ: Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Construction of Major Subject of Huadu District People's Hospital of Guangzhou (Grant no. YNZDXK202201, 2022–2025) and Huadu District Basic and Applied Basic Research Joint Funded Project (Grant no. 23HDQYLH06).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1496869/full#supplementary-material

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for Sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287

2. Esposito S, De Simone G, Boccia G, De Caro F, Pagliano P. Sepsis and septic shock: new definitions, new diagnostic and therapeutic approaches. *J Glob Antimicrob Resist*. (2017) 10:204–12. doi: 10.1016/j.jgar.2017.06.013

3. Chiu C, Legrand M. Epidemiology of sepsis and septic shock. *Curr Opin Anaesthesiol*. (2021) 34:71–6. doi: 10.1097/ACO.0000000000000958

4. Liu Z, Meng Z, Li Y, Zhao J, Wu S, Gou S, et al. Prognostic accuracy of the serum lactate level, the SOFA score and the qSOFA score for mortality among adults with Sepsis. *Scand J Trauma Resusc Emerg Med*. (2019) 27:51. doi: 10.1186/s13049-019-0609-3

5. Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. *JAMA*. (2017) 317:290–300. doi: 10.1001/jama.2016.20328

6. Liu X, Shen Y, Li Z, Fei A, Wang H, Ge Q, et al. Prognostic significance of APACHE II score and plasma suPAR in Chinese patients with sepsis: a prospective observational study. *BMC Anesthesiol*. (2016) 16:46. doi: 10.1186/s12871-016-0212-3

7. Barichello T, Generoso JS, Singer M, Dal-Pizzol F. Biomarkers for sepsis: more than just fever and leukocytosis-a narrative review. *Crit Care*. (2022) 26:14. doi: 10.1186/s13054-021-03862-5

8. Pierrakos C, Velissaris D, Bisdorff M, Marshall JC, Vincent J-L. Biomarkers of sepsis: time for a reappraisal. *Crit Care*. (2020) 24:287. doi: 10.1186/s13054-020-02993-5

9. Lee JH, Kim S-H, Jang JH, Park JH, Jo KM, No T-H, et al. Clinical usefulness of biomarkers for diagnosis and prediction of prognosis in sepsis and septic shock. *Medicine*. (2022) 101:e31895. doi: 10.1097/MD.0000000000031895

10. Xu C, Zheng L, Jiang Y, Jin L. A prediction model for predicting the risk of acute respiratory distress syndrome in sepsis patients: a retrospective cohort study. *BMC Pulm Med*. (2023) 23:78. doi: 10.1186/s12890-023-02365-z

11. Bhavani SV, Semler M, Qian ET, Verhoef PA, Robichaux C, Churpek MM, et al. Development and validation of novel sepsis subphenotypes using trajectories of vital signs. *Intensive Care Med*. (2022) 48:1582–92. doi: 10.1007/s00134-022-06890-z

12. Burzykowski T, Geubbelmans M, Rousseau A-J, Valkenborg D. Generalized linear models. *Am J Orthod Dentofacial Orthop*. (2023) 164:604–6. doi: 10.1016/j.ajodo.2023.07.005

13. Deo RC. Machine learning in medicine. *Circulation*. (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593

14. García-Gallo JE, Fonseca-Ruiz NJ, Celi LA, Duitama-Muñoz JF. A machine learning-based model for 1-year mortality prediction in patients admitted to an intensive care unit with a diagnosis of sepsis. *Med Intensiva*. (2020) 44:160–70. doi: 10.1016/j.medin.2018.07.016

15. Agnello L, Vidali M, Padoan A, Lucis R, Mancini A, Guerranti R, et al. Machine learning algorithms in sepsis. *Clin Chim Acta*. (2024) 553:117738. doi: 10.1016/j.cca.2023.117738

16. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of Sepsis in the ICU using machine learning: a systematic review. *Front Med*. (2021) 8:607952. doi: 10.3389/fmed.2021.607952

17. Islam MM, Nasrin T, Walther BA, Wu C-C, Yang H-C, Li Y-C. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Prog Biomed*. (2019) 170:1–9. doi: 10.1016/j.cmpb.2018.12.027

18. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. (2021) 11:e048008. doi: 10.1136/bmjopen-2020-048008

19. Kaiser I, Mathes S, Pfahlberg AB, Uter W, Berking C, Heppt MV, et al. Using the prediction model risk of Bias assessment tool (PROBAST) to evaluate melanoma prediction studies. *Cancers*. (2022) 14:33. doi: 10.3390/cancers14123033

20. Fleischmann-Struzek C, Mellhammar L, Rose N, Cassini A, Rudd KE, Schlattmann P, et al. Incidence and mortality of hospital- and ICU-treated sepsis: results from an updated and expanded systematic review and meta-analysis. *Intensive Care Med*. (2020) 46:1552–62. doi: 10.1007/s00134-020-06151-x

21. Font MD, Thyagarajan B, Khanna AK. Sepsis and septic shock – basics of diagnosis, pathophysiology and clinical decision making. *Med Clin North Am*. (2020) 104:573–85. doi: 10.1016/j.mcna.2020.02.011

22. Liu Y, Zhou S, Wei H, An S. A comparative study of forest methods for time-to-event data: variable selection and predictive performance. *BMC Med Res Methodol*. (2021) 21:193. doi: 10.1186/s12874-021-01386-8

23. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe Sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med*. (2019) 47:1485–92. doi: 10.1097/CCM.0000000000003891

24. Cantey JB, Lee JH. Biomarkers for the diagnosis of neonatal Sepsis. *Clin Perinatol*. (2021) 48:215–27. doi: 10.1016/j.clp.2021.03.012

25. Grondman I, Pirvu A, Riza A, Ioana M, Netea MG. Biomarkers of inflammation and the etiology of sepsis. *Biochem Soc Trans*. (2020) 48:1–14. doi: 10.1042/BST20190029

26. Opal SM, Wittebole X. Biomarkers of infection and Sepsis. *Crit Care Clin*. (2020) 36:11–22. doi: 10.1016/j.ccc.2019.08.002

27. Zhou G, Liu J, Zhang H, Wang X, Liu D. Elevated endothelial dysfunction-related biomarker levels indicate the severity and predict sepsis incidence. *Sci Rep*. (2022) 12:21935. doi: 10.1038/s41598-022-26623-y

28. Makkar N, Soneja M, Arora U, Sood R, Biswas S, Jadon RS, et al. Prognostic utility of biomarker levels and clinical severity scoring in sepsis: a comparative study. *J Investig Med*. (2022) 70:1399–405. doi: 10.1136/jim-2021-002276

29. Li X, Li T, Dong G, Wei Y, Xu Z, Yang J. Clinical value of serum Interleukin-18 in neonatal Sepsis diagnosis and mortality prediction. *J Inflamm Res*. (2022) 15:6923–30. doi: 10.2147/JIR.S393506

30. Rezar R, Mamandipoor B, Seelmaier C, Jung C, Lichtenauer M, Hoppe UC, et al. Hyperlactatemia and altered lactate kinetics are associated with excess mortality in sepsis: a multicenter retrospective observational study. *Wien Klin Wochenschr*. (2023) 135:80–8. doi: 10.1007/s00508-022-02130-y

31. Luo Y, Li L, Chen X, Gou H, Yan K, Xu Y. Effects of lactate in immunosuppression and inflammation: Progress and prospects. *Int Rev Immunol*. (2022) 41:19–29. doi: 10.1080/08830185.2021.1974856

32. Jagan N, Morrow LE, Walters RW, Plambeck RW, Patel TM, Moore DR, et al. Sympathetic stimulation increases serum lactate concentrations in patients admitted with sepsis: implications for resuscitation strategies. *Ann Intensive Care*. (2021) 11:24. doi: 10.1186/s13613-021-00805-9

33. Tongyoo S, Sutthipool K, Viarasilpa T, Permpikul C. Serum lactate levels in cirrhosis and non-cirrhosis patients with septic shock. *Acute Crit Care*. (2022) 37:108–17. doi: 10.4266/acc.2021.00332

34. Buonacera A, Stancanelli B, Colaci M, Malatino L. Neutrophil to lymphocyte ratio: An emerging marker of the relationships between the immune system and diseases. *Int J Mol Sci*. (2022) 23:636. doi: 10.3390/ijms23073636

35. Mortaz E, Alipoor SD, Adcock IM, Mumby S, Koenderman L. Update on neutrophil function in severe inflammation. *Front Immunol*. (2018) 9:2171. doi: 10.3389/fimmu.2018.02171

36. Zhang H, Wang Y, Qu M, Li W, Wu D, Cata JP, et al. Neutrophil, neutrophil extracellular traps and endothelial cell dysfunction in sepsis. *Clin Transl Med*. (2023) 13:e1170. doi: 10.1002/ctm2.1170

37. Sabouri E, Majdi A, Jangjui P, Rahigh Aghsan S, Naseri Alavi SA. Neutrophil-to-lymphocyte ratio and traumatic brain injury: a review study. *World Neurosurg*. (2020) 140:142–7. doi: 10.1016/j.wneu.2020.04.185

38. Lattanzi S, Brigo F, Trinka E, Cagnetti C, Di Napoli M, Silvestrini M. Neutrophil-to-lymphocyte ratio in acute cerebral hemorrhage: a system review. *Transl Stroke Res*. (2019) 10:137–45. doi: 10.1007/s12975-018-0649-4

39. Lin N, Li J, Yao X, Zhang X, Liu G, Zhang Z, et al. Prognostic value of neutrophil-to-lymphocyte ratio in colorectal cancer liver metastasis: a meta-analysis of results from multivariate analysis. *Int J Surg*. (2022) 107:106959. doi: 10.1016/j.ijsu.2022.106959

40. Corbeau I, Jacot W, Guiu S. Neutrophil to lymphocyte ratio as prognostic and predictive factor in breast Cancer patients: a systematic review. *Cancers*. (2020) 12:958. doi: 10.3390/cancers12040958

41. Bu X, Zhang L, Chen P, Wu X. Relation of neutrophil-to-lymphocyte ratio to acute kidney injury in patients with sepsis and septic shock: a retrospective study. *Int Immunopharmacol*. (2019) 70:372–7. doi: 10.1016/j.intimp.2019.02.043

42. Liu S, Li Y, She F, Zhao X, Yao Y. Predictive value of immune cell counts and neutrophil-to-lymphocyte ratio for 28-day mortality in patients with sepsis caused by intra-abdominal infection. *Burns Trauma*. (2021) 9:tkaa040. doi: 10.1093/burnst/tkaa040

43. Shi Y, Yang C, Chen L, Cheng M, Xie W. Predictive value of neutrophil-to-lymphocyte and platelet ratio in in-hospital mortality in septic patients. *Heliyon*. (2022) 8:e11498. doi: 10.1016/j.heliyon.2022.e11498

44. Drăgoescu AN, Pădureanu V, Stănculescu AD, Chiuțu LC, Tomescu P, Geormăneanu C, et al. Neutrophil to lymphocyte ratio (NLR)-a useful tool for the prognosis of Sepsis in the ICU. *Biomedicines*. (2021) 10:75. doi: 10.3390/biomedicines10010075

45. Rose-John S, Jenkins BJ, Garbers C, Moll JM, Scheller J. Targeting IL-6 trans-signalling: past, present and future prospects. *Nat Rev Immunol*. (2023) 23:666–81. doi: 10.1038/s41577-023-00856-y

46. Ríos-Toro J-J, Márquez-Coello M, García-Álvarez J-M, Martín-Aspas A, Rivera-Fernández R, Sáez de Benito A, et al. Soluble membrane receptors, interleukin 6, procalcitonin and C reactive protein as prognostic markers in patients with severe sepsis and septic shock. *PLoS One*. (2017) 12:e0175254. doi: 10.1371/journal.pone.0175254

47. Song W, Tian F, Wang Y, Sun Q, Guo F, Zhao G, et al. Predictive value of C-reactive protein, procalcitonin, and interleukin-6 on 30-day mortality in patients with bloodstream infections. *Med Clin*. (2023) 160:540–6. doi: 10.1016/j.medcli.2023.01.022

48. Barrett D. IL-6 blockade in cytokine storm syndromes. *Adv Exp Med Biol*. (2024) 1448:565–72. doi: 10.1007/978-3-031-59815-9_37

49. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform*. (2023) 24:bbad002. doi: 10.1093/bib/bbad002

50. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. *Int J Med Inform*. (2022) 159:104679. doi: 10.1016/j.ijmedinf.2021.104679

51. Rigatti SJ. Random forest. *J Insur Med*. (2017) 47:31–9. doi: 10.17849/insm-47-01-31-39.1

52. Xie D, Ying M, Lian J, Li X, Liu F, Yu X, et al. Serological indices and ultrasound variables in predicting the staging of hepatitis B liver fibrosis: a comparative study based on random forest algorithm and traditional methods. *J Cancer Res Ther*. (2022) 18:2049–57. doi: 10.4103/jcrt.jcrt_1394_22

53. Fan S, Lin J, Wu S, Mu X, Guo J. Random forest model can predict the prognosis of hospital-acquired *Klebsiella pneumoniae* infection as well as traditional logistic regression model. *PLoS One*. (2022) 17:e0278123. doi: 10.1371/journal.pone.0278123

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 January 2025
DOI 10.3389/fmed.2024.1467565

OPEN ACCESS

EDITED BY
Rahul Kashyap,
WellSpan Health, United States

REVIEWED BY
Priyal Mehta,
Saint Vincent Hospital, United States
Md. Enamul Hoq,
University of Arkansas for Medical Sciences, United States

*CORRESPONDENCE
Wei Shen
✉ shenweiijs@outlook.com
Ning Zhou
✉ ursula1a5bos@hotmail.com

†These authors have contributed equally to this work

RECEIVED 20 July 2024
ACCEPTED 16 December 2024
PUBLISHED 06 January 2025

CITATION
Liu Y, Zhao S, Du W, Shen W and Zhou N (2025) Predicting the risk of gastroparesis in critically ill patients after CME using an interpretable machine learning algorithm – a 10-year multicenter retrospective study. *Front. Med.* 11:1467565. doi: 10.3389/fmed.2024.1467565

Predicting the risk of gastroparesis in critically ill patients after CME using an interpretable machine learning algorithm – a 10-year multicenter retrospective study

Yuan Liu[1†], Songyun Zhao[2†], Wenyi Du[1†], Wei Shen[1*] and Ning Zhou[1*]

[1]Department of General Surgery, Wuxi People's Hospital Affiliated to Nanjing Medical University, Wuxi, China, [2]Department of Neurosurgery, Wuxi People's Hospital Affiliated to Nanjing Medical University, Wuxi, China

Background: Gastroparesis following complete mesocolic excision (CME) can precipitate a cascade of severe complications, which may significantly hinder postoperative recovery and diminish the patient's quality of life. In the present study, four advanced machine learning algorithms—Extreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Machine (SVM), and *k*-nearest neighbor (KNN)—were employed to develop predictive models. The clinical data of critically ill patients transferred to the intensive care unit (ICU) post-CME were meticulously analyzed to identify key risk factors associated with the development of gastroparesis.

Methods: We gathered 34 feature variables from a cohort of 1,097 colon cancer patients, including 87 individuals who developed gastroparesis post-surgery, across multiple hospitals, and applied a range of machine learning algorithms to construct the predictive model. To assess the model's generalization performance, we employed 10-fold cross-validation, while the receiver operating characteristic (ROC) curve was utilized to evaluate its discriminative capacity. Additionally, calibration curves, decision curve analysis (DCA), and external validation were integrated to provide a comprehensive evaluation of the model's clinical applicability and utility.

Results: Among the four predictive models, the XGBoost algorithm demonstrated superior performance. As indicated by the ROC curve, XGBoost achieved an area under the curve (AUC) of 0.939 in the training set and 0.876 in the validation set, reflecting exceptional predictive accuracy. Notably, in the *k*-fold cross-validation, the XGBoost model exhibited robust consistency across all folds, underscoring its stability. The calibration curve further revealed a favorable concordance between the predicted probabilities and the actual outcomes of the XGBoost model. Additionally, the DCA highlighted that patients receiving intervention under the XGBoost model experienced significantly greater clinical benefit.

Conclusion: The onset of postoperative gastroparesis in colon cancer patients remains an elusive challenge to entirely prevent. However, the prediction model developed in this study offers valuable assistance to clinicians in identifying key high-risk factors for gastroparesis, thereby enhancing the quality of life and survival outcomes for these patients.

KEYWORDS

colonic neoplasms, intensive care unit, gastroparesis, prognosis, risk factor, machine learning

Introduction

Colon cancer is a malignant neoplasm that originates from the epithelial cells of the colonic mucosa, with a notably poor prognosis for affected patients. A significant proportion of these individuals require admission to the intensive care unit (ICU) for close monitoring and treatment due to complex clinical conditions or postoperative complications. Globally, approximately 8 million new cases of colon cancer are diagnosed annually, accounting for more than one-tenth of all newly diagnosed malignant tumors. As the modern diet, characterized by high fat, high meat, and low fiber intake, becomes increasingly prevalent, the incidence of colon cancer is expected to rise steadily (1). The current approach to treating colorectal cancer is primarily determined by the cancer's stage, the patient's overall health, and other individualized factors. However, radical surgical intervention remains the cornerstone of treatment. Hohenberger et al. (2) were the first to introduce complete mesocolic excision (CME), a surgical technique that involves the meticulous removal of both the tumor and the surrounding lymph nodes by excising the entire colonic mesentery and the associated lymphatic tissue in the region of the tumor. This procedure aims to achieve a higher tumor resection rate while minimizing the risk of recurrence (3). Pedrazzani et al.'s (4) retrospective study affirmed that the CME procedure ensures the complete excision of all cancerous tissue, thereby preventing the spread of the tumor to surrounding healthy tissues. Adhering to this principle has notably contributed to a significant reduction in local recurrence rates following surgery. Despite the notable benefits of CME in enhancing the outcomes and survival rates of colon cancer patients, the procedure is not without its risks, which can lead to complications that impact recovery and subsequent treatment. In some cases, these complications may necessitate admission to the ICU for further management. Gastroparesis, a condition characterized by impaired gastric emptying, is a frequently overlooked and often misdiagnosed complication following radical colon cancer surgery. While its incidence is more commonly associated with gastric cancer surgeries, its occurrence after colon cancer surgery should not be underestimated. Gastroparesis results from dysfunction in the nerves or muscles of the stomach, leading to delayed gastric emptying. Symptoms, including nausea, vomiting, bloating, and loss of appetite, often mimic the typical recovery process post-surgery. As such, these symptoms are frequently mistaken for normal postoperative reactions or mild dyspepsia, delaying or hindering timely diagnosis (5). Numerous studies have demonstrated that the development of gastroparesis in postoperative patients significantly heightens the risk of tumor recurrence and metastasis (6–8). Moreover, akin to other postoperative complications, gastroparesis leads to extended hospital stays and has increasingly become a formidable public health challenge worldwide (9). Gastroparesis, a prevalent complication following radical colon cancer surgery, not only imposes direct health risks but also has a profound impact on the financial wellbeing of patients and their families. Affected individuals may endure significant quality-of-life challenges, including malnutrition and diminished ability to perform daily activities, which can lead to a decrease in family income and an increased financial burden. Consequently, accurately predicting the onset of gastroparesis following total mesocolic excision and identifying high-risk patients is of paramount importance.

Surgeons typically assess the risk of gastroparesis in surgical patients based on clinical experience and examination reports; however, this approach has its limitations. On one hand, surgeons often rely on their own professional judgment and clinical expertise, leading to varying assessments of the same condition, which can be somewhat subjective. In more complex or rare cases, exclusive reliance on experience may result in biased evaluations. On the other hand, while clinical examinations and laboratory tests (such as blood tests and gastric emptying scans) provide valuable supporting information, they typically reflect the patient's current status and may not offer an accurate prediction of postoperative gastroparesis risk. Some clinicians also employ traditional linear models and logistic regression for risk factor studies of postoperative gastroparesis in an effort to improve prediction accuracy (10). However, the development of postoperative gastroparesis is rarely attributable to a single factor; rather, it results from the interplay of multiple factors, such as the type of surgery, patient age, underlying comorbidities, intraoperative manipulations, and anesthesia techniques. Traditional regression models typically assume the independence of variables, disregarding the complex interactions among these factors. This limitation has prompted clinical researchers to acknowledge that regression models alone are insufficient for addressing challenges in clinical disease prediction. In recent years, with the rapid advancements in data science and machine learning, an increasing number of studies have shifted toward more sophisticated algorithms, such as random forests, support vector machines, and deep learning (11). These machine learning techniques excel at discerning the unique characteristics of different patient types within vast datasets, thereby facilitating the development of personalized medical solutions. Each patient's

condition, genetic makeup, lifestyle, and other factors are distinct, yet traditional medical treatments often adhere to a "one-size-fits-all" approach. In contrast, machine learning can analyze detailed patient data to identify specific treatment needs, enabling more tailored and effective healthcare strategies (12).

In this study, a machine learning model was developed to predict high-risk factors for gastroparesis following CME for colon cancer, by analyzing the clinical data of critically ill patients in ICU wards. This model is capable of identifying high-risk individuals at risk of developing gastroparesis after colon cancer CME, without the need for conventional imaging techniques, such as abdominal CT, thereby offering a potential means to reduce healthcare costs.

Materials and methods

Study subjects

In this study, we utilized clinical data from two medical institutions: Wuxi People's Hospital, affiliated with Nanjing Medical University, and Wuxi Second People's Hospital. The inclusion criteria for cases were as follows: (1) patients who underwent laparoscopic-assisted CME or traditional open CME; (2) all patients were transferred to the ICU due to postoperative complications; (3) the surgical team consisted of senior physicians skilled in independently performing CME; and (4) postoperative pathology confirmed a diagnosis of colorectal cancer. Exclusion criteria included: (1) patients with concurrent malignant tumors; (2) patients with distant metastasis of colon cancer confirmed through pathological examination or imaging; (3) patients with severe cardiovascular or respiratory conditions; (4) patients with significant organ dysfunction, such as liver or kidney disease; and (5) patients with incomplete clinical data, missing cases, or lost to follow-up. All patients were followed for a minimum of 3 years post-surgery. This retrospective study was approved by the Ethics Committees of Wuxi People's Hospital and Wuxi Second People's Hospital, and was conducted with patient consent, with all personal information anonymized. The ethical approval number for this study is KY22086.

Study design and data collection

The dataset included 34 preoperative variables (collected within 24 h before surgery), intraoperative variables, and postoperative variables (assessed 48 h after the initial surgery). Preoperative variables encompassed patient demographics (gender, age, smoking history, alcohol use, and body mass index), fundamental clinical characteristics (American Society of Anesthesiologists score, Nutrition Risk Screening 2002 score, history of prior surgeries, adjuvant chemotherapy, and adjuvant radiotherapy), medical history (anemia, diabetes, hypothyroidism, hypertension, chronic obstructive pulmonary disease, hyperlipidemia, and coronary artery disease), laboratory test results (albumin, carcinoembryonic antigen, and carbohydrate antigen 19-9), and tumor characteristics (T-stage, N-stage, peripheral nerve invasion, tumor size, and tumor number). Intraoperative variables included the type of surgery, surgical approach, surgery duration, intraoperative blood loss, blood transfusions, and percutaneous arterial oxygen saturation levels. Postoperative variables consisted of laboratory indices (procalcitonin, C-reactive protein, and serum amyloid A). The primary outcome of this study was the incidence of postoperative gastroparesis.

Diagnosis of gastroparesis

The diagnostic criteria for postoperative gastroparesis are as follows: (1) the presence of gastrointestinal symptoms, including nausea, vomiting, early satiety, bloating, or epigastric pain; (2) the exclusion of other conditions that may present with similar gastrointestinal symptoms, such as mechanical obstruction, drug-induced side effects, or metabolic disorders; (3) the elimination of confounding factors, such as the use of medications that may impair smooth muscle contraction; and (4) the confirmation of delayed gastric emptying through transgastric scintigraphy or magnetic resonance imaging (13, 14).

Development and evaluation of predictive models for machine learning algorithms

In the present study, statistical analyses were performed using SPSS and R software. The construction and evaluation of the clinical prediction models involved the following steps: (1) Data preprocessing: colon cancer patients from Wuxi People's Hospital between January 2010 and January 2020 were designated as the internal validation set, while patients from Wuxi Second People's Hospital during the same period served as the external validation set. The internal validation set was randomly divided into a training set (70%) and a test set (30%). This approach strikes a balance between evaluating model performance and generalization ability, allowing the model to be trained on a substantial portion of the data while reserving a portion for testing the model's predictive accuracy. Furthermore, given the moderate size of the dataset, this ratio ensures the training set contains a sufficient number of samples to capture key patterns and features, while the 30% test set provides an adequate sample for validating the model's generalizability. (2) Univariate and multivariate regression analyses were performed on the internal validation set data. The Chi-square test was applied to categorical variables, while the *t*-test was used for continuous variables with a normal distribution. For continuous variables that were not normally distributed, the rank sum test was employed. A *p*-value of less than 0.05 was considered statistically significant. Logistic regression analysis was conducted on variables identified as significant in univariate analysis to assess their independent effects on postoperative gastroparesis. Four models—Extreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Machine (SVM), and *k*-nearest neighbor (KNN)—were utilized to evaluate the importance of each factor and rank them accordingly. Variables that ranked in the top 10 across all 4 models and were deemed meaningful in both univariate and multivariate analyses were selected. These four models represent different types of machine learning algorithms: tree-based models (XGBoost and RF), KNN, and SVM. By

combining these diverse model types, the limitations of any single algorithm can be mitigated, providing a more comprehensive and objective evaluation of factor importance. Both XGBoost and RF are integrated decision-tree-based models that inherently produce feature importance scores. These models are well-suited to handle non-linear relationships and complex interactions, making them highly effective for analyzing datasets with numerous variables and identifying key factors. SVM, with its strong generalization capability, is particularly suited to situations with small sample sizes, allowing the model to remain sensitive to a few key variables while minimizing the risk of overfitting. KNN, despite being sensitive to data noise, offers valuable insights in small sample or local similarity analyses, providing an intuitive reflection of the relationship between variables and outcomes. By comparing these different algorithms, a more holistic assessment of each factor's importance can be made, and key variables that perform consistently well across multiple models can be identified, ensuring that the selected important factors possess greater applicability and robustness under varied prediction conditions. (3) Evaluate and build prediction models: the refined feature variables were used as input labels for the four machine learning algorithms—SVM, RF, XGBoost, and KNN. Differentiation, calibration, and clinical utility are key evaluation criteria for assessing predictive models. Each criterion highlights a distinct aspect of model performance and provides a comprehensive assessment of the model's quality and practical value. Discrimination measures the ability of a model to distinguish between positive and negative samples (e.g., diseased versus undiseased). A high discriminative power indicates that the model can effectively differentiate between distinct categories of individuals. The area under the curve (AUC) value was derived from plotting receiver operating characteristic (ROC) curves to evaluate the model's performance across various thresholds. The closer the AUC is to 1, the better the model's discriminatory ability. Calibration assesses the agreement between the predicted probabilities and the actual outcomes, reflecting the model's "reliability." A well-calibrated model precisely predicts the actual occurrence rate for a given probability, ensuring that the model's output aligns closely with real-world observations. We plotted calibration curves, grouping the probabilities predicted by the model and comparing them to the actual rates of occurrence. Ideally, the calibration curve should follow a 45 diagonal, representing perfect calibration. Deviations from this diagonal indicate discrepancies between the predicted probabilities and actual outcomes, which may manifest as overestimations or underestimations. Clinical utility assesses the real-world value of a model in clinical decision-making, specifically the benefit it brings to both patients and healthcare providers. It focuses on how a predictive model influences patient health outcomes across various thresholds. To analyze clinical utility, we used decision curve analysis (DCA), which evaluates the net benefit at different prediction thresholds. Net Benefit (NB) is calculated as the benefit derived from positive model predictions minus the cost of misclassification at a given threshold. DCA helps determine whether the model offers substantial clinical value at specific thresholds. Internal validation was conducted using *k*-fold cross-validation. In this method, the dataset is randomly divided into *k* subsets (or folds) of approximately equal size. Typically, *k* is set to 10 (i.e., 10-fold cross-validation), although this value can be adjusted based on dataset size and specific needs. Ten-fold cross-validation is commonly used as it strikes a balance between bias and variance while optimizing computational efficiency and model stability. In each iteration, one subset serves as the test set, while the remaining *k*-1 subsets are used for training. This process allows the model to be constructed and evaluated *k* times, with each subset serving as the test set once. Evaluation metrics such as AUC, accuracy, sensitivity, and specificity are recorded in each iteration. The final model performance is the average result of these *k* iterations, providing a comprehensive assessment of model stability across different data divisions. (4) External validation of the best model: the generalizability and predictive efficiency of the optimal model were assessed by applying it to an external validation set from an independent cohort. The model's performance was again evaluated using ROC curves to confirm its robustness and ability to accurately predict postoperative gastroparesis in patients outside the original training dataset. This step ensures that the model's performance is not limited to the internal dataset and that it can be effectively used in real-world clinical settings. (5) Model interpretation: to interpret the model's predictions and gain insight into the role of different features, Shapley Additive Explanations (SHAP) were employed. SHAP values provide a clear explanation of how each feature contributes to the model's predictions. The SHAP summary plot visualizes the importance of each feature, ranking them based on their impact on the model's decision. For individualized patient predictions, the SHAP force plot was used, which demonstrates the influence of each feature on the predicted risk of gastroparesis. The SHAP force diagram calculates and displays the contribution of each feature to the predicted value, showing which variables increase or decrease the likelihood of gastroparesis for an individual patient. This allows clinicians to pinpoint the key risk factors specific to each patient and make more informed, personalized clinical decisions.

Results

Basic clinical information of the patient

The study encompassed a total of 1,097 colon cancer patients, of which 87 patients (7.93%) were diagnosed with gastroparesis (Figure 1 and Table 1). The internal validation set consisted of 787 colon cancer patients, with 61 patients (7.75%) diagnosed with gastroparesis. The external validation set included 310 colon cancer patients, of whom 26 patients (8.39%) had gastroparesis. The comprehensive original dataset underpinning this study is provided in Supplementary Table 1. The code utilized in this research has been uploaded to NutCloud, accessible via the following link: https://www.jianguoyun.com/p/DWh9chMQl-GKDBjEj-sFIAA.

Screening for risk factors for postoperative gastroparesis

The results of univariate and multivariate analyses identified several independent factors influencing the occurrence of postoperative gastroparesis, including age, albumin (ALB) levels, history of anemia, history of diabetes mellitus, history of hypothyroidism, history of adjuvant radiotherapy, type of

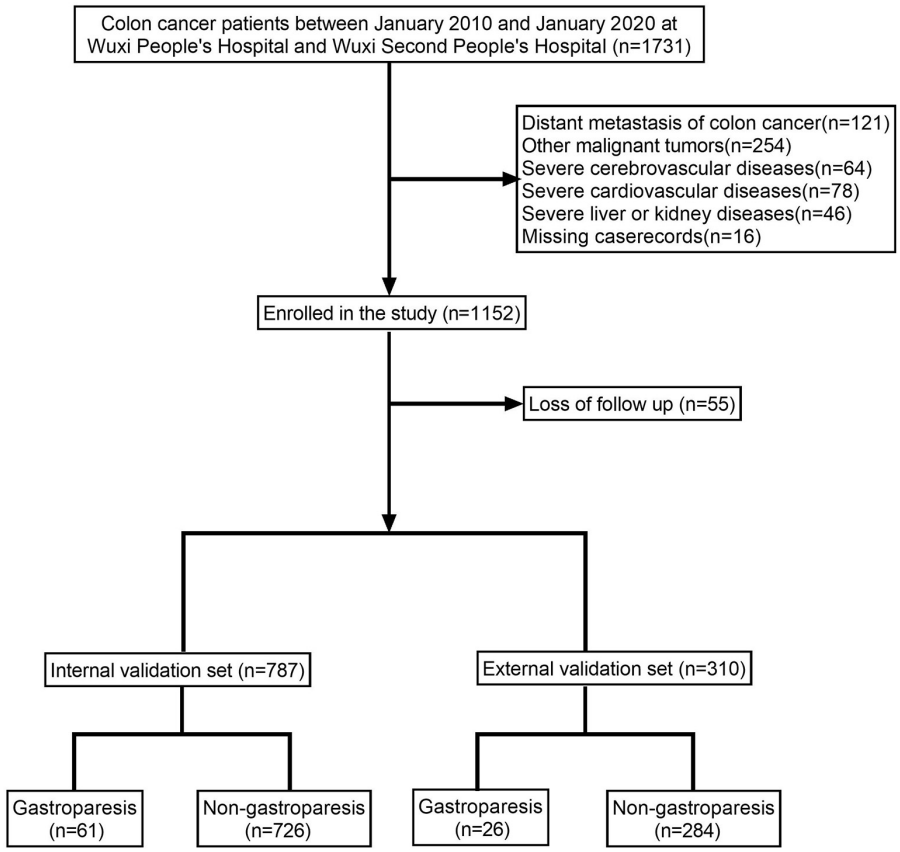

FIGURE 1
Flow diagram of patients included in the study.

surgery, duration of surgery, intraoperative bleeding, tumor size, and number of tumors ($p < 0.05$) (Table 2). The XGBoost, RF, SVM, and KNN models further identified key risk factors for postoperative gastroparesis, which included advanced age, hypoproteinemia, history of anemia, history of diabetes mellitus, history of hypothyroidism, open surgery, long operative time, and high intraoperative bleeding (Figures 2A–D). Based on a comprehensive analysis of these factors, the prediction model incorporated the following variables: age ≥65, hypoproteinemia, history of anemia, history of diabetes mellitus, history of hypothyroidism, open surgery, operative time ≥270 min, and intraoperative bleeding ≥100 ml.

Model building and evaluation

The ROC curve analysis demonstrated that the XGBoost model achieved the highest performance among the four models, with an AUC value of 0.939 in the training set and 0.876 in the validation set, outperforming the other three models (Table 3). The calibration curves of all models closely followed the ideal 45 diagonal, indicating a strong alignment between the predicted probabilities and the actual outcomes. Additionally, the DCA curves revealed that all four models provided a net clinical benefit when compared to both full treatment and no treatment scenarios (Figures 3A–D). The study evaluated the generalization ability of the four models using *k*-fold cross-validation. A total of 118 cases (15.00%) from the internal validation set were selected as the validation set, while the remaining samples were used as the training set. The models were subjected to 10-fold cross-validation. For the XGBoost algorithm, the AUC value in the validation set was 0.8735 ± 0.0764, and the AUC in the test set was 0.9247, with an overall accuracy of 0.8908 (Figures 4A–C). This underscores the exceptional discriminative power and robust generalizability of the XGBoost model, rendering it the most suitable choice for the present study. In contrast, the RF algorithm demonstrated an AUC value of 0.8321 ± 0.0415 in the validation set, with a corresponding AUC of 0.8566 in the test set, yielding an accuracy of 0.8113. The SVM algorithm exhibited an AUC of 0.8061 ± 0.0647 in the validation set and 0.7324 in the test set, with an accuracy of 0.8143. The KNN algorithm, on the other hand, recorded an AUC of 0.7852 ± 0.0654 in the validation set, and 0.7054 in the test set, achieving an accuracy of 0.8864. Following a thorough comparative analysis, the XGBoost algorithm was selected as the foundation for the predictive model in this investigation.

Model external validation

The disease prediction model demonstrated exceptional accuracy, as reflected by an AUC value of 0.788 in the external validation set (Figure 4D).

TABLE 1 Preoperation and intraoperative information.

Variables		All (n = 1,146)	Non-gastroparesis (n = 1,051)	Gastroparesis (n = 95)	p-Value
Sex	Female	362 (45.997)	331 (45.592)	31 (50.820)	0.431
	Male	425 (54.003)	395 (54.408)	30 (49.180)	
Age	65	604 (76.747)	576 (79.339)	28 (45.902)	<0.001
	≥65	183 (23.253)	150 (20.661)	33 (54.098)	
BMI	25 kg/m^2	521 (66.201)	488 (67.218)	33 (54.098)	0.037
	≥25 kg/m^2	266 (33.799)	238 (32.782)	28 (45.902)	
ASA	3	512 (65.057)	477 (65.702)	35 (57.377)	0.19
	≥3	275 (34.943)	249 (34.298)	26 (42.623)	
Drinking history	No	538 (68.361)	495 (68.182)	43 (70.492)	0.709
	Yes	249 (31.639)	231 (31.818)	18 (29.508)	
Smoking history	No	541 (68.742)	507 (69.835)	34 (55.738)	0.023
	Yes	246 (31.258)	219 (30.165)	27 (44.262)	
ALB	≥30 g/L	568 (72.173)	543 (74.793)	25 (40.984)	<0.001
	30 g/L	219 (27.827)	183 (25.207)	36 (59.016)	
NRS2002 score	3	543 (68.996)	499 (68.733)	44 (72.131)	0.582
	≥3	244 (31.004)	227 (31.267)	17 (27.869)	
Surgical history	No	483 (61.372)	459 (63.223)	24 (39.344)	<0.001
	Yes	304 (38.628)	267 (36.777)	37 (60.656)	
Anemia	No	581 (73.825)	553 (76.171)	28 (45.902)	<0.001
	Yes	206 (26.175)	173 (23.829)	33 (54.098)	
Hyperlipidemia	No	641 (81.449)	597 (82.231)	44 (72.131)	0.051
	Yes	146 (18.551)	129 (17.769)	17 (27.869)	
Hypertension	No	413 (52.478)	389 (53.581)	24 (39.344)	0.032
	Yes	374 (47.522)	337 (46.419)	37 (60.656)	
Diabetes	No	638 (81.067)	613 (84.435)	25 (40.984)	<0.001
	Yes	149 (18.933)	113 (15.565)	36 (59.016)	
Hypothyroidism	No	585 (74.333)	558 (76.860)	27 (44.262)	<0.001
	Yes	202 (25.667)	168 (23.140)	34 (55.738)	
COPD	No	651 (82.719)	607 (83.609)	44 (72.131)	0.023
	Yes	136 (17.281)	119 (16.391)	17 (27.869)	
CHD	No	681 (86.531)	629 (86.639)	52 (85.246)	0.76
	Yes	106 (13.469)	97 (13.361)	9 (14.754)	
Adjuvant radiotherapy	No	597 (75.858)	559 (76.997)	38 (62.295)	0.01
	Yes	190 (24.142)	167 (23.003)	23 (37.705)	
Adjuvant chemotherapy	No	570 (72.427)	531 (73.140)	39 (63.934)	0.122
	Yes	217 (27.573)	195 (26.860)	22 (36.066)	
Surgical procedure	Laparoscopic surgery	630 (80.051)	591 (81.405)	39 (63.934)	0.001
	Open surgery	157 (19.949)	135 (18.595)	22 (36.066)	
Emergency surgery	No	576 (73.189)	535 (73.691)	41 (67.213)	0.273
	Yes	211 (26.811)	191 (26.309)	20 (32.787)	
Surgery time	270 min	498 (63.278)	470 (64.738)	28 (45.902)	0.003
	≥270 min	289 (36.722)	256 (35.262)	33 (54.098)	

(Continued)

TABLE 1 (Continued)

Variables		All (n = 1,146)	Non-gastroparesis (n = 1,051)	Gastroparesis (n = 95)	p-Value
Intraoperative bleeding	100 ml	527 (66.963)	504 (69.421)	23 (37.705)	<0.001
	≥100 ml	260 (33.037)	222 (30.579)	38 (62.295)	
Blood transfusion	No	637 (80.940)	589 (81.129)	48 (78.689)	0.641
	Yes	150 (19.060)	137 (18.871)	13 (21.311)	
SPO_2	≥90%	633 (80.432)	584 (80.441)	49 (80.328)	0.983
	90%	154 (19.568)	142 (19.559)	12 (19.672)	
T-stage	T1∼T2	559 (71.029)	530 (73.003)	29 (47.541)	<0.001
	T3∼T4	228 (28.971)	196 (26.997)	32 (52.459)	
N-stage	N0	562 (71.410)	528 (72.727)	34 (55.738)	0.005
	N1∼N2	225 (28.590)	198 (27.273)	27 (44.262)	
PNI	No	705 (89.581)	654 (90.083)	51 (83.607)	0.112
	Yes	82 (10.419)	72 (9.917)	10 (16.393)	
Tumor number	2	595 (75.604)	570 (78.512)	25 (40.984)	<0.001
	≥2	192 (24.396)	156 (21.488)	36 (59.016)	
Tumor size	5 cm	536 (68.107)	510 (70.248)	26 (42.623)	<0.001
	≥5 cm	251 (31.893)	216 (29.752)	35 (57.377)	
CEA level	5 ng/ml	575 (73.062)	524 (72.176)	51 (83.607)	0.053
	≥5 ng/ml	212 (26.938)	202 (27.824)	10 (16.393)	
CA199 level	37 U/mL	583 (74.079)	539 (74.242)	44 (72.131)	0.718
	≥37 U/mL	204 (25.921)	187 (25.758)	17 (27.869)	
PCT level	0.05 ng/ml	571 (72.554)	526 (72.452)	45 (73.770)	0.825
	≥0.05 ng/ml	216 (27.446)	200 (27.548)	16 (26.230)	
CRP level	10 mg/L	530 (67.344)	497 (68.457)	33 (54.098)	0.022
	≥10 mg/L	257 (32.656)	229 (31.543)	28 (45.902)	
SAA level	10 mg/L	557 (70.775)	521 (71.763)	36 (59.016)	0.036
	≥10 mg/L	230 (29.225)	205 (28.237)	25 (40.984)	

OR, odds ratio; CI, confidence interval; BMI, body mass index; ASA, The American Society of Anesthesiologists; ALB, albumin; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 19-9; PCT, procalcitonin; CRP, C-reactive protein; SAA, serum amyloid A; NRS2002, nutrition risk screening 2002; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; SPO_2, percutaneous arterial oxygen saturation.

Model explanation

The SHAP summary plot revealed that multiple risk factors contribute to the development of gastroparesis following CME, with intraoperative bleeding exceeding 100 ml, a history of anemia, diabetes mellitus, hypoproteinemia, age ≥65, open surgery, operative duration ≥270 min, and a history of hypothyroidism emerging as the most influential determinants (Figure 5).

The SHAP force diagram illustrates the predictive analysis of the study model for four colon cancer patients with gastroparesis. For patient 1, the model predicted a gastroparesis probability of 0.15, with contributing factors including a history of anemia, open surgery, and an operative duration ≥270 min. For patient 2, the predicted probability was 0.94, influenced by a history of anemia, intraoperative bleeding ≥100 ml, a history of hypothyroidism, hypoproteinemia, a history of diabetes mellitus, and age ≥65. In patient 3, the predicted probability of gastroparesis was 0.52, with risk factors including a history of diabetes mellitus, intraoperative bleeding ≥100 ml, hypoproteinemia, and age ≥65. Lastly, patient 4 had a predicted probability of 0.05, with contributing factors such as hypoproteinemia, operative time ≥270 min, and a history of diabetes mellitus, while age ≥65 decreased the probability (Figures 6A–D).

Discussion

In the present study, SHAP analysis was employed to visualize and interpret the model, revealing that advanced age, prolonged surgical duration, excessive intraoperative bleeding, surgical approach, hypoproteinemia, anemia, and a history of diabetes mellitus and hypothyroidism are significant risk factors for gastroparesis following CME. Traditionally, imaging tests such as CT, MRI, and gastrointestinal fluoroscopy are utilized to diagnose postoperative gastroparesis. However, these diagnostic tools are not only costly but may also subject patients to discomfort

TABLE 2 Univariate and multivariate analysis of variables related to gastroparesis.

Variables		Univariate analysis		Multivariate analysis	
		OR, 95% CI	*p*-Value	OR, 95% CI	*p*-Value
Sex	Female	Reference			
	Male	0.81 [0.48, 1.37]	0.432		
Age	65	Reference		Reference	
	≥65	4.53 [2.65, 7.72]	<0.001	3.76 [1.73, 8.21]	<0.001
BMI	25 kg/m^2	Reference		Reference	
	≥25 kg/m^2	1.74 [1.03, 2.95]	0.039	1.34 [0.64, 2.83]	0.441
ASA	3	Reference			
	≥3	1.42 [0.84, 2.42]	0.192		
Drinking history	No	Reference			
	Yes	0.90 [0.51, 1.59]	0.71		
Smoking history	No	Reference		Reference	
	Yes	1.84 [1.08, 3.12]	0.024	1.10 [0.51, 2.38]	0.809
ALB	≥30 g/L	Reference		Reference	
	30 g/L	4.27 [2.50, 7.31]	<0.001	2.81 [1.30, 6.06]	0.009
NRS2002 score	3	Reference			
	≥3	0.85 [0.47, 1.52]	0.582		
Surgical history	No	Reference		Reference	
	Yes	2.65 [1.55, 4.53]	<0.001	1.47 [0.69, 3.14]	0.32
Anemia	No	Reference		Reference	
	Yes	3.77 [2.21, 6.41]	<0.001	3.67 [1.71, 7.89]	<0.001
Hyperlipidemia	No	Reference			
	Yes	1.79 [0.99, 3.23]	0.054		
Hypertension	No	Reference		Reference	
	Yes	1.78 [1.04, 3.04]	0.034	1.35 [0.63, 2.89]	0.442
Diabetes	No	Reference		Reference	
	Yes	7.81 [4.51, 13.52]	<0.001	5.12 [2.38, 11.02]	<0.001
Hypothyroidism	No	Reference		Reference	
	Yes	4.18 [2.45, 7.13]	<0.001	3.94 [1.85, 8.39]	<0.001
COPD	No	Reference		Reference	
	Yes	1.97 [1.09, 3.57]	0.025	2.19 [0.91, 5.27]	0.08
CHD	No	Reference			
	Yes	1.12 [0.54, 2.35]	0.76		
Adjuvant radiotherapy	No	Reference		Reference	
	Yes	2.03 [1.17, 3.50]	0.011	2.40 [1.06, 5.42]	0.036
Adjuvant chemotherapy	No	Reference			
	Yes	1.54 [0.89, 2.66]	0.125		
Surgical procedure	Laparoscopic surgery	Reference		Reference	
	Open surgery	2.47 [1.42, 4.30]	0.001	3.24 [1.45, 7.25]	0.004
Emergency surgery	No	Reference			
	Yes	1.37 [0.78, 2.39]	0.274		
Surgery time	270 min	Reference		Reference	
	≥270 min	2.16 [1.28, 3.66]	0.004	2.30 [1.08, 4.90]	0.032

(Continued)

TABLE 2 (Continued)

Variables		Univariate analysis		Multivariate analysis	
		OR, 95% CI	*p*-Value	OR, 95% CI	*p*-Value
Intraoperative bleeding	100 ml	Reference		Reference	
	≥100 ml	3.75 [2.18, 6.44]	<0.001	3.46 [1.61, 7.40]	0.001
Blood transfusion	No	Reference			
	Yes	1.16 [0.61, 2.21]	0.641		
SPO_2	≥90%	Reference			
	90%	1.01 [0.52, 1.94]	0.983		
T-stage	T1∼T2	Reference		Reference	
	T3∼T4	2.98 [1.76, 5.06]	<0.001	1.39 [0.65, 2.98]	0.393
N-stage	N0	Reference		Reference	
	N1∼N2	2.12 [1.25, 3.60]	0.006	1.89 [0.88, 4.03]	0.101
PNI	No	Reference			
	Yes	1.78 [0.87, 3.66]	0.116		
Tumor number	2	Reference		Reference	
	≥2	5.26 [3.07, 9.03]	<0.001	3.65 [1.76, 7.58]	<0.001
Tumor size	5 cm	Reference		Reference	
	≥5 cm	3.18 [1.87, 5.41]	<0.001	3.67 [1.74, 7.77]	<0.001
CEA level	5 ng/ml	Reference			
	≥5 ng/ml	0.51 [0.25, 1.02]	0.057		
CA199 level	37 U/ml	Reference			
	≥37 U/ml	1.11 [0.62, 2.00]	0.718		
PCT level	0.05 ng/ml	Reference			
	≥0.05 ng/ml	0.94 [0.52, 1.69]	0.825		
CRP level	10 mg/L	Reference		Reference	
	≥10 mg/L	1.84 [1.09, 3.12]	0.023	1.43 [0.66, 3.09]	0.365
SAA level	10 mg/L	Reference		Reference	
	≥10 mg/L	1.76 [1.03, 3.01]	0.037	1.00 [0.45, 2.19]	0.995

OR, odds ratio; CI, confidence interval; BMI, body mass index; ASA, The American Society of Anesthesiologists; ALB, albumin; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 19-9; PCT, procalcitonin; CRP, C-reactive protein; SAA, serum amyloid A; NRS2002, nutrition risk screening 2002; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; SPO_2, percutaneous arterial oxygen saturation.

and additional medical risks. By leveraging the predictive model developed in this study, clinicians can assess the risk of gastroparesis in advance, based on clinical data and patient-specific characteristics, thus minimizing the need for unnecessary imaging procedures. This approach enhances diagnostic and treatment efficiency while reducing the number of tests required during the diagnostic process. The machine learning model constructed here offers a precise method for identifying patients at high risk of developing gastroparesis after surgery, enabling early detection and the provision of personalized care, ultimately improving the effectiveness of clinical intervention.

The present study sought to assess the performance of four machine learning algorithms in developing risk prediction models. The XGBoost algorithm exhibited remarkable accuracy, distinguished by its efficiency, flexibility, and adaptability, making it an optimal choice for this analysis (15). In contrast to the RF algorithm, the XGBoost algorithm adopts a gradient boosting integration approach that emphasizes difficult-to-classify samples, thereby enhancing generalization performance and ensuring greater stability of the model (16). The SVM and KNN algorithms also exhibited high accuracy and effectively mitigated overfitting issues. However, in the context of managing multiple features and large datasets, the XGBoost algorithm utilizes both L1 and L2 regularization techniques to address overfitting more effectively. Moreover, XGBoost is capable of automatically handling missing values and offers valuable insights into the contribution of each feature to the prediction, enhancing its interpretability. This characteristic renders it particularly advantageous for complex, multidimensional studies. As a result, following a thorough comparison of the four machine learning algorithms, the XGBoost algorithm was selected to develop the predictive model for the occurrence of gastroparesis after CME.

Extreme Gradient Boosting is an ensemble method based on decision trees, specifically utilizing Gradient Boosting Trees. While it excels in numerous tasks, its inherent model complexity can still present the risk of overfitting. One key parameter

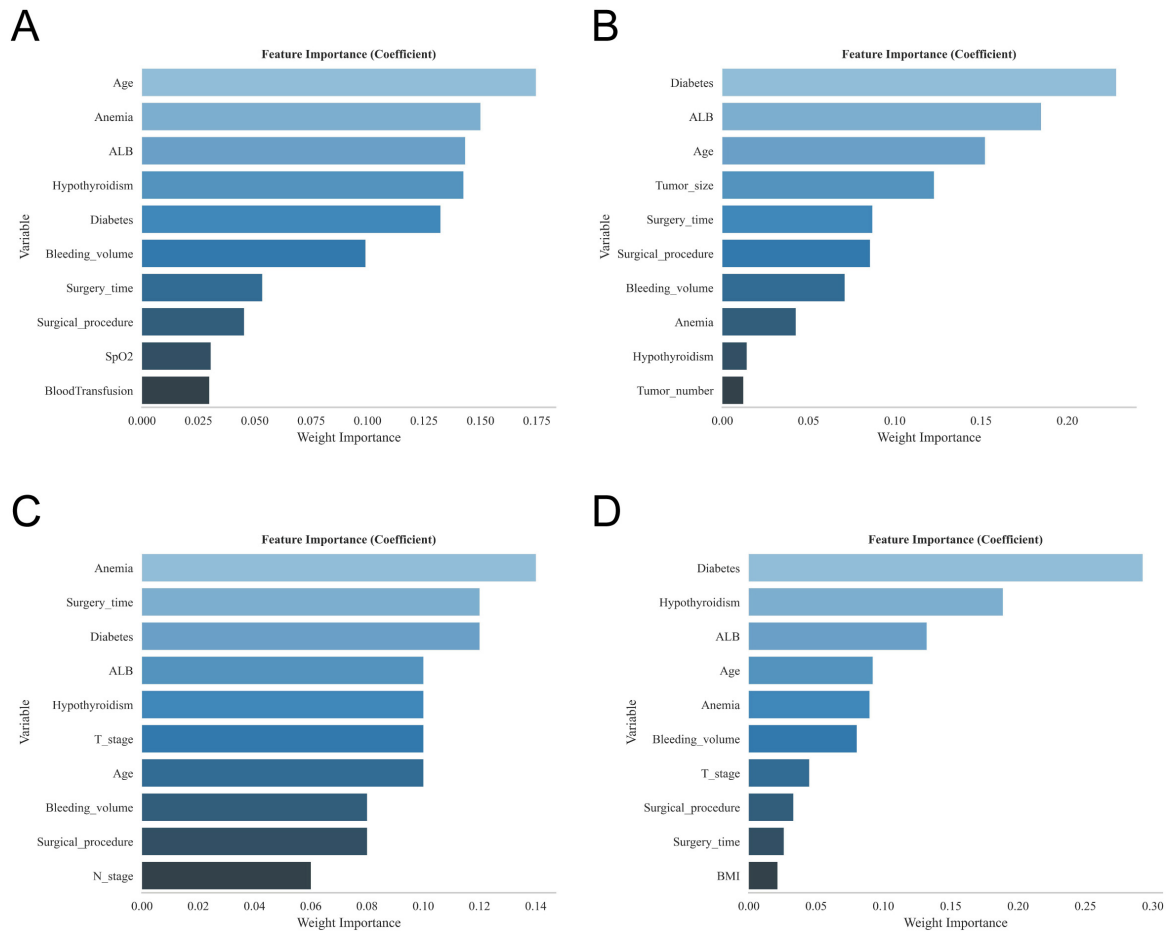

FIGURE 2
The variable ranking plots of the four models. **(A)** Variable importance ranking diagram of the XGBoost model. **(B)** Variable importance ranking diagram of the RF model. **(C)** Variable importance ranking diagram of the SVM model. **(D)** Variable importance ranking diagram of the KNN model.

TABLE 3 Evaluation of the performance of the four models in the internal validation set.

		AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 score (95% CI)
KNN	Training set	0.936 (0.897–0.975)	0.949 (0.949–0.949)	0.417 (0.383–0.451)	0.991 (0.989–0.992)	0.542 (0.517–0.568)
	Validation set	0.735 (0.604–0.866)	0.899 (0.886–0.911)	0.129 (0.121–0.137)	0.982 (0.962–1.003)	0.201 (0.181–0.220)
XGBoost	Training set	0.939 (0.901–0.978)	0.892 (0.876–0.907)	0.868 (0.865–0.871)	0.894 (0.877–0.910)	0.538 (0.496–0.580)
	Validation set	0.876 (0.796–0.957)	0.829 (0.817–0.842)	0.706 (0.498–0.915)	0.842 (0.807–0.878)	0.445 (0.374–0.516)
RF	Training set	0.887 (0.834–0.941)	0.862 (0.839–0.886)	0.78 (0.732–0.828)	0.869 (0.840–0.898)	0.452 (0.430–0.473)
	Validation set	0.852 (0.758–0.946)	0.813 (0.770–0.857)	0.74 (0.597–0.883)	0.821 (0.788–0.854)	0.439 (0.318–0.559)
SVM	Training set	0.932 (0.891–0.973)	0.924 (0.923–0.926)	0.791 (0.765–0.817)	0.935 (0.935–0.935)	0.602 (0.584–0.620)
	Validation set	0.858 (0.770–0.946)	0.864 (0.845–0.883)	0.677 (0.657–0.698)	0.884 (0.863–0.906)	0.494 (0.484–0.505)

CI, confidence interval; KNN, *k*-nearest neighbor; XGBoost, extreme gradient boosting; RF, random forest; SVM, support vector machine; AUC, area under the curve.

influencing this complexity is max_depth, which controls the depth of the decision tree. If max_depth is set too high, the model becomes capable of capturing intricate data patterns, including noise and outliers present in the training set. This can lead to overfitting, adversely impacting the model's performance on unseen data. Additionally, XGBoost operates through integrated learning, constructing multiple trees. If the number of trees is excessive, the individual fitting capacity of each tree increases, which may result in overfitting of the training data by the ensemble model. Furthermore, the learning rate plays a crucial role in determining the contribution of each tree to the overall model. A smaller learning rate causes the model to learn at a slower pace with each iteration, necessitating a larger number of trees to gradually fit the data. While this can enhance model stability, an overly large number of trees can cause the model to overfit the training set. To mitigate the risk of overfitting, we controlled

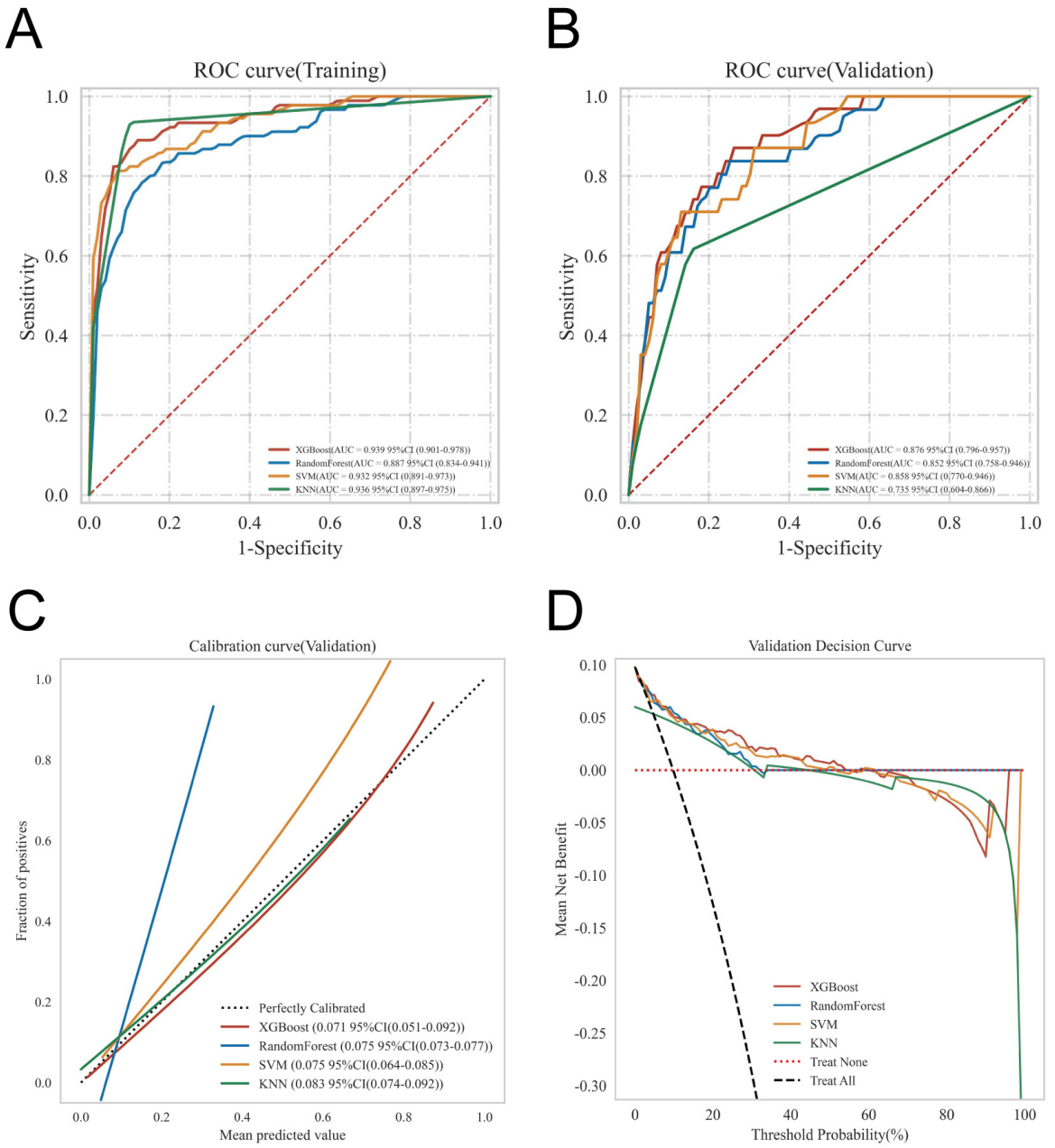

FIGURE 3
Evaluation of the four models for predicting gastroparesis. **(A)** ROC curves for the training set of the four models. **(B)** ROC curves for the validation set of the four models. **(C)** Calibration plots of the four models. The 45° dotted line on each graph represents the perfect match between the observed (*y*-axis) and predicted (*x*-axis) complication probabilities. A closer distance between two curves indicates greater accuracy. **(D)** DCA curves of the four models. The intersection of the red curve and the All curve is the starting point, and the intersection of the red curve and the None curve is the node within which the corresponding patients can benefit.

max_depth to limit the tree's depth, thereby preventing the model from fitting noise or irrelevant data. Additionally, reducing the number of trees (n_estimators) helps avoid overfitting. XGBoost also offers regularization parameters, such as gamma, lambda, and alpha, which can be adjusted to manage model complexity and further prevent overfitting. Finally, we adjusted the learning_rate and n_estimators to balance stability and generalization. Lower learning rates typically require more trees to fit the data, but they contribute to improved generalization, avoiding excessive overfitting. Through these practices, we effectively enhanced the stability and generalization ability of the model, enabling it to perform optimally in practical applications.

Studies (17, 18) have highlighted the efficacy of machine learning algorithms in clinical diagnosis and prognosis, revealing their superior ability to predict adverse outcomes in disease progression compared to traditional diagnostic methods. As individuals age, the digestive system undergoes a range of changes, including a decline in the secretion of digestive enzymes and a reduction in the peristaltic capacity of the smooth muscles within the gastrointestinal tract (19). Moreover, elderly patients often

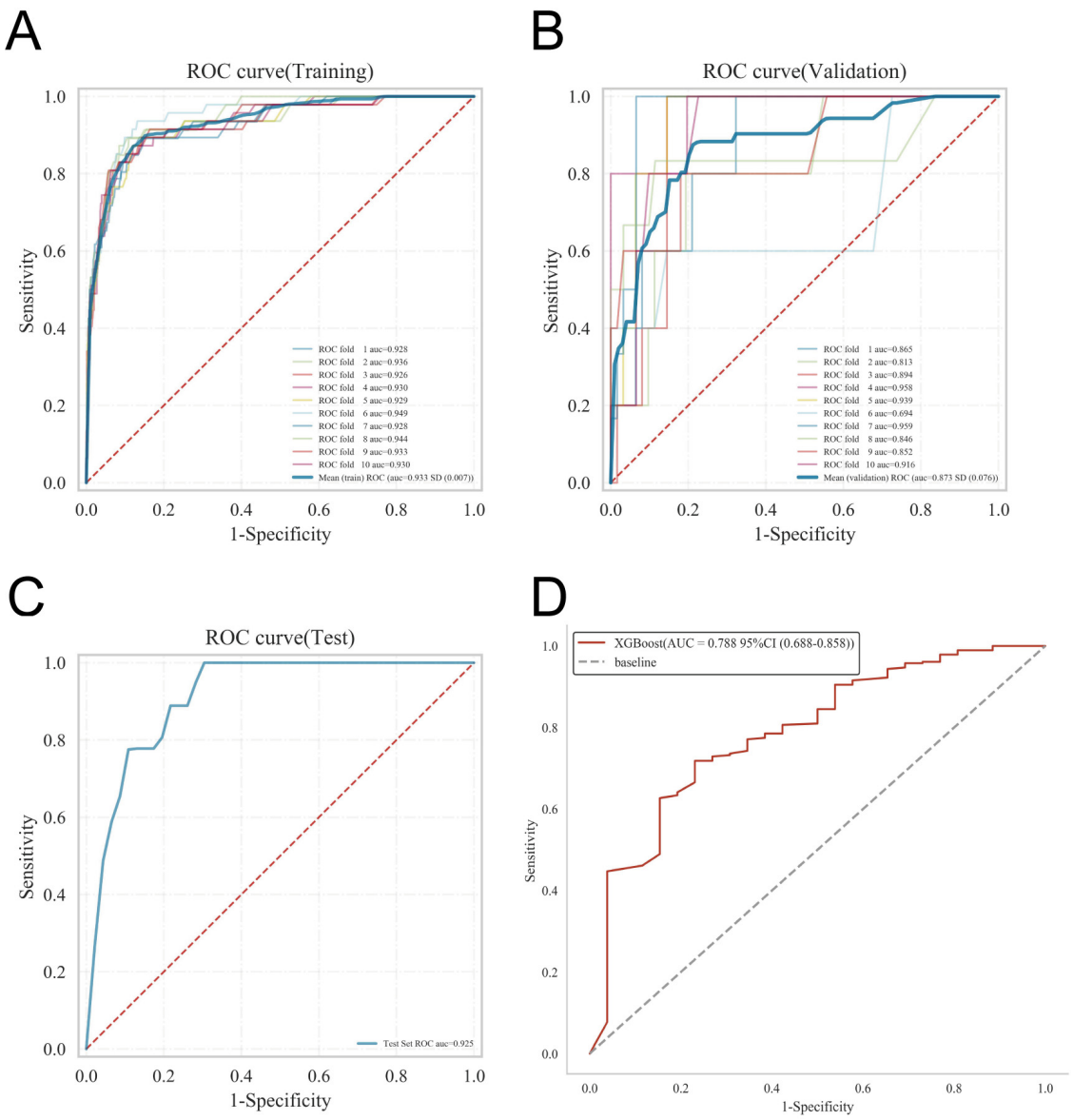

FIGURE 4
Internal validation of the XGBoost model. **(A)** ROC curve of the XGBoost model for the training set. **(B)** ROC curve of the XGBoost model for the validation set. **(C)** ROC curve of the XGBoost model for the test set. **(D)** External validation of the XGBoost model.

suffer from chronic cardiovascular and respiratory conditions that can affect the nerves and muscles controlling the digestive system, leading to delayed gastric emptying. With a diminished capacity for compensatory function, older individuals exhibit a reduced ability to withstand surgical stress, further exacerbating disruptions in gastrointestinal motility. As a result, postoperative gastroparesis in the elderly arises from a convergence of multiple factors. A study by Meng et al. (20), which included 563 oncology patients, revealed a strong correlation between advanced age and the development of postoperative gastroparesis following abdominal surgery, thereby reinforcing the conclusions of the present study.

The present study corroborates prior research by identifying surgery as a significant determinant of postoperative complications (21, 22). The surgical procedure itself can cause damage to the stomach's nerves and muscles, impairing both contraction and emptying functions. Traditional open radical colon cancer surgery, which involves an incision in the lower left abdomen, is associated with extended operation times, extensive resections, considerable postoperative pain, and slower patient recovery. In contrast, laparoscopic surgery offers superior precision and convenience, allowing the surgeon to intuitively evaluate lesion size and surrounding tissues, while performing intricate procedures such as tissue dissection and intestinal anastomosis. This approach minimizes damage to healthy gastrointestinal tissues, promoting faster recovery of postoperative gastrointestinal motility. Numerous studies have shown that patients undergoing laparoscopic surgery experience more rapid gastrointestinal recovery and a lower incidence of postoperative complications, including those related to gastric and intestinal function (23, 24). Consequently, we assert that the type of surgery is a pivotal

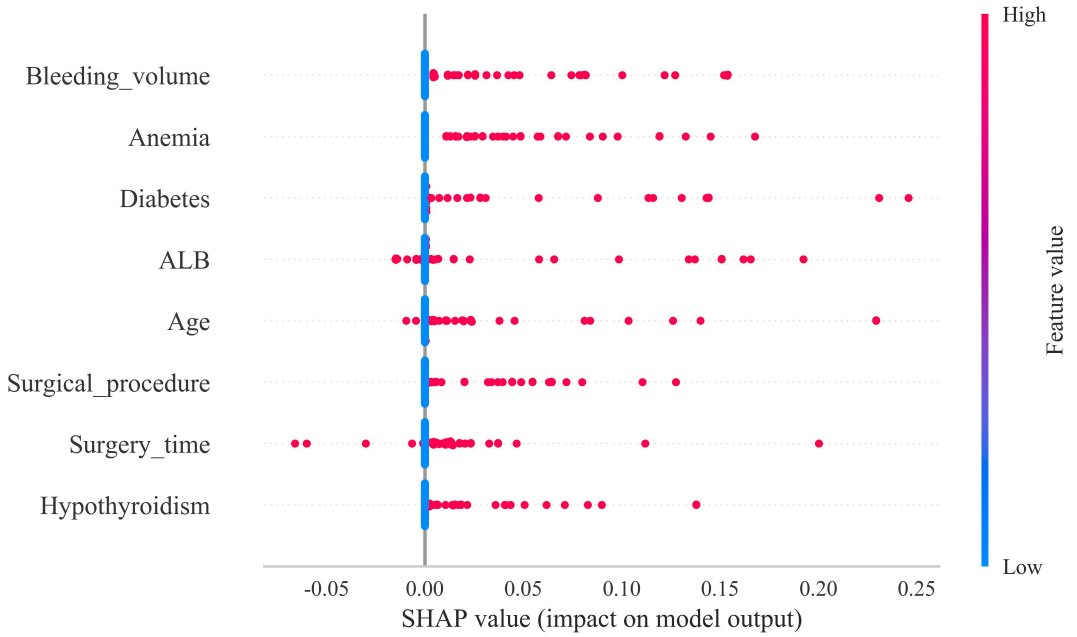

FIGURE 5
Shapley Additive Explanations (SHAP) summary plot. Risk factors are arranged along the *y*-axis based on their importance, which is given by the mean of their absolute Shapley values. The higher the risk factor is positioned in the plot, the more important it is for the model.

factor in the onset of postoperative gastroparesis. This study further identified prolonged operative duration and increased intraoperative bleeding as significant contributors to the development of gastroparesis following surgery. These factors are often associated with the complexity of the surgical procedure and the heightened risk of inadvertent injury to gastric omental vessels and lymph nodes, which can trigger an exacerbated inflammatory response and disrupt the normal motility of the gastrointestinal tract (25). Furthermore, excessive intraoperative bleeding can compromise hemodynamics within the gastric mucosa and other intestinal segments, thereby delaying the recovery of gastrointestinal tissues. Prolonged surgical durations often require increased administration of anesthetic agents, which inhibit sympathetic constrictor nerve fibers, causing vascular smooth muscle relaxation and a reduction in blood pressure. This drop in blood pressure can precipitate severe complications, including gastroparesis. Additionally, anesthetic drugs may interfere with vagal nerve function, leading to persistent spasm of the pyloric sphincter and impairing the efficiency of gastric emptying (26). Therefore, it is imperative for the surgical team to meticulously plan the procedure in advance and collaborate seamlessly during the operation to enhance surgical efficiency, reduce operative time, and minimize the risk of postoperative gastroparesis.

Furthermore, this study aimed to illuminate the underlying rationale behind the postoperative gastroparesis prediction model using four distinct samples. In the disease prediction analysis of the second sample, the patient's nutritional status emerged as a pivotal predictor. Plasma albumin, despite its modest molecular weight, plays an indispensable role in maintaining both plasma and tissue fluid osmolality. Hypoalbuminemia results in reduced plasma colloid osmotic pressure, heightening the risk of intestinal wall edema and disrupting gastrointestinal motility. Moreover, patients with hypoalbuminemia exhibit diminished numbers and activity of antibody synthase, thereby increasing their vulnerability to complications such as postoperative infections and anastomotic leakage, which further hinder gastrointestinal recovery. These findings emphasize the critical need for clinicians to carefully assess preoperative albumin levels and promptly manage postoperative complications. Enhancing parenteral nutrition for patients with hypoproteinemia, coupled with the supplementation of a high-protein diet when feasible, may help alleviate the risk of hypoalbuminemia and reduce the likelihood of postoperative gastroparesis (27). Similarly, patients with anemia are more susceptible to adverse outcomes, including postoperative gastroparesis. Right hemicolectomy patients frequently experience tumor infiltration that disrupts mucosal and submucosal vessels in the colon lining, contributing to malnutrition-related conditions such as iron deficiency anemia. Hemoglobin, vital for oxygen transport in the bloodstream, is indispensable for maintaining adequate oxygen supply to bodily tissues. Anemia leads to tissue hypoxia, a result of insufficient hemoglobin, which severely impairs the normal functioning of gastrointestinal smooth muscles. Additionally, the small perigastric vessels play a crucial role in nourishing the perigastric vagus nerve. Al-Saffar et al. (28) demonstrated that preoperative hemoglobin levels serve as an independent risk factor for delayed gastric emptying. The incidence of gastroparesis is markedly higher in anemic patients compared to their non-anemic counterparts, underscoring the imperative for vigilant monitoring of gastrointestinal function in anemic individuals and the prompt initiation of preventive strategies to mitigate gastrointestinal complications.

The present investigation further reveals that the underlying condition significantly influences the onset of postoperative gastroparesis. Diabetic patients often manifest neuropathic

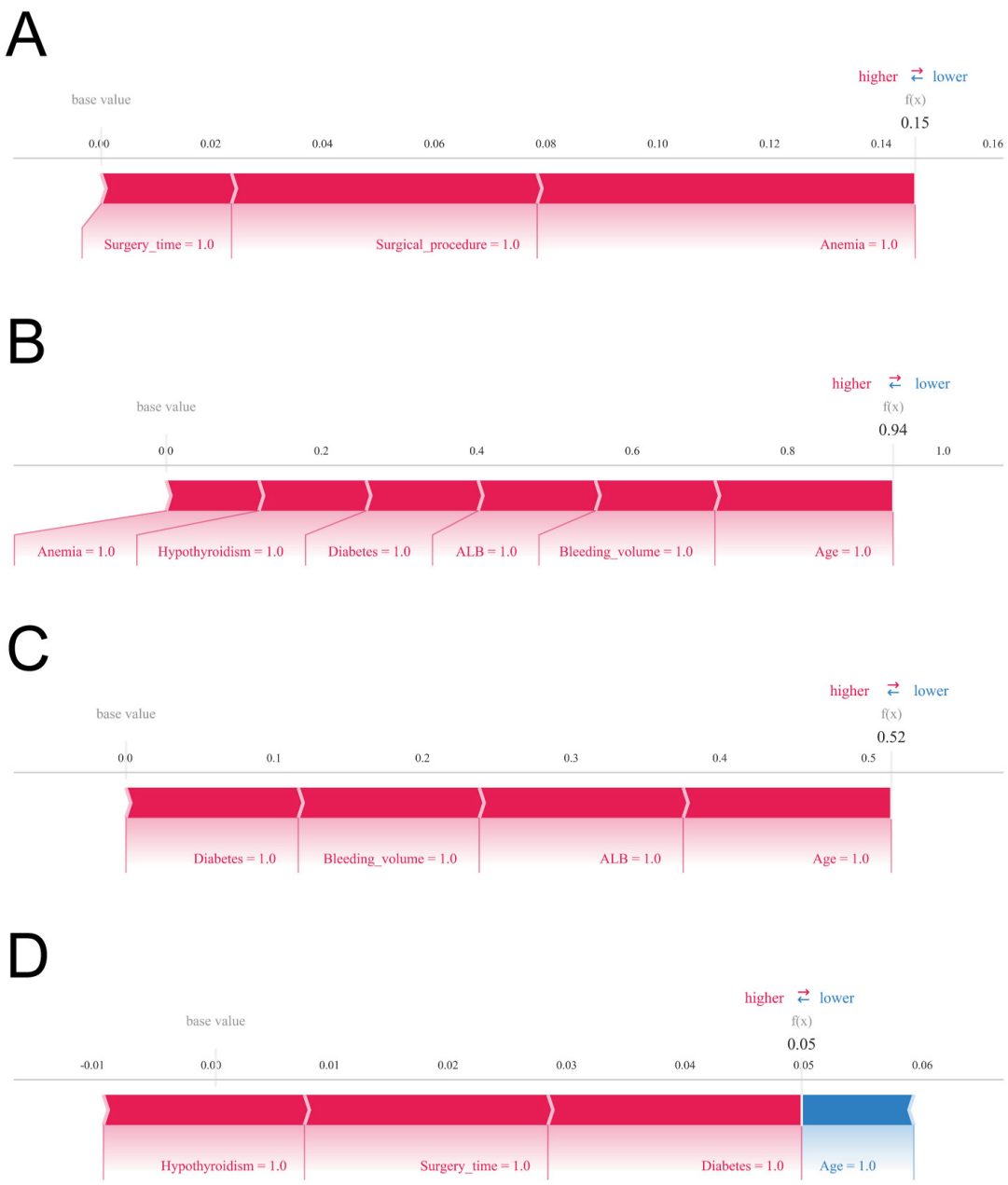

FIGURE 6
Shapley Additive Explanations (SHAP) force plot. The contributing variables are arranged in the horizontal line, sorted by the absolute value of their impact. Blue represents features that have a negative effect on disease prediction, with a decrease in SHAP values; red represents features that have a positive effect on disease prediction, with an increase in SHAP values. **(A)** Predictive analysis of patient 1. **(B)** Predictive analysis of patient 2. **(C)** Predictive analysis of patient 3. **(D)** Predictive analysis of patient 4.

alterations that impair both their autonomic and visceral nerves, thereby hindering the motility of the smooth muscles within the gastrointestinal tract (29). Farrugia (30) demonstrated that a reduction in Cajal mesenchymal cells, key regulators of gastrointestinal motility, constitutes a critical mechanism for delayed gastric emptying in an animal model, thereby providing further validation for the findings of the present study. Furthermore, we posit that fluctuations in blood glucose levels resulting from perioperative fasting serve as a significant trigger for the onset of gastroparesis in these patients. Postoperatively, many of these patients develop insulin resistance, which hampers the secretion and release of gastrin and disrupts the function of the autonomic nervous system, thereby impairing gastric emptying (31). The rate of gastric emptying was found to be closely correlated with diabetes in a comprehensive, multicenter study conducted over 30 years by clinical researchers worldwide (31, 32). Similarly, individuals with hypothyroidism exhibit comparable effects. Thyroxine plays a crucial role in sustaining the body's physiological functions, particularly in motor function. However, clinicians often focus primarily on limb movements in hypothyroid patients, neglecting the activity of glandular organs. Pathologists who conducted biopsies of gastric tissues from

hypothyroid patients observed edema and thickening of gastric mucosal cells, accompanied by surrounding inflammatory cell infiltration (33). Building on this, Ghoshal et al. (34) conducted a clinical study involving 60 hypothyroid patients and found a strong correlation between reduced thyroxine levels and impaired gastrointestinal motility. It is posited that enhanced vagal function in hypothyroid patients results in hyperexcitability, disrupting the normal coordination and rhythmicity of the gastrointestinal smooth muscles. Experimental findings by Khraisha et al. (35) in patients with chronic atrophic gastritis and gastroparesis further corroborate the therapeutic potential of thyroxine in addressing gastric motility disorders.

The present study thoroughly assessed the model's performance across differentiation, calibration, and clinical utility. However, several limitations warrant consideration. The predictive models were primarily based on clinical indicators and laboratory data, without incorporating imaging data such as CT, MRI, or ultrasound scans. While we acknowledge the significant role of imaging data in enhancing prediction accuracy and precision, the lack of imaging data in our analysis was due to constraints in available data sources. Future studies that integrate both imaging and clinical data could potentially further refine the model's predictive capability. Additionally, while machine learning algorithms offer superior accuracy, they tend to be more complex and less interpretable. The computational and decision-making processes of these models operate in a "black box," which diminishes the intuitiveness and transparency associated with more traditional approaches, such as logistic regression modeling (36). This study was retrospective in nature, relying on historical patient data or electronic health records. Due to the absence of real-time interventions and standardized treatment protocols, retrospective data may suffer from incomplete documentation or inherent biases, potentially affecting the reliability of the findings. Furthermore, the study sample was not randomized, which introduces selection bias and may limit the generalizability of the results. Additionally, because the data were sourced from a specific hospital or region, the patient population characteristics—such as age, gender, and underlying conditions—may differ from those in other regions or populations. This distributional bias could undermine the external validity of the model, particularly its applicability to other regions or patient groups. Variations in medical resources, treatment standards, and genetic backgrounds across different regions could also lead to differing disease presentations and postoperative recovery, further impacting the model's broader applicability. Moreover, we acknowledge that certain potential confounders, such as anastomotic leakage leading to severe infections and impaired nutrient absorption, may exacerbate the risk of gastroparesis. However, these variables were not included due to data limitations and study design constraints. In future research, we plan to conduct a multicenter, prospective study to enhance the external validation of our model, ensuring its predictive accuracy across diverse clinical settings.

Conclusion

Identifying high-risk patients for gastroparesis after CME enables the implementation of timely interventions to enhance patient outcomes. The predictive model exhibited exceptional accuracy and strong clinical utility, offering surgeons a valuable tool for early diagnosis and proactive management. The analysis highlighted that the development of postoperative gastroparesis in colon cancer patients is strongly associated with advanced age, extended operative time, substantial intraoperative bleeding, surgical approach, hypoproteinemia, anemia, and preexisting conditions such as diabetes mellitus and hypothyroidism.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in this article/Supplementary material.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Wuxi People's Hospital and Wuxi Second People's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing. SZ: Conceptualization, Formal analysis, Investigation, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. WD: Conceptualization, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. WS: Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Software, Validation, Visualization, Writing – review & editing. NZ: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Top Talent Support Program for young and middle-aged people of Wuxi Health Committee (Grant No. HB2020007).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1467565/full#supplementary-material

SUPPLEMENTARY TABLE 1
Raw data.

References

1. Alexander DD, Weed DL, Miller PE, Mohamed MA. Red meat and colorectal cancer: A quantitative update on the state of the epidemiologic science. *J Am Coll Nutr.* (2015) 34:521–43. doi: 10.1080/07315724.2014.992553

2. Hohenberger W, Weber K, Matzel K, Papadopoulos T, Merkel S. Standardized surgery for colonic cancer: Complete mesocolic excision and central ligation–technical notes and outcome. *Colorectal Dis.* (2009) 11:354–64. doi: 10.1111/j.1463-1318.2008.01735.x

3. Feng B, Sun J, Ling TL, Lu AG, Wang ML, Chen XY, et al. Laparoscopic Complete Mesocolic Excision (Cme) with medial access for right-hemi colon cancer: Feasibility and technical strategies. *Surg Endosc.* (2012) 26:3669–75. doi: 10.1007/s00464-012-2435-9

4. Pedrazzani C, Lauka L, Sforza S, Ruzzenente A, Nifosì F, Delaini G, et al. Management of nodal disease from colon cancer in the laparoscopic era. *Int J Colorectal Dis.* (2015) 30:303–14. doi: 10.1007/s00384-014-2075-8

5. Egboh SC, Abere S. Gastroparesis: A multidisciplinary approach to management. *Cureus.* (2022) 14:e21295. doi: 10.7759/cureus.21295

6. Baia M, Conti L, Pasquali S, Sarre-Lazcano C, Abatini C, Cioffi SPB, et al. Delayed gastric emptying after multivisceral resection for retroperitoneal sarcoma. *Ann Surg Oncol.* (2022) 29:3264–70. doi: 10.1245/s10434-021-11154-z

7. Maehata T, Sato Y, Nakamoto Y, Kato M, Kawashima A, Kiyokawa H, et al. Updates in the field of submucosal endoscopy. *Life (Basel).* (2022) 13:104. doi: 10.3390/life13010104

8. Tack J, Janssen P. Gastroduodenal motility. *Curr Opin Gastroenterol.* (2010) 26:647–55. doi: 10.1097/MOG.0b013e32833ece1e

9. Lacy BE, Crowell MD, Mathis C, Bauer D, Heinberg LJ. Gastroparesis: Quality of life and health care utilization. *J Clin Gastroenterol.* (2018) 52:20–4. doi: 10.1097/mcg.0000000000000728

10. Blackett JW, Benvenuto L, Leiva-Juarez MM, D'Ovidio F, Arcasoy S, Jodorkovsky D. Risk factors and outcomes for gastroparesis after lung transplantation. *Dig Dis Sci.* (2022) 67:2385–94. doi: 10.1007/s10620-021-07249-y

11. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: A brief primer. *Behav Ther.* (2020) 51:675–87. doi: 10.1016/j.beth.2020.05.002

12. Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. *J Transl Med.* (2022) 20:215. doi: 10.1186/s12967-022-03364-0

13. Sullivan A, Temperley L, Ruban A. Pathophysiology, aetiology and treatment of gastroparesis. *Dig Dis Sci.* (2020) 65:1615–31. doi: 10.1007/s10620-020-06287-2

14. Zheng T, Camilleri M. Management of gastroparesis. *Gastroenterol Hepatol (N Y).* (2021) 17:515–25.

15. Arezzo F, Cormio G, Mongelli M, Cazzato G, Silvestris E, Kardhashi A, et al. Machine learning applied to Mri evaluation for the detection of lymph node metastasis in patients with locally advanced cervical cancer treated with neoadjuvant chemotherapy. *Arch Gynecol Obstet.* (2022) 307:1911–9. doi: 10.1007/s00404-022-06824-6

16. Melinte-Popescu AS, Vasilache IA, Socolov D, Melinte-Popescu M. Predictive performance of machine learning-based methods for the prediction of preeclampsia-a prospective study. *J Clin Med.* (2023) 12:418. doi: 10.3390/jcm12020418

17. Ebrahimi M, Aghagolzadeh P, Shamabadi N, Tahmasebi A, Alsharifi M, Adelson DL, et al. Understanding the undelaying mechanism of ha-subtyping in the level of physic-chemical characteristics of protein. *PLoS One.* (2014) 9:e96984. doi: 10.1371/journal.pone.0096984

18. Lin Z, Chou WC. Machine learning and artificial intelligence in toxicological sciences. *Toxicol Sci.* (2022) 189:7–19. doi: 10.1093/toxsci/kfac075

19. Lermite E, Pessaux P, Brehant O, Teyssedou C, Pelletier I, Etienne S, et al. Risk factors of pancreatic fistula and delayed gastric emptying after pancreaticoduodenectomy with pancreaticogastrostomy. *J Am Coll Surg.* (2007) 204:588–96. doi: 10.1016/j.jamcollsurg.2007.01.018

20. Meng H, Zhou D, Jiang X, Ding W, Lu L. Incidence and risk factors for postsurgical gastroparesis syndrome after laparoscopic and open radical gastrectomy. *World J Surg Oncol.* (2013) 11:144. doi: 10.1186/1477-7819-11-144

21. Al-Ani HH, Sims JL, Niederer RL. Cataract surgery in uveitis: Risk factors, outcomes, and complications. *Am J Ophthalmol.* (2022) 244:117–24. doi: 10.1016/j.ajo.2022.08.014

22. Crane J, Hamed M, Borucki JP, El-Hadi A, Shaikh I, Stearns AT. Complete mesocolic excision versus conventional surgery for colon cancer: A systematic review and meta-analysis. *Colorectal Dis.* (2021) 23:1670–86. doi: 10.1111/codi.15644

23. Magistro C, Bertoglio CL, Giani A, Mazzola M, Rubicondo C, Maspero M, et al. Laparoscopic complete mesocolic excision versus conventional resection for right-sided colon cancer: A propensity score matching analysis of short-term outcomes. *Surg Endosc* (2022) 36:3049–58. doi: 10.1007/s00464-021-08601-z

24. Fukagawa T. Role of staging laparoscopy for gastric cancer patients. *Ann Gastroenterol Surg.* (2019) 3:496–505. doi: 10.1002/ags3.12283

25. Yang DD, He K, Wu XL, Yang LK, Nie SF. Risk factors of gastroparesis syndrome after abdominal non-gastroduodenal operation and its prevention. *Asian Pac J Trop Med.* (2013) 6:497–9. doi: 10.1016/s1995-7645(13)60082-6

26. van Berge Henegouwen MI, van Gulik TM, DeWit LT, Allema JH, Rauws EA, Obertop H, et al. Delayed gastric emptying after standard pancreaticoduodenectomy versus pylorus-preserving pancreaticoduodenectomy: An analysis of 200 consecutive patients. *J Am Coll Surg* (1997) 185:373–9. doi: 10.1016/s1072-7515(97)00078-1

27. Hayakawa T, Kaneko H, Konagaya T, Shinozaki K, Kasahara A, Funaki Y, et al. Enhanced somatostatin secretion into the gastric juice with recovery of basal acid output after *Helicobacter* pylori eradication in gastric ulcers. *J Gastroenterol Hepatol.* (2003) 18:505–11. doi: 10.1046/j.1440-1746.2003.03008.x

28. Al-Saffar A, Lennernäs H, Hellström PM. Gastroparesis, metoclopramide, and tardive dyskinesia: Risk revisited. *Neurogastroenterol Motil.* (2019) 31:e13617. doi: 10.1111/nmo.13617

29. Petri M, Singh I, Baker C, Underkofler C, Rasouli N. Diabetic gastroparesis: An overview of pathogenesis, clinical presentation and novel therapies, with a focus on ghrelin receptor agonists. *J Diabetes Complic.* (2021) 35:107733. doi: 10.1016/j.jdiacomp.2020.107733

30. Farrugia G. Interstitial cells of cajal in health and disease. *Neurogastroenterol Motil.* (2008) 20(Suppl. 1):54–63. doi: 10.1111/j.1365-2982.2008.01109.x

31. Hasler WL. Gastroparesis: Pathogenesis, diagnosis and management. *Nat Rev Gastroenterol Hepatol.* (2011) 8:438–53. doi: 10.1038/nrgastro.2011.116

32. Bharucha AE, Batey-Schaefer B, Cleary PA, Murray JA, Cowie C, Lorenzi G, et al. Delayed gastric emptying is associated with early and long-term hyperglycemia in type 1 diabetes mellitus. *Gastroenterology.* (2015) 149:330–9. doi: 10.1053/j.gastro.2015.05.007

33. Nassar Y, Richter S. Gastroparesis in non-diabetics: Associated conditions and possible risk factors. *Gastroenterology Res.* (2018) 11:340–5. doi: 10.14740/gr1060w

34. Ghoshal UC, Bhut B, Misra A. Patients with specific gastrointestinal motility disorders are commonly diagnosed as functional gi disorders in the early stage by community physicians due to lack of awareness. *Turk J Gastroenterol.* (2021) 32:336–48. doi: 10.5152/tjg.2021.20514

35. Khraisha OS, Al-Madani MM, Peiris AN, Paul TK. Gastroparesis - a novel cause of persistent thyroid stimulating hormone elevation in hypothyroidism. *J La State Med Soc.* (2015) 167:47–9.

36. Aromolaran O, Aromolaran D, Isewon I, Oyelade J. Machine learning approach to gene essentiality prediction: A review. *Brief Bioinform.* (2021) 22:bbab128. doi: 10.1093/bib/bbab128

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 07 January 2025
DOI 10.3389/fmed.2024.1473533

OPEN ACCESS

EDITED BY
Rahul Kashyap,
WellSpan Health, United States

REVIEWED BY
Abhinav Hoskote,
WellSpan Health, United States
Smitesh Padte,
WellSpan Health, United States

*CORRESPONDENCE
Yong-Goo Shin
✉ ygshin92@korea.ac.kr
Seung Park
✉ spark.cbnuh@gmail.com

†These authors have contributed equally to this work

RECEIVED 31 July 2024
ACCEPTED 17 December 2024
PUBLISHED 07 January 2025

CITATION
Kim J, Kim G-H, Kim J-W, Kim KH, Maeng J-Y, Shin Y-G and Park S (2025) Transformer-based model for predicting length of stay in intensive care unit in sepsis patients.
Front. Med. 11:1473533.
doi: 10.3389/fmed.2024.1473533

Transformer-based model for predicting length of stay in intensive care unit in sepsis patients

Jeesu Kim[1,2†], Geun-Hyeong Kim[1,2†], Jae-Woo Kim[1,2], Ka Hyun Kim[1,2], Jae-Young Maeng[1], Yong-Goo Shin[3*] and Seung Park[2,4*]

[1]Medical Artificial Intelligence Center, Chungbuk National University Hospital, Cheongju, Republic of Korea, [2]College of Medicine, Chungbuk National University, Cheongju, Republic of Korea, [3]Department of Electronics and Information Engineering, Korea University, Sejong, Republic of Korea, [4]Department of Biomedical Engineering, Chungbuk National University Hospital, Cheongju, Republic of Korea

Introduction: Sepsis, a life-threatening condition with a high mortality rate, requires intensive care unit (ICU) admission. The increasing hospitalization rate for patients with sepsis has escalated medical costs due to the strain on ICU resources. Efficient management of ICU resources is critical to addressing this challenge.

Methods: This study utilized the dataset collected from 521 patients with sepsis at Chungbuk National University Hospital between July 2020 and August 2023. A transformer-based deep learning model was developed to predict ICU length of stay (LOS). The model incorporated global and local input data analysis through classification and feature-wise tokens, based on sequential organ failure assessment (SOFA) criteria. Model performance was evaluated using four-fold cross-validation.

Results: The proposed model achieved a mean absolute error (MAE) of 2.05 days for predicting ICU LOS. The result demonstrates the ability of the proposed model to provide accurate and reliable predictions.

Discussion: The proposed model offers valuable insights for healthcare resource management by optimizing ICU resource allocation and potentially reducing medical expenses. These findings highlight the applicability of the proposed model to efficient healthcare cost management.

KEYWORDS

sepsis, intensive care unit, length of stay, sequential organ failure assessment, transformer, tabular data

1 Introduction

Sepsis, a life-threatening condition, arises when the body's response to infection induces widespread inflammation (1). This inflammatory response can damage multiple organ systems, leading to severe multi-organ failure (2). The rapid progression of sepsis can result in death without timely and appropriate treatment (3). Global guidelines often recommend intensive care for patients with sepsis (2). Despite advancements in medical technology, including early diagnostic methods, rapid antibiotic administration (4), and advanced supportive care such as mechanical ventilation and extracorporeal membrane oxygenation (5), sepsis continues to pose a significant healthcare challenge worldwide due to increasing mortality and morbidity rates (6–11).

Treating sepsis in the intensive care unit (ICU) incurs significantly higher costs than treating other diseases owing to the need for advanced life support measures, prolonged hospitalization, and the complexity of managing multi-organ failure (3, 12, 13). Recent statistics indicate that the annual cost of treating sepsis in the United States exceeds $24 billion, making it the most expensive treatment option for hospitals (13). This high cost is primarily driven by the length of stay (LOS) in the ICU because patients with sepsis often require prolonged intensive care. Sepsis accounts for 4.7–42.2% of global ICU utilization because ICU admission is recommended as an aggressive treatment regimen (14). Additionally, sepsis readmission rates are alarmingly high, with approximately 19% of survivors readmitted within 30 days, further escalating healthcare expenditures (15). Accurate prediction of ICU LOS for sepsis patients is crucial, as it enables healthcare facilities to optimize resource allocation, such as bed utilization, staffing, and equipment availability. By improving care efficiency, hospitals can reduce operational costs, enhance patient turnover rates, and ultimately contribute to cost savings for both healthcare providers and the broader system (16, 17).

.Efforts to predict ICU LOS have significantly advanced in recent years. In 2022, Wu et al. (18) demonstrated the utility of machine learning techniques by predicting ICU LOS (area under the receiver operating characteristic curve (AUROC) = 0.742) using gradient boosting decision trees (GBDTs). In the same year, Deng et al. (19) improved accuracy (AUC = 0.765) by utilizing temporal data and focusing on the changes in progression according to treatment stages using gated recurrent units (GRU) and long short-term memory (LSTM) networks. In 2023, the emphasis shifted to simpler, clinically interpretable models, such as linear regression models utilizing the sequential organ failure assessment (SOFA) score. Zangmo and Khwannimit (20) developed a model to classify sepsis patients with ICU LOS exceeding 3 days (AUC = 0.530), while Farimani et al. (21) proposed a model to predict ICU LOS in cardiac surgery patients (root mean square error (RMSE) = 5.181). Despite the advances, existing models have struggled to effectively capture the complex feature interactions inherent in structured data. GBDT emphasizes individual feature importance through splits (22), making it less effective at explicitly modeling complex interactions or high-dimensional relationships. On the other hand, GRU and LSTM models are optimized for processing sequential data, the models exhibit structural limitations in learning complex inter-variable relationships in structured datasets (19). Similarly, linear regression assumes linear relationships between variables and is, therefore, unable to capture nonlinear interactions (23). The limitations underscore the necessity for innovative methods capable of effectively learning complex and high-dimensional data structures. Hence, we present a transformer-based solution to address the limitations of structured data analysis by simultaneously capturing nonlinear feature interactions and learning global relationships through attention mechanisms.

Transformers have shown promise in structured data analysis by incorporating innovative mechanisms such as the classification (CLS) token, a functionality originally introduced in the bidirectional encoder representations from transformers (BERT) (24) in 2018. The CLS token serves as a global representation of the input sequence, summarizing overall patterns in structured data through self-attention, and allows Transformers to effectively capture global dependencies across features. In 2021, Models such as self-attention and intersample attention transformer (SAINT) (25) and feature tokenizer (FT)-Transformer (26) successfully leveraged CLS tokens, achieving performance improvements in tabular datasets. SAINT enables feature-to-feature and sample-to-sample interactions, while the FT-Transformer captures intricate inter-feature relationships and global patterns. However, while the CLS token excels at capturing global information, achieving a comprehensive analysis of attention mechanisms using CLS tokens can be challenging, as attention mechanisms may ignore important local feature details, particularly in datasets with complex interdependencies (27–29). The limitation of the CLS token underscores a persistent challenge in Transformer-based models when applied to highly intricate structured data.

This study focuses on predicting the ICU LOS for patients with sepsis using a transformer-based DL model applied to SOFA-based tabular data. The proposed model uses an attention mechanism and a skip-connected token process, integrating global information from a CLS token and local information from feature-wise tokens during the final classification. This approach adds to the growing body of work on applying DL techniques to tabular data in predicting ICU LOS for patients with sepsis.

2 Methods

2.1 Dataset information

2.1.1 Study population

To develop the DL model, we constructed a dataset from patients treated for sepsis at Chungbuk National University Hospital (Cheong-Ju, Korea) between July 3, 2020 and August 3, 2023. The study, conducted following the principles of the Declaration of Helsinki, received approval from the Institutional Review Board of Chungbuk National University Hospital (IRB no. CBNUH 2021-02-034-001). Patient information was anonymized and de-identified prior to analysis.

As shown in Figure 1, we initially identified patients meeting the Sepsis-3 guidelines for suspicion or diagnosis of sepsis, defined as a quick SOFA (qSOFA) score of ≥ 2. We sequentially excluded patients who met the following criteria: ICU admission post-surgery, readmission due to sepsis during treatment, ICU stays of less than 24 h, withdrawal of life-sustaining therapy, ICU discharge, admission with cardiogenic shock, hypovolemic shock, or acute stroke, procalcitonin level of ≤0.05, missing data, death, and ICU LOS outliers. This process resulted in a dataset comprising 521 patients.

We collected various clinical and SOFA-related features to construct a sepsis-specific ICU LOS prediction model. The features included: (1) Clinical features: age, sex, body mass index (BMI), lactate, atrial fibrillation (AF), systolic blood pressure(SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), partial pressure of arterial oxygen and fraction of inspired oxygen ratio (PaO2/FiO2, P/F ratio), Glasgow Coma Scale (GCS)]. (2) SOFA-related features: vasopressor (VASO), mechanical ventilator (MV), 24-h urinary excretion (UR), platelets (Plt), serum total bilirubin (Bil), serum creatinine (Cr)]. (3) Target feature: ICU LOS.

Table 1 presents detailed statistical information on the features used in this study. Numerical features are described using means, standard deviations, and min-max ranges, while categorical features are reported as frequencies and percentages. The Pearson correlation coefficient for numerical features and point-biserial correlation for categorical features were calculated to determine their correlation with the target feature. *p*-values in Table 1 test the null hypothesis that the correlation coefficient is zero.

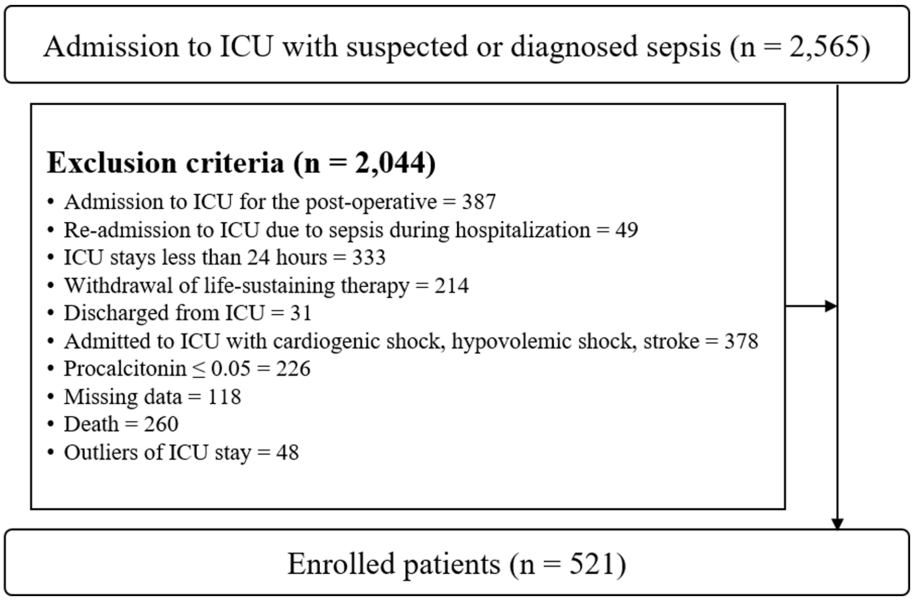

FIGURE 1
Flow diagram of the study inclusion and exclusion.

TABLE 1 Statistical information of features in our dataset.

Characteristics	Categories		Dataset (*n* = 521)	Correlation coefficient	*p*-value
Clinical features	Age	Years	69.19 ± 14.65 (19–95)	−0.09	0.05
	Sex	Female	239 (45.87%)	0.01	0.90
		Male	282 (54.13%)		
	BMI	kg/cm^2	22.41 ± 4.18 (11.55–43.94)	−0.07	0.11
	LACTATE	mmol/L	2.58 ± 2.56 (0–29)	0.16	< 0.01
	AF	yes	86 (16.51%)	0.06	0.19
		no	435 (83.49%)		
	SBP	mmHg	89.67 ± 19.59 (33–176)	−0.08	0.08
	DBP	mmHg	49.31 ± 11.46 (17–90)	−0.09	0.04
	MBP	mmHg	60.28 ± 11.83 (24–104)	−0.10	0.03
	PF	-	290.69 ± 152.05 (16–943)	−0.31	< 0.01
	GCS	-	11.08 ± 3.55 (3–15)	−0.42	< 0.01
SOFA-related features	VASO	Yes	298 (57.2%)	0.11	0.02
		No	223 (42.8%)		
	MV	Yes	171 (32.82%)	0.43	< 0.01
		No	350 (67.18%)		
	UR	cm^3	1803.52 ± 1332.96 (0–1,250)	−0.19	< 0.01
	Plt	×10^3/μl	154.16 ± 96.02 (4–657)	−0.06	0.19
	Bil	mg/dl	1.31 ± 2.09 (0.09–22.58)	−0.02	0.72
	Cr	mg/dl	2.14 ± 2.20 (0.11–16.60)	0.07	0.13
Target feature	ICU LOS	days	5.24 ± 3.57 (1.01–16.47)		

SOFA, sequential organ failure assessment; BMI, body mass index; AF, atrial fibrillation; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; P/F ratio, partial pressure of arterial oxygen and a fraction of inspired oxygen ratio (PaO_2/FiO_2); GCS, Glasgow Coma Scale; ICU LOS, length of stay in intensive care unit; VASO, vasopressor; MV, mechanical ventilator; UR, 24 h urinary excretion; Plt, platelets; Bil, serum total bilirubin; Cr, serum creatinine.

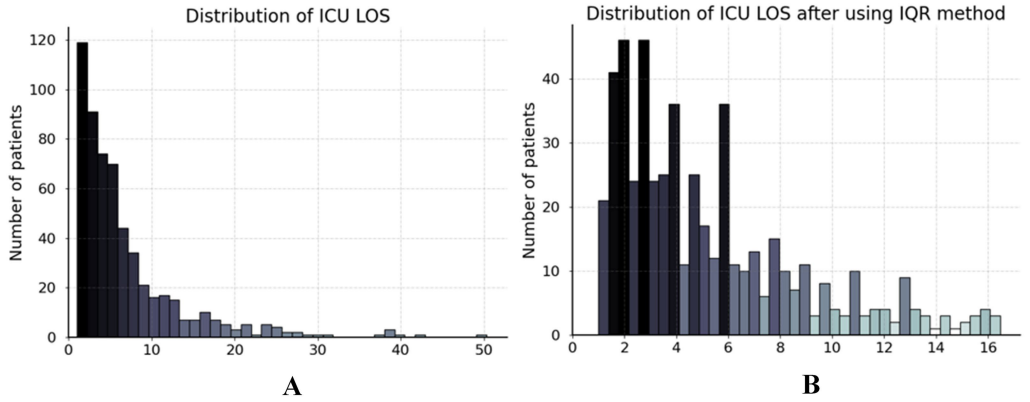

FIGURE 2
Distribution of ICU LOS in dataset. **(A)** The ICU LOS before data preprocessing and **(B)** the ICU LOS after handling outliers using the IQR method.

2.1.2 Data preprocessing

As depicted in Figure 2A, the ICU LOS distribution in our dataset exhibited a pronounced positive skew, with a concentration of values at the lower end and a long tail extending towards higher values. This necessitated the removal of outliers prior to analysis. The interquartile range (IQR) method was employed to handle outliers in ICU LOS (30). The IQR method effectively retains most data points within a reasonable range, excluding outliers that could potentially distort the analysis (31). Specifically, the IQR is the range between the first quartile (Q1) and third quartile (Q3) of the data, with outliers defined as points below Q1−1.5IQR or above Q3 + 1.5IQR (32). Our study identified patients with an ICU LOS > 16.52 days as outliers, excluding 48 patients as shown in Figure 2B. Furthermore, we standardized the dataset to ensure that all features contributed equally to the analysis and to prevent any single feature from disproportionately influencing the results due to scale differences. This procedure was applied exclusively to numerical features. The standardization formula is defined in Equation 1 as follows:

$$Z = \frac{\left(X - \bar{X}\right)}{s}, \tag{1}$$

where $\bar{X}$ represents the mean and s denotes the standard deviation (33).

2.2 Model architecture

We developed a transformer-based DL model using a CLS token to predict the ICU LOS. The architecture of the proposed model is depicted in Figure 3; it consists of three modules as shown in Figure 4.

2.2.1 Module of concatenating CLS tokens

The input data $x \in \mathbb{R}^{col}$, where col represents the number of input features, is batch normalized before entering the "fzCLSBlock." As represented in Figure 4A, x in the fzCLSBlock is concatenated with a trainable CLS token $\mathrm{CLS}_0 \in \mathbb{R}^1$, which is zero-initialized to ensure stable training (24). The CLS token is the first special token of every sequence and is widely used as an aggregate sequence representation for classification tasks (26). The concatenated vectors are embedded through a dense layer to achieve a representative embedding of the input data and capture the complex relationships. This process is expressed in Equation 2 as follows:

$$z = \mathrm{Dense}\left(\mathrm{CLS}_0 \,\|\, x\right) \in \mathbb{R}^{(col+1)\times d}, \tag{2}$$

where d represents the embedding dimension and || denotes the concatenate function.

2.2.2 Module of multi-head self-attention

Inspired by networks in several studies using transformers (25, 34–36), we employed the self-attention mechanism of the transformer encoder. The self-attention mechanism calculates model weights to assess the relevance of each feature and captures interactions between features or instances. Recent research have demonstrated superior prediction accuracy by incorporating self-attention mechanisms in new networks such as TabNet (35) and FT-transformers (26). These findings suggest the efficacy of self-attention mechanisms for analyzing tabular datasets.

The projected vector $z \in \mathbb{R}^{(col+1)\times d}$ (as defined above) is analyzed in the "AttnEncoderBlock" illustrated in Figure 4B. The input vector z is linearly transformed into query (Q), key (K), and value (V) matrices within the single attention head of multihead self-attention (MSA) (37). The attention weight is calculated by taking the dot product of Q and K, normalizing it by the square root of the dimension of K, and applying a softmax function. After that, the attention head outputs the dot product of the attention score and V, which are computed in parallel five times. Besides the MSA, the AttnEncoderBlock includes a fully connected feedforward network (FFN) composed of two linear transformations with a rectified linear unit (ReLU) activation in between (38).

2.2.3 Module of analyzing global and local information

Previous research indicates that CLS tokens often fail to adequately capture the semantic content of the input because they focus more on global information than local and low-level features (39). We designed a skip-connected token process, which comprehensively analyzes global and local data information, to address this issue. A

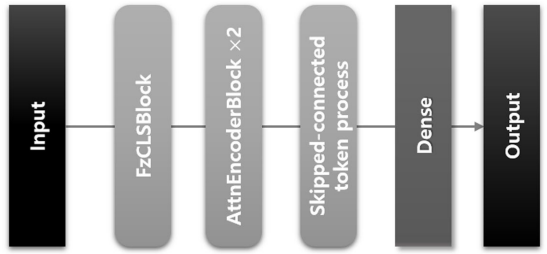

FIGURE 3
Architecture of the proposed model for predicting ICU LOS in patients with sepsis.

skip-connected token process addes token values representing both global and local information.

As presented in Figure 4C, the output of the AttnEncoderBlock is batch normalized and divided into the CLS $\in \mathbb{R}^{1\times d}$ token, summarizing all features, and the feature-wise token $f' \in \mathbb{R}^{col\times d}$, maintaining the unique information of each feature. The f', containing local information, passes through a dense layer to convert the local information into more abstract and high-level features. This layer captures complex dependencies and correlations between local features by identifying the interactions between various feature dimensions and learning appropriate weights. Additionally, the CLS token, containing global information, is added to f, enabling a comprehensive analysis of global and local information. These computations can be expressed in Equations 3 and 4 as follows:

$$f = \text{Dense}\left(\text{Flatten}\left(f'\right)\right) \in \mathbb{R}^{d}, \tag{3}$$

$$t_{sc} = \text{Dense}\left(\text{CLS} + f\right) \in \mathbb{R}^{d}. \tag{4}$$

The t_{sc} token is used for the final prediction of the proposed model, predicting the ICU LOS via a dense layer with one unit.

2.3 Implementation details

The proposed model was implemented using Python 3.9 on a workstation with an 11th Gen Intel(R) Core(TM) i7-11700K processor at 3.60 GHz and 64 GB of RAM. We applied exponential decay to control the learning rate during training, gradually reducing it to ensure stable convergence. The proposed model was configured with a batch size of 32 and the Adam optimizer at a learning rate of 1e-3. Learning rate decay was applied every 10 steps at a rate of 0.96. Furthermore, we compared the prediction accuracy of the proposed model with that of conventional ML and DL models. Hyperparameters for random forest (RF) (40), extreme gradient boosting (XGBoost) (40), support vector regression (SVR) (41), multiple linear regression (MLR) (42), and TabNet (35) were set to their respective default values.

2.4 Model performance evaluation

We conducted a four-fold cross-validation to verify the reliability and consistency of the predictions of the proposed model. Twenty percent of the dataset was allocated for testing, while the remaining dataset was divided into four folds. Each iteration of the four-fold validation consisted of one fold used for validation and the remaining folds used for training. We adopted the following three key metrics to quantitatively evaluate the performance of the proposed model because it performed a regression task: coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). Detailed descriptions of each metric are as follows:

The R^2 value measures the proportion of variance in the dependent feature that can be predicted from the independent features. The R^2 value ranges from 0 to 1, where 0 indicates that the model does not explain the variability in the response data around its mean, and 1 indicates that the model explains all the variability of the response data around its mean (43). R^2 value for an ideal model is close to 1 and is computed using Equation 5 as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}, \tag{5}$$

where n denotes the number of patients, y_i corresponds to the observed value, $\hat{y}_i$ represents the predicted value, and $\bar{y}$ is the average ICU LOS. The R^2 metric is crucial as it directly correlates with the proportion of the total variation in the target feature explained by the model. A high R^2 indicates that the model captures a significant portion of the variance, vital for predictive accuracy (44).

The MAE represents the average absolute difference between the predicted and observed values of the model (45). It provides a straightforward and interpretable measure of the average prediction error (46). Ideally, the MAE value approaches zero and is computed using Equation 6 as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \tag{6}$$

where n represents the number of data points used for model testing, $\hat{y}_i$ corresponds to the value predicted by the model for the i-th sample, and y_i denotes the corresponding observed value (47). The MAE is advantageous due to its reduced sensitivity to outliers compared to metrics such as RMSE, making it a more reliable indicator of the average performance of a model, particularly when handling datasets with noisy or extreme values (48).

The RMSE is the square root of the average squared difference between the predicted and actual observations. It is widely used due to its ability to penalize larger errors more heavily than MAE, highlighting significant deviations (48). The formula for RMSE is calculated using Equation 7:

$$\text{MSE} = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}. \tag{7}$$

This metric provides an aggregate measure of model accuracy, encompassing both bias and variance components of error. RMSE is

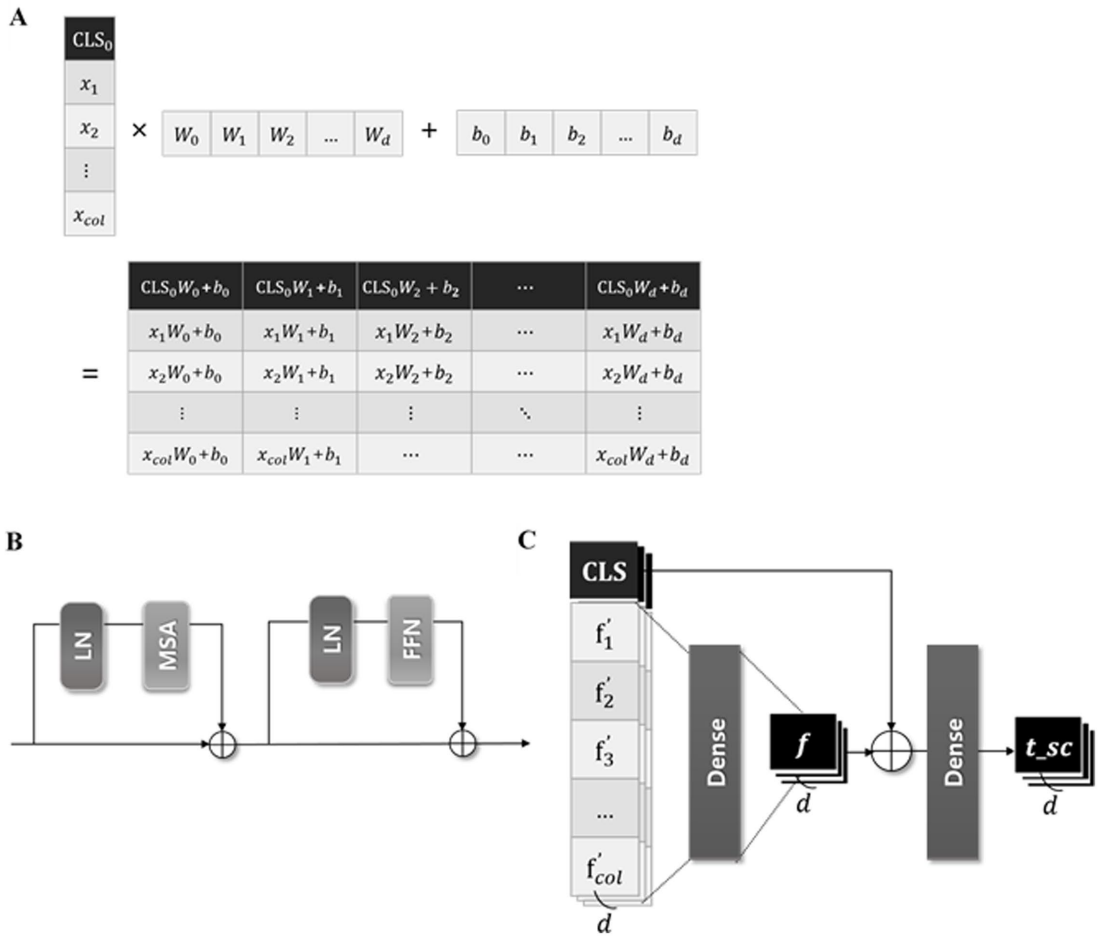

FIGURE 4
Three modules of the proposed model. **(A)** The illustration of the process concatenating CLS token in FzCLSBlock, **(B)** schematic diagram of the module of multi-head self-attention, **(C)** process of global and local information analysis.

valuable in applications where larger errors are more significant and must be minimized. Its sensitivity to large errors makes it essential for ensuring robustness and precision (49).

3 Results

3.1 Model performance comparison

We conducted a performance comparison of the proposed model using a four-fold cross-validation of the datasets. In Table 2, the proposed model demonstrated promising predictive performance, achieving an average R^2, MAE, and RMSE of 0.29 ± 0.01, 2.05 ± 0.03, and 2.72 ± 0.02, respectively. The average R^2 indicates the proposed model could explain approximately 29% of the variability in ICU LOS. Notably, the R^2, MAE and RMSE values showed minimal variation across folds, demonstrating the stability.

Figure 5 presents the calibration plot for each model, illustrating the agreement between predicted and observed ICU LOS. Darker points in these plots represent a better fit with actual values. The plots indicate that while the proposed model accurately predicts shorter ICU stays, it exhibits noticeable deviations for longer stays. This finding suggests that the proposed model demonstrated strong performance for shorter ICU stays; however, it may require further refinement to improve accuracy for longer stays, which are often associated with more complex and variable patient conditions.

Additionally, we compared the performance of the proposed model with conventional ML and DL models, as shown in Table 3. The proposed model leveraged the skip-connected token process to enhance its predictive power by capturing interactions within tabular data. Comparisons were made with other DL models using MSA, such as TabNet and FT-Transformer, as well as traditional models known for their strong performance on tabular data, including RF, XGBoost, and MLR. The proposed model demonstrated superior performance compared to the other models, indicating that it provides more accurate predictions.

3.2 Ablation study

We conducted an ablation study to evaluate the effectiveness of the proposed skip-connected token process. This study compared the information delivered to the ICU LOS output layer by altering specific components. The performance of models was compared across three

categories: models that used only local information analysis, only global information analysis, or a combination of both. Detailed configurations and corresponding performance indicators are provided in Table 4.

The first model, which utilized only global information analysis, demonstrated poor performance. In contrast, the second model, relying solely on local information analysis, exhibited improved results. Notably, the proposed model, integrating global and local information analysis, achieved an R^2, MAE, and RMSE of 0.29 ± 0.01, 2.05 ± 0.03, and 2.72 ± 0.02, respectively, outperforming the other two models.

These results indicate that the proposed model, which uses skip-connected token process, has the highest explanatory power and lowest prediction error, demonstrating a significant enhancement in overall model performance. This underscores the necessity of skip-connected token process in integrating local and global information for improved predictions.

TABLE 2 Performance of the proposed model evaluated using four-fold cross-validation.

Four-fold cross-validation	CBNUH (521 patients)		
	R^2	MAE	RMSE
Fold1	0.30	2.03	2.69
Fold2	0.28	2.06	2.73
Fold3	0.29	2.01	2.71
Fold4	0.27	2.08	2.74
Average	0.29 ± 0.01	2.05 ± 0.03	2.72 ± 0.02

3.3 Model interpretation

We employed Shapley additive descriptions (SHAP) to assess the impact of each feature on the model predictions. Figure 6A displays the mean absolute SHAP values for each feature, highlighting their importance in the model predictions. The top three most influential features were GCS, MV and PF. This ranking elucidates the primary factors that drive the predictions of the proposed model, offering valuable insights into which features most significantly affect ICU length of stay predictions.

Figure 6B presents the SHAP summary plots for four different percentiles (20th, 40th, 60th, and 80th) of ICU LOS. These plots visualize the SHAP values for individual predictions, indicating how each feature affects the predicted ICU LOS. The color scale shows the direction of the impact, with red and blue indicating an increase and decrease in the predicted LOS, respectively.

Sequentially across percentiles, the high GCS score and application of MV reduced predicted LOS, primarily at the 20th and 40th percentiles of ICU LOS (see blue section). Conversely, the low PF ratio increased LOS. The GCS represents a level of consciousness rating of 3–15 that assesses neurological status, with a lower score indicating worse status (50). The MV and PF are important indicators of respiratory function, reflecting the need for mechanical ventilation and the oxygen exchange capacity of the lungs, respectively (49). The impact of MV application and low PF was also evident at the 60th percentile, significantly increasing ICU LOS. On the other hand, an average level of Plt indicates a properly functioning coagulation system and reduces ICU LOS. The low GCS score, high Bil level, and MV application played a significant role in 80th percentile ICU LOS, with severe GCS score significantly increasing expected ICU LOS. Bil is another important predictor of ICU

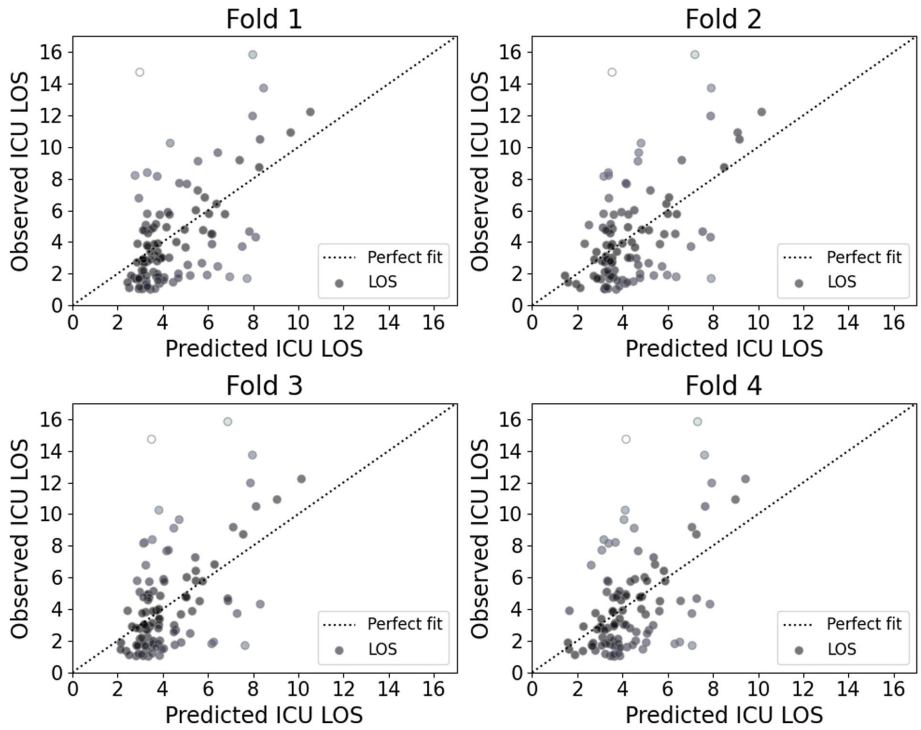

FIGURE 5
Calibration plot of each fold model. ICU, Intensive care unit; LOS, length of stay.

LOS, with elevated levels indicating liver dysfunction or hemolysis. However, contrarily in our study, elevated Bil was shown to reduce ICU LOS.

4 Discussion

This study demonstrated that the transformer-based DL model outperformed traditional ML and DL models in predicting ICU LOS for patients with sepsis using SOFA-based tabular data. The proposed model, leveraging a skip-connected token process to integrate global and local information, achieved an average R^2, MAE, and RMSE of 0.29, 2.05 days, and 2.72 days, respectively. Reliable predictions of ICU LOS are clinically and operationally impactful, as they enable better resource allocation and improve patient outcomes, particularly for critical conditions like sepsis (51, 52). The proposed model builds on these insights by providing an efficient tool that uses limited SOFA-based data to achieve practical predictions.

TABLE 3 Comparison of conventional model performance.

Model	CBNUH		
	R^2	MAE	RMSE
RF	0.09 ± 0.01	2.25 ± 0.01	3.07 ± 0.01
XGBoost	0.10 ± 0.02	2.23 ± 0.02	3.05 ± 0.04
MLR	0.22 ± 0.00	2.13 ± 0.01	2.84 ± 0.01
TabNet	−0.39 ± 0.21	2.77 ± 0.17	3.79 ± 0.29
FT-Transformer	0.26 ± 0.03	2.12 ± 0.02	2.76 ± 0.05
Proposed model	0.29 ± 0.01	2.05 ± 0.03	2.72 ± 0.02

TABLE 4 Ablation study on the proposed model.

Method	R^2	MAE	RMSE
Local information analysis	0.22 ± 0.02	2.11 ± 0.06	2.84 ± 0.04
Global information analysis	0.15 ± 0.02	2.27 ± 0.04	2.97 ± 0.03
Local and Global information analysis	0.29 ± 0.01	2.05 ± 0.03	2.72 ± 0.02

The strengths of the proposed model are manifold: First, the input features are based on the SOFA criteria, widely used in ICUs to assess organ dysfunction severity in critically ill patients. The model requires only 16 SOFA-related clinical features collected within 24 h of ICU admission, making it a convenient tool for predicting ICU LOS in patients with sepsis due to the accessibility of SOFA criteria data. Second, the proposed model was designed to work with tabular data, the most common structured data format, which requires less computational power than other data types and does not necessitate high-end hardware. Third, the model effectively captures comprehensive information from the features utilizing CLS and feature-wise tokens, analyzing global and local information. The proposed model employs MSA to capture global interactions between features, further analyze local information through dense layers and then integrates both in the final prediction to enhance performance.

Furthermore, the proposed model was interpreted using SHAP, providing valuable insights into the relative importance of various features in predicting ICU LOS. The top three influential features in this study were GCS, MV, and PF. The GCS score, underscored for its critical role in assessing neurological status, showed a positive correlation with ICU LOS. This finding is consistent with a previous

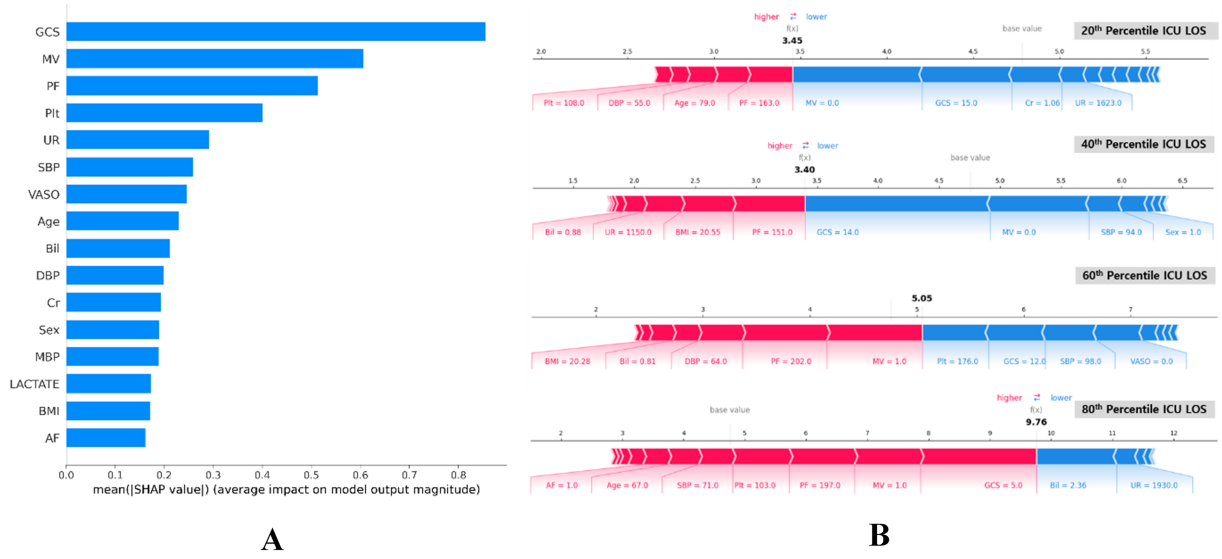

FIGURE 6
SHAP importance of features for predicting ICU LOS. **(A)** Summary plot of SHAP feature importance, represented by the mean absolute Shapley values. The plot illustrates the significance of each covariate in the final predictive model. **(B)** SHAP force plots for data instances with predicted ICU LOS at the 80th, 60th, 40th, and 20th percentiles (bottom). These plots provide an explanation for individual predictions made by the model. Note: the base value of 4.77 days is consistent across all plots. P/F, partial pressure of arterial oxygen and fraction of inspired oxygen ratio (PaO_2/FiO_2); GCS, Glasgow Coma Scale; AF, atrial fibrillation; Cr, serum creatinine; VASO, vasopressor; Plt, platelets; MV, mechanical ventilator; UR, 24 h urinary excretion; Bil, serum total bilirubin; DBP, diastolic blood pressure; BMI, body mass index; SBP, systolic blood pressure; MBP, mean blood pressure; ICU LOS, length of stay in intensive care unit.

study (53), and highlights the importance of GCS as the most significant predictor. Similarly, MV and PF are respiratory indices associated with ICU LOS prediction in this study. The results in our study are consistent with previous studies showing that MV use and lower PF increased ICU LOS (53, 54). Conversely, this study found that elevated bilirubin levels were associated with a shorter ICU LOS, which contrasts with a previous study where higher bilirubin levels prolonged the length of hospital stay (55). The correlation of these factors indicates that these may assist in determining ICU LOS.

However, this study has several limitations. The dataset was derived from a single institution, potentially limiting the generalizability of the findings. Future research should aim to validate the proposed model across diverse healthcare settings and larger multicenter datasets. Additionally, while the transformer-based model outperformed others in predicting ICU LOS, it showed an opportunity for improvement, particularly in predicting stays longer than 8 days. This result suggests the need for additional data on longer durations to improve the prediction of extended ICU LOS in real medical scenarios.

5 Conclusion

We developed a transformer-based DL model to predict ICU LOS in patients with sepsis using data collected within the first 24 h of ICU admission. The proposed model achieved an MAE of 2.05 days. The proposed model effectively captures complex feature interactions by integrating global and local information through a novel skip-connected token process. Additionally, the proposed model utilizes a set of SOFA-related features that are widely used to assess the severity of organ dysfunction in clinical practice. Such an approach ensures simplicity of data collection and wide applicability, making the proposed model practical for use in a variety of healthcare settings.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because the review board waived the requirement for informed consent, owing to the retrospective design of this study.

Author contributions

JK: Methodology, Software, Validation, Visualization, Writing – original draft. G-HK: Formal analysis, Methodology, Software, Writing – review & editing. J-WK: Data curation, Formal analysis, Methodology, Software, Writing – review & editing. KK: Data curation, Formal analysis, Validation, Writing – original draft. J-YM: Conceptualization, Data curation, Formal analysis, Writing – original draft. Y-GS: Methodology, Project administration, Supervision, Validation, Writing – review & editing. SP: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by a research grant from Chungbuk National University in 2024, and by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (Grant Number: RS-2023-00267328).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287

2. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Crit Care Med*. (2021) 49:e1063–143. doi: 10.1097/CCM.0000000000005337

3. Khwannimit B, Bhurayanontachai R. The direct costs of intensive care management and risk factors for financial burden of patients with severe sepsis and septic shock. *J Crit Care*. (2015) 30:929–34. doi: 10.1016/j.jcrc.2015.05.011

4. Kim HI, Park S. Sepsis: early recognition and optimized treatment. *Tuberc Respir Dis*. (2019) 82:6–14. doi: 10.4046/trd.2018.0041

5. Fortenberry JD, Paden ML. Extracorporeal therapies in the treatment of sepsis: experience and promise. Seminars in pediatric infectious diseases. Amsterdam: Elsevier (2006).

6. Dombrovskiy VY, Martin AA, Sunderram J, Paz HL. Facing the challenge: decreasing case fatality rates in severe sepsis despite increasing hospitalizations. *Crit Care Med*. (2005) 33:2555–62. doi: 10.1097/01.CCM.0000186748.64438.7B

7. van Gestel A, Bakker J, Veraart CP, van Hout BA. Prevalence and incidence of severe sepsis in Dutch intensive care units. *Crit Care*. (2004) 8:1–10. doi: 10.1186/cc2858

8. Stoller J, Halpin L, Weis M, Aplin B, Qu W, Georgescu C, et al. Epidemiology of severe sepsis: 2008-2012. *J Crit Care*. (2016) 31:58–62. doi: 10.1016/j.jcrc.2015.09.034

9. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med*. (2003) 348:1546–54. doi: 10.1056/NEJMoa022139

10. Martin GS. Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. *Expert Rev Anti-Infect Ther*. (2012) 10:701–6. doi: 10.1586/eri.12.50

11. Patel JJ, Taneja A, Niccum D, Kumar G, Jacobs E, Nanchal R. The association of serum bilirubin levels on the outcomes of severe sepsis. *J Intensive Care Med*. (2015) 30:23–9. doi: 10.1177/0885066613488739

12. Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and costs of sepsis in the United States—an analysis based on timing of diagnosis and severity level. *Crit Care Med*. (2018) 46:1889–97. doi: 10.1097/CCM.0000000000003342

13. Torio CM, Moore BJ. (2016). National inpatient hospital costs: The most expensive conditions by payer, 2013.

14. Raman V, Laupland KB. Challenges to reporting the global trends in the epidemiology of ICU-treated sepsis and septic shock. *Curr Infect Dis Rep*. (2021) 23:1–8. doi: 10.1007/s11908-021-00749-y

15. Fingar K, Washington R. (2016). Trends in hospital readmissions for four high-volume conditions, 2009–2013.

16. Yinusa A, Faezipour M. Optimizing healthcare delivery: a model for staffing, patient assignment, and resource allocation. *Appl Syst Innov*. (2023) 6:78. doi: 10.3390/asi6050078

17. Cosgrove SE. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. *Clin Infect Dis*. (2006) 42:S82–9. doi: 10.1086/499406

18. Wu J, Lin Y, Li P, Hu Y, Zhang L, Kong G. Predicting prolonged length of ICU stay through machine learning. *Diagnostics*. (2021) 11:2242. doi: 10.3390/diagnostics11122242

19. Deng Y, Liu S, Wang Z, Wang Y, Jiang Y, Liu B. Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. *Front Med*. (2022) 9:933037. doi: 10.3389/fmed.2022.933037

20. Zangmo K, Khwannimit B. Validating the APACHE IV score in predicting length of stay in the intensive care unit among patients with sepsis. *Sci Rep*. (2023) 13:5899. doi: 10.1038/s41598-023-33173-4

21. Farimani RM, Amini S, Bahaadini K, Eslami S. (2024). Predicting length of stay in intensive care units for cardiovascular surgery patients using APACHE II, APACHE IV, SAPS II and SOFA Scores.

22. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: a highly efficient gradient boosting decision tree. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems. Vol. 30. Long Beach, CA, USA: Curran Associates Inc. (2017). p. 3146–3154. Available at: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

23. Hope TM. Linear regression. Machine learning. Amsterdam: Elsevier; (2020). p. 67–81.

24. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv* preprint arXiv:181004805. (2018)

25. Somepalli G, Goldblum M, Schwarzschild A, Bruss CB, Goldstein T. Saint: improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv* preprint arXiv:210601342. (2021)

26. Gorishniy Y, Rubachev I, Khrulkov V, Babenko A. Revisiting deep learning models for tabular data. In: Advances in Neural Information Processing Systems. Vol. 34. Curran Associates Inc. (2021). p. 18932–43. Available at: https://proceedings.neurips.cc/paper/2021/file/01a1a84df7a0baaa88a1a9fa7848d42c-Paper.pdf

27. Chen C-FR, Fan Q, Panda R, editors. (2021). "Crossvit: cross-attention multi-scale vision transformer for image classification." in *Proceedings of the IEEE/CVF international conference on computer vision*.

28. Yuan L, Chen Y, Wang T, Yu W, Shi Y, Jiang Z-H, et al., editors. (2021). "Tokens-to-token vit: training vision transformers from scratch on imagenet." in *Proceedings of the IEEE/CVF international conference on computer vision*.

29. Wang J, Yu X, Gao Y. Feature fusion vision transformer for fine-grained visual categorization. *arXiv* preprint arXiv:210702341. (2021)

30. Mansoori A, Zeinalnezhad M, Nazarimanesh L. Optimization of tree-based machine learning models to predict the length of hospital stay using genetic algorithm. *J Healthc Eng*. (2023) 2023:3395. doi: 10.1155/2023/9673395

31. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. (2014) 14:1–13. doi: 10.1186/1471-2288-14-135

32. Walfish S. A review of statistical outlier methods. *Pharm Technol*. (2006) 30:82.

33. Raju VG, Lakshmi KP, Jain VM, Kalidindi A, Padma V, editors. (2020). "Study the influence of normalization/transformation process on the accuracy of supervised classification." in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT): IEEE*.

34. Hwang Y, Song J. Recent deep learning methods for tabular data. *Commun Stat Appl Methods*. (2023) 30:215–26. doi: 10.29220/CSAM.2023.30.2.215

35. Arik SÖ, Pfister T. Tabnet: attentive interpretable tabular learning. *Proc AAAI Conf Artif Intel*. (2021) 35:6679–87. doi: 10.1609/aaai.v35i8.16826

36. Song W, Shi C, Xiao Z, Duan Z, Xu Y, Zhang M, et al., editors. (2019). "Autoint: automatic feature interaction learning via self-attentive neural networks." in *Proceedings of the 28th ACM international conference on information and knowledge management*.

37. Choi Y, Lee Y, Cho J, Baek J, Kim B, Cha Y, et al., editors. (2020). "Towards an appropriate query, key, and value computation for knowledge tracing." in *Proceedings of the seventh ACM conference on learning@ scale*.

38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, CA, USA: Curran Associates Inc. (2017). p. 6000–6010. Available at: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

39. Choi J-W, Yang M, Kim J-W, Shin YM, Shin Y-G, Park S. Prognostic prediction of sepsis patient using transformer with skip connected token for tabular data. *Artif Intell Med*. (2024) 149:102804. doi: 10.1016/j.artmed.2024.102804

40. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. (2006) 63:3–42. doi: 10.1007/s10994-006-6226-1

41. Dou B, Zhu Z, Merkurjev E, Ke L, Chen L, Jiang J, et al. Machine learning methods for small data challenges in molecular science. *Chem Rev*. (2023) 123:8736–80. doi: 10.1021/acs.chemrev.3c00189

42. Tranmer M, Elliot M. Multiple linear regression. *Cathie Marsh Centre Census Surv Res*. (2008) 5:1–5.

43. Nagelkerke NJ. A note on a general definition of the coefficient of determination. *Biometrika*. (1991) 78:691–2. doi: 10.1093/biomet/78.3.691

44. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Heidelberg: Springer (2013).

45. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature. *Geosci Model Dev*. (2014) 7:1247–50. doi: 10.5194/gmd-7-1247-2014

46. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res*. (2005) 30:79–82. doi: 10.3354/cr030079

47. Li C, Chen L, Feng J, Wu D, Wang Z, Liu J, et al. Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator. *IEEE Access*. (2019) 7:110710–21. doi: 10.1109/ACCESS.2019.2934166

48. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci Model Dev Discuss*. (2014) 7:1525–34. doi: 10.5194/gmd-7-1525-2014

49. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast*. (2006) 22:679–88. doi: 10.1016/j.ijforecast.2006.03.001

50. Bastos PG, Sun X, Wagner DP, Wu AW, Knaus WA. Glasgow coma scale score in the evaluation of outcome in the intensive care unit: findings from the acute physiology and chronic health evaluation III study. *Crit Care Med*. (1993) 21:1459–65. doi: 10.1097/00003246-199310000-00012

51. Stone K, Zwiggelaar R, Jones P, Mac PN. A systematic review of the prediction of hospital length of stay: towards a unified framework. *PLoS Dig Health*. (2022) 1:e0000017. doi: 10.1371/journal.pdig.0000017

52. Yu Z, Ashrafi N, Li H, Alaei K, Pishgar M. Prediction of 30-day mortality for ICU patients with Sepsis-3. *BMC Med Inform Decis Mak*. (2024) 24:223. doi: 10.1186/s12911-024-02629-6

53. Peres IT, Hamacher S, Oliveira FLC, Thomé AMT, Bozza FA. What factors predict length of stay in the intensive care unit? Systematic review and meta-analysis. *J Crit Care*. (2020) 60:183–94. doi: 10.1016/j.jcrc.2020.08.003

54. Tobi K, Ekwere I, Ochukpe C. Mechanical ventilation in the intensive care unit: a prospective study of indications and factors that affect outcome in a tertiary hospital in Nigeria. *J Anesth Clin Res*. (2017) 8:2. doi: 10.4172/2155-6148.1000718

55. Yang Z-X, Lv X-L, Yan J. Serum total bilirubin level is associated with hospital mortality rate in adult critically ill patients: a retrospective study. *Front Med*. (2021) 8:697027. doi: 10.3389/fmed.2021.697027

frontiers | Frontiers in Medicine

TYPE Review
PUBLISHED 07 January 2025
DOI 10.3389/fmed.2024.1487271

OPEN ACCESS

EDITED BY
Rahul Kashyap,
WellSpan Health, United States

REVIEWED BY
Priyal Mehta,
Saint Vincent Hospital, United States
Muhammad Daniyal Waheed,
Maroof International Hospital, Pakistan

*CORRESPONDENCE
Senjun Jin
jinsj_2008@163.com

RECEIVED 27 August 2024
ACCEPTED 16 December 2024
PUBLISHED 07 January 2025

CITATION
Tan Z, Li G, Zheng Y, Li Q, Cai W, Tu J and Jin S (2025) Advances in the clinical application of machine learning in acute pancreatitis: a review.
Front. Med. 11:1487271.
doi: 10.3389/fmed.2024.1487271

Advances in the clinical application of machine learning in acute pancreatitis: a review

Zhaowang Tan, Gaoxiang Li, Yueliang Zheng, Qian Li, Wenwei Cai, Jianfeng Tu and Senjun Jin*

Emergency and Critical Care Center, Department of Emergency Medicine, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou, Zhejiang, China

Traditional disease prediction models and scoring systems for acute pancreatitis (AP) are often inadequate in providing concise, reliable, and effective predictions regarding disease progression and prognosis. As a novel interdisciplinary field within artificial intelligence (AI), machine learning (ML) is increasingly being applied to various aspects of AP, including severity assessment, complications, recurrence rates, organ dysfunction, and the timing of surgical intervention. This review focuses on recent advancements in the application of ML models in the context of AP.

KEYWORDS

artificial intelligence, machine-learning model, acute pancreatitis, severity, complications, recurrence, mortality

1 Introduction

Acute pancreatitis (AP) is an inflammatory disorder affecting the parenchyma and peripancreatic tissue, characterized by severe abdominal pain, elevated pancreatic enzymes, and pancreatitis-related changes on abdominal imaging. The incidence of AP has shown a rising trend globally, with an average occurrence rate of 34 cases per 100,000 individuals. Approximately 20% of patients progress to either moderately severe acute pancreatitis (MSAP, accompanied by transient [≤48 h] organ dysfunction and/or local complications such as necrosis of pancreatic or peripancreatic tissue) or severe acute pancreatitis (SAP, accompanied by persistent [>48 h] organ failure), the mortality rate can reach as high as 20–40% (1).

Machine learning (ML) is a category of artificial intelligence tools in which virtual agents learn an optimized set of rules through trial and error—a policy that maximizes expected returns (2). ML has many ideal characteristics that can help with medical decision-making, and these algorithms are able to infer the best decision from suboptimal training sets. ML has been successfully applied to medical problems in the past, such as diabetes and sepsis (3, 4).

Machine learning has demonstrated significant potential in the field of medicine, particularly in disease diagnosis and prognosis. Over the past decade, the utilization of ML algorithms based on databases for acute pancreatitis has become increasingly prevalent.

Numerous studies have employed ML algorithms to forecast AP mortality rates (5), severity (6–8), complications (9), recurrence rates (10), as well as surgical or intervention strategies (7), with ML exhibiting robust reliability in these domains.

In recent years, ML algorithms and the prediction models based on them have generated significant interest among researchers. A growing body of evidence indicates that ML plays a crucial role in predicting acute pancreatitis diagnosis and prognosis. This review aims to offer an overview of the specific applications of ML in AP, with the expectation that artificial intelligence can furnish more evidence-based support for clinical practice in the future.

2 The role of ML in predicting AP mortality

According to the 2012 revision of Atlanta classification (RAC) (11), SAP accompanied by persistent organ failure carries a high mortality rate, ranging from 20 to 40% (1). When complicated by late-stage infections, the mortality rate becomes exceedingly high. Traditional scoring systems for predicting mortality are complex and limited. A systematic review revealed that the positive predictive values of the APACHE II score (AcutePhysiology and Chronic Health Evaluation score, widely used in the classification of critically ill patients and prognosis prediction, which can make a quantitative evaluation of the patient's condition, a higher score indicates a more serious condition, a poorer prognosis, and a higher rate of mortality), Ranson score (one of the earliest scoring systems for predicting the severity of AP and is primarily used to predict the severity of biliary pancreatitis), and Glasgow criteria (emphasis on objective laboratory indicators, including 8 indicators, assessed in 48h of admission to the hospital) were only 69, 63, and 66% respectively (12). Although APACHE II provides the best predictive value for mortality, there is currently no single scoring system that can reliably predict the mortality rate of acute pancreatitis. Therefore, in recent years, numerous early prediction models based on ML algorithms have been developed. These models offer valuable insights for early intervention and potentially reducing the mortality rate of SAP.

Ding et al. initially developed an artificial neural network (ANN) model using the MIMIC-III database, achieving an area under the receiver operating characteristic curve (AUC) of 0.769, which outperformed logistic regression with an AUC of 0.607, Ranson score with 0.652, and SOFA score with 0.401 in predicting in-hospital mortality rate for AP patients (13). The ANN model demonstrated superior overall performance and early-stage risk stratification capability for high-risk AP patients. Building on this, Ren et al. identified 856 AP admitted to the intensive care unit (ICU) from the MIMIC-IV database and developed 9 ML models. Among these, they selected the Gaussian naive Bayes (GNB) model, which demonstrated an AUC, accuracy, sensitivity, and specificity of 0.840, 0.787, 0.839, and 0.792 respectively—making it the most effective among all models tested (14). The GNB model's ability to identify high mortality risk in AP patients admitted to the ICU was further validated using an external database. Similarly, ML models, especially support vector machine (SVM) models, play a crucial role in predicting 28-day all-cause mortality in patients with SAP and analyzing their risk factors (15). The superior attributes of these models compared to traditional scoring systems enhance their effectiveness in early identification of SAP patients and reducing their mortality risk.

However, when it comes to specific causal diagnosis of AP for predicting mortality rates, the predictive capability of gradient boosting machine (GBM) machine learning models appears to be insignificant. Luthra collected 97,027 patients with biliary pancreatitis from the Nationwide Readmission Database over a 4-year period and compared the differences in predicting AP patient mortality between the GBM machine learning model and multivariate logistic regression analysis, finding that the GBM machine learning model had a higher positive predictive value (47.3% vs 35.9%) and lower sensitivity (40.1% vs 46.7%) (16). Therefore, he believes that in a large national database, traditional analysis and GBM machine learning model are comparable and not inferior, and the application of machine learning in managing database-based models for predicting hospital mortality due to common disease states is limited. It is worth noting that after statistical analysis, he found that the inpatient mortality rate of biliary pancreatitis was 0.97%, and hospital stay, age, SAP, patient income quartile, and sepsis were determined as the main predictors of mortality in biliary pancreatitis after it was determined.

3 The role of ML in predicting AP severity

A recent study in Japan showed that the mortality rate of SAP is about 16.7% (17), and early identification and personalized precision treatment can reduce the mortality rate of SAP. Previous studies have shown that precision treatment within 48 h of admission can significantly reduce mortality from SAP (18).

Due to the severity of SAP, high mortality rate, and association with organ failure, early identification and intervention of SAP patients are crucial. However, traditional scoring systems often require more than 24 h to perform and have limited accuracy. To address this, Luo et al. constructed and compared the predictive performance of five different ML models in training and validation cohorts, concluding that the random forest (RF) model performed the best and could be used to guide treatment and improve clinical outcomes (19). The AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the RF model were 0.961, 86.0, 90.0, 81.5, 84.4, and 88.0% in the training cohort, and 0.969, 90.1, 88.6, 91.5, 91.2, and 89.0% in the validation cohort, which were significantly higher than those of other scoring systems (20, 21), the RF model has a higher accuracy in predicting SAP in the early stages of AP (22). Similarly, after developing and comparing different ML prediction models in terms of their effectiveness in predicting the severity of AP, Rahul et al. concluded that the extreme gradient boosting (XGBoost) model showed the best performance in predicting SAP, which can accurately predict SAP at an early stage and provide assistance to clinicians in identifying

and intervening in SAP earlier (8, 23, 24). XGBoost is a machine learning technique that integrates regression tree gradient lifting methods and has gained widespread recognition in the machine learning literature (25–27), data mining challenges, and disease outcome prediction. Given its ability to predict SAP by combining imaging findings and clinical indicators, as well as its capacity to effectively handle missing values commonly encountered in clinical settings (8), early classification and identification of AP can provide valuable guidance for improved integration of medical resources.

4 The role of ML in predicting AP complications

4.1 Organ failure (OF)

Approximately 20% of patients develop organ failure in AP (28), and the presence of persistent organ dysfunction is a key factor in distinguishing between MAP, MSAP and SAP. Once OF occurs, the mortality rate can be as high as 30% (29), while also increasing the risk of infected pancreatic necrosis. Therefore, early identification of AP complicated by OF has a crucial impact on the emergency management of AP patients and plays a vital role in improving survival rates.

Four studies designed ML models to predict OF (9, 30–32). Qiu established models based on SVM, logistic regression (LR), and ANN to predict multiple organ failure (MOF) (30). The area under receiver operating characteristic curve (AUROC) values of these three models were not significantly different at 0.840, 0.832, and 0.834 respectively. Additionally, there was no significant difference in the AUROC compared to the traditional APACHE II score with an AUC value of 0.814 where $P > 0.05$. He believes that the three ML models can all be effective prognostic tools for predicting MOF in MSAP and SAP, and recommends using ANN, which only requires hematocrit, kinetic-time, IL-6, and creatinine as four common parameters. A multicenter cohort study employed complete blood count, serum biochemical markers, and coagulation indicators to develop 6 ML-based algorithm models for predicting MOF (9). Among these, the Adaptive Boosting algorithm (AdaBoost) exhibited superior predictive performance with an AUC of 0.826, sensitivity of 0.805, and specificity of 0.733. IL-6, creatinine, and kinetic time in coagulation indicators were identified as the three most significant independent variables, and monitoring these features can aid in preventing AP-related MOF. Numerous studies have indicated that the conventional use of ANN models is superior to APACHE II scores and LR models in predicting disease severity, MOF, and mortality, and the ANN models can accurately classify 96.2% of patients (31, 33). Lin et al. collected data from 314 Hyperlipidemic acute pancreatitis (HLAP) patients and established LR, NB (Naive Bayes), KNN (K-Nearest Neighbors), DT (Decision Tree) and RF models (32). The AUC values were 0.838, 0.824, 0.853, 0.897, and 0.915 respectively, all significantly higher than those of traditional prediction scoring systems. Among them, the RF model exhibited the highest predictive AUC for OF in HLAP patients with a sensitivity of 0.828 and accuracy of 0.814 among the 5 models tested. They concluded that the RF model outperforms other models as well as clinical scoring systems in predicting the occurrence of OF in HLAP patients and is beneficial for early intervention in high-risk HLAP patients for OF prevention.

AP-related OF mainly involves respiratory, circulatory, and renal failure. Some retrospective clinical analyses have confirmed the role of ML in acute kidney injury (AKI) associated with AP (34–36). Zhang et al. developed an automated machine learning (AutoML) algorithm prediction model that intelligently selects from a range of algorithms and hyperparameters to tailor models for specific datasets (34), enabling early prediction of AKI in AP patients. It demonstrates superior performance compared to traditional LR, requiring less time and achieving higher accuracy, thus significantly improving work efficiency. This warrants its clinical application and promotion. Lin et al. extracted data from the MIMIC-IV database to build a predictive ML model for SAP-AKI using 1,235 cases of SAP patients (35). The models included GBM, GLM, KNN, NB, ANN, RF, and SVM with AUC values of 0.814, 0.812, 0.671, 0.812, 0.688, 0.809 and 0.810 respectively. This highlights the significant role of GBM in predicting SAP-AKI and can assist clinical practitioners in identifying high-risk patients and intervening promptly to reduce mortality rates in intensive care units. It is also worth noting that systemic inflammatory response is inherently associated with the process of AKI and may be caused by local inflammation within renal tissues (37).

Acute respiratory distress syndrome (ARDS) is a common complication of AP, with approximately 30% of SAP patients developing ARDS (38), resulting in a mortality rate of up to 37% (28). Two retrospective analysis studies exploring ML models for AP-associated ARDS have yielded positive results (39, 40), successfully establishing predictive models based on ML. Compared with other models, the Bayesian Classifier (BC) model achieving the highest AUC at 0.891 and demonstrating the best predictive performance (39). The Ensemble Decision Trees (EDT) showed good predictive capabilities, with the highest accuracy (0.891) and precision (0.800). It is noteworthy that lower PaO_2 and Ca^{2+} levels upon admission, as well as elevated CRP, Procalcitonin, Lactic Acid, Neutrophil-Lymphocyte Ratio, White Blood Cell Count, and Amylase levels are significantly associated with an increased risk of developing ARDS in AP patients; among these features, PaO_2 is identified as the most important predictor.

4.2 Sepsis

ML techniques also demonstrate significant advantages in predicting and evaluating septic shock. In a large retrospective cohort study (41), 1,672 AP from the MIMIC III and MIMIC IV databases were selected to construct six ML models, including SVM, KNN, Multilayer Perceptron (MLP), LR, Gradient Boosting Decision Tree (GBDT), and AdaBoost. The GBDT model demonstrated superior performance in predicting sepsis among AP patients with an AUC of 0.985 on the test set, outperforming LR, Systemic Inflammatory Response Syndrome (SIRS) score, Bedside Index for Severity in Acute Pancreatitis (BISAP) score, Sequential Organ Failure Assessment (SOFA) score, quick SOFA

TABLE 1 The clinical application of machine learning in acute pancreatitis.

References	Disease	Sample size	ML-based model	Contrast model	AUC (95%CI)
Ding et al. (13)	Mortality	337	ANN	LR	0.769
Ren et al. (14)	Mortality	856	GNB	XGBoost, RF, SVM, et al.	0.840
Cai et al. (15)	Mortality	534	SVM	LR, XGBoost, RF, et al.	0.877
Anjuli et al. (16)	Mortality	97,027	GBM	LR	0.96
Qiu et al. (30)	Multiple Organ Failure	263	SVM	LR, ANN	0.840
Zhang et al. (34)	Acute kidney Injury	437	AutoML	LR, DL	0.963
Liu et al. (41)	Sepsis	1,672	GBDT	LR, SVM	0.985
Zhang et al. (39)	Acute Respiratory Distress Syndrome	460	BC	SVM, EDTs	0.891
Xia et al. (42)	Septic Shock	604	AE	SVM, RF, AdaBoost, et al.	0.900
Xu et al. (9)	Multiple Organ Failure	455	AdaBoost	LR, et al.	0.826
Lin et al. (32)	Organ Failure	314	RF	LR, RF, et al.	0.915
Lin et al. (32)	Acute kidney Injury	667	GBM	NB, KNN, et al.	0.814
Chen et al. (48)	Recurrence	389	LR	SVM	0.941
Rahul et al. (8)	SAP	61,894	XGBoost	LR, ANN	0.921
Zhou et al. (23)	Severity of AP	441	XGBoost	LR, SVM, DT, RF	0.906
Lan et al. (5)	Surgical Intervention Strategy	223	RF	LR, SVM	0.78
Luo et al. (55)	Surgical Intervention Strategy	15,813	RNN	NA	0.70

ANN, artificial neural networks; LR, logistic regression; DL, deep learning; BC, Bayesian Classifier; EDTs, Ensembles of Decision Trees; AE, auto-encoder; AB, AdaBoost; RF, random forest; NB, naive Bayes; KNN, k-nearest neighbors; XGBoost, extreme gradient boosting model; RF, random forest; RNN, recurrent neural network.

(qSOFA), and APACHE II scores in sepsis prediction. Similarly, another retrospective study data established multiple ML models for early prediction of septic shock in AP with sepsis (42), with the final auto-encoder (AE) model achieving the highest AUC on the validation set (AUC 0.900, accuracy 0.868), while the AUC on the test set was 0.879 and the accuracy was 0.790. The AE model performed better than traditional scoring systems in predicting septic shock in AP with sepsis within 28 days after admission.

5 The role of ML in predicting AP recurrence rate

Recurrent acute pancreatitis (RAP) is defined as a history of at least two episodes of AP with no evidence of pancreatic tissue or functional abnormalities during the remission period. It represents a distinct subtype of pancreatitis, and statistics indicate that 17–22% of diagnosed AP patients will experience recurrence (43). RAP serves as a significant risk factor for the development of chronic pancreatitis (CP), with up to 36% of RAP patients ultimately progressing to CP (44). CP is often accompanied by comorbidities such as diabetes, malnutrition, steatorrhea, and weight loss. Long-term follow-up studies have revealed that 1.3% of CP patients may progress to pancreatic cancer over an 8-year period (45), significantly impacting their quality of life and prognosis. Therefore, early identification and timely intervention for individuals at risk for developing RAP following an episode of acute pancreatitis may mitigate the incidence rates of both RAP and pancreatic cancer while enhancing long-term quality of life.

Radiomics is an emerging field that optimizes existing imaging resources to extract high-throughput quantitative features from medical images (46, 47). These features are further analyzed using predefined algorithms to develop models for clinical decision-making. Currently, radiomics has been widely applied in the precise analysis of tumors and their metastases (46).

ML models based on radiomics research for predicting RAP are currently underutilized in clinical practice. Two retrospective analysis studies have confirmed the role of SVM models in predicting and distinguishing RAP (10, 48). The SVM model demonstrates a significantly higher AUC than traditional clinical models (0.941 vs. 0.712, $p = 0.000$), with similar conclusions observed in the validation dataset (0.929 vs. 0.671, $p = 0.000$) (48). The SVM model constructed using radiomic features can effectively differentiate between patients with functional abdominal pain, RAP, and CP, achieving an overall average accuracy of 82.1%. For patients diagnosed with RAP solely based on symptoms of abdominal pain and laboratory values or those for whom imaging studies during AP episodes are unavailable, radiomics may serve as a valuable diagnostic adjunct (10).

6 The role of ML in predicting AP surgical intervention strategy

Infected pancreatic necrosis (IPN) is the most severe local complication in the late stage of AP. Once IPN occurs, it

indicates SAP, with a mortality rate as high as 30% (49). Treatment often involves a series of surgical debridement procedures known as "Step-up" strategies, including percutaneous catheter drainage (PCD), endoscopic transgastric necrosectomy, video-assisted minimally invasive surgery, and open surgery (50). There has been significant debate regarding the timing of surgical intervention for IPN. Research suggests that early surgery results in a mortality rate exceeding 50% (51), while delaying surgery until 4 weeks after the onset of IPN can reduce both complications and mortality rates (52). With the advancement of modern minimally invasive techniques, early endoscopic drainage during the course of AP has also proven to be safe and effective (53, 54). Early, timely, and accurate prediction of IPN occurrence and determination of the optimal timing for surgical intervention are crucial factors guiding subsequent treatment decisions.

Lan et al. included 223 patients with IPN who underwent surgical treatment for AP (5). They classified IPN patients based on whether the surgery was performed within 4 weeks using LR, SVM, and RF models. The RF model demonstrated a higher classification accuracy (0.80) compared to SVM (0.78) and LR (0.71). Additionally, they identified IL-6, infectious necrosis, fever, and CRP as key factors in determining the timing of surgical intervention for IPN patients. The ML model can effectively predict the optimal timing for surgical intervention in IPN, providing valuable guidance for clinicians in developing personalized surgical strategies for IPN patients.

Another large-scale retrospective clinical study involving 15,813 patients with AP has developed a novel ML model based on recurrent neural network (RNN) to predict the timing of surgical intervention for IPN (55). This model, known as Phased Long Short-Term Memory (Phased-LSTM), achieved an AUC greater than 0.70 and demonstrated stronger interpretability, making it suitable for predicting the optimal timing for surgery. The developed model visualizes specific surgical timings and changes in laboratory indicators from onset to discharge for AP patients, enabling comprehensive monitoring of patients with necrotizing pancreatitis throughout their hospitalization. Due to the ability of LSTM to forget and update long-term states, its performance surpassed that of SVM and RF, highlighting the advantages of time series models in handling temporal data.

7 Discussion

With the improvement in living standards, the incidence of AP has been increasing annually in recent years, with a rise of approximately 2–5% per year (56). Concurrently, the proportion of SAP is also rising. SAP is closely associated with multiple organ failure and has a high mortality and recurrence rate. Once the condition progresses to IPN and OF, the mortality rate can reach up to 30% (1). Therefore, early prediction of the severity of AP, the occurrence of complications, and the timing of intervention is crucial for clinical decision-making and timely intervention. However, traditional clinical prediction models, which are often based on multivariable analysis, are challenging to construct within 24h widespread clinical application. Consequently, it is imperative to develop a simple, effective, and clinically implementable model for early prediction of AP progression.

Artificial Intelligence (AI) encompasses a range of subfields within computer science. In recent years, advancements in algorithms such as ML, statistical learning, deep learning, and cognitive computing have played a pivotal role in the diagnosis and treatment of diseases such as sepsis and cancer (4, 57). ML, a subset of AI, is a burgeoning interdisciplinary field that integrates statistics, computer science, and other areas. It is not only used for text mining and classification in computer science, but is also increasingly applied in clinical practice. Various ML algorithm models for disease prediction and diagnosis have been developed based on AI technologies and are now widely accepted in the medical field. Recently, ML has begun to be applied to areas such as the severity of AP, complications, recurrence rates, organ dysfunction, and the timing of surgical intervention. This review focuses on recent advancements in the application of ML models in the context of AP (refer to Table 1).

We have observed that the majority of current ML models do not account for several important factors, including the etiology of AP and the stratification of severity. Additionally, most of the data utilized are retrospective, although these models have been validated on test and validation sets, their reliability still requires confirmation through clinical practice. Many studies are single-center with small sample sizes and lack external validation. Most research focuses on binary classification of AP into SAP and non-SAP. To date, there have been no ML models that provide accurate prognostication based on the 2012 Atlanta classification, which includes SAP, MSAP, and mild acute pancreatitis (MAP). Notably, most ML models remain limited to predicting traditional severity and complications, with a significant gap in predictive models for recurrence rates, optimal timing for surgery, pancreatic necrosis accumulation, and local complications such as infectious pancreatic necrosis. Future research should address these areas.

8 Conclusion

In conclusion, ML has proven to be an excellent predictor of mortality, severity, complications, recurrence, organ dysfunction, and timing of surgical intervention in acute pancreatitis, and is superior to traditional scoring systems such as the APACHE II score, the BISAP score, the SOFA score, and other traditional systems. However, much more prospective clinical studies are needed to validate this idea.

Author contributions

ZT: Conceptualization, Data curation, Writing – original draft, Writing – review and editing. GL: Investigation, Writing – review and editing. YZ: Formal analysis, Project administration, Writing – review and editing. QL: Software, Validation, Writing – review and editing. WC: Resources, Writing – review and editing. JT: Methodology, Supervision, Writing – review and editing. SJ: Supervision, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Health Science and Technology Plan of Zhejiang Province, China (2023KY051) and the Clinical Research Project of Zhejiang Provincial Administration of Traditional Chinese Medicine, China (2023ZL253).

Acknowledgments

We acknowledge all the doctors and nurses in the Emergency Department of Zhejiang Provincial hospital. And we thank all the participators in this review.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Boxhoorn L, Voermans RP, Bouwense SA, Bruno MJ, Verdonk RC, Boermeester MA, et al. Acute pancreatitis. *Lancet.* (2020) 39610252:726–34.

2. Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. *Artif Intell Med.* (2013) 571:9–19. doi: 10.1016/j.artmed.2012.12.003

3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 31622:2402–10.

4. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med.* (2018) 2411:1716–20.

5. Lan L, Guo Q, Zhang Z, Zhao W, Yang X, Lu H, et al. Classification of infected necrotizing pancreatitis for surgery within or beyond 4 weeks using machine learning. *Front Bioeng Biotechnol.* (2020) 8:541.

6. Kiss S, Pinter J, Molontay R, Nagy M, Farkas N, Sipos Z, et al. Hungarian pancreatic study G: early prediction of acute necrotizing pancreatitis by artificial intelligence: a prospective cohort-analysis of 2387 cases. *Sci Rep.* (2022) 121:7827. doi: 10.1038/s41598-022-11517-w

7. Shi N, Lan L, Luo J, Zhu P, Ward TRW, Szatmary P, et al. Predicting the need for therapeutic intervention and mortality in acute pancreatitis: a two-center international study using machine learning. *J Pers Med.* (2022) 12:616. doi: 10.3390/jpm12040616

8. Thapa R, Iqbal Z, Garikipati A, Siefkas A, Hoffman J, Mao Q, et al. Early prediction of severe acute pancreatitis using machine learning. *Pancreatology.* (2022) 221:43–50.

9. Xu F, Chen X, Li C, Liu J, Qiu Q, He M, et al. Prediction of multiple organ failure complicated by moderately severe or severe acute pancreatitis based on machine learning: a multicenter cohort study. *Mediators Inflamm.* (2021) 2021:5525118.

10. Mashayekhi R, Parekh VS, Faghih M, Singh VK, Jacobs MA, Zaheer A. Radiomic features of the pancreas on CT imaging accurately differentiate functional abdominal pain, recurrent acute pancreatitis, and chronic pancreatitis. *Eur J Radiol.* (2020) 123:108778. doi: 10.1016/j.ejrad.2019.108778

11. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Acute pancreatitis classification working G: classification of acute pancreatitis–2012: revision of the Atlanta classification and definitions by international consensus. *Gut.* (2013) 621:102–11. doi: 10.1136/gutjnl-2012-302779

12. Gravante G, Garcea G, Ong SL, Metcalfe MS, Berry DP, Lloyd DM, et al. Prediction of mortality in acute pancreatitis: a systematic review of the published evidence. *Pancreatology.* (2009) 95:601–14.

13. Ding N, Guo C, Li C, Zhou Y, Chai X. An artificial neural networks model for early predicting in-hospital mortality in acute pancreatitis in MIMIC-III. *Biomed Res Int.* (2021) 2021:6638919. doi: 10.1155/2021/6638919

14. Ren W, Zou K, Huang S, Xu H, Zhang W, Shi X, et al. Prediction of in-hospital mortality of intensive care unit patients with acute pancreatitis based on an explainable machine learning algorithm. *J Clin Gastroenterol.* (2024) 586:619–26. doi: 10.1097/MCG.0000000000001910

15. Cai W, Wu X, Chen Y, Chen J, Lin X. Risk factors and prediction of 28-day-all cause mortality among critically ill patients with acute pancreatitis using machine learning techniques: a retrospective analysis of multi-institutions. *J Inflamm Res.* (2024) 17:4611–23. doi: 10.2147/JIR.S463701

16. Luthra AK, Porter K, Hinton A, Chao WL, Papachristou GI, Conwell DL, et al. A comparison of machine learning methods and conventional logistic regression for the prediction of in-hospital mortality in acute biliary pancreatitis. *Pancreas.* (2022) 5110:1292–9. doi: 10.1097/MPA.0000000000002208

17. Yasuda H, Horibe M, Sanui M, Sasaki M, Suzuki N, Sawano H, et al. Etiology and mortality in severe acute pancreatitis: a multicenter study in Japan. *Pancreatology.* (2020) 203:307–17.

18. Petrov MS, Pylypchuk RD, Uchugina AF. A systematic review on the timing of artificial nutrition in acute pancreatitis. *Br J Nutr.* (2009) 1016:787–93. doi: 10.1017/S0007114508123443

19. Luo Z, Shi J, Fang Y, Pei S, Lu Y, Zhang R, et al. Development and evaluation of machine learning models and nomogram for the prediction of severe acute pancreatitis. *J Gastroenterol Hepatol.* (2023) 383:468–75. doi: 10.1111/jgh.16125

20. Valverde-Lopez F, Matas-Cobos AM, Alegria-Motte C, Jimenez-Rosales R, Ubeda-Munoz M, Redondo-Cerezo E. BISAP, RANSON, lactate and others biomarkers in prediction of severe acute pancreatitis in a European cohort. *J Gastroenterol Hepatol.* (2017) 329:1649–56. doi: 10.1111/jgh.13763

21. Pando E, Alberti P, Mata R, Gomez MJ, Vidal L, Cirera A, et al. Early changes in blood urea Nitrogen (BUN) can predict mortality in acute pancreatitis: comparative study between BISAP score, APACHE-II, and other laboratory markers-A prospective observational study. *Can J Gastroenterol Hepatol.* (2021) 2021:6643595. doi: 10.1155/2021/6643595

22. Yuan L, Ji M, Wang S, Wen X, Huang P, Shen L, et al. Machine learning model identifies aggressive acute pancreatitis within 48 h of admission: a large retrospective study. *BMC Med Inform Decis Mak.* (2022) 221:312. doi: 10.1186/s12911-022-02066-3

23. Zhou Y, Han F, Shi XL, Zhang JX, Li GY, Yuan CC, et al. Prediction of the severity of acute pancreatitis using machine learning models. *Postgrad Med.* (2022) 1347:703–10.

24. Zhao Y, Wei J, Xiao B, Wang L, Jiang X, Zhu Y, et al. Early prediction of acute pancreatitis severity based on changes in pancreatic and peripancreatic computed tomography radiomics nomogram. *Quant Imaging Med Surg.* (2023) 133:1927–36. doi: 10.21037/qims-22-821

25. Li Q, Yang H, Wang P, Liu X, Lv K, Ye M. XGBoost-based and tumor-immune characterized gene signature for the prediction of metastatic status in breast cancer. *J Transl Med.* (2022) 201:177. doi: 10.1186/s12967-022-03369-9

26. Lu D, Peng J, Wang Z, Sun Y, Zhai J, Wang Z, et al. Dielectric property measurements for the rapid differentiation of thoracic lymph nodes using XGBoost in patients with non-small cell lung cancer: a self-control clinical trial. *Transl Lung Cancer Res.* (2022) 113:342–56. doi: 10.21037/tlcr-22-92

27. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med.* (2020) 181:462. doi: 10.1186/s12967-020-02620-5

28. Schepers NJ, Bakker OJ, Besselink MG, Ahmed Ali U, Bollen TL, Gooszen HG, et al. Dutch pancreatitis study G: impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. *Gut.* (2019) 686:1044–51. doi: 10.1136/gutjnl-2017-314657

29. Petrov MS, Shanbhag S, Chakraborty M, Phillips AR, Windsor JA. Organ failure and infection of pancreatic necrosis as determinants of mortality in patients with acute pancreatitis. *Gastroenterology.* (2010) 1393:813–20.

30. Qiu Q, Nian YJ, Guo Y, Tang L, Lu N, Wen LZ, et al. Development and validation of three machine-learning models for predicting multiple organ failure in moderately severe and severe acute pancreatitis. *BMC Gastroenterol.* (2019) 191:118.

31. Hong WD, Chen XR, Jin SQ, Huang QK, Zhu QH, Pan JY. Use of an artificial neural network to predict persistent organ failure in patients with acute pancreatitis. *Clinics (Sao Paulo).* (2013) 681:27–31.

32. Lin W, Huang Y, Zhu J, Sun H, Su N, Pan J, et al. Machine learning improves early prediction of organ failure in hyperlipidemia acute pancreatitis using clinical and abdominal CT features. *Pancreatology.* (2024) 243:350–6. doi: 10.1016/j.pan.2024.02.003

33. Mofidi R, Duff MD, Madhavan KK, Garden OJ, Parks RW. Identification of severe acute pancreatitis using an artificial neural network. *Surgery.* (2007) 1411:59–66.

34. Zhang R, Yin M, Jiang A, Zhang S, Xu X, Liu L. Automated machine learning for early prediction of acute kidney injury in acute pancreatitis. *BMC Med Inform Decis Mak.* (2024) 241:16.

35. Lin S, Lu W, Wang T, Wang Y, Leng X, Chi L, et al. Predictive model of acute kidney injury in critically ill patients with acute pancreatitis: a machine learning approach using the MIMIC-IV database. *Ren Fail.* (2024) 461:2303395. doi: 10.1080/0886022X.2024.2303395

36. Yang D, Zhao L, Kang J, Wen C, Li Y, Ren Y, et al. Development and validation of a predictive model for acute kidney injury in patients with moderately severe and severe acute pancreatitis. *Clin Exp Nephrol.* (2022) 268:770–87.

37. Yilmaz H, Cakmak M, Inan O, Darcin T, Akcay A. Can neutrophil-lymphocyte ratio be independent risk factor for predicting acute kidney injury in patients with severe sepsis? *Ren Fail.* (2015) 372:225–9.

38. Fei Y, Gao K, Li WQ. Prediction and evaluation of the severity of acute respiratory distress syndrome following severe acute pancreatitis using an artificial neural network algorithm model. *HPB (Oxford).* (2019) 217:891–7.

39. Zhang M, Pang M. Early prediction of acute respiratory distress syndrome complicated by acute pancreatitis based on four machine learning models. *Clinics (Sao Paulo).* (2023) 78:100215. doi: 10.1016/j.clinsp.2023.100215

40. Yang D, Kang J, Li Y, Wen C, Yang S, Ren Y, et al. Development of a predictive nomogram for acute respiratory distress syndrome in patients with acute pancreatitis complicated with acute kidney injury. *Ren Fail.* (2023) 452:2251591. doi: 10.1080/0886022X.2023.2251591

41. Liu F, Yao J, Liu C, Shou S. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis. *BMC Surg.* (2023) 231:267.

42. Xia Y, Long H, Lai Q, Zhou Y. Machine learning predictive model for septic shock in acute pancreatitis with sepsis. *J Inflamm Res.* (2024) 17:1443–52.

43. Sankaran SJ, Xiao AY, Wu LM, Windsor JA, Forsmark CE, Petrov MS. Frequency of progression from acute to chronic pancreatitis and risk factors: a meta-analysis. *Gastroenterology.* (2015) 1496:1490–1500.e1491.

44. Ahmed Ali U, Issa Y, Hagenaars JC, Bakker OJ, Van Goor H, Nieuwenhuijs VB, et al. Dutch pancreatitis study G: risk of recurrent pancreatitis and progression to chronic pancreatitis after a first episode of acute pancreatitis. *Clin Gastroenterol Hepatol.* (2016) 145:738–46. doi: 10.1186/s13054-016-1208-6

45. Hao L, Zeng XP, Xin L, Wang D, Pan J, Bi YW, et al. Incidence of and risk factors for pancreatic cancer in chronic pancreatitis: a cohort of 1656 patients. *Dig Liver Dis.* (2017) 4911:1249–56. doi: 10.1016/j.dld.2017.07.001

46. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures. They are data. *Radiology.* (2016) 2782:563–77.

47. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 484:441–6.

48. Chen Y, Chen TW, Wu CQ, Lin Q, Hu R, Xie CL, et al. Radiomics model of contrast-enhanced computed tomography for predicting the recurrence of acute pancreatitis. *Eur Radiol.* (2019) 298:4408–17. doi: 10.1007/s00330-018-5824-1

49. De-Madaria E, Buxbaum JL. Advances in the management of acute pancreatitis. *Nat Rev Gastroenterol Hepatol.* (2023) 2011:691–2.

50. Hollemans RA, Bakker OJ, Boermeester MA, Bollen TL, Bosscha K, Bruno MJ, et al. Dutch pancreatitis study G: superiority of step-up approach vs open necrosectomy in long-term follow-up of patients with necrotizing pancreatitis. *Gastroenterology.* (2019) 1564:1016–26. doi: 10.1053/j.gastro.2018.10.045

51. Mowery NT, Bruns BR, Macnew HG, Agarwal S, Enniss TM, Khan M, et al. Surgical management of pancreatic necrosis: a practice management guideline from the Eastern Association for the Surgery of Trauma. *J Trauma Acute Care Surg.* (2017) 832:316–27. doi: 10.1097/TA.0000000000001510

52. Leppaniemi A, Tolonen M, Tarasconi A, Segovia-Lohse H, Gamberini E, Kirkpatrick AW, et al. 2019 WSES guidelines for the management of severe acute pancreatitis. *World J Emerg Surg.* (2019) 14:27.

53. Van Grinsven J, Van Santvoort HC, Boermeester MA, Dejong CH, Van Eijck CH, Fockens P, et al. Dutch pancreatitis study G: timing of catheter drainage in infected necrotizing pancreatitis. *Nat Rev Gastroenterol Hepatol.* (2016) 135:306–12.

54. Huang D, Li Q, Lu Z, Jiang K, Wu J, Gao W, et al. From "step-up" to "step-jump": a leap-forward intervention for infected necrotizing pancreatitis. *Chin Med J (Engl).* (2021) 1353:285–7. doi: 10.1097/CM9.0000000000001877

55. Luo J, Lan L, Peng L, Li M, Lu H, Yang D, et al. Predicting timing of surgical intervention using recurrent neural network for necrotizing pancreatitis. *IEEE Access.* (2020) 8:207905-207913.

56. Tenner S, Vege SS, Sheth SG, Sauer B, Yang A, Conwell DL, et al. American college of gastroenterology guidelines: management of acute pancreatitis. *Am J Gastroenterol.* (2024) 1193:419–37.

57. Liu Y, Chen PC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. *JAMA.* (2019) 32218:1806–16. doi: 10.1001/jama.2019.16489

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 11 April 2025
DOI 10.3389/fmed.2025.1501025

OPEN ACCESS

EDITED BY
Qinghe Meng,
Upstate Medical University, United States

REVIEWED BY
Nurettin Oner,
Ankara University, Türkiye
Hongsheng Wu,
Southern Medical University, China

*CORRESPONDENCE
Yongping You
✉ YYPL9@njmu.edu.cn
Liqing Bi
✉ biliqing@163.com

RECEIVED 24 September 2024
ACCEPTED 14 February 2025
PUBLISHED 11 April 2025

CITATION
Pan J, Yue Z, Ji J, You Y, Bi L, Liu Y, Xiong X, Gu G, Chen M and Zhang S (2025) Predicting nosocomial pneumonia of patients with acute brain injury in intensive care unit using machine-learning models.
Front. Med. 12:1501025.
doi: 10.3389/fmed.2025.1501025

Predicting nosocomial pneumonia of patients with acute brain injury in intensive care unit using machine-learning models

Junchen Pan[1], Zhen Yue[2], Jing Ji[2], Yongping You[2]*, Liqing Bi[2]*, Yun Liu[3], Xinglin Xiong[4], Genying Gu[1], Ming Chen[5] and Shen Zhang[6]

[1]Department of Neurosurgery, The Affiliated BenQ Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [2]Department of Neurosurgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [3]Department of Information, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [4]Department of Nursing, The Affiliated BenQ Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [5]Department of Rehabilitation, The Affiliated BenQ Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [6]Engineering Research Center of Health Service System Based on Ubiquitous Wireless Networks, Ministry of Education, Nanjing, China

Introduction: The aim of this study is to construct and validate new machine learning models to predict pneumonia events in intensive care unit (ICU) patients with acute brain injury.

Methods: Acute brain injury patients in ICU of hospitals from January 1, 2020, to December 31, 2021 were retrospective reviewed. Patients were divided into training, and validation sets. The primary outcome was nosocomial pneumonia infection during ICU stay. Machine learning models including XGBoost, DecisionTree, Random Forest, Light GBM, Adaptive Boost, BP, and TabNet were used for model derivation. The predictive value of each model was evaluated using accuracy, precision, recall, F1-score, and area under the curve (AUC), and internal and external validation was performed.

Results: A total of 280 ICU patients with acute brain injury were included. Five independent variables for nosocomial pneumonia infection were identified and selected for machine learning model derivations and validations, including tracheotomy time, antibiotic use days, blood glucose, ventilator-assisted ventilation time, and C-reactive protein. The training set revealed the superior and robust performance of the XGBoost with the highest AUC value (0.956), while the Random Forest and Adaptive Boost had the highest AUC value (0.883) in validation set.

Conclusion: Machine learning models can effectively predict the risk of nosocomial pneumonia infection in patients with acute brain injury in the ICU. Despite differences in populations and algorithms, the models we constructed demonstrated reliable predictive performance.

KEYWORDS

intensive care unit, area under the curve, acute brain injury, nosocomial pneumonia, machine-learning models

Introduction

Acute brain injury (ABI) encompasses a range of neurological disorders that can result in acute functional deficits, including ischemic or hemorrhagic stroke, subarachnoid hemorrhage from aneurysms, and traumatic brain injury, causing approximately 12 million deaths annually (1, 2). Due to the uncertain long-term functional prognosis of ABI, patients often have prolonged hospital stays, and the majority require endotracheal intubation for airway protection, along with mechanical ventilation and intracranial pressure reduction. Additionally, ABI is associated with immune system alterations mediated through inflammation and the autonomic nervous system, which may increase susceptibility to infections during and after hospitalization (3–5). In ABI patients who suffer from nosocomial infections, a common source of infection is the respiratory system, including ventilator-associated pneumonia and hospital-acquired pneumonia. Studies have shown that nosocomial pneumonia significantly prolongs patients' hospital stay and increases mortality and disability rates (3, 6).

Studies have indicated a significant association between the severity of ABI, chest trauma, smoking history, substance abuse, as well as interventions such as transfusion, sedation, and the need for tracheostomy, with the risk of nosocomial pneumonia (7–11). Given that nosocomial pneumonia can be considered a continuum of a single disease, describing its epidemiology and influencing factors is clinically valuable (12, 13). This can aid in better formulating preventive and management measures for nosocomial pneumonia in clinical practice. Existing pneumonia prediction scoring tools in clinical practice, such as CPIS score, A2DS2 score, and AIS-APS score, have certain limitations, primarily due to underutilization of classification information, leading to information loss (14–16).

Recently, artificial intelligence (AI) has rapidly developed in the field of medicine, with machine learning being the most widely used AI method. Machine learning models generate personalized probabilities of events for patients. Additionally, ML models can capture complex non-linear relationships in medical data, making full use of clinical information (17). However, there is currently no effective tool for rapidly predicting the occurrence of nosocomial pneumonia in ICU patients with ABI. This study aims to collect factors related to the occurrence of nosocomial pneumonia in neurosurgical ICU patients with ABI through retrospective analysis, and ultimately construct a predictive model for nosocomial pneumonia in ICU patients with ABI.

Materials and methods

Study setting

The patients admitted to the Neurosurgery Departments of Jiangsu Provincial People's Hospital and Benq Hospital from January 1, 2020, to December 31, 2021 were selected as the training set. To validate the model's generalization ability, patients from three healthcare systems, including Suqian First People's Hospital, Nanjing Jiangning Hospital, and Jiangsu Provincial People's Hospital, were selected as the external validation set. This study obtained approval from the Institutional Review Board of the research center. Given the retrospective design of this study, the requirement for obtaining informed consent from patients was waived. This study was reported in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline (18).

Inclusion and exclusion criteria

Patients were included if they met the following criteria: (1) aged between 18 and 80 years old; (2) admitted to the Neurosurgery ICU with ABI, including cerebral hemorrhage, trauma, vascular disease, or tumor; (3) hospital stay longer than 48 h; (4) received standard treatment during hospitalization, including surgical and conservative treatment. Patients with the following characteristics were excluded: (1) occurrence of pulmonary infection within 48 h of ICU admission or hospitalization; (2) patients who did not receive standardized postoperative treatment; (3) patients with immunodeficiency diseases; (4) patients with severe chronic heart, lung, kidney, or other organ diseases. Severe chronic heart disease included end-stage heart failure, severe cardiomyopathies, and complex congenital heart diseases that demand continuous medical support and substantially affect the patient's physiological function and prognosis; (5) patients with concomitant malignant tumors in other organs; (6) patients for whom data were unavailable.

Data collection

Collect demographic characteristics, vital signs, ventilator parameters, sputum culture results, blood test results, imaging findings, treatment, and prognosis data from the hospital electronic medical record system for ICU patients with acute brain injury, encompassing 61 dimensions. Nosocomial infections are defined as those that occur more than 48 h after hospital admission, and the diagnosed pneumonia based on clinical presentation, blood-related examinations, vital signs, radiological examinations, and sputum culture.

Statistical analysis

We used SPSS 26.0 and Python 3.7 for data analysis. Data cleaning was performed, including the removal of variables not included in the statistical analysis, missing values (with a missing rate exceeding 80%), and continuous values containing character groups. Subsequently, duplicate data entries were removed, and the cleaned data were exported. For variables with a missing rate of less than 80%, we used the Multivariate Imputation by Chained Equations (MICE) algorithm for imputation. MICE is a multiple imputation method that iteratively models each variable with missing values as a function of the other variables in the dataset. This approach allows us to capture the complex relationships between variables, which is essential in our medical dataset. By using MICE, we aimed to generate more accurate imputations and preserve the data's underlying structure. After imputation, the dataset was further processed for subsequent analysis. The training set were randomly divided into training subset and internal validation subset at a ratio of 7:3. Out of the total 280 patients, 196 patients were assigned to the training set, and 84 patients were

TABLE 1 The baseline characteristics of included patients.

Variables	Nosocomial pneumonia		*P*-value
	Yes (*n* = 67)	No (*n* = 213)	
Age	60.15	52.16	0.899
Sex (%)			0.176
Male	44	158	
Female	23	55	
Smoking history			0.046
Yes	16	79	
No	51	134	
Diabetes			<0.001
Yes	14	12	
No	53	201	
Heart disease			0.208
Yes	5	8	
No	62	205	
Stroke			0.012
Yes	8	8	
No	59	205	
Chemotherapy or immunosuppressive therapies			0.074
Yes	1	0	
No	66	213	

included in the validation set. Continuous variables that followed a normal distribution were presented as mean ± standard deviation, while those not following a normal distribution were presented as median and interquartile range. Categorical variables were presented as counts and proportions. Independent sample t-tests were used to compare normally distributed continuous variables, Mann–Whitney *U* tests were used for non-normally distributed continuous variables, and chi-square tests were used for comparisons between categorical variables. A p-value <0.05 was considered statistically significant.

Subsequently, the selected variables were used to build seven machine learning models, including the XGBoost model, DecisionTree model, Random Forest model, Light GBM model, Adaptive Boost model, BP model, and TabNet model. Selecting these models owing to their unique advantages in handling complex non-linear relationships within medical data. Ensemble models are well-equipped to capture interactions among multiple variables, which is crucial when predicting nosocomial pneumonia in ICU patients with ABI. Receiver operating characteristic (ROC) curves were plotted, and the area under the curve (AUC) was calculated. Furthermore, the performance of each model was evaluated by comparing their accuracy, precision, recall, and F1-score on both the internal validation subset and external validation set.

SPSS 26.0 was employed for initial data exploration and basic statistical tests, while Python 3.7 was used for data cleaning, machine-learning model construction, and performance evaluation. The *pandas* library in Python was crucial for data manipulation tasks. For building the machine-learning models, we relied on several libraries. The *scikit-learn* library was extensively used. The XGBoost model was implemented using the *xgboost* library, which offers efficient algorithms for gradient-boosting. The DecisionTree model was built using the *DecisionTreeClassifier* class from *scikit-learn*, and the Random Forest model was created using the *RandomForestClassifier* class in the same library. The Adaptive Boost model was implemented using the *AdaBoostClassifier* class from *scikit-learn*. The Light GBM model was built with the *lightgbm* library. The BP model was implemented using the *Keras* library, which is a high-level neural network API running on top of *TensorFlow*. The TabNet model was created using the *pytorch-tabnet* library designed for tabular data. For performance evaluation, functions from *scikit-learn* were used to assess performance evaluation.

Results

This study included a total of 280 ABI patients, and the baseline characteristics of the patients are shown in Table 1. A total of 67 patients with nosocomial pneumonia, while the remaining 213 patients without nosocomial pneumonia. There were no significant differences between groups for age ($p = 0.899$), sex ($p = 0.176$), history of heart disease ($p = 0.208$), and patients received chemotherapy or immunosuppressive therapies ($p = 0.074$). However, we noted significant differences between groups for smoking history ($p = 0.046$), diabetes ($p < 0.001$), and stroke history ($p = 0.012$).

A total of 2,953 data points were collected from 280 patients, including 24 qualitative features and 35 quantitative features (Appendix). Fifty-nine variables were analyzed through correlation heatmap analysis and random forest-based feature selection. The most significant predictors were subsequently entered into multiple logistic

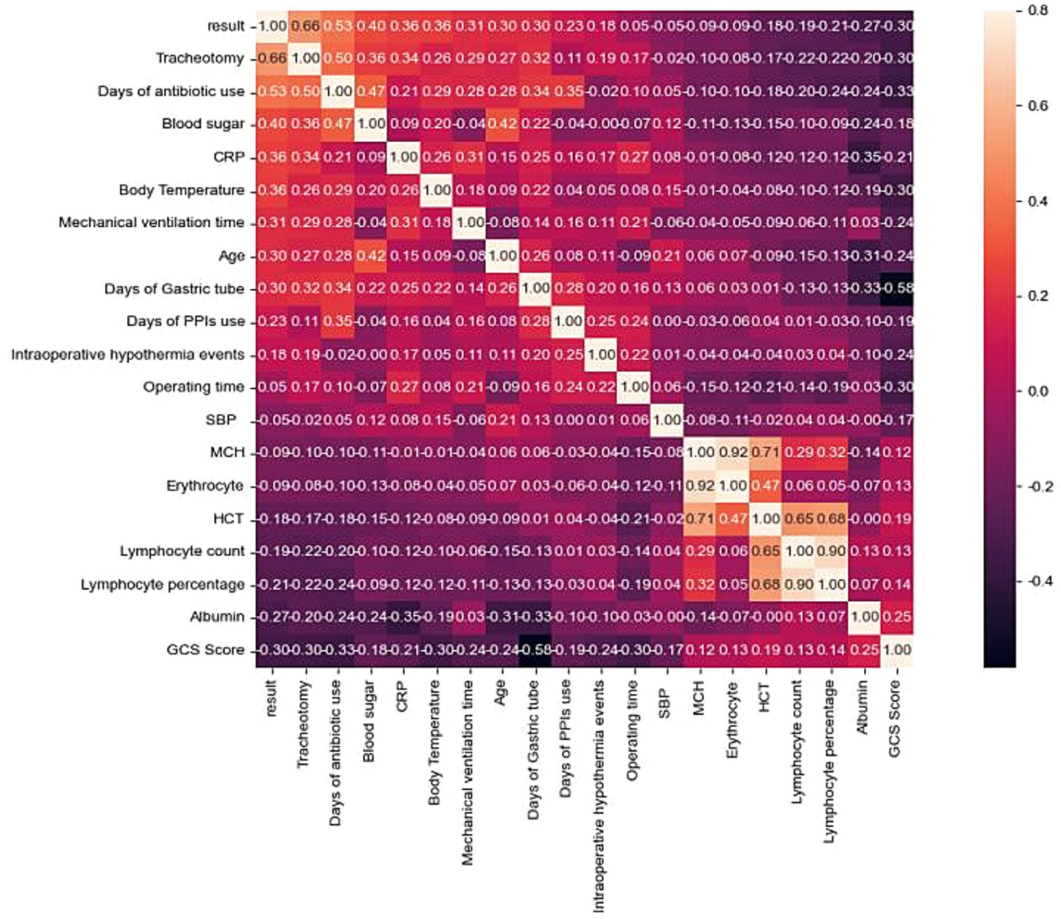

FIGURE 1
Pearson correlation heatmap between variable.

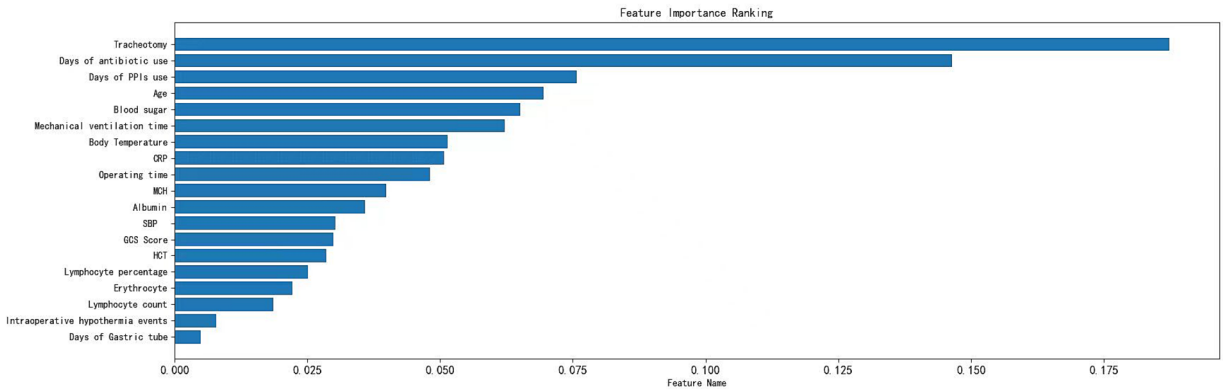

FIGURE 2
Variable importance of features included in the machine learning algorithm for prediction of nosocomial pneumonia in ABI patients.

regression modeling. Pearson correlation heatmap analysis was used for correlation analysis, revealing that the duration of tracheostomy was the most significant associated factor for pneumonia occurrence in ABI patients (Figure 1). In addition, the results of the Light Gradient Boosting Machine algorithm showed that the top 20 important indicators were tracheostomy duration, duration of antibiotic use, blood sugar, age, duration of mechanical ventilation, CRP, GCS score, body temperature, lymphocyte percentage, lymphocyte count, gastric tube, albumin, PPi, intraoperative hypothermia, total protein, erythrocyte count, average hemoglobin concentration, hematocrit, operation time, and mean corpuscular volume (Figure 2).

Through the cross-validation accuracy curve, it can be observed that when the number of features is 5. It should be noted that while the out-of-Bag (OOB) error score was relatively high at this point, our objective was to find the optimal number of features for overall model performance. The OOB error is just one of the many factors to consider. By examining the OOB error in combination with other performance metrics such as accuracy, precision, recall, F1-score, and AUC, we aimed to select a feature set that would provide the best generalization and predictive ability for the model. A slightly higher OOB error might be tolerated if the model shows superior performance in other aspects, which is crucial for practical application in predicting nosocomial pneumonia in ICU patients with ABI. Employing a stepwise method, with 5 features including tracheostomy time, duration of antibiotic use, blood sugar level, duration of mechanical ventilation, and CRP were applied to establish prediction model (Figure 3). Among the seven machine-learning algorithms evaluated, the XGBoost and Light GBM models demonstrated relatively high AUC values, indicating their strong discriminatory power in predicting nosocomial pneumonia in ICU patients with ABI (Table 2 and Figure 4). The external validation set found XGBoost showed highest precision (0.96), while Random Forest and Adaptive Boost models showed highest Light GBM (AUC: 0.883) (Table 2 and Figure 5). These high-performing models based on AUC values can potentially play a crucial role in clinical decision-making. A model with a high AUC can assist clinicians in early identification of patients at high risk of nosocomial pneumonia, enabling timely intervention and potentially improving patient outcomes.

Model visualization

Leveraging an XGBoost-based diagnostic model, our data platform conducts daily predictions of infection risk (Figure 6). Although we cannot provide a direct link to the platform due to privacy and security reasons, the platform functions in a manner similar to the interactive capabilities of the '*shiny*' package in R software. Users input the five key parameters: tracheostomy time, duration of antibiotic use, blood glucose levels, the length of time on ventilator support, and the CRP value. The platform then processes these inputs using the underlying XGBoost-based algorithm. The output is presented in a dynamic way. A pie chart shows the real – time distribution of incidence probabilities, giving users an immediate understanding of the patient's risk status. Additionally, a trend chart is available for each patient, which can be used to track the progression of the predicted risk over time. This visualization aims to assist clinicians in making more informed decisions regarding patient care.

Discussion

Currently, clinicians typically rely on clinical, radiological, and laboratory indicators to diagnose pneumonia and initiate empirical antibiotic therapy (19). However, early pneumonia is insensitive, radiological radiation causes some harm to patients, and clinical diagnosis of pathogenic microorganisms has some lag. Existing scoring systems do not cover all individual factors of patients and have low accuracy. Considering the impracticality of existing prediction models and the flexibility of machine learning methods in selecting predictive variables and transformation algorithms. This study employed machine learning to construct a predictive model for nosocomial pneumonia infection in ABI patients in the ICU. The study included 280 patients from multiple hospitals for model construction and validation. The results revealed that among the machine learning models constructed

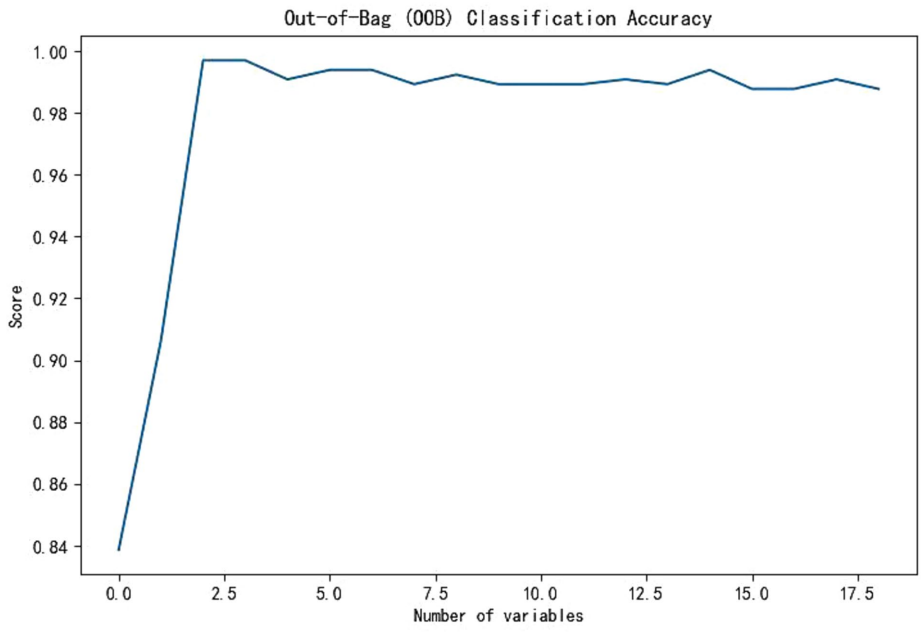

FIGURE 3
Accuracy chart of out-of-pocket data.

TABLE 2 The predictive performance among the constructed models.

Algorithmic	Training set				Validation set			
	Precision	Recall rate	F1-score	AUC	Precision	Recall rate	F1-score	AUC
XGBoost	0.98	0.96	0.97	0.956	0.96	0.96	0.96	0.820
Decision tree	0.91	0.88	0.89	0.878	0.92	0.92	0.92	0.700
Random Forest	0.96	0.92	0.94	0.918	0.94	0.94	0.94	0.883
Light GBM	0.98	0.95	0.96	0.950	0.95	0.95	0.96	0.758
Adaptive Boost	0.96	0.91	0.93	0.913	0.93	0.93	0.93	0.883
BP network	0.86	0.81	0.83	0.815	0.84	0.84	0.85	0.758
TabNet	0.90	0.84	0.86	0.841	0.88	0.88	0.89	0.674

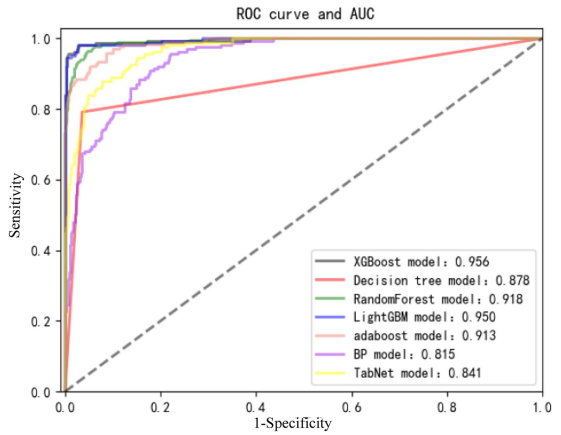

FIGURE 4
Receiver operating characteristic (ROC) curve and AUC among the 7 algorithm models.

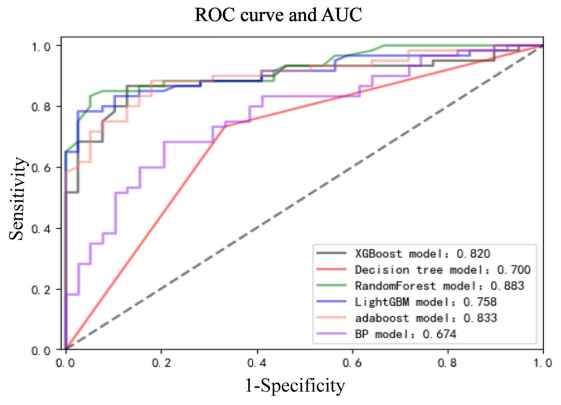

FIGURE 5
Receiver operating characteristic curve and AUC among the external validation set.

based on tracheostomy time, duration of antibiotic use, blood sugar level, duration of mechanical ventilation, and CRP, the XGBoost model exhibited the best overall performance. It achieved a precision of 98% and an AUC of 95.6% in predicting postoperative pneumonia in ABI patients. Similarly, in the external validation set, we observed the highest precision with the XGBoost model.

Several studies have already constructed a prediction model for pneumonia in patients with brain injury using a machine learning approach (20, 21). Zheng et al. (20) identified 468 patients with spontaneous intracerebral hemorrhage (sICH) and identified six independent variables, including nasogastric feeding, airway support, unconscious onset, surgery for external ventricular drainage, larger sICH volume, and ICU stay, and the prediction model constructed based on these variables could effectively predict stroke-associated pneumonia in patients with sICH. Lee et al. (21) identified 5,754 hospitalized stroke patients, and found random survival forest model showed superior discriminative ability for predicting post-stroke pneumonia. However, there are currently no studies that have constructed predictive models for nosocomial pneumonia infection in ABI populations in the ICU. Therefore, we adopted a machine learning approach to construct a predictive model for nosocomial pneumonia infection in ABI populations in the ICU, which can manage missing information without the need for imputation or preprocessing and has strong clinical applicability.

To further evaluate the performance of our proposed machine-learning models, we compared them with existing clinical tools, namely CPIS, A2DS2, and AIS-APS scores. With a cut-off value of CPIS ≥ 3, in critically ill patients, the AUC of CPIS for predicting ventilator-associated pneumonia was found to be 59% (22). The AUC of the A2DS2 model for predicting stroke-associated pneumonia was 85% (23). At the same time, the AUC of the AIS-APS score for predicting ischemic stroke-associated pneumonia was 87% (24). The optimal model constructed in this study had a higher AUC than the previous prediction models in both the training set and the validation set, highlighting the potential advantages of our machine-learning-based approach in predicting nosocomial pneumonia in ICU patients with ABI.

The constructed prediction model based on tracheostomy time, duration of antibiotic use, blood sugar level, duration of mechanical ventilation, and CRP showed high predictive performance. The potential reasons for this could explained by: (1) long-term endotracheal intubation and mechanical ventilation may lead to respiratory mucosal injury and inflammatory responses, thereby affecting local immune function, weakening the clearance ability against pathogenic microorganisms, and increasing the likelihood of infection occurrence (25); (2) long-term use of antibiotics may lead to bacterial resistance to drugs and cause a series of adverse reactions, such as disruption of intestinal flora balance, liver and kidney damage,

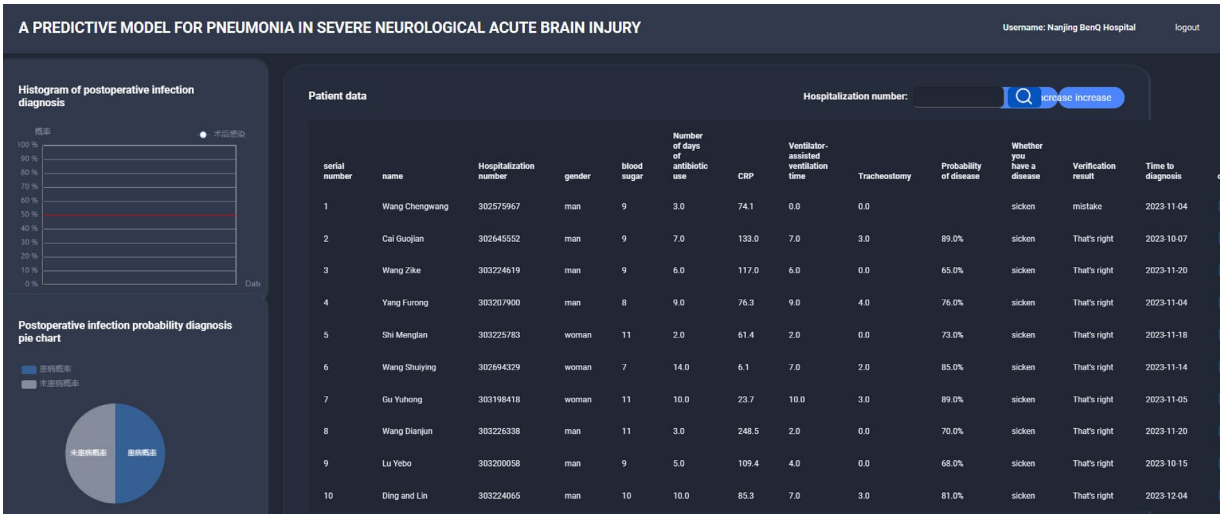

FIGURE 6
The diagnosis model of XGBoost is used to predict the infection risk.

thereby increasing the risk of infection (26); (3) high blood sugar may lead to immune suppression and exacerbate inflammation, which can reduce the body's ability to clear pathogens, trigger the release of inflammatory mediators, increase inflammation in lung tissues, and thus provide an optimal environment for bacterial infection (27); (4) various factors related to mechanical ventilation, such as positive end-expiratory pressure (PEEP), tidal volume, and FiO2 levels, can contribute to lung injury and inflammation, further predisposing patients to ventilator-associated pneumonia (28); and (5) elevated levels of CRP often reflect the inflammatory status of the body. The exacerbation of pulmonary inflammation may lead to tissue damage in the lungs and provide a favorable environment for the growth of pathogenic microorganisms, thereby increasing the risk of nosocomial pneumonia (29).

The results of this study found that the XGBoost model had the best performance in predicting nosocomial pneumonia infection among ABI patients in the ICU. The XGBoost model is commonly used for data mining. It is less prone to overfitting on limited datasets and has lower processing requirements compared to deep learning methods, yet it performs well under various variable conditions (30). Compared to deep learning models, our study objectives and dataset do not require the extraction of larger datasets. This also explains why the XGBoost model is most suitable. Combining the results of this study, a data platform based on the XGBoost diagnostic model can be constructed. Patient basic information, along with parameters including tracheostomy time, duration of antibiotic use, blood sugar level, duration of mechanical ventilation, and CRP, are input into the platform card. Through the platform calling the algorithm interface, diagnostic results can be displayed on the platform, showing the probability of occurrence in real-time, making the diagnosis more intuitive and better guiding clinical treatment.

Several shortcoming of this study should be mentioned. Firstly, this study was retrospectively designed, and the research results may be affected by recall bias and confounding bias. Secondly, the definition of nosocomial pneumonia varies across different research centers, which may affect the predictive ability of the constructed prediction models. Thirdly, not all variables are balanced across all research centers, which may introduce bias into the results. Although consistent results were obtained based on these unbalanced variables, the impact of heterogeneity should not be underestimated.

Conclusion

This study derived a predictive model for nosocomial pneumonia infection in ICU patients with ABI using machine learning techniques from multiple centers, and conducted multiple validations to obtain effective and robust confirmation. The results indicate that machine learning-based models can more accurately predict the risk of nosocomial pneumonia infection in ICU patients with ABI, aiding in the early identification and intervention of nosocomial pneumonia infection. It should be noted that our study has limitations related to the relatively small dataset size in the context of having over 50 variables. A small dataset may increase the risk of overfitting, as the machine-learning models may adapt too closely to the specific features of this limited sample, leading to poor generalization to new data. Moreover, it may not fully represent the entire spectrum of variability in the population of ICU patients with ABI, thus potentially limiting the generalizability of our findings. To address these limitations, future research could consider expanding the sample size. Multi-center studies could be conducted to gather a larger and more heterogeneous dataset, which would likely improve the stability and generalizability of the predictive models.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by First Affiliated Hospital of Nanjing Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JP: Data curation, Investigation, Validation, Writing – original draft. ZY: Formal analysis, Investigation, Methodology, Writing – review & editing. JJ: Investigation, Methodology, Validation, Writing – review & editing. YY: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. LB: Investigation, Methodology, Writing – review & editing. YL: Methodology, Validation, Writing – review & editing. XX: Methodology, Supervision, Writing – review & editing. GG: Investigation, Writing – review & editing. MC: Methodology, Supervision, Writing – review & editing. SZ: Investigation, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Jiangsu Provincial Hospital Association and Jiangsu Modern Hospital Management Research Center's 2023 Hospital Management Innovation Research Project (Grant no. JSYGY-3-2023-403), National Natural Science Foundation of China (Grant nos. 81972153 and 82120108018) and The First Affiliated Hospital with Nanjing Medical University Clinical Capacity Enhancement Project (Grant no. JSPH-MB-2023-1).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1501025/full#supplementary-material

References

1. Krishnamoorthy V, Hough CL, Vavilala MS, Komisarow J, Chaikittisilpa N, Lele AV, et al. Tracheostomy after severe acute brain injury: trends and variability in the USA. *Neurocrit Care*. (2019) 30:546–54. doi: 10.1007/s12028-019-00697-5

2. Gao L, Chang Y, Lu S, Liu X, Yao X, Zhang W, et al. A nomogram for predicting the necessity of tracheostomy after severe acute brain injury in patients within the neurosurgery intensive care unit: a retrospective cohort study. *Heliyon*. (2024) 10:e27416. doi: 10.1016/j.heliyon.2024.e27416

3. Meisel C, Schwab JM, Prass K, Meisel A, Dirnagl U. Central nervous system injury-induced immune deficiency syndrome. *Nat Rev Neurosci*. (2005) 6:775–86. doi: 10.1038/nrn1765

4. Sharma R, Shultz SR, Robinson MJ, Belli A, Hibbs ML, O'Brien TJ, et al. Infections after a traumatic brain injury: the complex interplay between the immune and neurological systems. *Brain Behav Immun*. (2019) 79:63–74. doi: 10.1016/j.bbi.2019.04.034

5. Kourbeti IS, Vakis AF, Papadakis JA, Karabetsos DA, Bertsias G, Filippou M, et al. Infections in traumatic brain injury patients. *Clin Microbiol Infect*. (2012) 18:359–64. doi: 10.1111/j.1469-0691.2011.03625.x

6. Harrison-Felix C, Whiteneck G, Devivo MJ, Hammond FM, Jha A. Causes of death following 1 year postinjury among individuals with traumatic brain injury. *J Head Trauma Rehabil*. (2006) 21:22–33. doi: 10.1097/00001199-200601000-00003

7. Hazeldine J, Lord JM, Belli A. Traumatic brain injury and peripheral immune suppression: primer and prospectus. *Front Neurol*. (2015) 6:235. doi: 10.3389/fneur.2015.00235

8. Zygun DA, Zuege DJ, Boiteau PJ, Laupland KB, Henderson EA, Kortbeek JB, et al. Ventilator-associated pneumonia in severe traumatic brain injury. *Neurocrit Care*. (2006) 5:108–14. doi: 10.1385/NCC:5:2:108

9. Bartolo M, Zucchella C, Aabid H, Valoriani B, Copetti M, Fontana A, et al. Impact of healthcare-associated infections on functional outcome of severe acquired brain injury during inpatient rehabilitation. *Sci Rep*. (2022) 12:5245. doi: 10.1038/s41598-022-09351-1

10. Glance LG, Stone PW, Mukamel DB, Dick AW. Increases in mortality, length of stay, and cost associated with hospital-acquired infections in trauma patients. *Arch Surg*. (2011) 146:794–801. doi: 10.1001/archsurg.2011.41

11. Kalil AC, Metersky ML, Klompas M, Muscedere J, Sweeney DA, Palmer LB, et al. Management of Adults with Hospital-acquired and Ventilator-associated Pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis*. (2016) 63:e61–e111. doi: 10.1093/cid/ciw353

12. Anderson D, Kutsogiannis DJ, Sligl WI. Sepsis in traumatic brain injury: epidemiology and outcomes. *Can J Neurol Sci*. (2020) 47:197–201. doi: 10.1017/cjn.2019.320

13. Cardozo Júnior LC, Silva RR. Sepsis in intensive care unit patients with traumatic brain injury: factors associated with higher mortality. *Rev Bras Ter Intensiva*. (2014) 26:148–54. doi: 10.5935/0103-507x.20140022

14. Shan J, Chen HL, Zhu JH. Diagnostic accuracy of clinical pulmonary infection score for ventilator-associated pneumonia: a meta-analysis. *Respir Care*. (2011) 56:1087–94. doi: 10.4187/respcare.01097

15. Kishore AK, Vail A, Bray BD, Chamorro A, Napoli MD, Kalra L, et al. Clinical risk scores for predicting stroke-associated pneumonia: a systematic review. *Eur Stroke J*. (2016) 1:76–84. doi: 10.1177/2396987316651759

16. Ni J, Shou W, Wu X, Sun J. Prediction of stroke-associated pneumonia by the A2DS2, AIS-APS, and ISAN scores: a systematic review and meta-analysis. *Expert Rev Respir Med*. (2021) 15:1461–72. doi: 10.1080/17476348.2021.1923482

17. Hamilton AJ, Strauss AT, Martinez DA, Hinson JS, Levin S, Lin G, et al. Machine learning and artificial intelligence: applications in healthcare epidemiology. *Antimicrob Steward Healthc Epidemiol*. (2021) 1:e28. doi: 10.1017/ash.2021.192

18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. (2015) 350:g 7594. doi: 10.1136/bmj.g7594

19. Koulenti D, Zhang Y, Fragkou PC. Nosocomial pneumonia diagnosis revisited. *Curr Opin Crit Care*. (2020) 26:442–9. doi: 10.1097/MCC.0000000000000756

20. Zheng Y, Lin YX, He Q, Zhuo LY, Huang W, Gao ZY, et al. Novel machine learning models to predict pneumonia events in supratentorial intracerebral hemorrhage populations: an analysis of the Risa-MIS-ICH study. *Front Neurol*. (2022) 13:955271. doi: 10.3389/fneur.2022.955271

21. Lee CC, Su SY, Sung SF. Machine learning-based survival analysis approaches for predicting the risk of pneumonia post-stroke discharge. *Int J Med Inform*. (2024) 186:105422. doi: 10.1016/j.ijmedinf.2024.105422

22. Liang Y, Zhu C, Tian C, Lin Q, Li Z, Li Z, et al. Early prediction of ventilator-associated pneumonia in critical care patients: a machine learning model. *BMC Pulm Med*. (2022) 22:250. doi: 10.1186/s12890-022-02031-w

23. Huang J, Liu M, He W, Liu F, Cheng J, Wang H. Use of the A2DS2 scale to predict morbidity in stroke-associated pneumonia: a systematic review and meta-analysis. *BMC Neurol*. (2021) 21:33. doi: 10.1186/s12883-021-02060-8

24. Tu TM, Phua SS, Acharyya S, Ng WM, Oh DC. Predicting pneumonia in acute Ischaemic stroke: comparison of five prediction scoring models. *Ann Acad Med Singap*. (2017) 46:237–44. doi: 10.47102/annals-acadmedsg.V46N6p237

25. Dai Y, Qiao J, Ye QP, Li XY, Hu JH, Dou ZL. Exploring the influence of dysphagia and tracheostomy on pneumonia in patients with stroke: a retrospective cohort study. *Brain Sci*. (2022) 12:1664. doi: 10.3390/brainsci12121664

26. Westendorp WF, Vermeij JD, Hilkens NA, Brouwer MC, Algra A, van der Worp HB, et al. Development and internal validation of a prediction rule for post-stroke infection and post-stroke pneumonia in acute stroke patients. *Eur Stroke J*. (2018) 3:136–44. doi: 10.1177/2396987318764519

27. Muscari A, Falcone R, Recinella G, Faccioli L, Forti P, Pastore Trossello M, et al. Prognostic significance of diabetes and stress hyperglycemia in acute stroke patients. *Diabetol Metab Syndr*. (2022) 14:126. doi: 10.1186/s13098-022-00896-9

28. Wu VKS, Fong C, Walters AM, Lele AV. Prevalence, clinical characteristics, and outcomes related to ventilator-associated events in Neurocritically ill patients. *Neurocrit Care*. (2020) 33:499–507. doi: 10.1007/s12028-019-00910-5

29. Kalra L, Smith CJ, Hodsoll J, Vail A, Irshad S, Manawadu D. Elevated C-reactive protein increases diagnostic accuracy of algorithm-defined stroke-associated pneumonia in afebrile patients. *Int J Stroke*. (2019) 14:167–73. doi: 10.1177/1747493018798527

30. Fillinger S, de la Garza L, Peltzer A, Kohlbacher O, Nahnsen S. Challenges of big data integration in the life sciences. *Anal Bioanal Chem*. (2019) 411:6791–800. doi: 10.1007/s00216-019-02074-9

Frontiers in Medicine

Translating medical research and innovation into improved patient care

A multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact