Interpretive machine learning predicts short-term mortality risk in elderly sepsis patients

Xing-Yu Zhu[1,2], Zhi-Meng Jiang[1], Xiao- Li[1], Zi-Wen Lv[1], Jian-Wei Tian[2] and Fei-Fei Su[2]*

[1]Graduate School of Hebei North University, Zhangjiakou, Hebei, China, [2]Department of Cardiovascular Medicine, Chinese People's Liberation Army Air Force Medical Center, Beijing, China

Backgrounds: Sepsis is a leading cause of in-hospital mortality. However, its prevalence is increasing among the elderly population. Therefore, early identification and prediction of the risk of death in elderly patients with sepsis is crucial. The objective of this study was to create a machine learning model that can predict short-term mortality risk in elderly patients with severe sepsis in a clear and concise manner.

Methods: Data was collected from the MIMIC-IV (2.2). It was randomly divided into a training set and a validation set using a 7:3 ratio. Mortality predictors were determined through Recursive Feature Elimination (RFE). A prediction model for 28 days of ICU stay was built using six machine-learning algorithms. To create a comprehensive and nuanced model resolution, Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) were used to systematically interpret the models at both a global and detailed level.

Results: The study involved the analysis of 4,056 elderly patients with sepsis. A feature recursive elimination algorithm was utilized to select eight variables out of 49 for model development. Six machine learning models were assessed, and the Extreme Gradient Boosting (XGBoost) model was found to perform the best. The validation set achieved an AUC of 0.88 (95% CI: 0.86−0.90) and an accuracy of 0.84 (95% CI: 0.81−0.86) for this model. To examine the roles of the eight key variables in the model, SHAP analysis was employed. The global ranking order was made evident, and through the use of LIME analysis, the weights of each feature range in the prediction model were determined.

Conclusion: The study's machine learning prediction model is a dependable tool for forecasting the prognosis of elderly patients with severe sepsis.

KEYWORDS

sepsis, machine learning, shapley additive explanations, local interpretable model-agnostic explanations, XGBoost

Introduction

Sepsis has the potential to lead to multiple organ dysfunction syndrome (MODS) and in severe cases, even death. It is a major factor contributing to mortality and morbidity worldwide (Rudd et al., 2020; Singer et al., 2016). In the late 1970s, it was estimated that approximately 164,000 new cases of sepsis occurred annually in the United States (Martin et al., 2003). It is worth noting that the incidence of sepsis has been

increasing globally, not just in the United States (Walkey et al., 2013; Fleischmann et al., 2016; Rhee et al., 2017). Estimates suggest that sepsis mortality rates can vary between 10 and 52 percent (Martin et al., 2003; Kaukonen et al., 2014; Li A. et al., 2022; Epstein et al., 2016). It has been observed that the age group of 65 years and over is more vulnerable to sepsis, with a significant percentage of cases, occurring in this demographic. Considering the ongoing demographic shift towards an aging population, it is anticipated that the incidence of sepsis will rise in the future (Martin et al., 2003; Kaukonen et al., 2014; Angus et al., 2001; Giacomello and Toniolo, 2021). Identifying and predicting mortality risk in elderly septicaemic patients at an early stage is of utmost importance. This information can provide a valuable reference for clinicians to improve patient survival and prognosis through timely and effective therapeutic measures. Although some machine learning-based predictive models have been developed, their limited interpretability has hindered their application in clinical practice.

Rapid advancements in machine learning (ML) have provided powerful tools for extracting complex patterns, assisting clinical decision-making, and improving patient prognosis prediction in medicine. This section briefly outlines several prevalent ML models and their applications in medical prognosis. Logistic regression, a classical statistical learning method, is suitable for binary classification problems. Its model is concise and efficient, offering high interpretability and widespread application in areas such as cardiovascular disease risk assessment (Zabor et al., 2022); however, it struggles with complex non-linear relationships. Support vector machines (SVM), employing kernel methods to map data into high-dimensional space, excel at handling non-linear classifications (Zhou, 2022). While demonstrating superior performance in predicting breast cancer patient survival, they entail complex hyperparameter tuning and high computational costs (Sarkar and Mali, 2023). Neural networks, leveraging multiple layers of neurons to simulate non-linear mappings, are widely used in EHR data analysis and medical image processing (Manimegalai et al., 2022). Nevertheless, their "black box" nature and reliance on large annotated datasets remain challenges. Multilayer perceptrons (MLP), a foundational neural network architecture, exhibit strong performance in image segmentation due to their flexibility; however, their efficiency diminishes when confronted with multimodal data (Gao et al., 2023; Zhang et al., 2023). Naive Bayes (NB), based on Bayes' theorem and assuming feature independence, boasts high computational efficiency. Demonstrating excellent performance in genomic data analysis and cancer patient stratification prediction, its performance is nevertheless affected by feature correlations (Das and Dutta, 2020; Lu et al., 2024). Extreme Gradient Boosting (XGBoost), employing ensemble learning to construct robust predictive models, shows superior performance in predicting mortality risk in heart failure patients (Li J. et al., 2022). While its efficient feature handling capabilities are beneficial for imbalanced datasets, the risk of overfitting in small datasets necessitates careful consideration (Takefuji, 2025). These models, each with its unique strengths, provide valuable support for data analysis and decision-making in medical prognosis. Future advancements, focusing on enhancing model explainability and effectively integrating multimodal data, promise to further unlock their clinical potential.

Research in critical care medicine has been exploring prognostic modeling for sepsis patients. It has been found that commonly used serological indices, such as calcitonin, platelet, and lactate levels, may have limited prognostic value in assessing the effect of sepsis (Brunkhorst et al., 2000; Gattinoni et al., 2019). Sepsis is a clinical syndrome that presents a wide range of biological characteristics. As a result, it can be challenging to fully reflect the patient's condition based on individual indicators alone (Hernandez et al., 2019). The efficacy of emerging machine learning techniques in comparison to traditional means of prediction is contingent on the nature of the dataset and the field in which they are employed. It has been hypothesized that machine learning techniques may be more appropriate for highly innovative fields with large amounts of data (Rajula et al., 2020). By meticulously designing and optimizing algorithms, they are capable of learning from large and intricate datasets to uncover more profound associations between patients' clinical indicators and prognostic outcomes. Nevertheless, it is important to acknowledge that machine learning models in clinical practice may have limitations stemming from the lack of clarity and intuition in the decision-making process (Karim et al., 2023).

When constructing algorithmic models, it is common to use precision and recall of the test set as benchmarks for measuring model strength. However, when communicating with non-professionals, relying on a single value may not be practical. Therefore, it is important to demonstrate the internal logic and principles of the model to enhance its credibility. It is worth noting that certain models are not always constructed based on clear rules. Some black-box models, despite their higher predictive accuracy, may be opaque and unable to provide a specific decision-making basis. This can make it challenging for the general public to comprehend and have faith in their predictions. Thankfully, there are now tools available such as SHAP and lime that can provide valuable insights into the decision-making processes of machine learning models (Hilton et al., 2020; Lundberg et al., 2020). These tools can facilitate a deeper comprehension of how machine learning models operate, which could lead to their expanded implementation in clinical practice.

This investigation aims to develop a machine learning model that can predict the short-term risk of death in elderly patients with sepsis with accuracy. Advanced SHAP and lime technical tools will be utilized to comprehensively analyze and explore the model at both the holistic and local levels.

Methods

Database

This retrospective study utilized data from the publicly available electronic health record dataset MIMIC-IV(2.2) (Johnson et al., 2023). The database was created by the Computational Physiology Laboratory of the Massachusetts Institute of Technology (MIT), the Beth Israel Deaconess Medical Centre (BIDMC) at Harvard Medical School, United States, and Philips Healthcare. Patient information was collected and research resources were created with approval from the Beth Israel Deaconess Medical Centre's Institutional Review Board, which waived the need for informed consent and supported the data-sharing initiative

(Johnson et al., 2023; Goldberger et al., 2000). The study adheres rigorously to the tenets set forth in the Declaration of Helsinki. Zi-wen Lv, the author of this study, has completed the CITI course and passed the 'Conflict of Interest' and 'Data or Specimen Research Only' examinations (ID: 55,109,354). As a result, we are authorized to use this database.

Study population

This study leverages the advantages of the heterogeneous and multisource clinical data contained in the MIMIC-IV (version 2.2) database, which provides comprehensive electronic health records (EHRs) and longitudinal follow-up information to support the refined analytical requirements of prognostic modeling. Using Structured Query Language (SQL) and the Navicat 16.3.8 software platform, a systematic screening and extraction of raw datasets meeting the inclusion criteria was conducted to construct an association framework between high-dimensional features and clinical outcomes. During data processing, patients with septicemia in the MIMIC database were accurately identified and classified using specific codes, such as R6520, R6521, and 99,592. Data on patients, including socio-demographic characteristics, vital signs, laboratory parameters, and complications, was extracted using structured query language (Yang et al., 2020). To determine the study sample, strict inclusion and exclusion criteria were applied. Inclusion criteria required patients to be admitted to the ICU with a confirmed diagnosis of sepsis and to be 65 years old or older. Exclusion criteria were applied to patients who were not first-time ICU admissions and those with more than 30% of missing data variables.

Approaching the issue of missing data

Data loss is a known issue in clinical trials, which may impact the integrity of the original data set and potentially weaken the robustness and validity of study findings (Sterne et al., 2009). Multiple imputation is a commonly used method for dealing with missing data in interpolation. This method provides valid estimates while also accounting for the uncertainty associated with missing data (Pedersen et al., 2017; Heymans and Twisk, 2022). To address the issue of missing data, a multiple imputation approach was employed. Variables such as triglycerides and total cholesterol, which had a missing rate of over 30%, were excluded from the analysis. The remaining variables were refined using multiple imputation techniques. The data-filling task was completed using the 'mice' package in R4.2.3. Through engagement with clinical experts specializing in sepsis management, we systematically assessed the relevance and potential impact of excluded variables. Their expert consensus confirmed that the removal of these variables would not compromise the prognostic accuracy or introduce significant bias in our analysis of sepsis outcomes.

Statistical analyses

The data were analyzed using both R languages (version 4.2.3). For continuous variables that follow a normal distribution, the mean and standard deviation ($\bar{x}\pm s$) were used to describe their concentration and dispersion trends. For continuous variables that do not follow a normal distribution, the median and interquartile range (M(Q1, Q3)) were used to characterize their distributions. Differences between groups for normally distributed variables were compared using the t-test, while the Mann-Whitney U test was used for non-normally distributed variables. The distributions of categorical variables were visualized as percentages. The chi-square test was used to determine significant associations between categories. Additionally, we analyzed the non-linear relationship between characteristic variables and the risk of short-term mortality in elderly patients with sepsis using restricted cubic bars.

Development and validation of predictive models

While models with rich features often outperform those with fewer features in terms of accuracy, this is not always the practice case. In clinical applications, it is important to carefully consider the number of features used in a model, as simply increasing the number of features does not always lead to improved performance. This is because irrelevant or redundant features may negatively impact the model's accuracy, and an excessive number of features may result in overfitting. The study employed a Recursive Feature Elimination (RFE) algorithm to develop a model based on 49 variables that were strongly associated with short-term mortality outcomes.

The RFE algorithm is a model-based strategy for selecting features. It screens the best feature combinations by iteratively training the model and gradually eliminating the features with the lowest weights. To ensure a more accurate measurement of performance fluctuation during the feature selection process, this study incorporates a layer of resampling (10-fold cross-validation) outside the RFE algorithm. The algorithm identified eight features that had a significant impact on predicting short-term risk of mortality in elderly patients with severe sepsis. This study utilized six machine learning models, specifically Logistic Regression, Neural Network, Support Vector Machine, Multilayer Perceptron, Naive Bayes, and Extreme Gradient Boosting, to create effective prediction models. The patient dataset was randomly sampled and split into training and validation sets in a 7:3 ratio during the development of the prediction model. To ensure accurate parameter tuning and model resilience, a tenfold cross-validation technique was employed. The models' predictive performance was assessed using the area under the working characteristic curve of the subjects as the primary index. Furthermore, a detailed evaluation of each model was conducted, including sensitivity, specificity, positive predictive value, negative predictive value, recall, precision, and F1 score. To provide a more comprehensive representation of the optimal models' performance, we have included standard curves and decision analysis curves (DCA) (Van Calster et al., 2018).

Tools for interpreting machine learning

This study employs two model interpretation methods, namely, SHAP and lime, to conduct a thorough analysis of the risk prediction model that we have developed. We extensively investigate the

contribution of each clinical variable in the model. It is worth noting that SHAP is a commonly used tool for evaluating the impact of features in machine learning models (Li J. et al., 2022; Fahmy et al., 2022). The main concept is to measure the impact of each feature on the model's final output, providing a comprehensive understanding of the 'black box model' from both a global and local perspective (Lundberg and Lee, 2017). The SHAP value provides an accurate representation of the positive and negative impact that each predictor has on the target variable (Lundberg and Lee, 2017). Lime is an advanced algorithm that can generate interpretable models for any classifier or regressor, enabling accurate interpretation of predicted results through local approximation. This is accomplished by training a local substitution model to provide a detailed interpretation of a single instance (Peng and Menzies, 2021). Lime can provide a comprehensive understanding of the impact of various characteristic variables on the predictive model. This can help to ensure that the model is accurate and reliable.

Result

Patient characteristics

The study involved a total of 4,056 elderly patients diagnosed with sepsis, out of which 1,259 patients passed away within 28 days. These patients were randomly assigned to either the training set (2,865 patients) or the validation set (1,191 patients) in a 7:3 ratio. Table 1 presents a detailed comparison of the baseline characteristics between the group that passed away within 28 days and the non-death group in the training set. During the study, we employed a recursive feature elimination algorithm to determine the eight most closely related variables to short-term mortality risk. These variables, listed in order of importance, are Bicarbonate, GCS, Lactate, Platelets, PaO2/FiO2, SBP, Sodium, and upon analyzing Table 1, it becomes evident that these eight core indicators are statistically significant when comparing the fatal and non-fatal groups in the training set. The study examined the correlation between eight core indicators and the risk of death in elderly patients with sepsis, utilizing restricted cubic spline analysis. The findings suggest a significant non-linear correlation between these metrics and the risk of death in elderly patients with sepsis (p-non-linear <0.001). For further details, please refer to Figure 1.

Model performance and interpretation

Six distinctive machine-learning models were constructed on the training dataset, such as logistic regression, neural network, support vector machine, multilayer perceptron, Naive Bayes, and extreme gradient boosting. Table 2, Table 3, and Figures 2A, B present the key metrics of these models on the validation dataset, including sensitivity, specificity, positive predictive value, negative predictive value, recall, F1 score, AUC value, and precision. Based on a comprehensive trade-off comparison, the extreme gradient boosting model demonstrated the optimal prediction performance, achieving an AUC value of 0.88 (0.86.0.90) and a high accuracy rate of 0.84 (0.81.0.86). To investigate the performance and calibration capability of the Extreme Gradient Boosting machine learning

model, the DCA curve and calibration curve were plotted for the training and test sets, respectively. In clinical contexts, the term 'net benefit' refers to the probability of a patient's disease occurring being minimized when further medical intervention is deemed necessary (Lee et al., 2021). Figures 2C, 3D display three lines: orange, black, and blue. The orange line represents the net benefit of no treatment for all individuals, which is naturally zero. The black line represents the net benefit of treating all individuals, which decreases as the threshold probability increases. The graph displays the change in net benefit of our decision model at different threshold probabilities. If the blue line closely follows the orange and black reference lines, it may suggest that the model has limited practical application value. However, if the blue line consistently exceeds the reference lines across a wide range of threshold intervals, it may demonstrate that the model has a higher net benefit. The ML prediction model constructed demonstrated good net gains in both the training set and the validation set, as shown in Figures 2C, D. The calibration curve illustrates the model's predicted probability compared to the actual observation probability in the data (refer to Figures 2E, F). The results indicate that the machine learning model developed in this study exhibits exceptional calibration ability.

Figure 3A illustrates the global feature importance mapping generated by inputting the SHAP value matrix into the bar graph function. The plot represents the average absolute value of each feature's global importance across all samples, while the feature's criticality in the model is visualized on the Y-axis. Figure 3B presents each patient as a point, with their X-axis coordinates corresponding to the predicted value given by the prediction model. The colors of the dots are changed to represent the values that are predicted by the model. For example, it is worth noting that the Glasgow Coma Scale (GCS) has a significant impact on predicting outcomes. As the GCS value increases, so does the mortality rate. This effect is determined by the level of the feature and is visualized by the yellow and purple dots on the graph. Yellow dots represent high GCS values on the positive side of the X-axis, while purple dots represent low GCS values on the negative side of the X-axis. The model indicated a negative correlation with Platelets, PaO2/FiO2, SBP, and Sodium. Notably, PaO_2/FiO_2 exhibited a bimodal distribution, reflecting the differential contributions of distinct patient subgroups, whereas other variables, such as lactate and platelet count, displayed unimodal distributions. This visualization elucidates the relative importance of each biomarker in predicting mortality risk, highlighting potential threshold effects (e.g., the PaO_2/FiO_2 threshold for delineating the severity of hypoxemia) and underscoring the complexity of oxygenation indices in sepsis prognosis. Such analysis enhances the interpretability of machine learning models and informs the development of precision medicine strategies tailored to the specific physiological characteristics of individual patients.

Figures 3C, D display the personalized trait attribution analysis for two randomly selected patients. The mean effect value, Ef(x), across all sample data is 0.311. Each characteristic variable's contribution is visualized as an arrow, with its direction indicating its effect on the probability of the outcome - either decreasing the likelihood of a negative outcome or increasing the chance of a positive outcome. To distinguish between positive (purple) and negative (yellow) effects, they are color-coded. The diagram presents the arrows in an order based on their magnitude of

TABLE 1 Baseline characteristics of the training set.

Variables	Total	Survival within 28 days	Death within 28 days	p
Number (sample size)	2865	1975	890	
Sex, n (%)				0.204
Female	1316 (46)	891 (45)	425 (48)	
Male	1549 (54)	1084 (55)	465 (52)	
Ethnicity, n (%)				<0.001
Black	259 (9)	183 (9)	76 (9)	
White	2041 (71)	1445 (73)	596 (67)	
Others	565 (20)	347 (18)	218 (24)	
Age, Median (Q1, Q3)	77 (71, 85)	77 (71, 84)	78 (72, 85)	0.006
BMI (kg/m2), Median (Q1,Q3)	27.02 (24.16, 31.25)	27.16 (24.35, 31.57)	26.7 (23.68, 30.75)	0.003
Weight (kg), Median (Q1, Q3)	75.5 (63, 90.6)	76.1 (63.5, 91)	73.75 (62.12, 89.57)	0.019
Height (cm), Median (Q1, Q3)	168 (162.69, 173.18)	168.07 (163, 173.61)	166.75 (160.57, 173)	<0.001
SBP (mmHg), Median (Q1,Q3)	150 (135, 167)	153 (138, 169)	143 (130, 160)	<0.001
PaO2/FiO2, Median (Q1,Q3)	340.84 (283.33, 419.05)	359.15 (293.33, 422.79)	293.33 (283.33, 379.46)	<0.001
GCS, Median (Q1, Q3)	1 (1, 3)	1 (1, 2)	2 (1, 4)	<0.001
Heart failure, n (%)	1255 (44)	857 (43)	398 (45)	0.534
Respiratory failure, n (%)	1419 (50)	847 (43)	572 (64)	<0.001
Hypertension, n (%)	1110 (39)	808 (41)	302 (34)	<0.001
AMI, n (%)	431 (15)	288 (15)	143 (16)	0.331
Atrial fibrillation, n (%)	1341 (47)	910 (46)	431 (48)	0.26
AKI, n (%)	1976 (69)	1308 (66)	668 (75)	<0.001
Diabetes, n (%)	1188 (41)	816 (41)	372 (42)	0.841
RBC (m/uL), Median (Q1,Q3)	3.71 (3.3, 4.15)	3.75 (3.33, 4.16)	3.62 (3.23, 4.11)	<0.001
Lactate (mmol/L), Median (Q1,Q3)	2.8 (1.9, 4.8)	2.4 (1.7, 3.7)	4.4 (2.6, 8.5)	<0.001
WBC (K/uL), Median (Q1,Q3)	18.8 (13.4, 26.1)	18.1 (13.15, 25.1)	20.2 (14.62, 29)	<0.001
Sodium (mEq/L), Median (Q1,Q3)	144 (141, 148)	144 (141, 148)	143 (139, 148)	<0.001
Platelets (K/uL), Median (Q1,Q3)	273 (186, 386)	296 (205, 415)	230 (144, 318)	<0.001
Total_Calcium (mg/dL), Median (Q1, Q3)	8.8 (8.4, 9.3)	8.9 (8.4, 9.3)	8.8 (8.2, 9.4)	<0.001
Hemoglobin (g/L), Median (Q1,Q3)	11 (9.8, 12.3)	11.1 (9.8, 12.3)	10.8 (9.6, 12.2)	0.01
RDW (%), Median (Q1,Q3)	16.8 (15.1, 19.1)	16.6 (15, 18.8)	17.4 (15.62, 19.7)	<0.001
ALT (IU/L, Median (Q1,Q3)	49 (22, 136)	46 (22, 117.5)	56 (23, 252.75)	<0.001
Lymphocytes (%), Median (Q1,Q3)	11.1 (7, 16.6)	11.7 (7.5, 17)	10 (6, 15.2)	<0.001

TABLE 1 (*Continued*) Baseline characteristics of the training set.

Variables	Total	Survival within 28 days	Death within 28 days	p
ALP (IU/L), Median (Q1,Q3)	131 (85, 215)	130 (84, 213.5)	135.73 (86, 220.75)	0.188
Potassium (mEq/L), Median (Q1,Q3)	4.9 (4.4, 5.5)	4.8 (84, 213.5)	5.1 (4.6, 5.8)	<0.001
Neutrophils (%), Median (Q1,Q3)	86 (80.4, 90.5)	86 (81, 90.5)	86.17 (79, 90.5)	0.054
INR, Median (Q1,Q3)	1.7 (1.4, 2.7)	1.6 (1.3, 2.4)	2 (1.5, 3.1)	<0.001
HCT (%), Median (Q1,Q3)	34.2 (30.8, 38.1)	34.3 (30.9, 38)	33.9 (30.5, 38.3)	0.278
Chloride (mEq/L), Median (Q1,Q3)	110 (106, 114)	110 (106, 114)	109 (104, 114)	<0.001
BUN (mg/dL), Median (Q1,Q3)	45 (28, 72)	41 (26, 65)	56.5 (36, 83)	<0.001
Albumin (g/dL), Median (Q1,Q3)	2.9 (2.6, 3.2)	2.9 (2.68, 3.2)	2.8 (2.4, 3.2)	<0.001
Glucose (mg/dL), Median (Q1, Q3)	191 (148, 266)	186 (144.5, 258.5)	204.5 (158, 282.75)	<0.001
Scr (mg/dL), Median (Q1,Q3)	1.9 (1.2, 3.2)	1.6 (1.1, 2.7)	2.6 (1.6, 3.9)	<0.001
TBIL (mg/dL), Median (Q1,Q3)	1.03 (0.53, 2.3)	1 (0.5, 2)	1.3 (0.6, 3.2)	<0.001
PT (S), Median (Q1,Q3)	18.4 (15, 28.7)	17.2 (14.5, 25.5)	21.45 (16.5, 32.3)	<0.001
Monocytes (%), Median (Q1,Q3)	6.72 (4.6, 9.1)	7 (4.7, 9.31)	6.1 (4, 8.8)	<0.001
Bicarbonate (mEq/L), Median (Q1,Q3)	28 (25, 32)	29 (26, 32)	25 (20, 29)	<0.001
PTT (S), Median (Q1,Q3)	45.2 (34, 92.2)	41.7 (33.1, 85.52)	55.25 (37.5, 106.72)	<0.001
Basophils (%), Median (Q1,Q3)	0.3 (0.1, 0.5)	0.3 (0.2, 0.5)	0.2 (0, 0.4)	<0.001
MCV (%), Median (Q1,Q3)	96 (91, 101)	95 (91, 100)	97 (92, 102)	<0.001
AG (mEq/L), Median (Q1,Q3)	19 (16, 23)	18 (16, 21)	21 (18, 26)	<0.001
pH value, Median (Q1, Q3)	7.37 (7.3, 7.42)	7.37 (7.31, 7.42)	7.35 (7.27, 7.42)	<0.001
SOFA, Median (Q1, Q3)	8 (5, 11)	7 (5, 10)	11 (8, 13)	<0.001

BMI: body mass index; SBP: systolic blood pressure; GCS: glasgow coma scale; AMI: acute myocardial infarction; AKI: acute kidney injury; RBC: red blood cell; WBC: white blood cell; RDW: red blood cell distribution width; ALT: alanine aminotransferase; ALP: alkaline phosphatase; INR: international normalized ratio; HCT: hematocrit; BUN: blood urea nitrogen; Scr: Serum creatinine; TBIL: total bilirubin; PT: prothrombin time; PTT: partial thromboplastin time; MCV: mean corpuscular volume; AG: anion gap.

influence on the results. The length of each arrow accurately reflects the strength of influence of each feature, proportional to the SHAP value of the corresponding feature. Figure 3C presents the characteristic attributions of a patient who unfortunately passed away. In contrast, Figure 3D shows the characteristic attributions of a patient who successfully recovered. The identified high-risk factors that increased the risk of mortality were high lactate levels, low PaO2/FiO2, high centrocytes, low platelet count, and low bicarbonate. Although a low PaO2/FiO2 ratio is still considered a risk factor for mortality, it is important to note that this patient's high bicarbonate levels, low GCS, and high platelet count were found to significantly reduce the risk of death. This comparative analysis highlights the complex interplay of various factors in risk assessment.

To further elaborate on our constructed models, we have also employed lime, an interpretable tool commonly used in the field of machine learning (Peng and Menzies, 2021). Figure 4 demonstrates how physicians can improve their decision-making process with the assistance of the model, provided that clear and comprehensible explanations are given. The blue color in the figure represents features that contribute to the predicted outcome, while the red color indicates features that detract from it. The graph illustrates the weight assigned to supporting (blue) or not supporting (red) on the horizontal axis, while the importance of features is ranked on the vertical axis. In the case of the patient who survived, it can be concluded that bicarbonate in the range of 28–32 and lactate in the range of 1.9–2.8 are supportive, as well as a blood pressure between 150–167. The prediction accuracy for this case is high at 0.89, which demonstrates the excellent performance of our prediction model.

Figure 5 presents a global analysis of 2,500 randomly selected patients, demonstrating the range of features and their
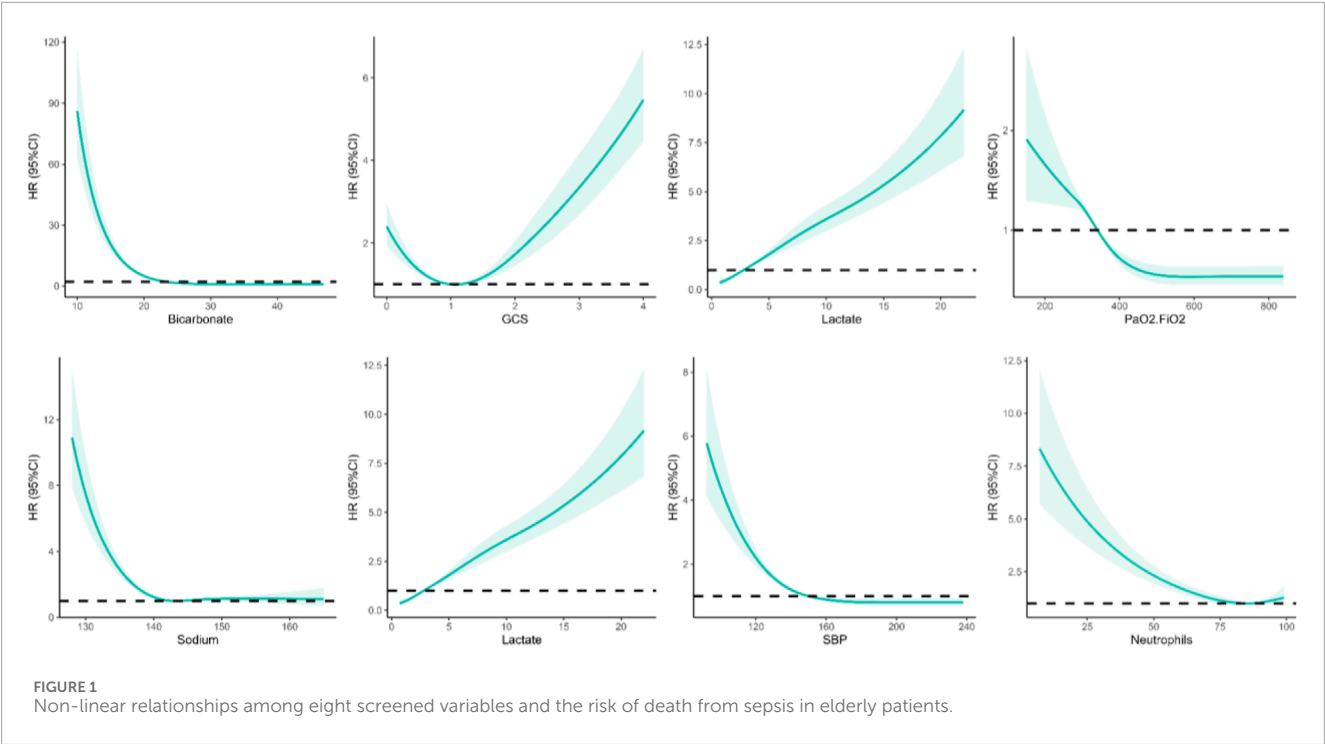
FIGURE 1
Non-linear relationships among eight screened variables and the risk of death from sepsis in elderly patients.

TABLE 2 Predictive power of eight machine learning models.

CLS	Sensitivity	Specificity	Pos pred value	Neg pred value
XGBoost	0.64 (0.59.0.69)	0.93 (0.91.0.94)	0.80 (0.75.0.84)	0.85 (0.83.0.87)
Neural Network	0.62 (0.57.0.67)	0.93 (0.91.0.94)	0.79 (0.74.0.84)	0.84 (0.82.0.87)
SVM	0.62 (0.57.0.67)	0.91 (0.89.0.93)	0.76 (0.71.0.81)	0.84 (0.82.0.87)
Naive Bayes	0.80 (0.76.0.84)	0.71 (0.68.0.74)	0.55 (0.51.0.60)	0.89 (0.86.0.91)
Multilayer Perceptron	0.61 (0.56.0.66)	0.92 (0.89.0.93)	0.76 (0.71.0.81)	0.84 (0.81.0.86)
Logistic Regression	0.59 (0.54.0.64)	0.92 (0.89.0.93)	0.76 (0.71.0.81)	0.83 (0.81.0.86)

TABLE 3 Predictive power of eight machine learning models.

CLS	F1 score	Recall	AUC(95%CI)	Accuracy
XGBoost	0.71	0.64	0.88 (0.86.0.90)	0.84 (0.81.0.86)
Neural Network	0.70	0.62	0.86 (0.83.0.88)	0.83 (0.81.0.85)
SVM	0.69	0.62	0.85 (0.82.0.87)	0.82 (0.80.0.84)
Naive Bayes	0.65	0.80	0.84 (0.82.0.87)	0.74 (0.71.0.76)
Multilayer Perceptron	0.68	0.61	0.83 (0.80.0.86)	0.82 (0.80.0.84)
Logistic Regression	0.66	0.59	0.82 (0.79.0.84)	0.82 (0.79.0.84)

CLS, classifiers; AUC, area under the curve; CI, confidence interval; SVM, support vector machine; XGBoost: Extreme Gradient Boosting.
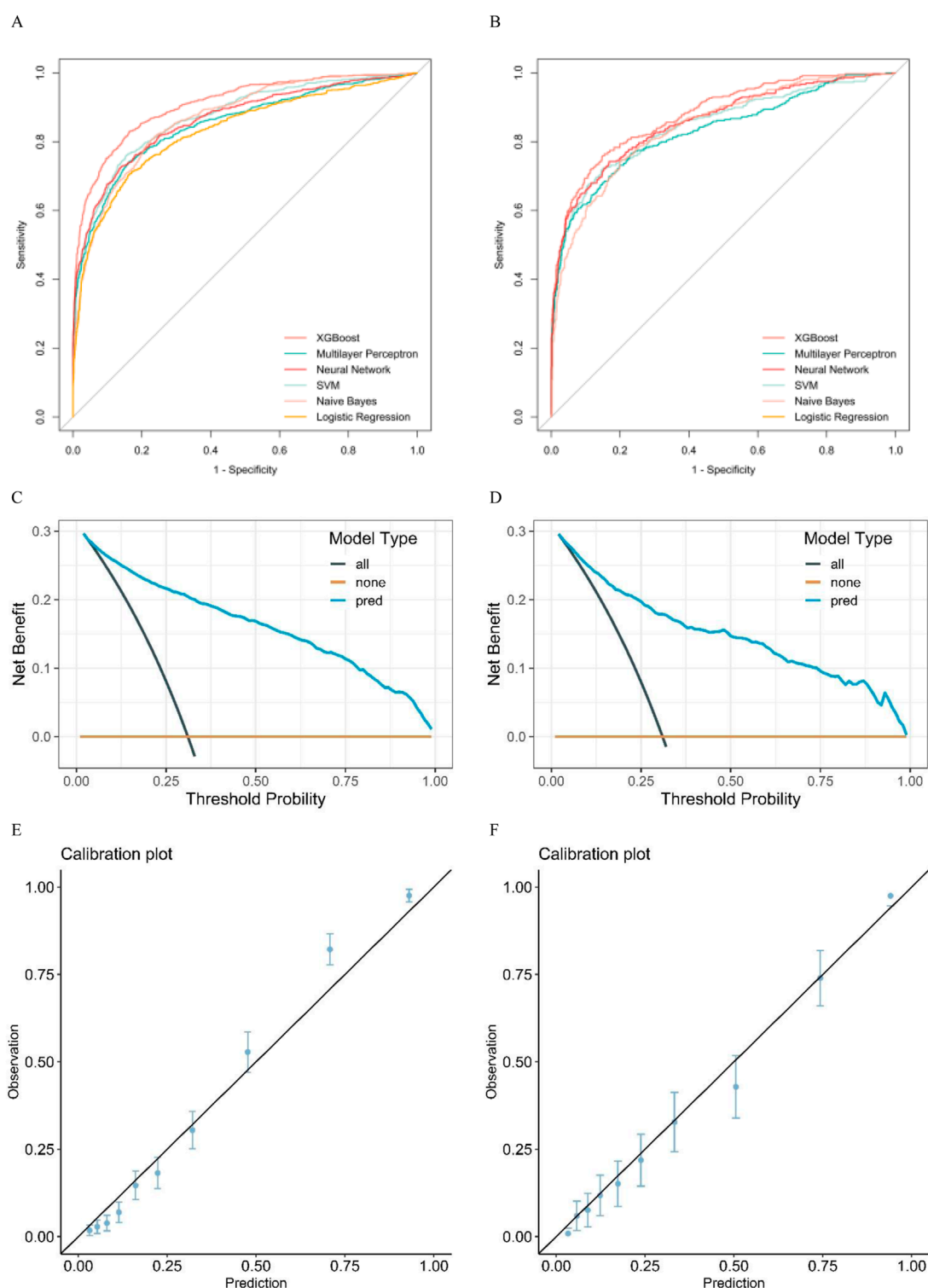
FIGURE 2
(A) pertains to the training set, while **(B)** pertains to the validation set. **(A, B)** display the ROC curves of eight machine learning models used to predict the risk of death in elderly patients with sepsis. Additionally, **(C, D)** show the DCA curves of the XGBboost machine learning model used to predict the risk of death in elderly sepsis patients. **(C)** represents the training set, while **(D)** represents the validation set. Figures **(E, F)** display the calibration curves of the XGBboost machine learning model that was used to predict the risk of death in elderly sepsis patients. Figure E pertains to the training set, whereas Figure F pertains to the validation set.
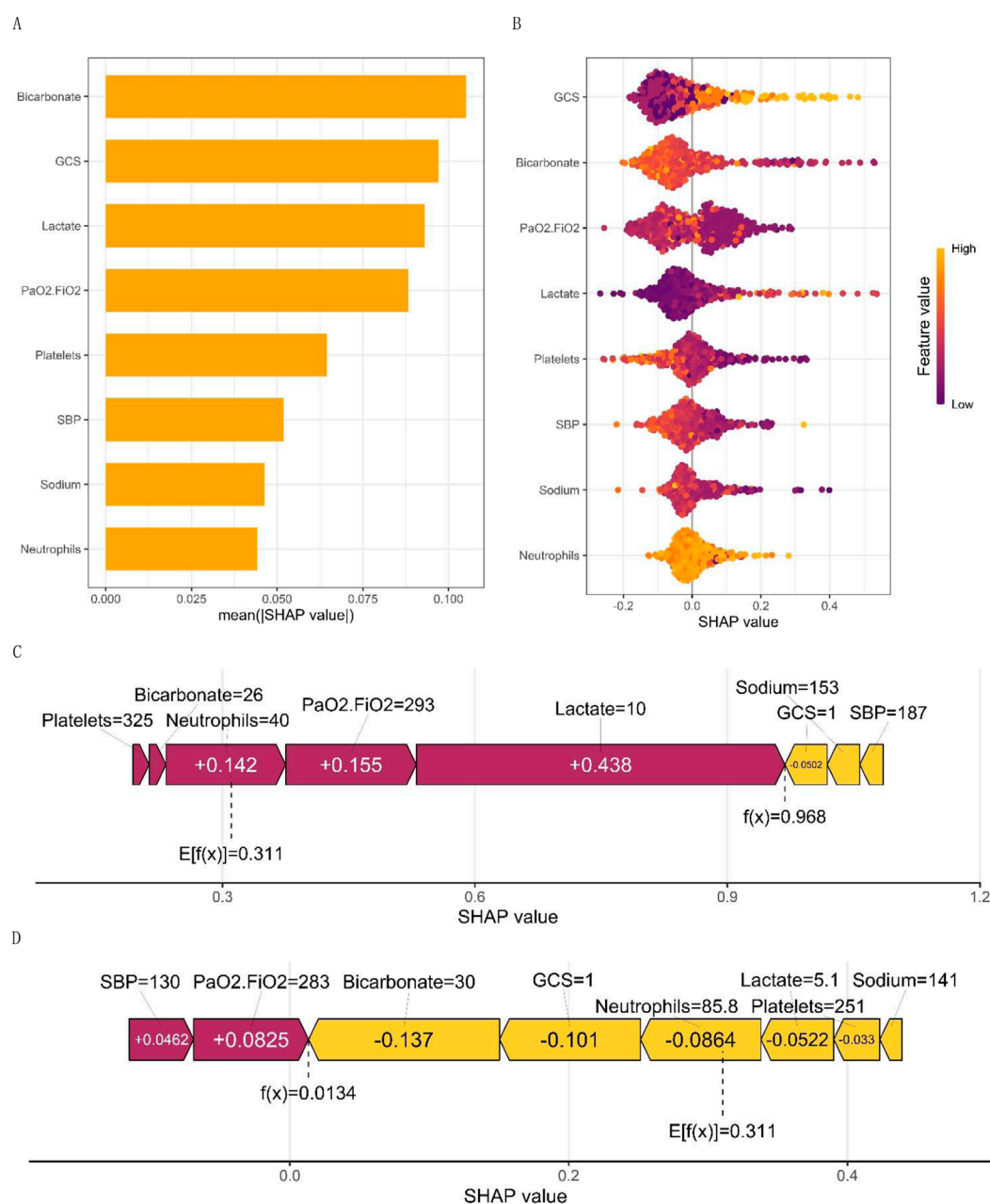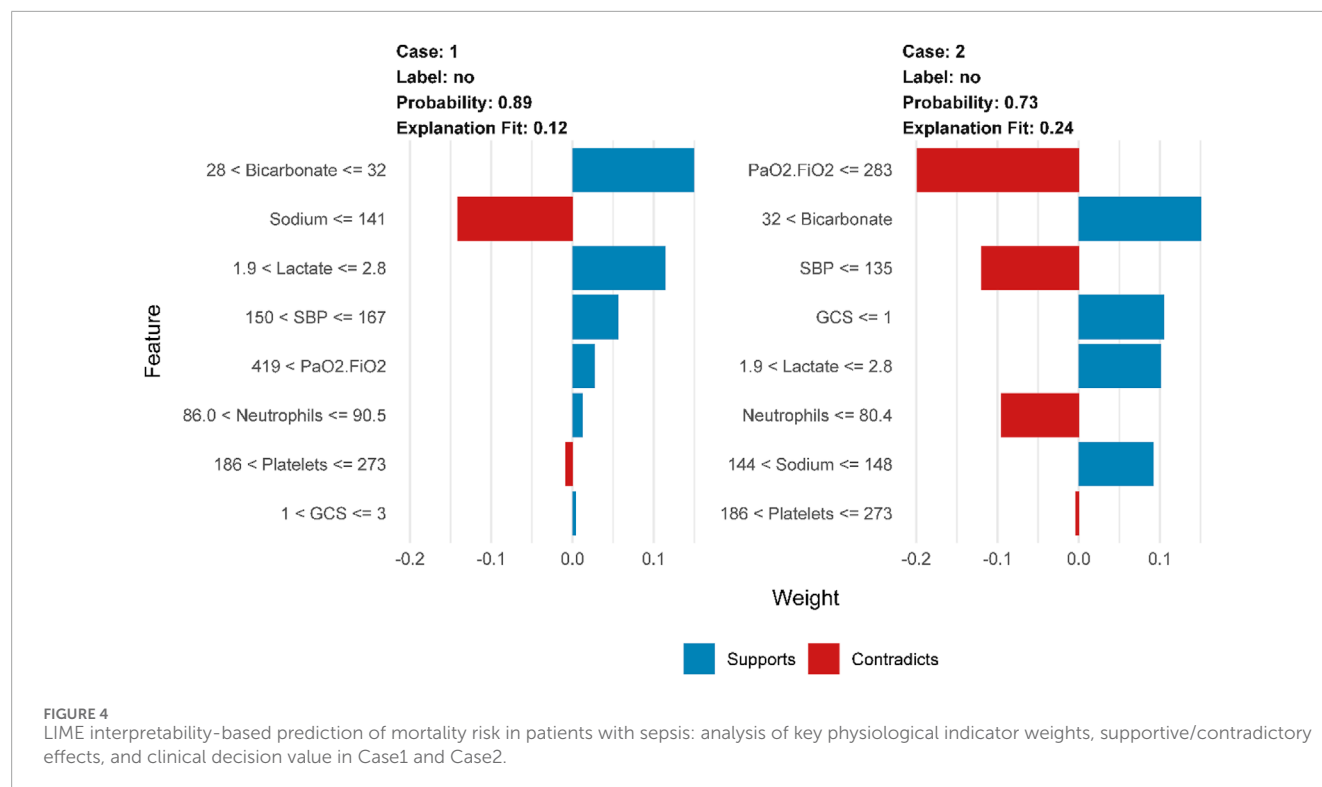
FIGURE 3
(A) displays the SHAP values for macro feature importance, while **(B)** presents a scatterplot of macro feature density. Each row of the plot represents a feature, with the SHAP value as the horizontal coordinate. Features are ranked according to the average absolute value of SHAP, the most important feature of the model. Wide areas indicate a large number of samples clustered together. The diagram employs a dot to represent a sample, with yellow indicating a higher value of the feature and purple indicating a lower value. This generates a ranking graph of the feature's significance. **(C, D)** show the micro single-sample feature influence diagram.

corresponding weight assignments for these cases. The gradient from red to blue in the figure shows the dynamic evolution of the feature weights. The dark blue region indicates strong support for the predicted outcome, while the increasingly reddish regions suggest a gradual weakening of the support for the predicted outcome. The construction of the prediction model involves feature ranges represented by vertical coordinates. The shade of their color directly maps the importance and influence of the feature in the model. These feature range values may be more suitable for clinicians to make decisions and may be more easily understood.
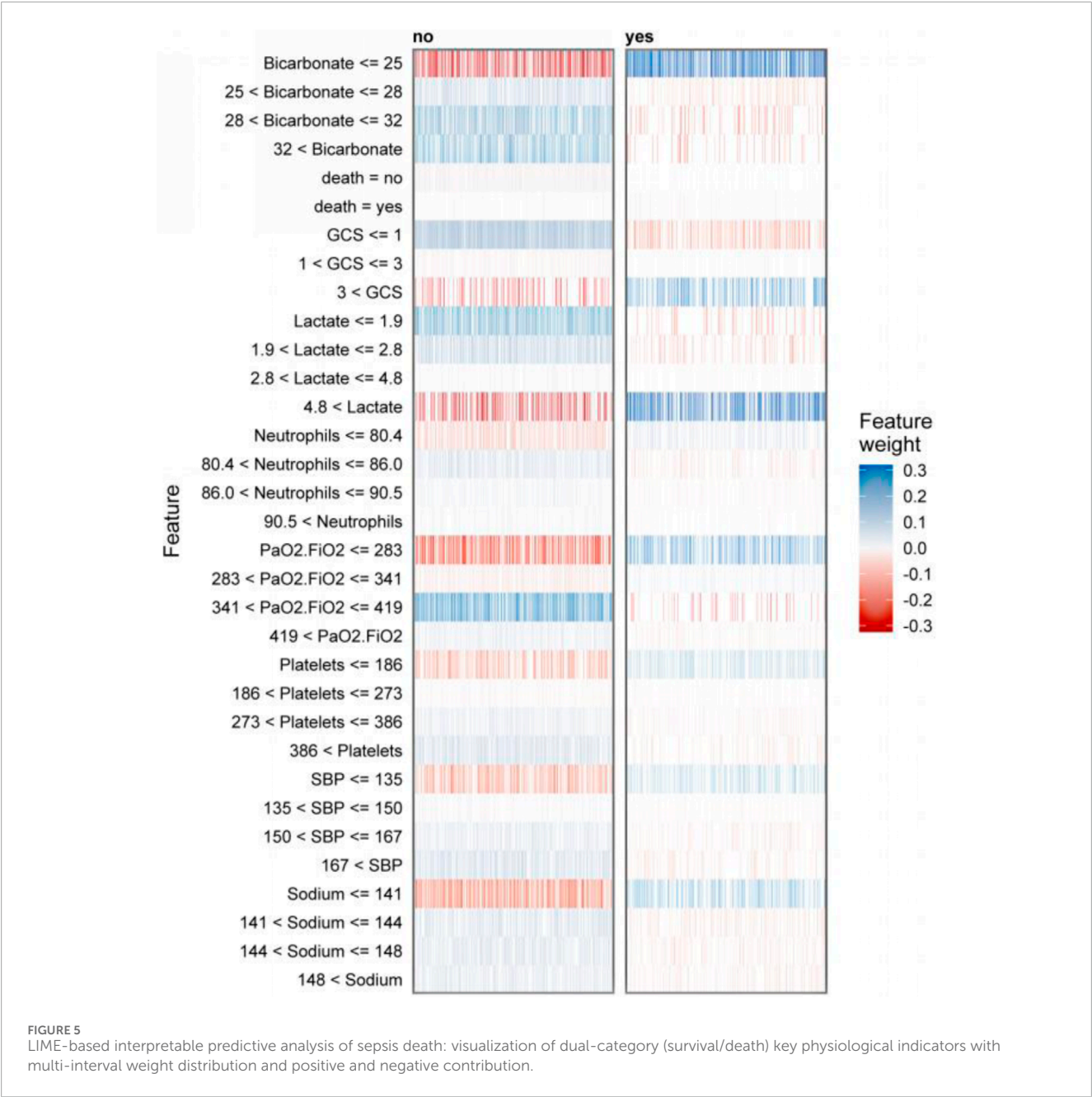
Discussion

In this study, a machine learning model was developed and validated to predict short-term mortality risk in elderly sepsis patients. The model is designed to be highly interpretable and easy

FIGURE 4
LIME interpretability-based prediction of mortality risk in patients with sepsis: analysis of key physiological indicator weights, supportive/contradictory effects, and clinical decision value in Case1 and Case2.

to understand. This study analyzed 49 variables of elderly sepsis patients in detail after hospital admission, including demographic data, vital signs, and laboratory test indices. A feature recursive elimination algorithm was used to accurately select the eight key feature variables with the highest association with the risk of death from a large number of feature variables for use in constructing the model. The study suggests that the XGBoost algorithm may be more effective than other machine learning algorithms in predicting short-term death risk in elderly sepsis patients. XGBoost is known for its efficiency, flexibility, and widespread use in data mining, medicine, and other fields (Li J. et al., 2022; Hou et al., 2020).

In recent years, machine learning-based predictive models have become increasingly prevalent. However, it is important to acknowledge that these models can be difficult to interpret, often resembling opaque 'black boxes' that hinder understanding of the decision-making process, even for those who comprehend the underlying mathematical algorithms (Fleuren et al., 2020; Giannini et al., 2019). We aim to establish trust and encourage the use of our machine-learning predictive models among physicians. To achieve this, we utilized two advanced interpretable analysis techniques, SHAP and lime, to provide a detailed analysis of our XGBoost machine learning predictive model. Our analysis systematically explored the associations between characteristic variables and the risk of death from sepsis in the elderly. We are confident that our approach will provide physicians with the necessary information to make informed decisions (Azodi et al., 2020). The lime interpretable analysis technique has the advantage of representing the weighting of the predictive model accounted for by the ranges of the characteristic variables. This approach is clinically applicable and provides physicians with clear and intuitive decision support.

Figure 5 displays the interpreter that was constructed after screening 2,500 samples. This demonstrates the full range of feature values and their corresponding feature importance weights in the prediction. The figure displays negative values in red and positive values in blue, with the shade of blue indicating the strength of support for the conclusion. It is worth noting that the right side of the graph represents cases where deaths occurred, while the left side represents cases where no deaths occurred. According to the graph, it can be observed that when bicarbonate levels are 25 mEq/L or lower, there is a significant increase in the risk of death for patients. This finding highlights the importance of monitoring bicarbonate levels in patient care. Acid-base imbalances are a common occurrence in critically ill patients, and it is important to address them promptly and effectively (Achanti and Szerlip, 2023). According to a cohort study, there is an association between low bicarbonate levels and increased mortality (Mitra et al., 2020). According to the study, patients with a GCS score greater than three were found to have an increased risk of death. It has been previously suggested by research that hypotension accompanying an abnormal GCS can be a crucial indicator for identifying patients at high risk of sepsis infection (Lane et al., 2020). The study suggests that a lactate value exceeding 4.8 mmol/L is a significant high-risk factor for sepsis patients facing mortality risk. Furthermore, Liu and Yang et al.'s study provides evidence of a strong association between plasma lactate levels and poor prognosis and mortality prediction in sepsis patients (Liu et al., 2019; Yang et al., 2022). In elderly patients with sepsis, a PaO2/FiO2 ratio below 283 is a significant risk indicator for life-threatening conditions. It is crucial to intervene rapidly when the ratio decreases to prevent patient mortality, as confirmed by previous studies (Annane et al., 2017). Platelet count is a critical factor in thrombosis, and thrombocytopenia is

FIGURE 5
LIME-based interpretable predictive analysis of sepsis death: visualization of dual-category (survival/death) key physiological indicators with multi-interval weight distribution and positive and negative contribution.

a prevalent condition in sepsis patients. Therefore, platelet count serves as a crucial indicator for assessing sepsis severity (Iba and Levy, 2018; Lyons et al., 2018).

There are, of course, limitations to this study. The data used in this study were obtained from publicly available databases, which may not include all the necessary variables. Moreover, the study sample is predominantly from Western countries, which may restrict the applicability of our model to other ethnic groups. Additionally, the retrospective and observational nature of this study may be susceptible to selection bias. We are confident that the model developed in this study can accurately predict the short-term risk of mortality in elderly patients with sepsis.

Conclusion

We have developed an XGBoost model that is easy to understand and accurately predicts the risk of death in elderly sepsis patients. Our interpretable machine learning tools have helped to identify the risk factors for elderly patients with sepsis, which has increased the confidence of healthcare providers in the predictions. The model's variable influence ranges and weight assignments are easily understandable, which makes it more practical for clinical applications. This provides physicians with a clear and intuitive basis for decision support. It is worth noting that this feature enhances the model's credibility and reliability.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://physionet.org/content/mimiciv/2.2/. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Patient information was collected and research resources were created with approval from the Beth Israel Deaconess Medical Centre's Institutional Review Board, which waived the need for informed consent and supported the data-sharing initiative. 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 participant's legal guardians/next of kin because Patient information was collected and research resources were created with approval from the Beth Israel Deaconess Medical Centre's Institutional Review Board, which waived the need for informed consent and supported the data-sharing initiative.

Author contributions

X-YZ: Data curation, Software, Validation, Visualization, Writing–original draft. Z-MJ: Investigation, Validation, Writing–original draft. X-L: Methodology, Visualization, Writing–original draft. Z-WL: Methodology, Validation, Writing–original draft. J-WT: Investigation, Methodology, Writing–review and editing. F-FS: Conceptualization, Investigation, Methodology, Writing–review and editing.

Funding

Acknowledgments

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

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

Achanti, A., and Szerlip, H. M. (2023). Acid-base disorders in the critically ill patient. *Clin. J. Am. Soc. Nephrol. CJASN* 18 (1), 102–112. Epub 2022/08/24. doi:10.2215/cjn.04500422

Angus, D. C., Linde-Zwirble, W. T., Lidicker, J., Clermont, G., Carcillo, J., and Pinsky, M. R. (2001). Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit. care Med.* 29 (7), 1303–1310. Epub 2001/07/11. doi:10.1097/00003246-200107000-00002

Annane, D., Pastores, S. M., Rochwerg, B., Arlt, W., Balk, R. A., Beishuizen, A., et al. (2017). Guidelines for the diagnosis and management of critical illness-related corticosteroid insufficiency (circi) in critically ill patients (Part I): society of critical care medicine (sccm) and European society of intensive care medicine (esicm) 2017. *Crit. care Med.* 45 (12), 2078–2088. Epub 2017/09/25. doi:10.1097/ccm.0000000000002737

Azodi, C. B., Tang, J., and Shiu, S. H. (2020). Opening the black box: interpretable machine learning for geneticists. *Trends Genet. TIG* 36 (6), 442–455. Epub 2020/05/13. doi:10.1016/j.tig.2020.03.005

Brunkhorst, F. M., Wegscheider, K., Forycki, Z. F., and Brunkhorst, R. (2000). Procalcitonin for early diagnosis and differentiation of sirs, sepsis, severe sepsis, and septic shock. *Intensive care Med.* 26 (Suppl. 2), S148–S152. Epub 2008/05/13. doi:10.1007/bf02900728

Das, B. K., and Dutta, H. S. (2020). Gfnb: gini index-based fuzzy naive Bayes and blast cell segmentation for leukemia detection using multi-cell blood smear images. *Med. and Biol. Eng. and Comput.* 58 (11), 2789–2803. Epub 2020/09/16. doi:10.1007/s11517-020-02249-y

Epstein, L., Dantes, R., Magill, S., and Fiore, A. (2016). Varying estimates of sepsis mortality using death certificates and administrative codes–United States, 1999-2014. *MMWR Morb. Mortal. Wkly. Rep.* 65 (13), 342–345. Epub 2016/04/08. doi:10.15585/mmwr.mm6513a2

Fahmy, A. S., Csecs, I., Arafati, A., Assana, S., Yankama, T. T., Al-Otaibi, T., et al. (2022). An explainable machine learning approach reveals prognostic significance of right ventricular dysfunction in nonischemic cardiomyopathy. *JACC Cardiovasc. imaging* 15 (5), 766–779. Epub 2022/01/17. doi:10.1016/j.jcmg.2021.11.029

Fleischmann, C., Scherag, A., Adhikari, N. K., Hartog, C. S., Tsaganos, T., Schlattmann, P., et al. (2016). Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am. J. Respir. Crit. care Med.* 193 (3), 259–272. Epub 2015/09/29. doi:10.1164/rccm.201504-0781OC

Fleuren, L. M., Klausch, T. L. T., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., et al. (2020). Machine learning for the prediction of sepsis: a systematic review

and meta-analysis of diagnostic test accuracy. *Intensive care Med.* 46 (3), 383–400. Epub 2020/01/23. doi:10.1007/s00134-019-05872-y

Gao, S., Yang, W., Xu, M., Zhang, H., Yu, H., Qian, A., et al. (2023). U-mlp: mlp-based ultralight refinement network for medical image segmentation. *Comput. Biol. Med.* 165, 107460. Epub 2023/09/14. doi:10.1016/j.compbiomed.2023.107460

Gattinoni, L., Vasques, F., Camporota, L., Meessen, J., Romitti, F., Pasticci, I., et al. (2019). Understanding lactatemia in human sepsis. Potential impact for early management. *Am. J. Respir. Crit. care Med.* 200 (5), 582–589. Epub 2019/04/16. doi:10.1164/rccm.201812-2342OC

Giacomello, E., and Toniolo, L. (2021). Nutrition, diet and healthy aging. *Nutrients* 14 (1), 190. Epub 2022/01/12. doi:10.3390/nu14010190

Giannini, H. M., Ginestra, J. C., Chivers, C., Draugelis, M., Hanish, A., Schweickert, W. D., et al. (2019). A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit. care Med.* 47 (11), 1485–1492. Epub 2019/08/08. doi:10.1097/ccm.0000000000003891

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 101 (23), E215–E220. Epub 2000/06/14. doi:10.1161/01.cir.101.23.e215

Hernandez, G., Bellomo, R., and Bakker, J. (2019). The ten pitfalls of lactate clearance in sepsis. *Intensive care Med.* 45 (1), 82–85. Epub 2018/05/14. doi:10.1007/s00134-018-5213-x

Heymans, M. W., and Twisk, J. W. R. (2022). Handling missing data in clinical research. *J. Clin. Epidemiol.* 151, 185–188. Epub 2022/09/24. doi:10.1016/j.jclinepi.2022.08.016

Hilton, C. B., Milinovich, A., Felix, C., Vakharia, N., Crone, T., Donovan, C., et al. (2020). Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. *NPJ Digit. Med.* 3, 51. Epub 2020/04/15. doi:10.1038/s41746-020-0249-z

Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., et al. (2020). Predicting 30-days mortality for mimic-iii patients with sepsis-3: a machine learning approach using xgboost. *J. Transl. Med.* 18 (1), 462. Epub 2020/12/09. doi:10.1186/s12967-020-02620-5

Iba, T., and Levy, J. H. (2018). Inflammation and thrombosis: roles of neutrophils, platelets and endothelial cells and their interactions in thrombus formation during sepsis. *J. thrombosis haemostasis JTH* 16 (2), 231–241. Epub 2017/12/02. doi:10.1111/jth.13911

Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., et al. (2023). Mimic-iv, a freely accessible electronic health record dataset. *Sci. data* 10 (1), 1. Epub 2023/01/04. doi:10.1038/s41597-022-01899-x

Karim, M. R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., et al. (2023). Explainable ai for bioinformatics: methods, tools and applications. *Briefings Bioinforma.* 24 (5), bbad236. Epub 2023/07/21. doi:10.1093/bib/bbad236

Kaukonen, K. M., Bailey, M., Suzuki, S., Pilcher, D., and Bellomo, R. (2014). Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. *Jama* 311 (13), 1308–1316. Epub 2014/03/19. doi:10.1001/jama.2014.2637

Lane, D. J., Wunsch, H., Saskin, R., Cheskes, S., Lin, S., Morrison, L. J., et al. (2020). Epidemiology and patient predictors of infection and sepsis in the prehospital setting. *Intensive care Med.* 46 (7), 1394–1403. Epub 2020/05/30. doi:10.1007/s00134-020-06093-4

Lee, C., Light, A., Alaa, A., Thurtle, D., van der Schaar, M., and Gnanapragasam, V. J. (2021). Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the surveillance, epidemiology, and end results (seer) database. *Lancet Digital health* 3 (3), e158–e165. Epub 2021/02/08. doi:10.1016/s2589-7500(20)30314-9

Li, A., Ling, L., Qin, H., Arabi, Y. M., Myatra, S. N., Egi, M., et al. (2022a). Epidemiology, management, and outcomes of sepsis in icus among countries of differing national wealth across asia. *Am. J. Respir. Crit. care Med.* 206 (9), 1107–1116. Epub 2022/06/29. doi:10.1164/rccm.202112-2743OC

Li, J., Liu, S., Hu, Y., Zhu, L., Mao, Y., and Liu, J. (2022b). Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. *J. Med. Internet Res.* 24 (8), e38082. Epub 2022/08/10. doi:10.2196/38082

Liu, Y., Zheng, J., Zhang, D., and Jing, L. (2019). Neutrophil-lymphocyte ratio and plasma lactate predict 28-day mortality in patients with sepsis. *J. Clin. laboratory analysis* 33 (7), e22942. Epub 2019/07/03. doi:10.1002/jcla.22942

Lu, R., Dumonceaux, T., Anzar, M., Zovoilis, A., Antonation, K., Barker, D., et al. (2024). Mnbc: a multithreaded minimizer-based naïve Bayes classifier for improved metagenomic sequence classification. *Bioinforma. Oxf. Engl.* 40 (10), btae601. Epub 2024/10/13 21:10. doi:10.1093/bioinformatics/btae601

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable ai for trees. *Nat. Mach. Intell.* 2 (1), 56–67. Epub 2020/07/02. doi:10.1038/s42256-019-0138-9

Lundberg, S. M., and Lee, S.-I. (2017). "A unified approach to interpreting model predictions," in *Proceedings of the 31st international conference on neural information processing systems*. Long Beach, California, USA: Curran Associates Inc., 4768–4777.

Lyons, P. G., Micek, S. T., Hampton, N., and Kollef, M. H. (2018). Sepsis-associated coagulopathy severity predicts hospital mortality. *Crit. care Med.* 46 (5), 736–742. Epub 2018/01/27. doi:10.1097/ccm.0000000000002997

Manimegalai, P., Suresh Kumar, R., Valsalan, P., Dhanagopal, R., Vasanth Raj, P. T., and Christhudass, J. (2022). 3d convolutional neural network framework with deep learning for nuclear medicine. *Scanning* 2022, 9640177. Epub 2022/08/05. doi:10.1155/2022/9640177

Martin, G. S., Mannino, D. M., Eaton, S., and Moss, M. (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *N. Engl. J. Med.* 348 (16), 1546–1554. Epub 2003/04/18. doi:10.1056/NEJMoa022139

Mitra, B., Roman, C., Charters, K. E., O'Reilly, G., Gantner, D., and Cameron, P. A. (2020). Lactate, bicarbonate and anion gap for evaluation of patients presenting with sepsis to the emergency department: a prospective cohort study. *Emerg. Med. Australasia EMA* 32 (1), 20–24. Epub 2019/06/12. doi:10.1111/1742-6723.13324

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., et al. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clin. Epidemiol.* 9, 157–166. Epub 2017/03/30. doi:10.2147/clep.S129785

Peng, K., and Menzies, T. (2021). "Documenting evidence of a reuse of '"Why should I trust you?": explaining the predictions of any classifier," in *Proceedings of the 29th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering*xx.

Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., and Fanos, V. (2020). Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Med. Kaunas. Lith.* 56 (9), 455. Epub 2020/09/12. doi:10.3390/medicina56090455

Rhee, C., Dantes, R., Epstein, L., Murphy, D. J., Seymour, C. W., Iwashyna, T. J., et al. (2017). Incidence and trends of sepsis in us hospitals using clinical vs claims data, 2009-2014. *Jama* 318 (13), 1241–1249. Epub 2017/09/14. doi:10.1001/jama.2017.13836

Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., et al. (2020). Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study. *Lancet London, Engl.* 395 (10219), 200–211. Epub 2020/01/20. doi:10.1016/s0140-6736(19)32989-7

Sarkar, S., and Mali, K. (2023). Firefly-svm predictive model for breast cancer subgroup classification with clinicopathological parameters. *Digit. health* 9, 20552076231207203. Epub 2023/10/20. doi:10.1177/20552076231207203

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama* 315 (8), 801–810. Epub 2016/02/24. doi:10.1001/jama.2016.0287

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ Clin. Res. ed* 338, b2393. Epub 2009/07/01. doi:10.1136/bmj.b2393

Takefuji, Y. (2025). Beyond xgboost and shap: unveiling true feature importance. *J. Hazard. Mater.* 488, 137382. Epub 2025/01/30. doi:10.1016/j.jhazmat.2025.137382

Van Calster, B., Wynants, L., Verbeek, J. F. M., Verbakel, J. Y., Christodoulou, E., Vickers, A. J., et al. (2018). Reporting and interpreting decision curve analysis: a guide for investigators. *Eur. Urol.* 74 (6), 796–804. Epub 2018/09/23. doi:10.1016/j.eururo.2018.08.038

Walkey, A. J., Wiener, R. S., and Lindenauer, P. K. (2013). Utilization patterns and outcomes associated with central venous catheter in septic shock: a population-based study. *Crit. care Med.* 41 (6), 1450–1457. Epub 2013/03/20. doi:10.1097/CCM.0b013e31827caa89

Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., et al. (2020). Brief introduction of medical database and data mining Technology in big data era. *J. evidence-based Med.* 13 (1), 57–69. Epub 2020/02/23. doi:10.1111/jebm.12373

Yang, K., Fan, M., Wang, X., Xu, J., Wang, Y., Tu, F., et al. (2022). Lactate promotes macrophage Hmgb1 lactylation, acetylation, and exosomal release in polymicrobial sepsis. *Cell death Differ.* 29 (1), 133–146. Epub 2021/08/08. doi:10.1038/s41418-021-00841-9

Zabor, E. C., Reddy, C. A., Tendulkar, R. D., and Patil, S. (2022). Logistic regression in clinical studies. *Int. J. Radiat. Oncol. Biol. Phys.* 112 (2), 271–277. Epub 2021/08/21. doi:10.1016/j.ijrobp.2021.08.007

Zhang, M., Wen, G., Zhong, J., Chen, D., Wang, C., Huang, X., et al. (2023). Mlp-like model with convolution complex transformation for auxiliary diagnosis through medical images. *IEEE J. Biomed. health Inf.* 27 (9), 4385–4396. Epub 2023/07/19. doi:10.1109/jbhi.2023.3292312

Zhou, S. (2022). Sparse svm for sufficient data reduction. *IEEE Trans. pattern analysis Mach. Intell.* 44 (9), 5560–5571. Epub 2021/04/24. doi:10.1109/tpami.2021.3075339