Journal of Translational Medicine

RESEARCH

Open Access

Developing a rapid screening tool for high-risk ICU patients of sepsis: integrating electronic medical records with machine learning methods for mortality prediction in hospitalized patients—model establishment, internal and external validation, and visualization

Songchang Shi[1†], Lihui Zhang[1†], Shujuan Zhang[1], Jinyang Shi[4], Donghuang Hong[3], Siqi Wu[5], Xiaobin Pan[1] and Wei Lin[2*]

Abstract

Objectives To develop a machine learning-based prediction model using clinical data from the first 24 h of ICU admission to enable rapid screening and early intervention for sepsis patients.

Methods This multicenter retrospective cohort study analyzed electronic medical records of sepsis patients using machine learning methods. We evaluated model performance in predicting sepsis outcomes within the first 24 h of ICU admission across US and Chinese healthcare settings.

Results From 31 clinical features, machine learning models demonstrated significantly better predictive performance than traditional approaches for sepsis outcomes. While linear regression achieved low test scores (0.25), machine learning methods reached scores of 0.78 and AUCs above 0.8 in testing. Importantly, these models maintained robust performance (scores 0.63–0.77) in external validation.

Conclusions The application of machine learning-based prediction models for sepsis could significantly improve patient outcomes through early detection and timely intervention in the critical first 24 h of ICU admission, supporting clinical decision-making.

Keywords Sepsis, Machine learning, Mortality, Prediction, Visualization

†Songchang Shi and Lihui Zhang have contributed equally to this work.

*Correspondence:
Wei Lin
caolalin0929@163.com
Full list of author information is available at the end of the article

Introduction

Sepsis remains a leading cause of mortality worldwide [1]. Recent studies indicate that approximately 49 million people globally continue to battle sepsis, accounting for 20% of all deaths internationally [2].

Despite the success of quality improvement programs in reducing sepsis mortality rates in high-income countries [1], low- and middle-income nations continue to shoulder a disproportionate burden of this condition, with persistently high mortality rates [3]. Furthermore, current evidence suggests that the mortality rate reductions achieved through existing sepsis quality improvement measures may not be sustainable. Consequently, the quest for novel and more effective quality improvement strategies remains a critical and urgent challenge to address.

The delayed recognition of sepsis severity often results in postponed interventions and heightened mortality, presenting a major challenge in sepsis treatment [4]. Early identification of clinical indicators related to sepsis prognosis is now acknowledged as a key strategy for enhancing the quality of sepsis care [4–6]. Machine learning has demonstrated significant strengths in the development of early prediction models and has found extensive application across various clinical domains [7–9]. The establishment of an early prediction model for sepsis using machine learning techniques enables clinicians to identify high-risk patients at initial stages, focus on clinical features that are pertinent to outcomes, and engage in timely intervention during early management [10].

This research involved extracting clinical data from electronic medical records and employing machine learning techniques to develop a predictive model. This model was based on the clinical data of patients during the initial 24 h following their admission to the Intensive Care Unit (ICU). The study is crucial for the early detection of high-risk patients and for identifying effective clinical data from an individual standpoint, which can be pivotal of developing a Rapid Screening Tool for prompt and early intervention.

Methods

Database and definition

This retrospective study utilized the Medical Information Mart for Intensive Care (MIMIC)-IV database (version 2.2, available at https://mimic.mit.edu/iv/). This database encompasses the medical records of all patients admitted to the ICU at Beth Israel Deaconess Medical Center between 2008 and 2019 [11]. Access to this database was granted following the author (S.S.) successfully completing the Protecting Human Research Participants exam (Record ID: 44174677).

Sepsis patients were selected from the MIMIC-IV 2.2 database, adhering to the sepsis 3.0 definition for criteria [12]. When sepsis was identified, it was marked by an acute change in the sequential organ failure assessment (SOFA) score of 2 points or more, stemming from an infection [13, 14]. The inclusion criteria are as follows: (1) Patients aged between 18 and 80 years; (2) An ICU stay exceeding 24 h; (3) First-time ICU admissions. The exclusion criteria include: (1) Individuals with malignant tumors; (2) Those suffering from chronic kidney disease; (3) Pregnant patients (Fig. 1).

This study was conducted in compliance with the Helsinki Declaration and its subsequent amendments. MIMIC-IV is an anonymized public database. Approval for this project was granted by the institutional review boards of both the Massachusetts Institute of Technology (MIT) and Beth Israel Deaconess Medical Center (BIDMC), and it received a waiver of informed consent.

The external validation dataset was sourced from the department of Critical Care Medicine of the Fujian Provincial Hospital Jinshan Branch, in the southeast of China. Patient inclusion spanned from January 2023 to November 2023, adhering to the same inclusion and exclusion criteria as previously established. This study received approval from the Ethics Review Committee of Fujian Provincial Hospital (K2023-12-010).

Characteristic variable

The demographic data collected includes age, gender, race, height, weight, insurance category, language, and marital status. Laboratory indicators were extracted as average values recorded within the first 24 h of ICU admission. These include complete blood count, blood biochemistry tests, arterial blood gas analysis, coagulation function, lipid profiles, liver and kidney function tests, and myocardial enzyme spectrum. The analysis also considered the presence and stage of Acute Kidney Injury (AKI) [15]. Among these, any with a missing value ratio exceeding 20% were excluded.

Statistical analysis

PgAdmin4 was utilized for data extraction from the MIMIC-IV database, enabling the execution of structured query language (SQL) commands. The R software package (version 4.2.3) is used for data preprocessing, while Python (version 3.7) is applied both for the establishment and validation of the machine learning model, and for the visualization display of the model. Statistical significance is defined as a P value <0.05.

Initially, the get_dummies function in Python is utilized to process categorical variables in the dataset. Following a missing value analysis, the gaps are filled using the KNN (K-Nearest Neighbors) method (n_neighbors = 5) [9, 10,

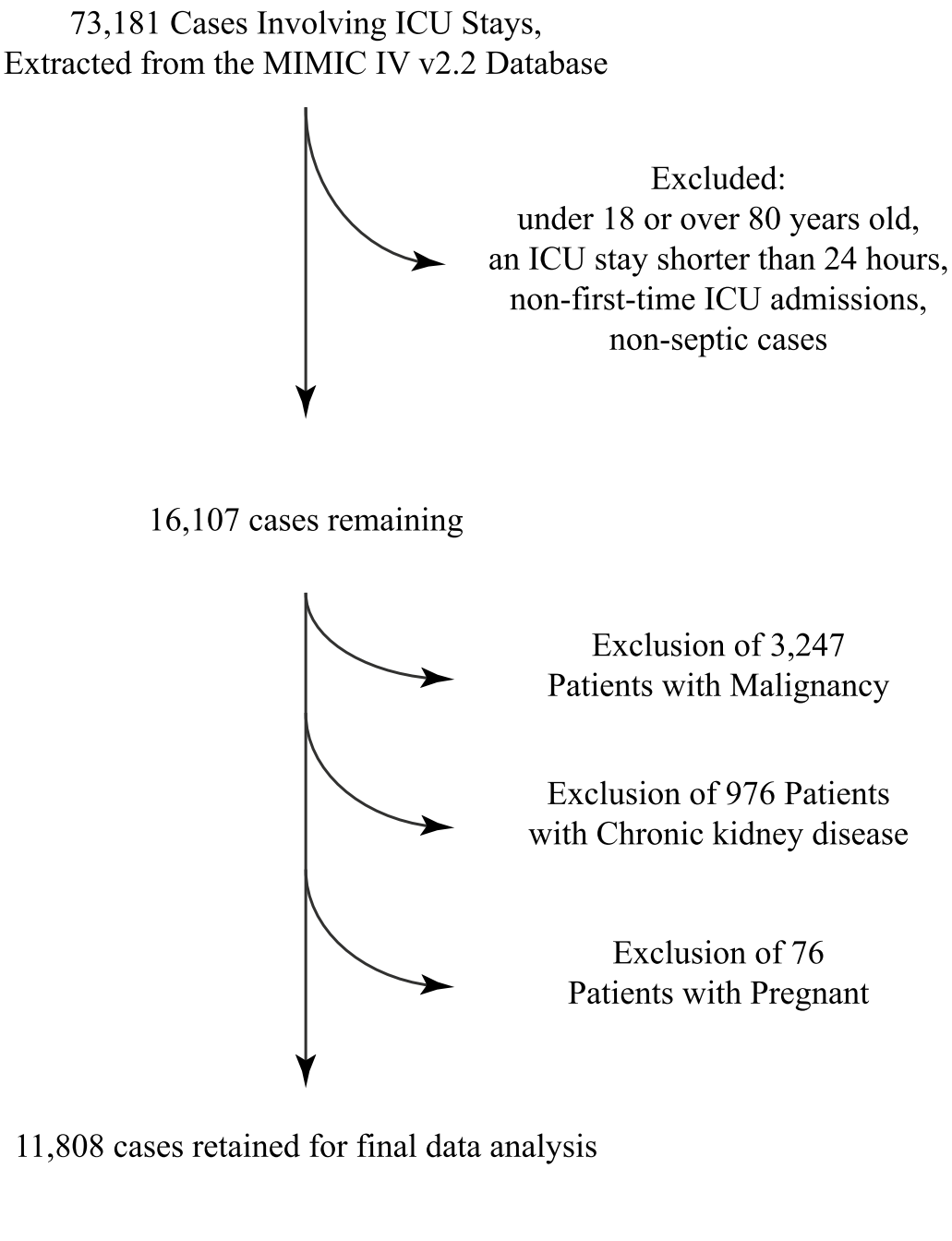

Fig. 1 Flowchart of selection

16]. The dataset is then split into a training set (8,265 cases) and a test set (3,543 cases) in a 7:3 ratio using a machine learning algorithm. Due to the frequent issue of class imbalance in clinical data, this study implemented two preprocessing methods to mitigate overfitting and underfitting concerns. Firstly, it transformed all variables into a normal distribution by normalizing the dataset. Secondly, it utilized logistic regression with an L2 penalty term (penalty = "l2", C = 0.05) within the machine learning framework to identify characteristic variables. To establish the prediction model, two traditional algorithms, linear regression and lasso regression, are employed. Subsequently, nine commonly used machine learning algorithms [17, 18]—logistic regression, GNB (Gaussian

Naïve Bayes), KNN, SVM (Support Vector Machine), ANN (Artificial Neural Network), decision tree, random forest, GBM (Gradient Boosting Machine), and CatBoost (Categorical Boosting)—are also applied for model development. The efficacy of the established prediction model is evaluated using various methods such as model scoring, confusion matrix, classification report, ROC (Receiver Operating Characteristic), Five-Fold Cross-Validation, DCA (Decision Curve Analysis), and probability curves. The external validation set underwent a detailed analysis of missing values, followed by the application of KNN (n_neighbors = 5) for missing value interpolation. Post-standardization, the dataset was imported into the pre-constructed machine learning model for comprehensive validation. Models are evaluated using various methods such as model scoring, confusion matrix, classification report, ROC, DCA through external validation set. Finally, the model is interpreted through SHAP (SHapley Additive exPlanations) visualization and demonstrated through individualized instances, enhancing its potential application in clinical settings.

Results

All characteristic variables are presented in Supplemental 1. A total of 11,808 patients were included in the study, of which 3,419 died in the hospital, while 8,389 survived. The average age of the patients included was 59 years. Among them, 7,305 (61.9%) were male, and 4,503 (38.1%) were female. CRRT was administered to 690 patients (5.8%), and 7,371 patients (62.4%) received treatment with invasive mechanical ventilation. Analysis of missing values is presented in Supplemental 2.

In comparing model performance for sepsis mortality prediction, traditional approaches (linear and LASSO regression) showed limited ability (test scores: 0.25–0.26). Machine learning models achieved significantly better performance, with GBM reaching the highest test score (0.78), followed by Random Forest and Catboost (0.77). While decision trees and random forests achieved perfect training scores (1.0), their lower test scores (0.70–0.77) indicated overfitting. Machine learning models showed strong performance for survival predictions (precision: 0.75–0.80, F1-scores: 0.80–0.85), but lower accuracy for death predictions (precision: 0.48–0.69, F1-scores: 0.47–0.56). Table 1 presents detailed metrics.

Figure 2 displays the test performance of each machine learning prediction model in the test set. Figure 2A presents the test outcomes with varying results, while Fig. 2B illustrates the overall accuracy of the test set. The findings indicate that the machine learning prediction models developed in this study's training set were effective in predicting the mortality outcomes of sepsis patients

Table 1 Evaluation of model performance: Score, Precision, Recall, F1-score

	Train set score	Test set score	Validation set score		Test set			Validation set		
					Precision	Recall	F1-score	Precision	Recall	F1-score
Linear Regression	0.26	0.25								
Lasso Regression	0.26	0.25								
Logistic	0.79	0.77	0.68	Live	0.80	0.91	0.85	0.71	0.93	0.80
				Die	0.67	0.46	0.54	0.40	0.11	0.17
GNB	0.76	0.75	0.77	Live	0.80	0.87	0.83	0.78	0.93	0.85
				Die	0.60	0.47	0.53	0.70	0.39	0.50
KNN	0.82	0.75	0.63	Live	0.78	0.89	0.83	0.69	0.88	0.77
				Die	0.61	0.40	0.49	0.17	0.06	0.08
SVM	0.78	0.77	0.67	Live	0.79	0.93	0.85	0.71	0.88	0.79
				Die	0.69	0.40	0.51	0.38	0.17	0.23
ANN	0.90	0.76	0.68	Live	0.80	0.86	0.83	0.76	0.81	0.78
				Die	0.61	0.50	0.55	0.47	0.39	0.42
Decision Tree	1.0	0.70	0.63	Live	0.77	0.80	0.79	0.75	0.71	0.73
				Die	0.48	0.45	0.47	0.40	0.44	0.42
Random Forest	1.0	0.77	0.70	Live	0.80	0.91	0.85	0.71	0.98	0.82
				Die	0.68	0.44	0.54	0.50	0.06	0.10
GBM	0.82	0.78	0.68	Live	0.80	0.91	0.85	0.69	0.98	0.81
				Die	0.69	0.47	0.56	0.00	0.00	0.00
Catboost	0.91	0.77	0.72	Live	0.80	0.90	0.85	0.72	0.98	0.83
				Die	0.67	0.47	0.55	0.67	0.11	0.19

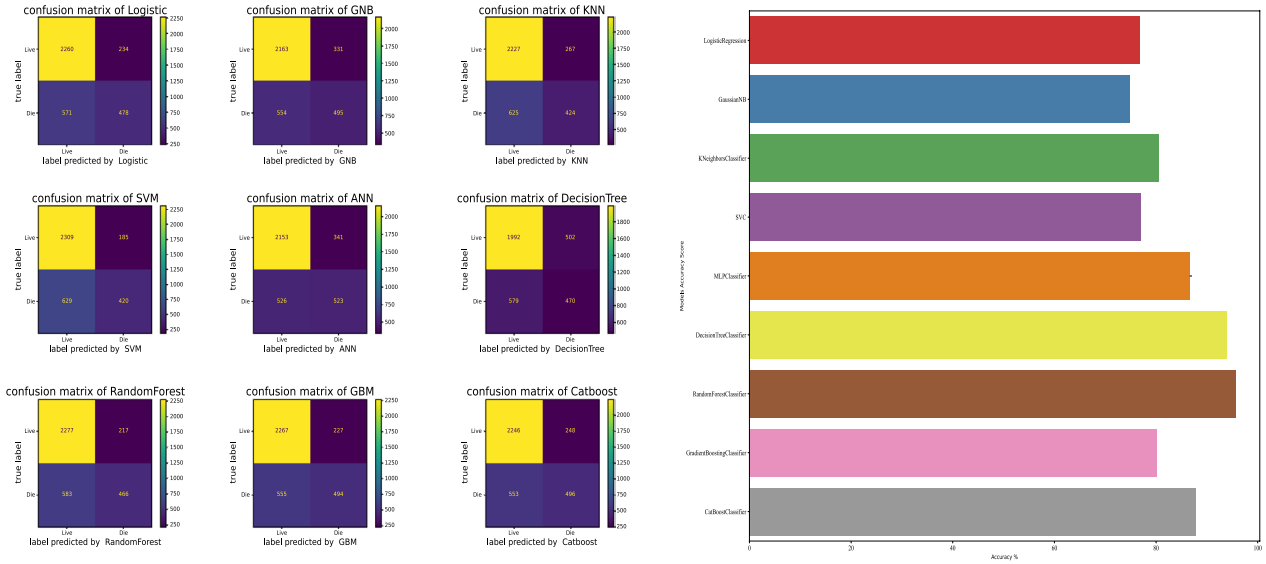

Fig. 2 Performance evaluation of predictive models. **A** Illustration of the confusion matrix to eexhibit the predictive effecacy of machine learning models in the test set; **B** Illustration of the accureacy of machine learning models

in the test set, achieving an accuracy rate of over 70% (excluding decision trees). Notably, the accuracy rates of Logistic, SVM, ANN, RF (Random Forest), GBM, and Catboost models all exceeded 75%.

Beyond accuracy, the model's performance was also rigorously evaluated. Figure 3A illustrates the ROC results in both training and testing sets, demonstrating that Logistic, SVM, RF, GBM, and Catboost models achieve impressive effectiveness with AUCs (Area Under the Curves) greater than 0.8 in the testing set. Figure 3B presents the outcomes of cross-validation, highlighting the stability of Logistic, SVM, RF, GBM, and Catboost with AUCs exceeding 0.8 in the test set. In Fig. 3C, DCA (Decision Curve Analysis) reveals that, in the majority of models, a substantial net benefit was evident across a broad spectrum of threshold probabilities, particularly in ANN, RF, GBM, and Catboost models. Figure 3D visualizes the probability curves, indicating minimal deviations for models other than the decision tree. Figure 3E offers a detailed visualization of the weights of the 31 key features, showcasing the specific significance of each important feature.

Table 1 and Fig. 4 display the results of the validation using external datasets. All characteristic variables are presented in Supplemental 3. Analysis of missing values is presented in Supplemental 4. The prediction model, developed through machine learning methods, achieved a score range of 0.63–0.77 on the external dataset, comparable to the test set results. Figure 4A details the predicted outcomes for various outcome variables. Figure 4B illustrates the ROC curves of prediction models, developed using diverse machine learning methods across training, testing, and validation sets. The AUC values for logistic regression, GNB, ANN, Decision Tree, Random Forest, and Catboost all exceed 0.60. Figure 4C presents the DCA for the external validation set, indicating that logistic regression, GNB, ANN, Decision Tree, and Catboost demonstrate potential benefits at certain thresholds.

The black-box properties of machine learning methodologies often obscures the interpretability of the models they generate [19]. To address this, we have incorporated SHAP values into our machine learning model, enhancing its interpretability from both a global and local standpoint. This enhancement facilitates more practical applications in clinical settings. Figure 5A, named Summary Plot, depicts a line consisting of numerous discrete points, where each point signifies an individual's SHAP value, with values greater than zero indicating an increased risk. The color spectrum, ranging from red to blue, represents the magnitude of feature variables—with deeper reds indicating higher values and darker blues indicating lower values. The line's length correlates with the variable's influence on the outcome. This visualization highlights the significant influence of factors such as the Charlson score, red blood cell distribution width within the first 24 h, APSIII score, and body weight on the predicted outcomes. Specifically, a higher Charlson score, wider red blood cell distribution in the initial 24 h, and an elevated APSIII score are associated with an increased risk of mortality. Figure 5B, named Dependence Plot, illustrates a dependency plot, revealing the relationship

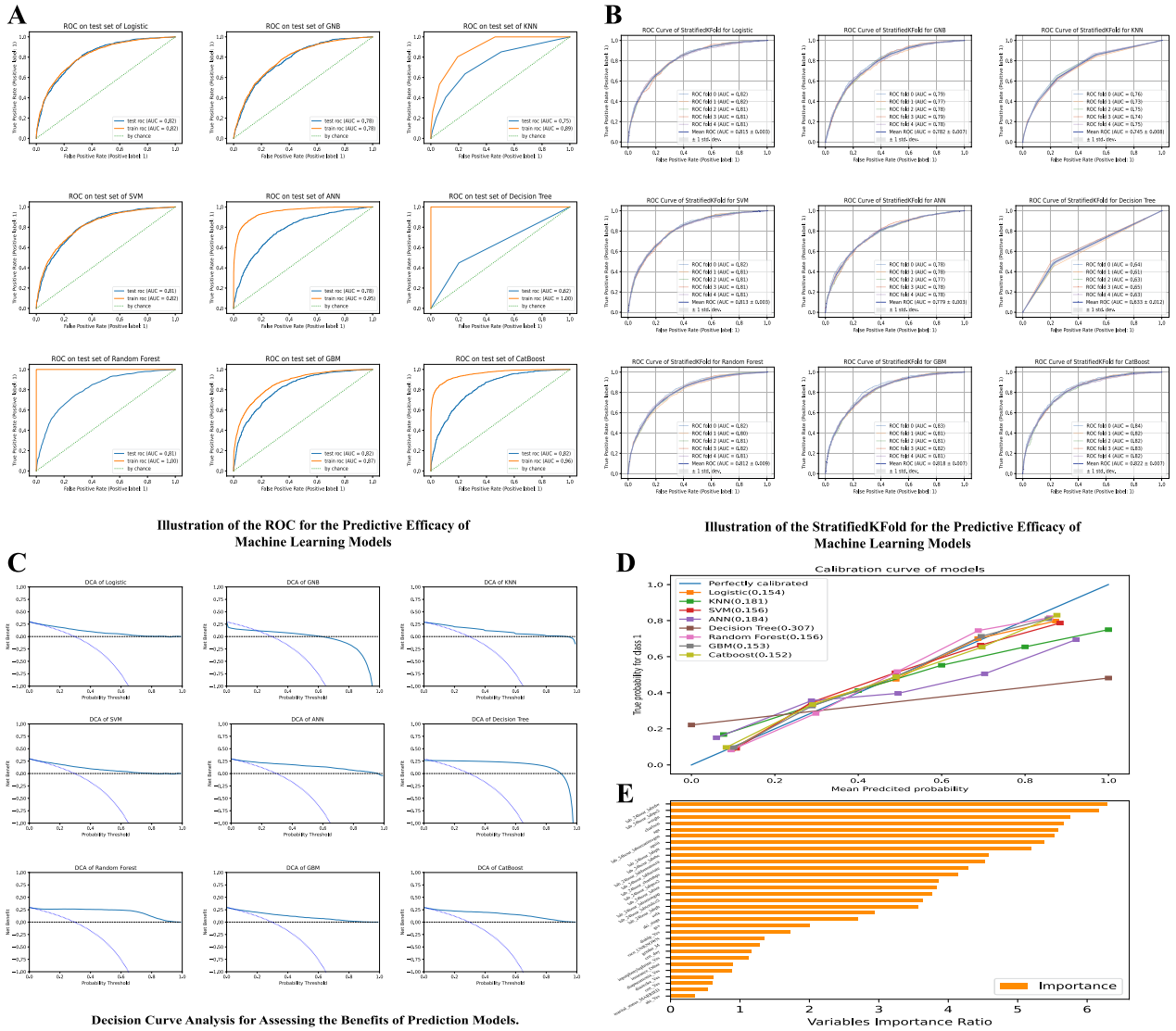

Fig. 3 Validation of predictive model performance and visualization of feature variables. **A** The ROC for predictive efficacy of models; **B** The stratifiedKFold for the predictive efficacy of models; **C** DCA; **D** Calibration curve of models; **E** The visualization of feature variables

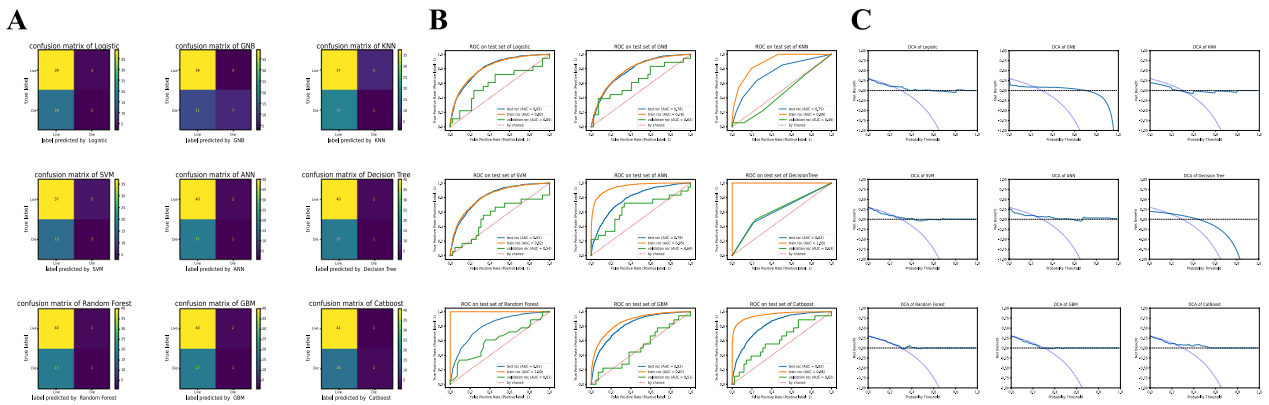

Fig. 4 External validation of predictive model performance. **A** The confusion matrix for the external validation of predictive model; **B** The ROC for the external validation of predictive model; **C** The DCA for the external validation of predictive model

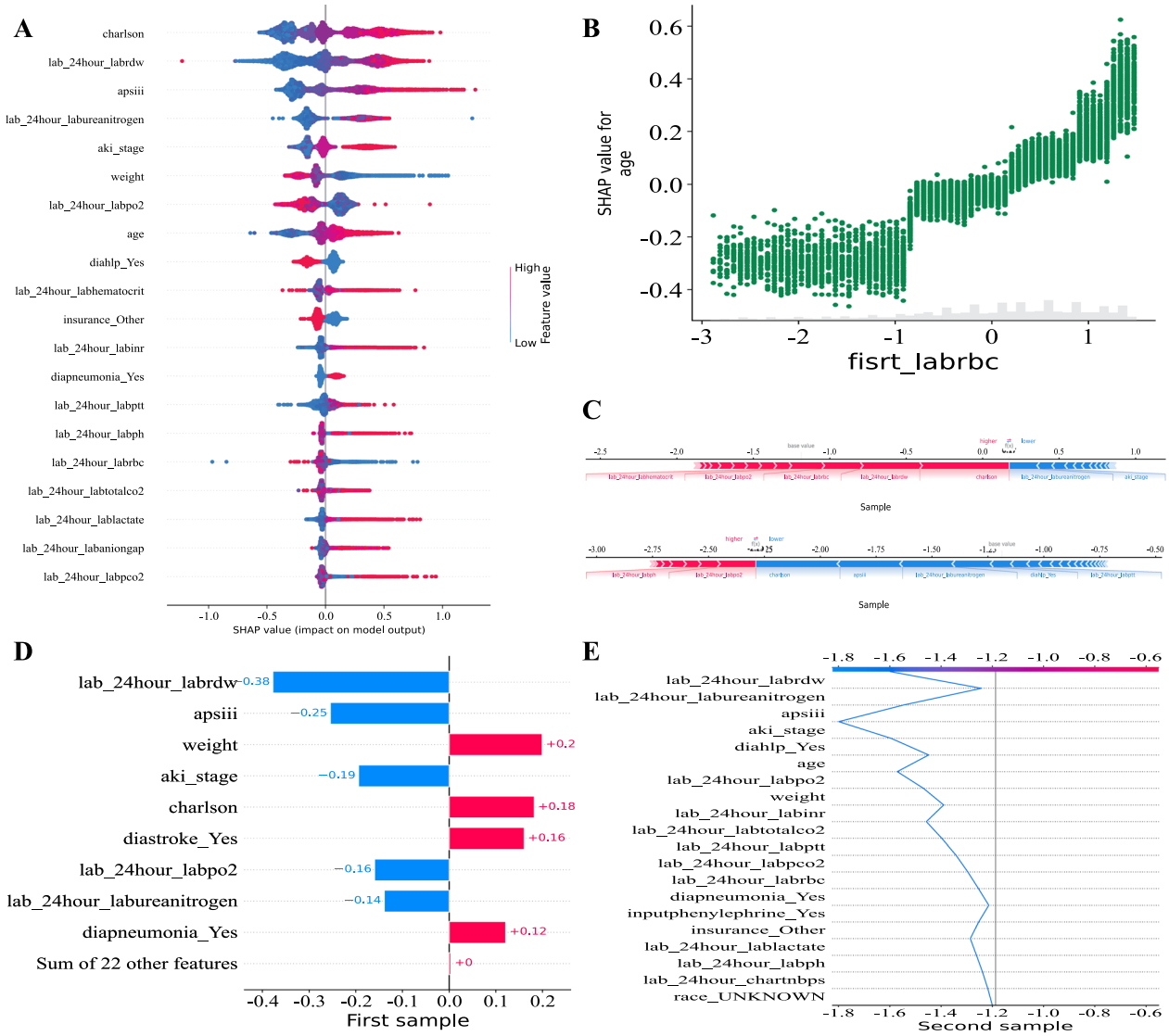

Fig. 5 Visual representation of predictive mode. **A** Summary plot; **B** Dependence plot; **C** SHAP force plot; **D** Waterfall plot; **E** Decision plot

between the Charlson score and SHAP values within the model. Notably, higher Charlson scores correspond with SHAP values situated in the high-risk zone. Additionally, this plot allows for the examination of interactions between average carbon dioxide partial pressure in the initial 24 h and the Charlson and SHAP values, discernible through color variations.

The implementation of SHAP not only allows for a global interpretation of machine learning models but also enables individual-specific analysis and personalized guidance. Figure 5C, referred to as the SHAP Force Plot, illustrates the aggregate SHAP value denoted as f(x). In the first example, a positive f(x) value signals a heightened risk of mortality for the patient, with factors such as the Charlson score, average width of red blood cells in the first 24 h, red blood cell count, and oxygen partial pressure acting as contributing risk elements. Conversely, in the second example, a negative f(x) value denotes a lower risk profile for the patient, although average pH and oxygen partial pressure measured every 24 h remain as risk factors. Figure 5D, known as a Waterfall Plot of SHAP, delineates the influence of each variable on the predicted outcome, displaying risk factors in red and protective factors in blue. This plot focuses on the first patient in the test set, highlighting their weight, Charlson score, and comorbid conditions like stroke and pneumonia as risk contributors, with specific weight values presented within the figure. Lastly, Fig. 5E, also termed

the decision plot of SHAP, visualizes the SHAP values of significant characteristic variables for individuals. It features the second patient from the test set, emphasizing that their average red blood cell width and APSIII score within the first 24 h serve as protective factors, reflected in their low SHAP values.

Discussion

Sepsis represents a complex and heterogeneous syndrome shaped by diverse host and environmental factors. Its clinical trajectory is often marked by rapid deterioration, underscoring the necessity for early and accurate identification. Epidemiological evidence suggests that sepsis disproportionately affects older adults, males, and individuals with preexisting chronic conditions. Consistent with these findings, the majority of patients in our cohort were aged over 60, with a substantial proportion diagnosed with comorbidities such as hypertension and diabetes. These observations underscore the need for customized predictive tools that account for the demographic and clinical complexities of the target population [5]. However, existing screening tools exhibit diminished sensitivity in predicting mortality [4, 20]. Relying on these tools with lower sensitivity could result in delayed diagnoses and overlooked cases [4, 20].

This study involved the extraction and analysis of 31 clinical features from the electronic health records of patients during the initial 24 h of ICU admission and discharge, culminating in the development of a predictive model for sepsis mortality. The model underwent rigorous performance testing, internal validation, and external validation, confirming its effectiveness in managing sepsis patients across two different medical centers. The incorporation of model visualization enhances its usability and practicality.

Timely identification of high-risk patients in cases of sepsis is crucial for reducing mortality associated with the condition. In light of this, the third international consensus definition for sepsis and septic shock has established the qSOFA (Quick Sequential Organ Failure Assessment) criteria, with the hope of simplifying and improving the practicality of early sepsis screening. However, the role of qSOFA in early sepsis detection remains a topic of debate [21, 22]. Machine learning, a subset of artificial intelligence (AI) algorithms, excels in rapidly processing multidimensional clinical data [9, 22]. This includes extensive patient information, laboratory findings, and imaging outcomes. AI efficiently automates data collection according to specific parameters and processes it at high velocity, enabling the rapid prediction of data trends. This assists clinicians in making quicker, more accurate diagnoses, facilitating early intervention, and potentially lowering mortality rates.

Previously, sepsis risk models based on conventional regression approaches lacked test and validation set verification, leading to subpar reproducibility. Moreover, traditional statistical methods fell short in their ability to automate data collection for validation purposes. In contrast, machine learning methodologies enable the testing of established models through both internal and external validation, thus more effectively verifying the model's predictive capabilities [23]. Consequently, machine learning approaches are gaining increasing recognition and are being widely implemented in early warning systems for diseases [24]. They facilitate the early detection of dynamic changes in clinical features and the investigation of how these changes influence prediction outcomes, thereby aiding in early warning and intervention strategies [25].

The evolution of artificial intelligence in the realm of sepsis management is ongoing [26]. Currently, significant advancements have been made in developing early prediction models for sepsis or septic shock using machine learning techniques. In recent years, a majority of machine learning models designed for the early prediction of sepsis have achieved an accuracy rate exceeding 80% [27]. Furthermore, these models typically boast an AUC higher than 0.8, reflecting their robust predictive capability in this critical area of medical diagnosis [6]. However, research on mortality prediction models for sepsis is relatively nascent and limited. While relying solely on AUC may not comprehensively represent a model's effectiveness [9, 28, 29], many existing machine learning-based prediction models primarily undergo ROC analysis. The prediction model developed by this research institute undergoes thorough validation using a complete suite of machine model assessment techniques, including prediction accuracy, AUC, cross-validation, DCA, and calibration curves, to ensure the stability and reliability of its performance.

External validation is an essential phase in the transition of predictive models from development to practical application, and thus, it receives considerable emphasis. Presently, the vast majority of machine learning-based prediction models for sepsis remain in the internal validation stage. Ying et al. developed a mortality prediction model for sepsis using the RuleFit method and proceeded with external validation [30]. However, only the AUC from the external validation was reported. This study carried out external validation employing third-party, real-world data, achieving a maximum accuracy of over 75%. This result further affirms the model's generalizability and applicability in diverse settings.

Although the application of machine learning across various medical fields has garnered increasing recognition for enhancing clinical practice and facilitating

personalized treatment, its inherent lack of explainability remains a challenge that needs addressing. The inability to effectively visualize models developed through machine learning can significantly hinder their promotion and application. Currently, machine learning prediction models for sepsis have not adequately resolved the issue of model visualization. This study integrates SHAP into the machine learning-based prediction model. SHAP's Summary Plot and Dependence Plot elucidate the influence of key clinical features in model construction and their impact on outcomes. Furthermore, SHAP's Force Plot and Waterfall Plot enable monitoring of each newly admitted ICU patient for high-risk status, identifying specific clinical features and their relative weights that influence their prognosis. This aids clinical practitioners in making informed, targeted decisions based on these insights.

This study represents the most comprehensive analysis of sepsis prediction using machine learning, encompassing data preprocessing, model development, validation, and visualization. Our approach transformed advanced machine learning techniques into a practical clinical tool, establishing groundwork for future multicenter studies. Specifically, the algorithm was designed and validated for the first 24 h of ICU admission—a critical timeframe where early intervention significantly improves survival. Our analysis demonstrates optimal performance when applied at ICU admission, enabling prompt identification of high-risk patients. While our current focus is this 24-h window, future research could explore the algorithm's utility in emergency department settings or for longer-term predictions.

Building on this foundation, we validated the algorithm's performance across different healthcare systems. Originally developed using US patient data, testing in Chinese healthcare settings revealed consistent predictive accuracy despite variations in clinical practices (such as sepsis bundle implementation), ICU admission criteria, available medical resources, and patient characteristics. Notable population differences included varying prevalence of comorbidities like diabetes, chronic kidney disease, and COPD—conditions our model identified as significant mortality risk factors. While these results are promising, additional multicenter validation studies are needed to further strengthen our findings.

To optimize model performance and ensure reliable validation results, we implemented careful patient selection criteria. Specifically, we excluded patients with incomplete medical records to maintain data quality, and those receiving end-of-life care or with terminal illnesses to focus on patients most likely to benefit from early sepsis detection and intervention.

Conclusion

This study demonstrates that machine learning-based prediction models for sepsis can enhance clinical outcomes through accurate risk assessment during the critical first 24 h of ICU admission. Despite healthcare system variations between US and Chinese populations, our algorithm maintained reliable predictive performance, suggesting broad applicability. These findings establish a foundation for larger multicenter clinical trials and the practical implementation of AI-assisted sepsis management in diverse healthcare settings.

Abbreviations

AI	Artificial Intelligence
AKI	Acute Kidney Injury
ANN	Artificial Neural Network
APS-III	Acute Physiology Score III
AUCs	Area Under the Curves
BIDMC	Beth Israel Deaconess Medical Center
Catboost	Categorical Boosting
CRRT	Continuous Renal Replacement Therapy
DCA	Decision Curve Analysis
EMRs	Electronic Medical Records
GBM	Gradient Boosting Machine
GCS	Glasgow Coma Scale
GNB	Gaussian Naïve Bayes
ICU	Intensive Care Unit
KNN	K-Nearest Neighbors
MIMIC	Medical Information Mart for Intensive Care
MIT	Massachusetts Institute of Technology
OASIS	Oxford Acute Severity of Illness Score
qSOFA	Quick Sequential Organ Failure Assessment
RF	Random Forest
ROC	Receiver Operating Characteristic
SAPS II	Simplified Acute Physiology Score II
SHAP	Shapley Additive exPlanations
SIRS	Systemic Inflammatory Response Syndrome
SOFA	Sequential Organ Failure Assessment
SQL	Structured Query Language
SVM	Support Vector Machine

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06102-4.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Acknowledgements

We thank all participants who provided data for this study.

Author contributions

SS and LZ: Formal analysis, methodology, original draft. SZ, JS, DH, SW: Data curation, resources. XP, WL: Conceptualization, formal analysis, original draft, project administration. SS, LZ and WL: Manuscript revision. WL: Study guarantor with full data access, responsible for data integrity and analysis accuracy.

Funding

This study was supported by Fujian Research and Training Grants for Young and Middle-aged Leaders in Healthcare (Grant number: (2023)417#), Fujian Joint Fund for Scientific and Technological Innovation Projects (Grant No. 2023Y9295, Natural Science Foundation of Fujian Province (Grant number:

2022J011017), Innovation Project of Fujian Provincial Health Commission (Grant number: 2021CXA003), and Startup Fund for scientific research, Fujian Medical University (Grant number: 2022QH1282).

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
This study was conducted in accordance with the Declaration of Helsinki and approved by The Ethics Committee of Fujian Provincial Hospital (approval number: K2023-12-010). Written informed consent was obtained from all participants.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Critical Care Medicine, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital South Branch, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fuzhou 350001, People's Republic of China. [2]Department of Endocrinology, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, No 134 Dongjie Street, Gulou District, Fuzhou, Fujian 350001, People's Republic of China. [3]Department of Critical Care Medicine, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fuzhou 350001, People's Republic of China. [4]Fujian Medical University, Fuzhou 350001, People's Republic of China. [5]Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fuzhou 350001, People's Republic of China.

Received: 9 February 2024 Accepted: 8 January 2025
Published online: 21 January 2025

References

1. Machado FR, Ferreira EM, Schippers P, de Paula IC, Saes LSV, de Oliveira FI, et al. Implementation of sepsis bundles in public hospitals in Brazil: a prospective study with heterogeneous results. Crit Care. 2017;21(1):268.
2. 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(10219):200–11.
3. Baykara N, Akalin H, Arslantas MK, Hanci V, Caglayan C, Kahveci F, et al. Epidemiology of sepsis in intensive care units in Turkey: a multicenter, point-prevalence study. Crit Care. 2018;22(1):93.
4. Machado FR, Cavalcanti AB, Monteiro MB, Sousa JL, Bossa A, Bafi AT, et al. Predictive accuracy of the quick sepsis-related organ failure assessment score in Brazil. A prospective multicenter study. Am J Respir Crit Care Med. 2020;201(7):789–98.
5. Nunnally ME, Ferrer R, Martin GS, Martin-Loeches I, Machado FR, De Backer D, et al. The surviving sepsis campaign: research priorities for the administration, epidemiology, scoring and identification of sepsis. Intensive Care Med Exp. 2021;9(1):34.
6. 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 Health. 2021;9:754348.
7. Lee R, Leighton SP, Thomas L, Gkoutos GV, Wood SJ, Fenton SH, et al. Prediction models in first-episode psychosis: systematic review and critical appraisal. Br J Psychiatry. 2022;220(Spec Iss 4 Themed Iss Precision Medicine and Personalised Healthcare in Psychiatry):1–13.
8. Lin W, Shi S, Huang H, Wang N, Wen J, Chen G. Development of a risk model for predicting microalbuminuria in the Chinese population using machine learning algorithms. Front Med. 2022;9:775275.
9. Shi S, Pan X, Zhang L, Wang X, Zhuang Y, Lin X, et al. An application based on bioinformatics and machine learning for risk prediction of sepsis at first clinical presentation using transcriptomic data. Front Genet. 2022;13:979529.
10. Pan X, Xie J, Zhang L, Wang X, Zhang S, Zhuang Y, et al. Evaluate prognostic accuracy of SOFA component score for mortality among adults with sepsis by machine learning method. BMC Infect Dis. 2023;23(1):76.
11. 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):1.
12. 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(3):304–77.
13. 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(11):1181–247.
14. He L, Yang D, Ding Q, Su Y, Ding N. Association between lactate and 28-day mortality in elderly patients with sepsis: results from MIMIC-IV database. Infect Dis Ther. 2023;12(2):459–72.
15. van der Slikke EC, Star BS, van Meurs M, Henning RH, Moser J, Bouma HR. Sepsis is associated with mitochondrial DNA damage and a reduced mitochondrial mass in the kidney of patients with sepsis-AKI. Crit Care. 2021;25(1):36.
16. Lin W, Shi S, Lan H, Wang N, Huang H, Wen J, et al. Identification of influence factors in overweight population through an interpretable risk model based on machine learning: a large retrospective cohort. Endocrine. 2023.
17. Tanaka T. [[Fundamentals] 5. Python+scikit-learn for machine learning in Medical Imaging]. Nihon Hoshasen Gijutsu Gakkai Zasshi. 2023;79(10):1189–93.
18. Lin W, Shi S, Huang H, Wen J, Chen G. Predicting risk of obesity in overweight adults using interpretable machine learning algorithms. Front Endocrinol. 2023;14:1292167.
19. Yang Y, Yuan Y, Han Z, Liu G. Interpretability analysis for thermal sensation machine learning models: an exploration based on the SHAP approach. Indoor Air. 2022;32(2):e12984.
20. 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 Inform Assoc. 2020;27(3):437–43.
21. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the emergency department. Am J Emerg Med. 2019;37(8):1490–7.
22. Kim T, Tae Y, Yeo HJ, Jang JH, Cho K, Yoo D, et al. Development and validation of deep-learning-based sepsis and septic shock early prediction system (DeepSEPS) using real-world ICU data. J Clin Med. 2023;12(22).
23. Layeghian Javan S, Sepehri MM, Layeghian Javan M, Khatibi T. An intelligent warning model for early prediction of cardiac arrest in sepsis patients. Comput Methods Programs Biomed. 2019;178:47–58.
24. Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. JMIR Med Inform. 2020;8(7):e18599.
25. Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P. Accessing artificial intelligence for clinical decision-making. Front Digit Health. 2021;3:645232.
26. 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 Inform Assoc. 2023;30(7):1349–61.
27. 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.
28. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic t, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.

29. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. Spine J. 2021;21(10):1643–8.
30. Wu Y, Huang S, Chang X. Understanding the complexity of sepsis mortality prediction via rule discovery and analysis: a pilot study. BMC Med Inform Decis Mak. 2021;21(1):334.

Publisher's Note