Prospective Cohort Study

INTERNATIONAL JOURNAL OF SURGERY

OPEN

Prediction of sepsis among patients with major trauma using artificial intelligence: a multicenter validated cohort study

Baisheng Sun, MD[a,b], Mingxing Lei, MD[a,d,f], Li Wang, MD[b], Xiaoli Wang, MM[b], Xiaoming Li, MM[g], Zhi Mao, MD[b], Hongjun Kang, MD, PhD[b], Hui Liu, MD[b], Shiying Sun, MD[c], Feihu Zhou, MD, PhD[b,e,*]

Background: Sepsis remains a significant challenge in patients with major trauma in the ICU. Early detection and treatment are crucial for improving outcomes and reducing mortality rates. Nonetheless, clinical tools for predicting sepsis among patients with major trauma are limited. This study aimed to develop and validate an artificial intelligence (AI) platform for predicting the risk of sepsis among patients with major trauma.

Patients and methods: This study involved 961 patients, with a prospective analysis of data from 244 patients with major trauma at our hospital and a retrospective analysis of data from 717 patients extracted from a database in the United States. The patients from our hospital constituted the model development cohort, and the patients from the database constituted the external validation cohort. The patients in the model development cohort were randomly divided into a training cohort and an internal validation cohort at a ratio of 8:2. The machine-learning algorithms used to train models included logistic regression, decision tree, extreme gradient boosting machine (eXGBM), neural network (NN), random forest, and light gradient boosting machine (LightGBM).

Results: The incidence of sepsis for the model development cohort was 43.44%. Twelve predictors, including gender, abdominal trauma, open trauma, red blood cell count, heart rate, respiratory rate, injury severity score, sequential organ failure assessment score, Glasgow coma scale, smoking, total protein concentrations, and hematocrit, were used as features in the final model. Internal validation showed that the NN model had the highest area under the curve (AUC) of 0.932 (95% CI: 0.917–0.948), followed by the LightGBM and eXGBM models with AUCs of 0.913 (95% CI: 0.883–0.930) and 0.912 (95% CI: 0.880–0.935), respectively. In the external validation cohort, the eXGBM model (AUC: 0.891, 95% CI: 0.866–0.914) had the highest AUC value, followed by the LightGBM model (AUC: 0.886, 95% CI: 0.860–0.906), and the AUC value of the NN model was only 0.787 (95% CI: 0.751–0.829). Considering the predictive performance for both the internal and external validation cohorts, the LightGBM model had the highest score of 82, followed by the eXGBM (81) and NN (76) models. Thus, the LightGBM has emerged as the optimal model, and it was deployed online as an AI application.

Conclusions: This study develops and validates an AI application to effectively assess the susceptibility of patients with major trauma to sepsis. The AI application equips healthcare professionals with a valuable tool to promptly identify individuals at high risk of developing sepsis. This will facilitate clinical decision-making and enable early intervention.

Keywords: external validation, ICU, machine learning, major trauma, prediction models, sepsis

[a]Chinese PLA Medical School, [b]Department of Critical Care Medicine, The First Medical Centre, Chinese PLA General Hospital, [c]State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, [d]National Clinical Research Center for Orthopedics, Sports Medicine and Rehabilitation, Chinese PLA General Hospital, [e]Medical Engineering Laboratory of Chinese PLA General Hospital, Beijing, [f]Department of Orthopedics, Hainan Hospital of Chinese PLA General Hospital, Hanan and [g]Department of Critical Care Medicine, Chongqing University Cancer Hospital, Chongqing, People's Republic of China

Baisheng Sun and Mingxing Lei contributed equally to the work.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

**Corresponding authors. Address: Department of Critical Care Medicine, The First Medical Centre, Chinese PLA General Hospital, Beijing 100853, People's Republic of China. Tel.: +86 010 668 755 48. E-mail: feihuzhou301@126.com (F. Zhou)*

International Journal of Surgery (2025) 111:467–480

Received 26 April 2024; Accepted 8 June 2024

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.lww.com/international-journal-of-surgery.

Published online 26 June 2024

http://dx.doi.org/10.1097/JS9.0000000000001866

Background

Sepsis is a life-threatening condition that occurs when the body's response to infection damages its tissues and organs[1]. Patients with trauma who are in the ICU have a higher risk of developing sepsis due to their weakened immune system and exposure to severe strikes. Studies have shown that sepsis is a common complication in patients with critical trauma in the ICU, with its prevalence ranging from 30.4%[2] to 33.3%[3]. Septic trauma can lead to organ dysfunction and failure, septic shock, and early death[4,5]. Its mortality rate among inpatients is high, ranging from 19.5% to 26.7%[6], and its associated healthcare cost remains high[7]. The epidemiology and financial burden of sepsis in patients with trauma in the ICU highlight the need for early detection and effective treatment of this condition.

The risk factors for sepsis among patients with trauma in the ICU include age, abnormal vital signs, comorbidities, severe injury, immunosuppression, and invasive procedures[2,8,9]. These factors could increase the risk of infection and subsequent sepsis. While identifying these risk factors is important for developing effective prevention and treatment strategies, the complexity of their interactions limits predictive accuracy. In addition, there was no single, universally applicable biomarker or algorithm that could accurately identify patients who will develop sepsis. To overcome these issues, several clinical prediction models have been developed to identify patients at a high risk of developing sepsis, such as the predictive score for sepsis risk in patients with trauma proposed in 2019[2] and the model for predicting the risk of sepsis in ICU-admitted patients with trauma in 2023[10]. The models facilitate prompt identification of patients at risk of sepsis, enabling personalized intervention and early treatment. Nonetheless, the predictive performances of these models need to be improved due to limited accuracy[2]. Further, they were not externally validated[2,10], and their generalizability may be limited. In addition, traditional statistical models may not fully capture these complexities, thus leading to suboptimal predictions and less effective clinical decision-making[11].

Artificial intelligence (AI) has emerged as a powerful tool for medical prediction, and it offers several advantages in healthcare. By leveraging machine-learning algorithms and big data analysis, AI has the potential to revolutionize medical prediction and improve patient outcomes. One key advantage of AI in medical prediction is its ability to process and analyze vast amounts of complex data quickly and accurately[11,12]. This enables AI to make accurate predictions and identify potential risks or outcomes of various medical conditions. Another advantage of AI is its potential to enhance clinical decision-making[13]. AI models can be trained on large datasets for diverse patient populations, allowing for more personalized and precise predictions. By considering several factors and variables, AI can provide clinicians with valuable insights and recommendations to guide treatment plans and interventions. In addition, AI may assist healthcare professionals in the early detection and diagnosis of diseases. By analyzing data from various sources, AI algorithms can identify subtle patterns or indicators of diseases at an early stage and enable timely intervention and treatment.

Therefore, the development of an AI platform for sepsis among patients with major trauma in the ICU is essential for improving their outcomes and reducing the burden of sepsis. The purpose of this study is to develop and validate a mobile application to evaluate the risk of sepsis among patients with major trauma in the ICU using machine-learning algorithms.

HIGHLIGHTS

- The light gradient boosting machine (LightGBM) model demonstrated superior prediction performance with high area under the curve (AUC) values.
- Web-based artificial intelligence (AI) application provides healthcare professionals with a valuable tool for early sepsis detection.
- A comprehensive scoring system incorporated nine evaluation metrics to assess prediction performance.
- The external validation cohort confirmed the reliability of the LightGBM model.
- Free and accessible online AI application facilitates prompt identification of high-risk individuals and intervention strategies.

Patients and methods

Patients and study design

This study included 961 patients. The data of 244 patients admitted to the ICU due to major trauma [injury severity score (ISS) ≥ 16] at our hospital between March 2017 and June 2021 were prospectively analyzed, while those of 717 patients were extracted from a publicly available database in the United States[14]. The 244 patients from our hospital constituted the model development cohort, and the patients from the database constituted the external validation cohort for the models. Patients were excluded if they (1) had an ISS of less than 16, (2) had trauma for more than 1 week before ICU admission, (3) only had a trauma site, (4) were younger than 18 years, (5) were pregnant, and (6) were discharged from ICU or died within 24 h after admission. If a patient had multiple admissions to ICU, the first admission was analyzed in this study. The patients from our hospital were randomly divided into the training and internal validation groups at a ratio of 8:2. The data of the training group were used to train and optimize the models, whereas the data of the internal validation group were used to internally validate the models. The data of the 717 patients with critical trauma extracted from the database were used for external validation of the model. Patients who were admitted to the ICU due to trauma were enrolled for analysis. The study design and patient selection flowchart is provided in Figure 1.

The study protocol was approved by the Ethics Committee Board of our hospital, and informed consent was obtained from each patient to review their medical records. This study was conducted in accordance with the Declaration of Helsinki. The use of the database was approved by the medical center[14], and we successfully completed the relevant courses on accessing the database and obtained a certificate. The work has been reported according to the STROCSS criteria[15] (Supplemental Digital Content 1, http://links.lww.com/JS9/C871).

Primary outcome

The primary outcome of the study was the diagnosis of sepsis during ICU hospitalization. In addition, the occurrence of sepsis

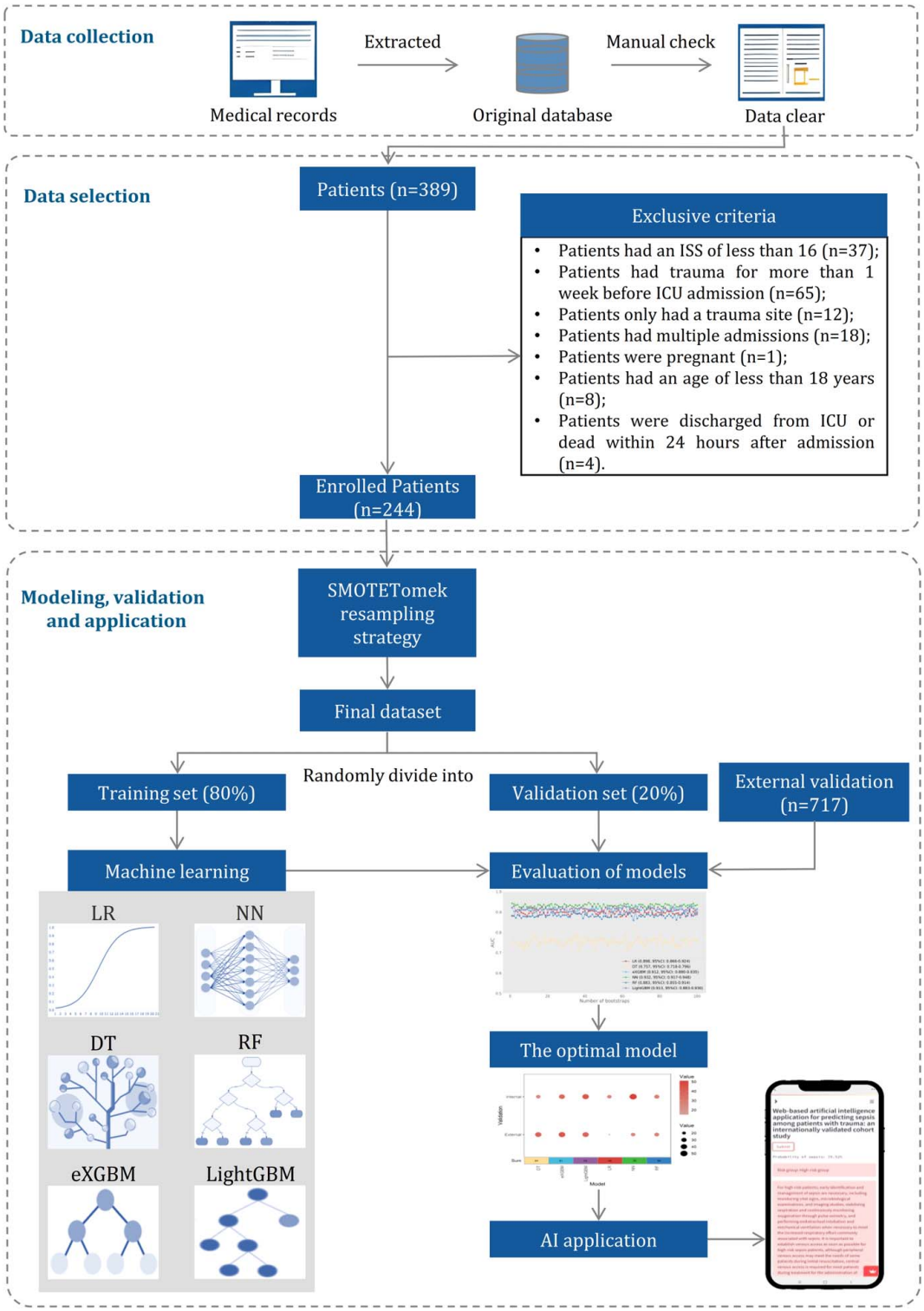

Figure 1. Study design and patient selection flowchart. DT, decision tree; eXGBM, extreme gradient boosting machine; LightGBM, light gradient boosting machine; LR, logistic regression; NN, neural network; RF, random forest; ISS, injury severity score.

Table 1

Patient's baseline clinical characteristics and comparison of clinical characteristics stratified by sepsis.

Characteristics	Overall	Sepsis		*P*
		No	Yes	
n	244	138	106	
Demographics				
Gender (female/male, %)	53/191 (21.7/78.3)	38/100 (27.5/72.5)	15/91 (14.2/85.8)	0.018
Age [years, mean (SD)]	45.09 (13.68)	45.01 (14.09)	45.20 (13.20)	0.918
BMI [kg/m^2, mean (SD)]	24.38 (3.79)	24.22 (4.09)	24.59 (3.38)	0.445
Smoking (no/yes, %)	124/120 (50.8/49.2)	78/60 (56.5/43.5)	46/60 (43.4/56.6)	0.057
Trauma information				
Trauma mechanism (%)				0.762
High fall	55 (22.5)	28 (20.3)	27 (25.5)	
Traffic collision	145 (59.4)	83 (60.1)	62 (58.5)	
Penetrating trauma	8 (3.3)	5 (3.6)	3 (2.8)	
Others	36 (14.8)	22 (15.9)	14 (13.2)	
Head trauma (no/yes, %)	165/79 (67.6/32.4)	87/51 (63.0/37.0)	78/28 (73.6/26.4)	0.108
Thoracic trauma (no/yes, %)	128/116 (52.5/47.5)	77/61 (55.8/44.2)	51/55 (48.1/51.9)	0.288
Abdominal trauma (no/yes, %)	196/48 (80.3/19.7)	120/18 (87.0/13.0)	76/30 (71.7/28.3)	0.005
Brain or spinal cord trauma (no/yes, %)	133/111 (54.5/45.5)	72/66 (52.2/47.8)	61/45 (57.5/42.5)	0.480
Open trauma (no/yes, %)	188/56 (77.0/23.0)	117/21 (84.8/15.2)	71/35 (67.0/33.0)	0.002
Comorbidities				
Hypertension (no/yes, %)	205/39 (84.0/16.0)	111/27 (80.4/19.6)	94/12 (88.7/11.3)	0.117
Diabetes (no/yes, %)	226/18 (92.6/7.4)	130/8 (94.2/5.8)	96/10 (90.6/9.4)	0.406
Coronary heart disease (no/yes, %)	239/5 (98.0/2.0)	134/4 (97.1/2.9)	105/1 (99.1/0.9)	0.540
Cerebral vascular disease (no/yes, %)	236/8 (96.7/3.3)	133/5 (96.4/3.6)	103/3 (97.2/2.8)	1.000
Chronic pulmonary disease (no/yes, %)	238/6 (97.5/2.5)	136/2 (98.6/1.4)	102/4 (96.2/3.8)	0.456
Liver disease (no/yes, %)	236/8 (96.7/3.3)	132/6 (95.7/4.3)	104/2 (98.1/1.9)	0.479
Renal disease (no/yes, %)	240/4 (98.4/1.6)	135/3 (97.8/2.2)	105/1 (99.1/0.9)	0.809
Medication use				
Aspirin (no/yes, %)	238/6 (97.5/2.5)	133/5 (96.4/3.6)	105/1 (99.1/0.9)	0.356
Warfarin (no/yes, %)	243/1 (99.6/0.4)	137/1 (99.3/0.7)	106/0 (100.0/0.0)	1.000
Beta blockers (no/yes, %)	241/3 (98.8/1.2)	135/3 (97.8/2.2)	106/0 (100.0/0.0)	0.346
Laboratory examination				
Total protein [g/l, mean (SD)]	50.50 (9.03)	51.40 (9.59)	49.34 (8.16)	0.078
Serum sodium [mmol/l, mean (SD)]	140.52 (4.89)	140.09 (4.70)	141.09 (5.10)	0.114
INR [mean (SD)]	1.29 (0.31)	1.27 (0.33)	1.31 (0.27)	0.361
Hematocrit [mean (SD)]	0.30 (0.17)	0.32 (0.21)	0.28 (0.06)	0.073
Red blood cell count [$\times 10^{12}$/l, mean (SD)]	3.29 (0.73)	3.39 (0.72)	3.17 (0.73)	0.021
Amylase [U/l, mean (SD)]	126.50 (154.50)	125.07 (179.66)	128.37 (114.55)	0.869
Lipase [U/l, mean (SD)]	172.98 (511.78)	166.45 (588.21)	181.48 (393.16)	0.821
Lactate dehydrogenase [U/l, mean (SD)]	424.57 (518.74)	385.70 (630.35)	475.18 (315.81)	0.182
Vital signs				
PaO_2 [mmHg, mean (SD)]	136.78 (56.46)	136.40 (56.63)	137.28 (56.51)	0.904
$PaCO_2$ [mmHg, mean (SD)]	38.73 (6.57)	38.74 (6.68)	38.72 (6.47)	0.979
pH [unit, mean (SD)]	7.41 (0.06)	7.41 (0.06)	7.41 (0.07)	0.830
Systolic blood pressure [mmHg, mean (SD)]	86.87 (13.76)	86.60 (12.97)	87.22 (14.78)	0.729
Diastolic blood pressure [mmHg, mean (SD)]	71.77 (12.56)	70.96 (12.15)	72.82 (13.07)	0.254
Heart rate [BPM, mean (SD)]	88.30 (18.34)	85.75 (17.33)	91.63 (19.16)	0.013
Respiratory rate [BPM, mean (SD)]	17.71 (3.86)	16.69 (3.38)	19.05 (4.05)	< 0.001
Severity of disease				
ISS [mean (SD)]	25.91 (7.73)	23.29 (6.30)	29.31 (8.12)	< 0.001
SOFA [mean (SD)]	5.37 (3.61)	3.76 (2.55)	7.47 (3.72)	< 0.001
GCS [mean (SD)]	13.07 (3.27)	14.17 (2.13)	11.63 (3.88)	< 0.001
Length of ICU stay [days, mean (SD)]	15.77 (28.90)	8.72 (9.09)	24.93 (40.92)	< 0.001

BMI, body mass index; BPM, beats per minute; GCS, Glasgow coma scale; INR, international normalized ratio; ISS, injury severity score; SD, standard deviation; SOFA, sequential organ failure assessment.

was explicitly documented in medical records. According to the Sepsis-3.0 guideline of the international consensus on sepsis (third edition, 2016), sepsis refers to life-threatening organ dysfunction resulting from an inflammatory response imbalance caused by infection[4]. The diagnostic criteria for sepsis are a quick sequential organ failure assessment (qSOFA) score of at least 2 or a sequential organ failure assessment (SOFA) score of at least 2 due to infection-induced rapid sequential organ dysfunction[16]. More detailed information is summarized in Supplementary File 1 (Supplemental Digital Content 2, http://links.lww.com/JS9/C872).

Data collection

This study used 39 variables, including demographic characteristics [gender, age, body mass index (BMI), and smoking], trauma information (trauma mechanism, head trauma, thoracic trauma, abdominal trauma, brain or spinal cord trauma, and open trauma), comorbidities (hypertension, diabetes, coronary heart disease, cerebral vascular disease, chronic pulmonary disease, live disease, and renal disease), medications use (aspirin, warfarin, and beta blockers), laboratory examination findings [total protein, serum sodium, international normalized ratio (INR), hematocrit, red blood cell count, amylase, lipase, and lactate dehydrogenase], vital signs (PaO_2, $PaCO_2$, pH, systolic blood pressure, diastolic blood pressure, heart rate, and respiratory rate), severity of disease (injury severity score (ISS), sequential organ failure assessment (SOFA) score[17,18], and Glasgow coma scale (GCS)[19]), and length of ICU stay. The first examination was performed after admission to the ICU. Using data from the first examinations in the ICU to predict the risk of sepsis offers significant advantages; early intervention is a key advantage, and healthcare providers can identify individuals at higher risk of sepsis at an early stage. The severity of disease score was used by independent clinicians to evaluate the trauma severity of each patient. Basic characteristics and clinical data were extracted from the electronic medical records.

Data preparation

In this study, we employed the SMOTETomek resampling strategy to mitigate the effects of imbalanced data distribution and produce a robust model[20]. SMOTETomek is a hybrid resampling technique that combines the Synthetic Minority Oversampling Technique and Tomek Links Undersampling. Furthermore, we used a data preprocessing pipeline to ensure that the data were transformed in a consistent and reproducible manner to enhance the accuracy and reliability of our machine-learning models. Data preprocessing pipelines were used to prepare the data for machine-learning algorithms in scikit-learn. These pipelines integrate multiple data preprocessing steps into a single object, which can transform the data in a uniform and reproducible manner. To ensure the consistency of the outcome classes (sepsis vs. no sepsis) in the sub-datasets and datasets, we adopted a stratified strategy. Preprocessing was carried out based on the features of the training set to prevent data leakage.

Modeling and validation

Binary classification problems are commonly encountered in clinical settings and addressed using various models. In this study, we used several models, including logistic regression (LR), decision tree (DT)[21,22], extreme gradient boosting machine (eXGBM)[23], neural network (NN)[24,25], random forest (RF)[26], and light gradient boosting machine (LightGBM)[20]. The details of the machine-learning algorithms used in the present study are summarized in Supplementary Table S1 (Supplemental Digital Content 3, http://links.lww.com/JS9/C873). To ensure consistency, each model was provided with the same input features. This study also used grid and random hyperparameter searches to identify the optimal hyperparameters for each model, using the area under the curve (AUC) as the optimization metric. To prevent overfitting and underfitting, we set wide upper and lower bounds for the hyperparameters during the search, such as a decision tree depth between 2 and 100. The algorithms were implemented using Python (version 3.9.7), and hyperparameter tuning was conducted using scikit-learn (version 1.2.2).

To assess the predictive performance of the models, this study used accepted and commonly used metrics, including AUC with 100 bootstraps, calibration curve, density curve, accuracy, precision, recall (sensitivity), F1 score, Brier score, and log loss. We used decision curve analysis to evaluate the clinical net benefit of each model. Density curves[27] can have various shapes and sizes, and they provide a rapid visual representation of the probability of sepsis among patients with and without sepsis. Accuracy is the proportion of correct predictions, precision is the proportion of true positives out of the total number of positive predictions, recall is the proportion of true positives out of the total number of actual positives, and the F1 score is the harmonic mean of the precision and recall. The confusion matrix[28] was used to calculate the accuracy, precision, recall, and F1 score. In addition, the Brier score[29,30] and log loss value[29] were used to evaluate the quality of classification model predictions. Feature importance was also used to analyze the contribution of each feature to the model output using SHAP (SHapley Additive exPlanations)[20,31,32]. More detailed information is summarized in Supplementary File 1 (Supplemental Digital Content 2, http://links.lww.com/JS9/C872).

Development of web-based artificial intelligence application

To deploy the machine-learning model as a web-based artificial intelligence application, a suitable cloud platform selection, virtual machine setup with appropriate specifications, installation of necessary software and dependencies, loading of the trained model, and web-based interface development were essential. We developed a user-friendly application using GitHub (https://github.com/) to estimate the risk of sepsis among patients with major trauma admitted to the ICU. Streamlit (https://share.streamlit.io/) was the most suitable cloud platform for hosting the application. The code for the AI application is available at https://github.com/Starxueshu/sepsispredictionshap. More detailed information on the design of the AI application and comparison of human and machine performance in predicting sepsis is summarized in Supplementary File 1 (Supplemental Digital Content 2, http://links.lww.com/JS9/C872).

Statistical analysis

The continuous variables are summarized as the mean and standard deviation (SD), while the categorical variables are presented as proportions. The categorical variables were compared using the Chi-squared test, while the continuous variables were compared using the Student *t*-test or Wilcoxon rank test. All statistical analyses were carried out using R (version 4.1.2). Two-sided *P* values less than 0.05 denoted statistical significance.

Results

Basic characteristics of the patients

The model development cohort included 244 patients with a mean age of 45.09 years; 78.3% were male. The most common trauma mechanism was traffic collision (59.4%), and patients had a heavy trauma burden, with thoracic trauma and brain or spinal cord trauma accounting for 47.5% and 45.5%,

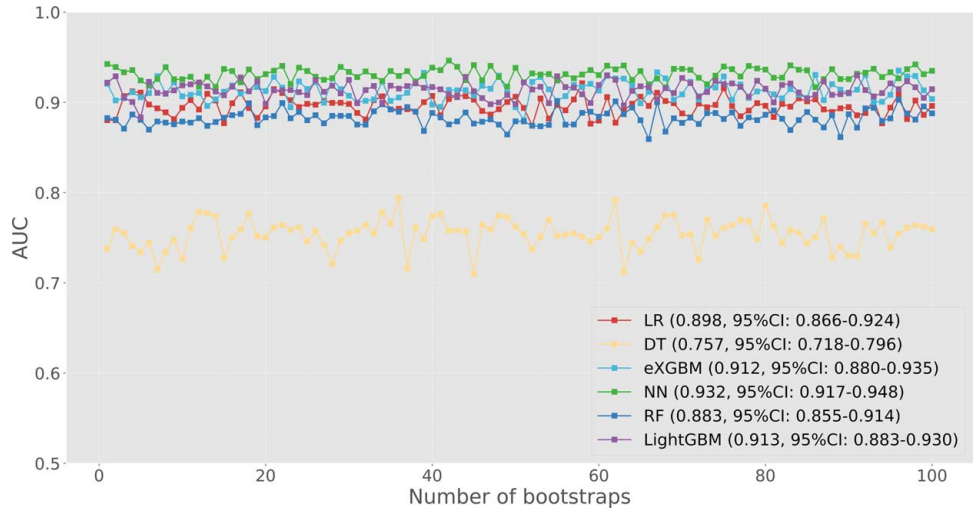

Figure 2. The area under the curve for each model with 100 bootstraps. DT, decision tree; eXGBM, extreme gradient boosting machine; LightGBM, light gradient boosting machine; LR, logistic regression; NN, neural network; RF, random forest.

respectively. Open trauma accounted for 23.0% of the cases. The most common comorbidity was hypertension (16.0%), followed by diabetes (7.4%). However, the comorbidity burden was not heavy, as the prevalence of most comorbidities was below 10.0%. Despite this, the severity of the disease was significant, with a mean ISS of 25.91, a mean SOFA score of 5.37, and a GCS of 13.07. The incidence of sepsis was 43.44% (106/244) for the patients with major trauma in the ICU. Additional details on patient medication use, laboratory examination findings, and vital signs are summarized in Table 1. Based on subgroup analysis (Supplementary File 1, Supplemental Digital Content 2, http://links.lww.com/JS9/C872), 12 predictors, including gender, abdominal trauma, open trauma, red blood cell count, heart rate, respiratory rate, ISS, SOFA score, GCS, smoking, total protein concentrations, and hematocrit, were used as features in the final model.

Model evaluation in the internal validation cohort

Among the six models, the NN model showed the best AUC value (0.932, 95% CI: 0.917–0.948), followed by the LightGBM model (0.913, 95% CI: 0.883–0.930) and eXGBM model (0.912, 95% CI: 0.880–0.935) (Fig. 2 and Table 2). The NN model also had the best accuracy (0.837), precision (0.864), recall (0.792), F1 score (0.826), Brier score (0.104), and log loss (0.326), indicating favorable predictive performance in the internal validation (Fig. 3). The LightGBM model had the second-best AUC value (0.913 [95% CI: 0.883–0.930]), accuracy (0.816), precision (0.826), F1 score (0.809), Brier score (0.125), and log loss (0.388). The eXGBM model exhibited a predictive performance comparable to that of the LightGBM model. The calibration curves demonstrated that the models, especially the NN, eXGBM, RF, and DT models, were sufficiently optimized; their curves were close to the perfectly calibrated line (Supplementary Fig. 1A, Supplemental Digital Content 4, http://links.lww.com/JS9/C874). Supplementary Figure 1B (Supplemental Digital Content 4, http://links.lww.com/JS9/C874) shows the histogram with the *x*-axis indicating the mean predicted probability and the *y*-axis indicating count. To obtain a quick visual understanding of the distributions of the mean predicted probabilities for the patients with and without sepsis, a density curve was plotted for each model (Fig. 4). It showed that the NN, LightGBM, and eXGBM models had favorable discrimination; the graph for the patients without sepsis was left skewed, and that for the patients with sepsis was right-skewed. Figure 5 shows a plot of the net benefit of each model against the threshold probabilities. The net benefit was calculated as the difference between the true positive and false positive rates and weighted by the threshold probability. All the models showed clinical utility, with the NN, eXGBM, and LightGBM models showing greater net benefit than the other models. The NN model had the highest score (51), followed by the LightGBM (43) and eXGBM (37), in the internal validation

Table 2

Prediction performance of all machine-learning models in the internal validation group.

Models	AUC (95% CI)	Accuracy	Precise	Recall	F1 score	Brier score	Log loss	Calibration slope	Intercept-in-large
LR	0.898 (95% CI: 0.866–0.924)	0.776	0.842	0.667	0.744	0.168	0.518	2.797	0.067
DT	0.757 (95% CI: 0.718–0.796)	0.776	0.810	0.708	0.756	0.173	0.531	0.861	0.009
eXGBM	0.912 (95% CI: 0.880–0.935)	0.816	0.826	0.792	0.809	0.131	0.415	1.199	0.262
NN	0.932 (95% CI: 0.917–0.948)	0.837	0.864	0.792	0.826	0.104	0.326	1.384	-0.008
RF	0.883 (95% CI: 0.855–0.914)	0.796	0.818	0.750	0.783	0.156	0.486	1.947	0.088
LightGBM	0.913 (95% CI: 0.883–0.930)	0.816	0.826	0.792	0.809	0.125	0.388	0.966	0.096

AUC, area under the curve; CI, confident interval; DT, decision tree; eXGBM, extreme gradient boosting machine; LightGBM, light gradient boosting machine; LR, logistic regression; NN, neural network; RF, random forest.

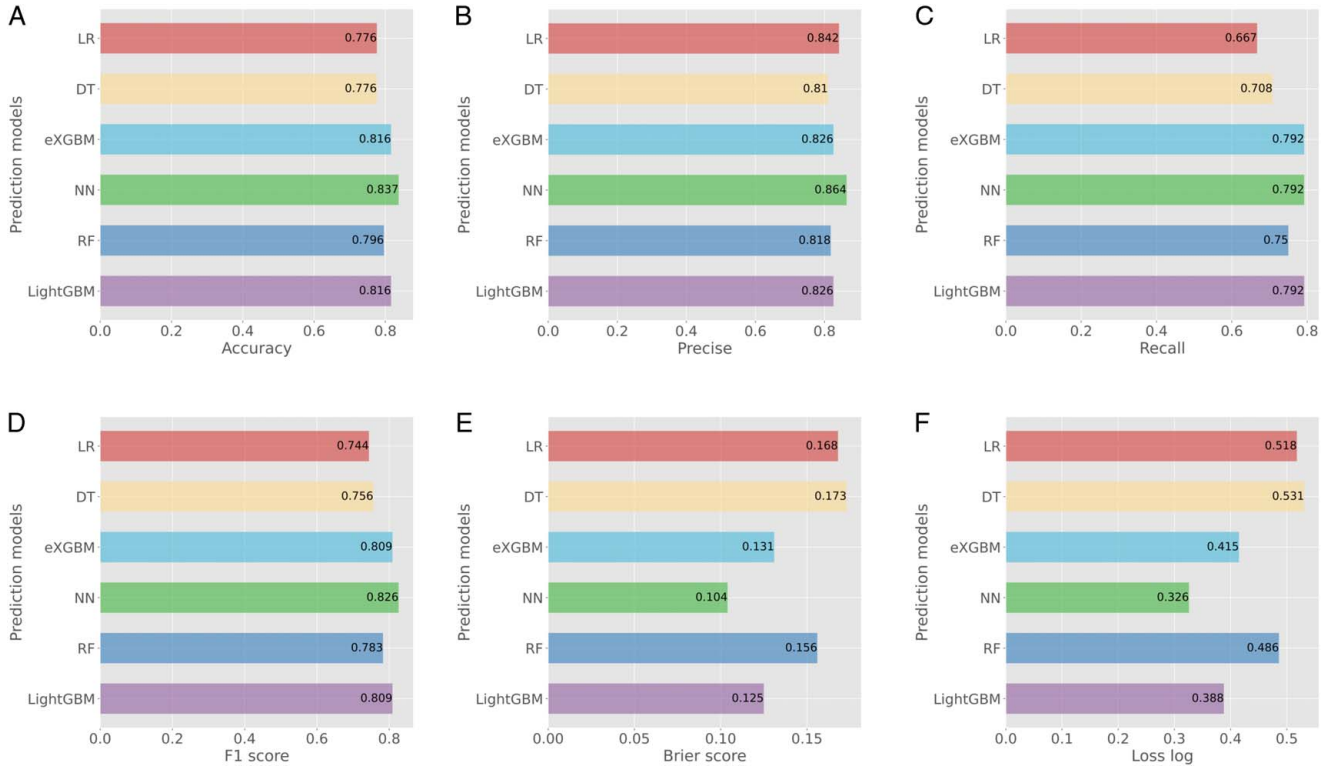

Figure 3. Evaluation of the predictive performance for each model. A. Accuracy; B. Precise; C. Recall; D. F1 score; E. Brier score; F. Log loss. DT, decision tree; eXGBM, extreme gradient boosting machine; LightGBM, light gradient boosting machine; LR, logistic regression; NN, neural network; RF, random forest.

(Supplementary Fig. 2A, Supplemental Digital Content 5, http://links.lww.com/JS9/C875).

Model evaluation during the external validation

A total of 717 patients with major trauma were included for analysis from the database, and their characteristics are summarized in Supplementary Table 2 (Supplemental Digital Content 6, http://links.lww.com/JS9/C876). The incidence of sepsis for this cohort was 43.51%, which was similar to that of the model development cohort (43.44%). The three models with the highest AUCs were eXGBM (AUC: 0.891, 95% CI: 0.866–0.914), LightGBM (AUC: 0.886, 95% CI: 0.860–0.906), and RF (AUC: 0.881, 95% CI: 0.841–0.902) (Supplementary Table 3, Supplemental Digital Content 7, http://links.lww.com/JS9/C877). The eXGBM model outperformed other models in terms of accuracy, F1 score, log loss, and intercept-in-large value. The eXGBM model also had the highest score (44), followed by the DT model (39) and the LightGBM model (39), for the external validation (Supplementary Fig. 2B, Supplemental Digital Content 5, http://links.lww.com/JS9/C875), whereas the NN model only had a score of 25.

On assessing the predictive performance for the internal and external validation cohorts, the LightGBM model had the highest score (82), followed by the eXGBM model (81) and NN model (76) (Fig. 6). Based on the internal and external validation results, the LightGBM model was identified as the optimal model for predicting sepsis in patients with major trauma in this study (Supplementary Table 4, Supplemental Digital Content 8, http://links.lww.com/JS9/C878). The SHAP showed that the most important five features were respiratory rate, SOFA, GCS, ISS, and total protein in the training group (Fig. 7A) and the internal validation group (Fig. 7B).

Deployment of the AI application

The optimal predictive model developed in this study has been deployed online and is accessible at the following link: https://sepsispredictionshap-j3zbv3hfkttrgr7vvezvv7.streamlit.app/. Users can freely access the application and utilize the model to predict the probability of sepsis by inputting the relevant information and selecting appropriate model parameters (Fig. 8A, B). For instance, a male patient who is a smoker and has no abdominal or open trauma has an ISS of 24, SOFA score of 4, GCS score of 6, red blood cell count of 2.28×10^{12}/Ll heart rate of 73 beats per minute, respiratory rate of 20 breaths per minute, hematocrit of 0.23, and total protein concentration of 38.93 g/l. Based on this information, the model categorizes this patient into the high-risk group, with a predicted probability of 76.52% of developing sepsis. Additionally, the application provides corresponding interventional strategies based on the predicted risk. Figure 8C shows the risk report for the patient, and Figure 8D provides the model introduction for the AI application. A comparison of human and machine performance in predicting sepsis is summarized in Supplementary File 1 (Supplemental Digital Content 2, http://links.lww.com/JS9/C872). A detailed video tutorial on how to use the AI application is available in Supplementary File 2 (Supplemental Digital Content 9, http://links.lww.com/JS9/C879).

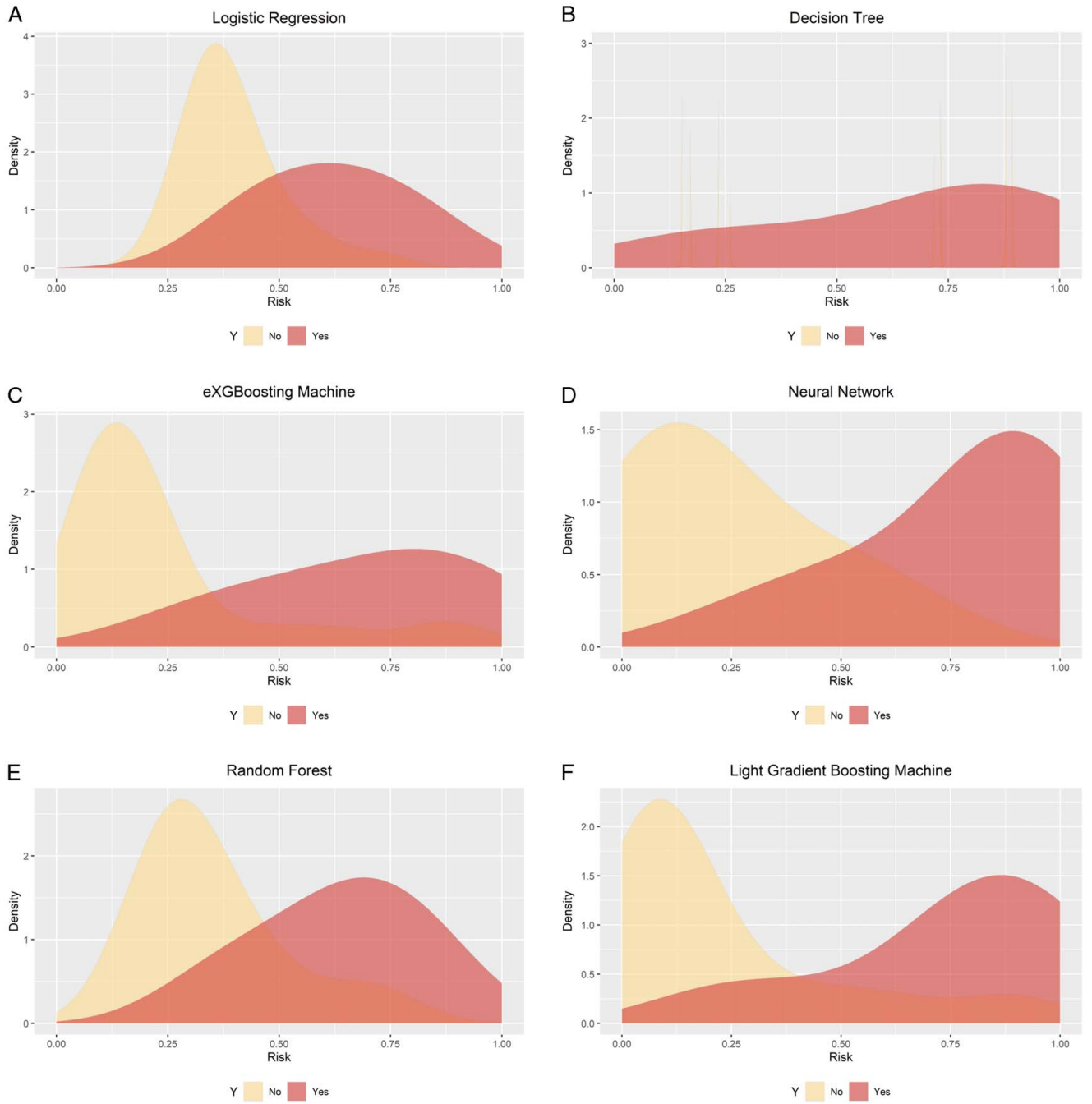

Figure 4. Density curve for each model. A. logistic regression; B. decision tree; C. extreme gradient boosting machine; D. neural network; E. random forest; F. light gradient boosting machine. The red indicates patients with sepsis, and the yellow indicates patients without sepsis.

Discussions

Principal findings

This study developed and validated a mobile AI application for sepsis in patients with major trauma. The application used 12 variables as model predictors, and the LightGBM model showed good reliability and predictive performance according to the comprehensive scoring system during the internal and external validation. The machine-learning algorithms enabled the development of accurate prediction models, and the international external validation of the model further strengthened its validity and applicability to other settings. Thus, the AI application developed in this study may be useful in guiding clinicians to identify patients with a high risk of sepsis and implement early intervention strategies to improve their outcomes.

Risk factors for sepsis among patients in the ICU

Sepsis is a major cause of morbidity and mortality among patients in the ICU, and patients with trauma are susceptible to sepsis. The

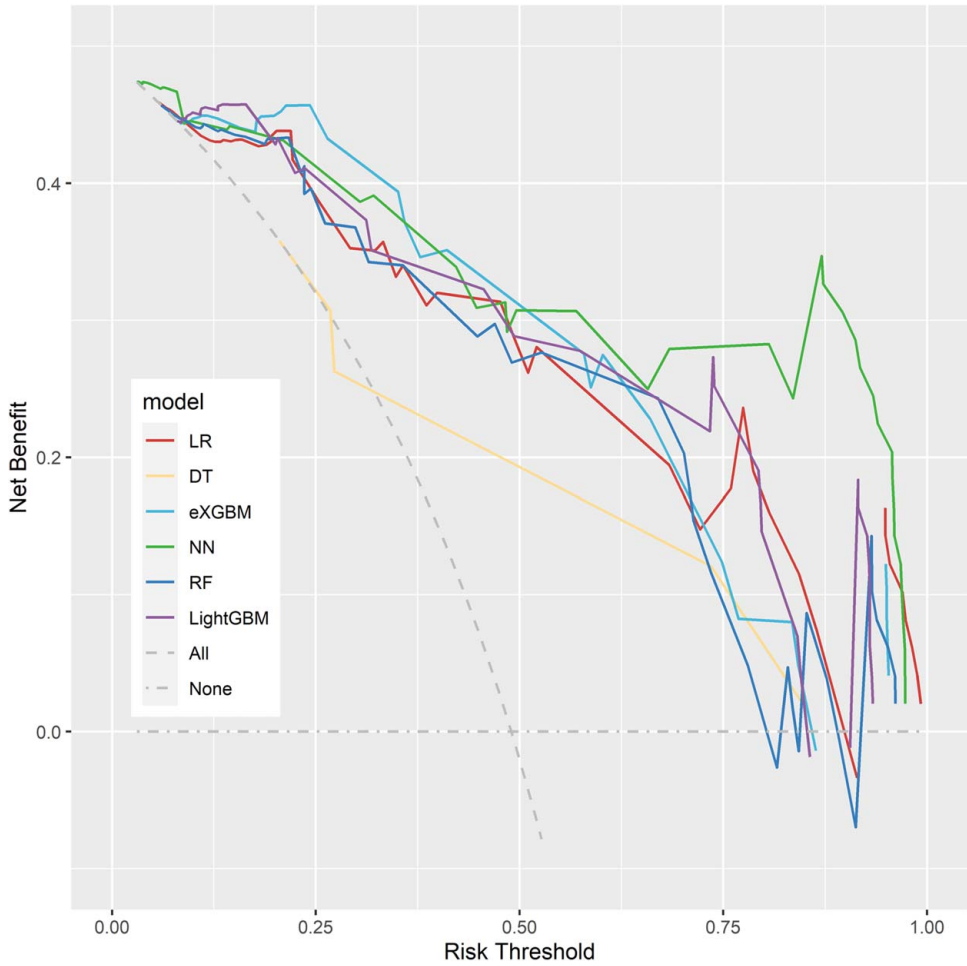

Figure 5. Decision curve analysis for each model. DT, decision tree; eXGBM, extreme gradient boosting machine; LightGBM, light gradient boosting machine; LR, logistic regression; NN, neural network; RF, random forest.

incidence of sepsis among patients in the ICU is relatively high; the incidence was 43.44% among patients with major trauma in the ICU. This was consistent with the results of previous studies ranging from 30.4%[2] to 33.3%[3] for patients with trauma in the ICU. The heterogeneity of the population may explain the variation because our study enrolled patients with major trauma.

Understanding the risk factors is important for developing effective prevention and treatment strategies. Some studies investigated the risk factors for sepsis in the ICU[33–35]. Manandhar *et al.*[33] conducted a prospective observational cohort study of 142 cases of neonatal sepsis in the neonatal ICU and reported that a high number of IV cannula insertion days and C-reactive protein (CRP) concentration were risk factors for the development of sepsis based on multivariate analysis. Shibata *et al.*[34] found that older age, vital signs prognosticating sepsis, and the presence of some comorbidities, such as diabetes mellitus, ischemic heart disease, and chronic kidney disease, were potential risk factors of sepsis after retrospectively analyzing the data of patients with qSOFA scores of less than 2. For patients with trauma in the ICU, the risk factors identified for sepsis include age, comorbidities, severity of injury, emergency surgery, and immunosuppression[2,8]. A national population-based prospective study identified some independent risk factors for posttraumatic sepsis, including male gender, advanced age, pre-existing medical condition, GCS, ISS, abbreviated injury scale score, number of injuries, number of red blood cell units transfused, and number of operative procedures[9]. A study involving 140 patients showed good diagnostic values of soluble interleukin-2 receptor, tumor necrosis factor-α, procalcitonin, and their combination for sepsis in patients with a closed abdominal injury and severe multiple abdominal injuries[36].

Our study demonstrated that male gender, abdominal trauma, open trauma, a lower red blood cell count, higher heart and respiratory rates, higher ISS and SOFA scores, lower GCS, smoking, lower total protein, and lower hematocrit were associated with severe sepsis, especially among patients with major trauma in the ICU. The findings were consistent with previous studies, as gender[3,9], ISS[2,8,9], GCS[2,3,9], heart rate[2], albumin[2], respiratory rate[10], and number of red blood cell units transfused[9] have been identified to be associated with sepsis among patients with trauma. Our study showed that abdominal trauma, open trauma, and smoking were relevant to sepsis. Patients in the ICU with open trauma or abdominal injuries had a heightened risk of developing sepsis, possibly due to the

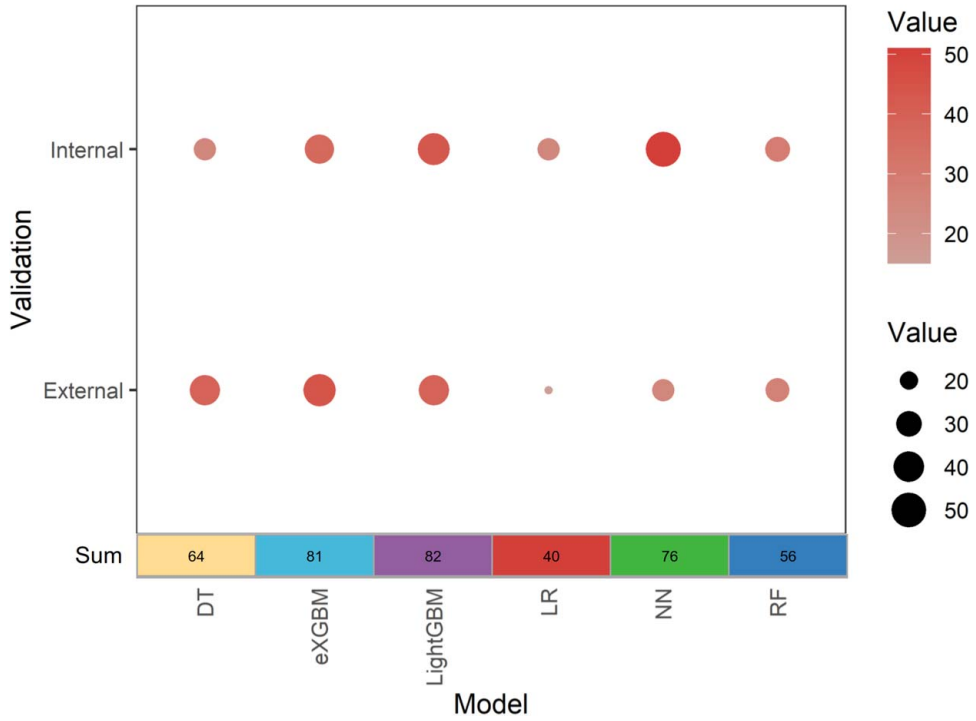

Figure 6. Bubble map for the predictive performances for internal and external validation cohorts. DT, decision tree; eXGBM, extreme gradient boosting machine; LightGBM, light gradient boosting machine; LR, logistic regression; NN, neural network; RF, random forest.

increased likelihood of bacterial contamination and infection associated with open wounds. The impaired skin barrier function caused by the wound can facilitate the entry of pathogens into the bloodstream, leading to the development of sepsis. Furthermore, the use of invasive procedures and broad-spectrum antibiotics can further disrupt the normal microbiome, creating an environment conducive to the growth of opportunistic pathogens. Smoking can compromise the immune system, making it more difficult for the body to fight off infections. Additionally, smoking can damage the respiratory tract and impair lung function, increasing the risk of pneumonia and other respiratory infections.

The previous experimental findings strongly supported the notion that female sex hormones, namely estrogen, confer beneficial effects in the aftermath of trauma-related hemorrhage, while male sex steroids, such as 5alpha-dihydrotestosterone, have suppressive effects on immune and cardiac functions in the same setting[37]. This may explain why male patients are more susceptible to sepsis following trauma than females. These striking sex-specific differences in response to trauma-related hemorrhage underscore the pivotal role of sex hormone concentrations in shaping the host response to injury. Specifically, female sex hormones appear to enhance the resilience of the organism in the face of trauma-related hemorrhage[38], while male sex steroids may impair the ability of the organism to mount a robust response to the insult. These findings are consistent with the reports of recent studies highlighting the importance of sex differences in the pathophysiology and treatment of trauma and shock. Furthermore, they suggest that therapeutic interventions targeting sex hormone pathways may hold promise for improving outcomes in trauma and critical care settings[37].

Prediction model of ICU sepsis among trauma patients

According to previous studies, there were only two prediction models that have been developed for predicting sepsis among patients with trauma admitted to the ICU. Lu *et al.*[2] developed a predictive score for sepsis risk in 2019 for patients with trauma using electronic medical record data. Clinical and laboratory variables of 684 patients with trauma were collected, and the model was constructed using a logistic regression model based on seven variables, including ISS, GCS, temperature, heart rate, albumin, INR, and CRP with an AUC of 0.799 for the training group and 0.790 for the validation group. More recently, Li *et al.*[10] developed a machine-learning model to predict the risk of sepsis in ICU-admitted patients with trauma from the MIMIC-IV database on an hourly scale. The model utilized data from up to 42 variables, including demographics, vital signs, arterial blood gas, and laboratory tests, and achieved good discrimination and calibration performance with the AUC ranging from 0.83 to 0.88 within the 4–24-h prediction window in the test set. The models could help identify patients at risk of sepsis promptly, enabling personalized intervention and early treatment. Nonetheless, the predictive performance of these models still needs improvement because none of the models can obtain excellent predictive performance as the AUC was not as high as 0.90. Next, external validation of the models was not implemented; thus, their validity and applicability to other settings were compromised. Finally, our study enrolled patients with multiple injuries and an ISS above 16, and the population was markedly heterogeneous. It should also be noted that our study selected the optimal model among six, and external validation was performed. Machine-learning algorithms have shown great potential in predicting sepsis among patients with trauma admitted to the

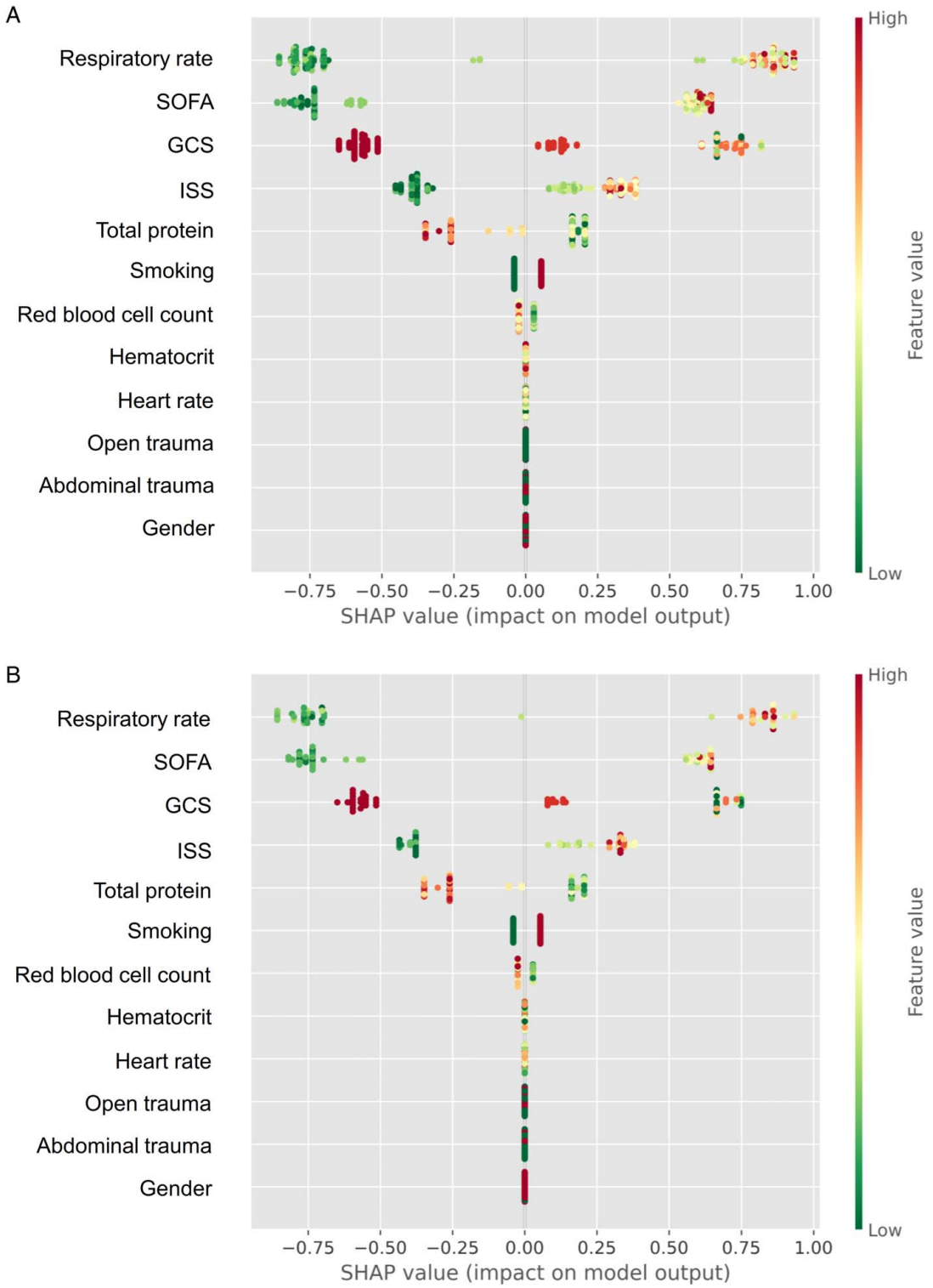

Figure 7. Feature importance analysis based on SHAP. A. Training cohort; B. Internal validation cohort. GCS, Glasgow coma scale; ISS, injury severity score; SHAP, SHapley Additive explanation; SOFA, sequential organ failure assessment.

ICU. Internal validation showed an AUC of 0.913 for the LightGBM model; external validation also showed good performance as the AUC value was still up to 0.886, indicating good predictive performance. While the AUC value of the NN model was up to 0.932, it decreased to only 0.787 for the external validation, indicating an unstable model. Thus, the LightGBM model was identified as the most optimal for predicting sepsis among patients with major trauma based on internal and external validation.

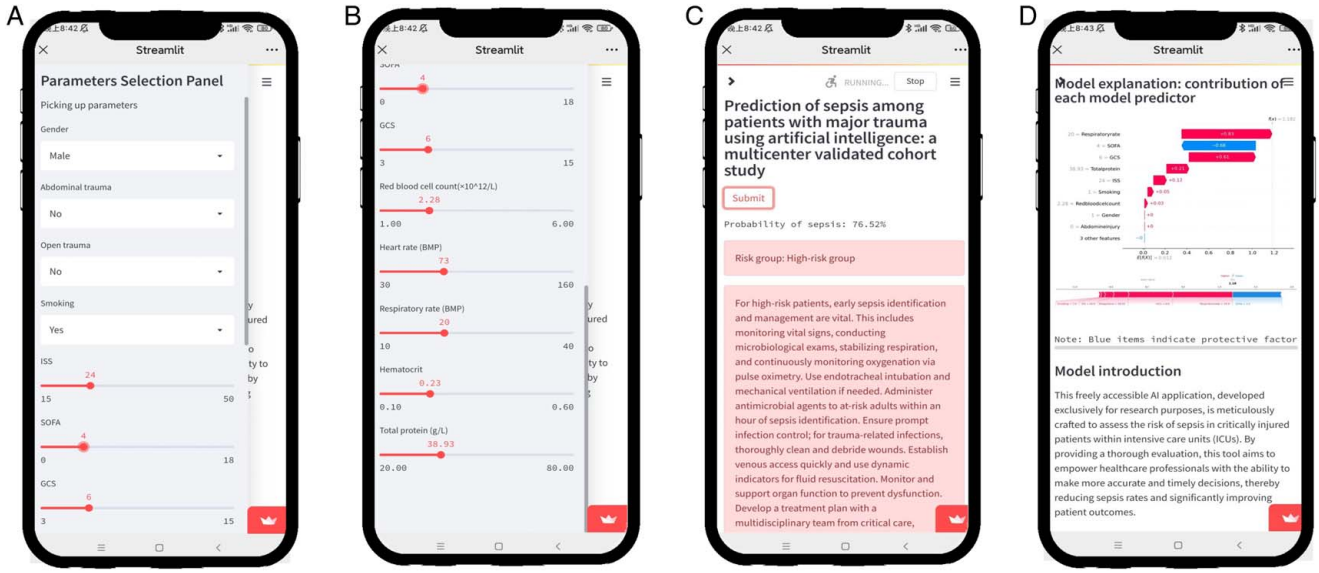

Figure 8. The application for estimating the risk of sepsis among patients with major trauma. A, B. Model input features panel; C. predicted risk of developing sepsis for the specific case; D. the section of the artificial intelligence (AI) application showing risk report for the patient and the model introduction. Blue items indicate protective factors and red items indicate risk factors.

Utilizing the AI application for clinical treatment guidance

Doctors, nurses, and other medical providers can access the application by clicking on the URL provided in the study, and all that is required is medical equipment connected to the Internet. Upon accessing the application, users can obtain individualized risk assessments for patients by simply inputting information and submitting the data. Additionally, patients are stratified into high-risk and low-risk groups for sepsis based on the optimal threshold.

For high-risk patients, the following aspects are crucial. (1) Early identification and management of sepsis, including monitoring vital signs, conducting microbiological examinations, stabilizing respiration, and continuously monitoring oxygenation through pulse oximetry[5]. In addition, endotracheal intubation and mechanical ventilation should be performed when necessary to meet the increased respiratory effort commonly associated with sepsis[39]. (2) For adult patients at risk of septic shock or with a high likelihood of sepsis, it is recommended to administer antimicrobial agents immediately[40,41], ideally within 1 h of sepsis identification. (3) In cases of infection caused by trauma, thorough cleaning and debridement of the wound should be performed to remove all contaminants, foreign bodies, and dead tissue. (4) It is important to establish venous access as soon as possible for high-risk sepsis patients. It is recommended to use dynamic indicators to guide fluid resuscitation rather than relying solely on physical examination or static indicators[42]. (5) Supporting and monitoring organ function in critically injured patients to prevent organ dysfunction. (6) Finally, a treatment plan should be developed collaboratively by a multidisciplinary team composed of specialists from critical care, emergency medicine, surgery, and infectious diseases.

For patients with low-risk sepsis, (1) it is also essential to closely monitor and evaluate all aspects of their condition, including performing neurological examinations, assessing consciousness levels, and regularly checking vital signs[5]. (2) In addition, although the risk of developing sepsis is not high, preventing other potential complications, such as deep vein thrombosis and pneumonia, remains crucial. (3) Prompt surgical interventions are necessary to address trauma sites, which may involve the removal of cranial hematomas, initial external fixation of pelvic fractures, and treatment of abdominal injuries. (4) Maintaining the patient's nutritional status through appropriate nutritional support is vital for their recovery[42]. (5) Furthermore, once the patient's condition stabilizes, early initiation of rehabilitation therapy is important to promote functional recovery.

Furthermore, to enhance the utility of the AI application in guiding clinical treatment decisions, we developed an individualized risk report for patients using the SHAP method. For example, if the risk report indicates that total protein, open trauma, and GCS are the most important contributors to sepsis, measures to improve total protein, address open trauma, and control GCS are especially crucial for these individuals.

Limitations

Despite the strengths of this study, some limitations need to be considered. First, some other variables, such as biological data, were not included in the analysis, which may affect the accuracy of the model. Second, our current prediction model remains static, focusing solely on individual variables at a given point in time, and it would be more advantageous to explore the dynamic relationship between the time series variables and risk of sepsis. The addition of new data may further optimize the model. We also plan to regularly update and validate the AI model with new data to ensure its ongoing effectiveness and relevance. Additionally, our team will routinely maintain the AI application to ensure its stability. Third, selection bias in comparing human and machine predictive performance may be present, making further extensive validation of the model essential.

Conclusions

This study develops and validates an AI application to effectively assess the susceptibility of patients with major trauma to sepsis. The LightGBM model shows superior predictive performance, and it may be a promising machine-learning algorithm for predicting sepsis. The AI application equips healthcare professionals with a valuable tool to promptly identify individuals at high risk of developing sepsis. This will facilitate clinical decision-making and enable the early implementation of intervention strategies. Ultimately, this development holds immense potential in elevating outcomes for critically injured patients.

Ethical approval

This study was approved by the Ethics Committee of the Chinese PLA General Hospital (Judgement's reference number: S2017-054-01). All data were analyzed anonymously, and this study abided by the Helsinki Declaration.

Consent

Written informed consent was obtained from the patient for publication and any accompanying images. A copy of the written consent is available for review by the Editor-in-Chief of this journal on request.

Source of funding

Independent Research Project of Medical Engineering Laboratory of Chinese PLA General Hospital (No. 2022SYSZZKY12).

Author contribution

All authors contributed to the study design, conducted the data collection and analyses, and drafted the paper. In addition, all authors have read and approved the manuscript.

Conflicts of interest disclosure

The author declares no conflict of interest.

Research registration unique identifying number (UIN)

1. Name of the registry: ChiCTR: Chinese Clinical Trial Registry.
2. Unique identifying number or registration ID: ChiCTR-OOC-17012108.
3. Hyperlink to your specific registration (must be publicly accessible and will be checked). https://www.chictr.org.cn/showproj.html?proj=20676.

Guarantor

Feihu Zhou.

Data availability statement

The datasets of the current study are available under reasonable request.

Provenance and peer review

Not commissioned, externally peer-reviewed.

References

[1] Nazer L, Abusara A, Aloran B, *et al*. Patient diversity and author representation in clinical studies supporting the Surviving Sepsis Campaign guidelines for management of sepsis and septic shock 2021: a systematic review of citations. BMC Infect Dis 2023;23:751.
[2] Lu HX, Du J, Wen DL, *et al*. Development and validation of a novel predictive score for sepsis risk among trauma patients. World J Emerg Surg 2019;14:11.
[3] Zhou J, Luo XY, Chen GQ, *et al*. Incidence, risk factors and outcomes of sepsis in critically ill post-craniotomy patients: a single-center prospective cohort study. Front Public Health 2022;10:895991.
[4] Seymour CW, Liu VX, Iwashyna TJ, *et al*. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315:762–74.
[5] Mas-Celis F, Olea-Lopez J, Parroquin-Maldonado JA. Sepsis in trauma: a deadly complication. Arch Med Res 2021;52:808–16.
[6] Kari KA, Wan Muhd Shukeri WF, Yaacob NM, *et al*. Prevalence and outcome of sepsis: mortality and prolonged intensive care unit stay among sepsis patients admitted to a tertiary centre in Malaysia. Malays J Med Sci 2023;30:120–32.
[7] Eguia E, Bunn C, Kulshrestha S, *et al*. Trends, cost, and mortality from sepsis after trauma in the United States: an evaluation of the national inpatient sample of hospitalizations, 2012–2016. Crit Care Med 2020;48:1296–303.
[8] Park JH, Choi SH, Yoon YH, *et al*. Risk factors for sepsis in Korean trauma patients. Eur J Trauma Emerg Surg 2016;42:453–8.
[9] Wafaisade A, Lefering R, Bouillon B, *et al*. Trauma Registry of the German Society for Trauma Surgery: Epidemiology and risk factors of sepsis after multiple trauma: an analysis of 29,829 patients from the Trauma Registry of the German Society for Trauma Surgery. Crit Care Med 2011;39:621–8.
[10] Li J, Xi F, Yu W, *et al*. Real-time prediction of sepsis in critical trauma patients: machine learning-based modeling study. JMIR Form Res 2023;7:e42452.
[11] Song X, Liu X, Liu F, *et al*. Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. Int J Med Inform 2021;151:104484.
[12] Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care 2019;23:112.
[13] Aldhoayan MD, Aljubran Y. Prediction of ICU patients' deterioration using machine learning techniques. Cureus 2023;15:e38659.
[14] Johnson AEW, Bulgarelli L, Shen L, *et al*. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023;10:1.
[15] Mathew G, Agha R, Albrecht J, *et al*. STROCSS 2021: Strengthening the reporting of cohort, cross-sectional and case–control studies in surgery. Int J Surg 2021;96:106165.
[16] Lauritsen SM, Thiesson B, Jorgensen MJ, *et al*. The framing of machine learning risk prediction models illustrated by evaluation of sepsis in general wards. NPJ Digit Med 2021;4:158.
[17] Do SN, Dao CX, Nguyen TA, *et al*. Sequential Organ Failure Assessment (SOFA) Score for predicting mortality in patients with sepsis in Vietnamese intensive care units: a multicentre, cross-sectional study. BMJ Open 2023;13:e064870.
[18] Moreno R, Rhodes A, Piquilloud L, *et al*. The Sequential Organ Failure Assessment (SOFA) Score: has the time come for an update? Crit Care 2023;27:15.
[19] Teasdale G, Maas A, Lecky F, *et al*. The Glasgow Coma Scale at 40 years: standing the test of time. Lancet Neurol 2014;13:844–54.
[20] Zhu C, Xu Z, Gu Y, *et al*. Prediction of post-stroke urinary tract infection risk in immobile patients using machine learning: an observational cohort study. J Hosp Infect 2022;122:96–107.

[21] Lei M, Han Z, Wang S, *et al*. Biological signatures and prediction of an immunosuppressive status-persistent critical illness-among orthopedic trauma patients using machine learning techniques. Front Immunol 2022; 13:979877.
[22] Lei M, Han Z, Wang S, *et al*. A machine learning-based prediction model for in-hospital mortality among critically ill patients with hip fracture: an internal and external validated study. Injury 2023;54:636–44.
[23] Hou N, Li M, He L, *et al*. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. J Transl Med 2020;18:462.
[24] Amiri P, Abbasi H, Derakhshan A, *et al*. Potential prognostic markers in the heart rate variability features for early diagnosis of sepsis in the pediatric intensive care unit using Convolutional Neural Network classifiers. Annu Int Conf IEEE Eng Med Biol Soc 2020;2020:5627–30.
[25] Scherpf M, Grasser F, Malberg H, *et al*. Predicting sepsis with a recurrent neural network using the MIMIC III database. Comput Biol Med 2019; 113:103395.
[26] Zhou J, Dong S, Wang P, *et al*. Identification of nine mRNA signatures for sepsis using random forest. Comput Math Methods Med 2022;2022: 5650024.
[27] Gao L, Cao Y, Cao X, *et al*. Machine learning-based algorithms to predict severe psychological distress among cancer patients with spinal metastatic disease. Spine J 2023;23:1255–69.
[28] Yi M, Cao Y, Zhou Y, *et al*. Association between hospital legal constructions and medical disputes: a multi-center analysis of 130 tertiary hospitals in Hunan Province, China. Front Public Health 2022;10: 993946.
[29] Shi X, Cui Y, Wang S, *et al*. Development and validation of a web-based artificial intelligence prediction model to assess massive intraoperative blood loss for metastatic spinal disease using machine learning techniques. Spine J 2024;24:146–60.
[30] Tong C, Niu Z, Zhu H, *et al*. Development and external validation of a novel model for predicting new clinically important atrial fibrillation after thoracoscopic anatomical lung cancer surgery: a multi-center retrospective cohort study. Int J Surg 2024;110:1645–52.
[31] Cui Y, Shi X, Wang S, *et al*. Machine learning approaches for prediction of early death among lung cancer patients with bone metastases using routine clinical characteristics: an analysis of 19,887 patients. Front Public Health 2022;10:1019168.
[32] Nohara Y, Matsumoto K, Soejima H, *et al*. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed 2022;214:106584.
[33] Manandhar S, Amatya P, Ansari I, *et al*. Risk factors for the development of neonatal sepsis in a neonatal intensive care unit of a tertiary care hospital of Nepal. BMC Infect Dis 2021;21:546.
[34] Shibata J, Osawa I, Ito H, *et al*. Risk factors of sepsis among patients with qSOFA < 2 in the emergency department. Am J Emerg Med 2021;50: 699–706.
[35] Kou YQ, Yang YP, Du SS, *et al*. Risk factors for the development of sepsis in patients with cirrhosis in intensive care units. Clin Transl Sci 2023;16: 1748–57.
[36] Zhai GH, Zhang W, Xiang Z, *et al*. Diagnostic value of sIL-2R, TNF-alpha and PCT for sepsis infection in patients with closed abdominal injury complicated with severe multiple abdominal injuries. Front Immunol 2021;12:741268.
[37] Choudhry MA, Bland KI, Chaudry IH. Gender and susceptibility to sepsis following trauma. Endocr Metab Immune Disord Drug Targets 2006;6:127–35.
[38] Bosch F, Angele MK, Chaudry IH. Gender differences in trauma, shock and sepsis. Mil Med Res 2018;5:35.
[39] Gavelli F, Castello LM, Avanzi GC. Management of sepsis and septic shock in the emergency department. Intern Emerg Med 2021;16: 1649–61.
[40] Kanjee Z, Berg K, Bell TD, *et al*. How would you resuscitate this patient with septic shock?: Grand Rounds Discussion From Beth Israel Deaconess Medical Center. Ann Intern Med 2023;176:253–9.
[41] Zampieri FG, Bagshaw SM, Semler MW. Fluid therapy for critically ill adults with sepsis: a review. JAMA 2023;329:1967–80.
[42] Ma XY, Tian LX, Liang HP. Early prevention of trauma-related infection/sepsis. Mil Med Res 2016;3:33.