Journal of Intensive Medicine 5 (2025) 193–201

Contents lists available at ScienceDirect

Journal of Intensive Medicine

journal homepage: www.elsevier.com/locate/jointm

Original Article

Predicting multiple organ dysfunction syndrome in trauma-induced sepsis: Nomogram and machine learning approaches

Jinyu Peng[1,2], Yun Li[1,2], Chao Liu[2], Zhi Mao[2], Hongjun Kang[2], Feihu Zhou[1,2,*]

[1] *Medical School of Chinese PLA, Beijing, China*
[2] *Department of Critical Care Medicine, The First Medical Centre, Chinese PLA General Hospital, Beijing, China*

ARTICLE INFO

Managing Editor: Jingling Bao/Zhiyu Wang

Keywords:
Trauma
Sepsis
Multiple organ dysfunction syndrome
Machine learning
Prediction model

ABSTRACT

Background: Multiple organ dysfunction syndrome (MODS) is a critical complication in trauma-induced sepsis patients and is associated with a high mortality rate. This study aimed to develop and validate predictive models for MODS in this patient population using a nomogram and machine learning approaches.

Methods: This retrospective cohort study utilized data from the Medical Information Mart for Intensive Care-IV 2.2 database, focusing on trauma patients diagnosed with sepsis within the first day of intensive care unit (ICU) admission. Predictive variables were extracted from the initial 24 h of ICU data. The dataset (2008–2019) was divided into a training set (2008–2016) and a temporal validation set (2017–2019). Feature selection was conducted using the Boruta algorithm. Predictive models were developed and validated using a nomogram and various machine learning techniques. Model performance was evaluated based on discrimination, calibration, and decision curve analysis.

Results: Among 1295 trauma patients with sepsis, 349 (26.95%) developed MODS. The 28-day mortality rates were 11.21% for non-MODS patients and 23.82% for MODS patients. Key predictors of MODS included the simplified acute physiology score II score, use of mechanical ventilation, and vasopressor administration. In temporal validation, all models significantly outperformed traditional scoring systems (all P <0.05). The nomogram achieved an area under the curve (AUC) of 0.757 (95% confidence interval [CI]: 0.700 to 0.814), while the random forest model demonstrated the highest performance with an AUC of 0.769 (95% CI: 0.712 to 0.826). Calibration plots showed excellent agreement between predicted and observed probabilities, and decision curve analysis indicated a consistently higher net benefit for the newly developed models.

Conclusion: The nomogram and machine learning models provide enhanced predictive accuracy for MODS in trauma-induced sepsis patients compared to traditional scoring systems. These tools, accessible via web-based applications, have the potential to improve early risk stratification and guide clinical decision-making, ultimately enhancing outcomes for trauma patients. Further external validation is recommended to confirm their generalizability.

Introduction

Trauma represents a significant global public health challenge, contributing to millions of deaths and disabilities annually. Among trauma patients, sepsis is a prevalent and life-threatening complication, ranking as a leading cause of mortality.[1–4] According to the Sepsis 3.0 definition, sepsis is characterized by life-threatening organ dysfunction resulting from a dysregulated host response to infection.[5] Severe trauma compromises the body's defense mechanisms, increasing susceptibility to bacterial and microbial invasion and heightening the risk of infection. Furthermore, trauma-induced immune dysfunction exacerbates this vulnerability, significantly raising the likelihood of sepsis.[6] Without timely diagnosis and treatment, trauma-induced sepsis can rapidly progress to multiple organ dysfunction syndrome (MODS), defined as the acute, potentially reversible failure of two or more organ systems due to various etiologies.[7–9]

Despite advancements in clinical treatments for MODS, its mortality rate remains alarmingly high, reaching up to 30% in severely injured patients.[10] Mortality risk correlates strongly with the number of affected organs and the severity

* Corresponding author: Feihu Zhou, Department of Critical Care Medicine, The First Medical Centre, Chinese PLA General Hospital, Beijing, China.
E-mail address: feihuzhou301@126.com (F. Zhou).

https://doi.org/10.1016/j.jointm.2024.12.008
Received 13 October 2024; Received in revised form 23 November 2024; Accepted 12 December 2024
Available online 8 February 2025

of dysfunction.[11] Thus, early identification and management of risk factors for MODS are crucial for improving patient outcomes.

Currently, there is a paucity of studies investigating the risk factors for MODS in patients with trauma-induced sepsis, particularly those supported by large-scale datasets. To address this research gap, our study utilizes the Medical Information Mart for Intensive Care (MIMIC-IV, version 2.2) database to develop and validate predictive models for MODS in trauma-induced sepsis patients. These models are constructed using nomogram and machine learning methodologies.

Methods

Study design and data source

This study employed a retrospective cohort design to develop and validate predictive models for MODS in patients with trauma-induced sepsis using nomogram and machine learning approaches. Data were sourced from the publicly available MIMIC-IV database (version 2.2),[12] comprising comprehensive clinical information from Beth Israel Deaconess Medical Center, including vital signs, laboratory test results, medications, and diagnoses. The dataset from 2008 to 2016 was used for model development, while data from 2017 to 2019 were reserved for temporal validation. Access to the MIMIC-IV database (No. 49639059) was granted, and due to the dataset's de-identified nature, the requirement for written informed consent was waived.

Definitions and case selection criteria

Patients were identified using the "icdpicr" package, which calculates the injury severity score (ISS) based on ICD-9 and ICD-10 codes.[13] Trauma patients were defined as those with an ISS >0. From this cohort, patients diagnosed with sepsis on the first day of intensive care unit (ICU) admission, according to Sepsis-3.0 criteria,[5] were included. MODS was defined as a sequential organ failure assessment (SOFA) score ≥2 in two or more organ systems.[14] Exclusion criteria included patients with prior ICU admissions, ICU stays of <1 day, or those meeting MODS criteria on the first day of ICU admission.

Data processing

Predictive variables were collected from the first 24 h following ICU admission, while the outcome variable (MODS) was assessed daily from day 2 to day 7 of the ICU stay. MODS was defined as dysfunction in two or more organ systems during this period, ensuring that predictive variables preceded the outcome assessment. Data preprocessing involved cleaning, outlier management, and feature selection. Variables with >30% missing data or deemed irrelevant to the study were excluded. Missing values were imputed using the random forest (RF) algorithm implemented through scikit-learn's IterativeImputer with RandomForestRegressor (scikit-learn version 1.1.1).

Variable selection

The Boruta algorithm, an all-relevant feature selection method based on RFs, was used to identify key predictors of MODS.[15] This algorithm compares the importance of original variables with that of randomly permuted copies, ensuring robust feature selection. Initially, 42 potential predictors were considered, including demographic data, comorbidities, injury severity indicators, and first-day clinical interventions. The simplified acute physiology score II (SAPS II) score was included among the candidate features, while hemoglobin, platelets, and international normalized ratio (INR) were added as additional severity indicators to minimize redundancy.

Model development

A nomogram was constructed using logistic regression with the selected features. Due to the limited non-linear processing capacity of logistic regression, certain variables were categorized into clinically relevant intervals based on expert knowledge and established scoring systems.[16,17]

Additionally, several machine learning models were implemented, including K-nearest neighbor (KNN), support vector classifier (SVC), RF, extreme gradient boosting (XGB), and multilayer perceptron (MLP). These models were chosen for their ability to capture complex, non-linear relationships within the data. Model hyperparameters were optimized using Bayesian optimization with 5-fold cross-validation. To enhance predictive performance, the stacking method was applied to ensemble the four best-performing models, creating an integrated model for validation. Model development followed the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) principles.[18]

Statistical analysis

Continuous variables with non-normal distributions were expressed as medians with interquartile ranges, while categorical variables were reported as counts and percentages. Model performance was evaluated using receiver operating characteristic (ROC) curves, with the area under the ROC curve (AUROC) calculated to assess classification accuracy. The optimal threshold for prediction was determined using the maximum Youden index, and a confusion matrix was generated accordingly. Model evaluation metrics, including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value, were calculated at various probability thresholds in the validation set. Variable contributions to the model output were assessed using the SHapley Additive exPlanations (SHAP) method.[19] Statistical comparisons between non-normally or unequally distributed data were performed using the Kruskal–Wallis test, while the chi-squared test was used for categorical data analysis. A P-value <0.05 was considered statistically significant. Machine learning modeling and statistical analyses were conducted using Python (version 3.7.8) and R (version 4.0.2).

Results

Study population

A total of 1295 trauma-induced sepsis patients were initially identified from the MIMIC-IV cohort (Figure 1). Among these, 349 patients developed MODS. Due to ambiguous admission times in the database, 36 patients whose admission dates could

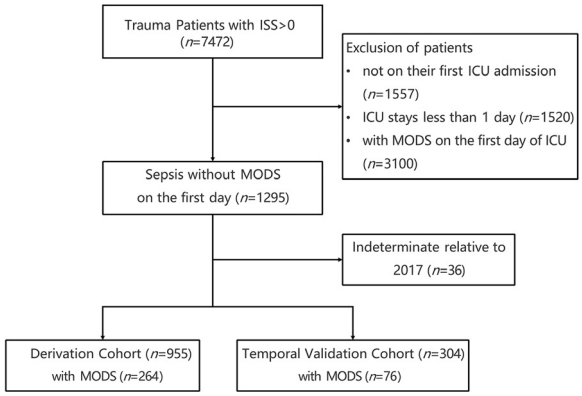

Figure 1. Process flow of the study.
ICU: Intensive care unit; ISS: Injury severity score; MODS: Multiple organ dysfunction syndrome.

not be classified as before or after 2017 were excluded. This resulted in 955 patients in the derivation cohort and 304 in the temporal validation cohort. The 28-day mortality rates for patients without and with MODS were 11.2% and 23.8%, respectively, while the 90-day mortality rates were 15.3% and 29.7%, respectively (P <0.001). Baseline cohort characteristics are summarized in Table 1.

Variables selected

The Boruta algorithm identified eight significant predictors of MODS development: age, weight, presence of invasive lines on the first ICU day, vasopressor use on the first ICU day, SAPS II score, maximum abbreviated injury scale (AIS) score for the chest region, mechanical ventilation, and platelet count.

Table 1
Baseline characteristics of the cohort.

Variables	Non-MODS (n = 919)	MODS (n = 340)	P-value
Age (years)	65.3 (45.8–80.9)	63.4 (45.9–76.5)	0.098
Male	562 (61.2)	236 (69.4)	0.008
Race/ethnicity			
White	597 (65.0)	193 (56.8)	0.009
Asian	14 (1.5)	5 (1.5)	
Black	37 (4.0)	14 (4.1)	
Hispanic	31 (3.4)	11 (3.2)	
Other	55 (6.0)	14 (4.1)	
Unknown	185 (20.1)	103 (30.3)	
Weight (kg)	76.0 (65.0–89.0)	80.0 (67.2–95.5)	<0.001
Surgical	440 (47.9)	143 (42.1)	0.076
Insurance type			
Medicaid	70 (7.6)	25 (7.4)	0.966
Medicare	358 (39.0)	135 (39.7)	
Other	491 (53.4)	180 (52.9)	
Emergency/urgent admission	812 (88.4)	301 (88.5)	1.000
Invasive line*	491 (53.4)	270 (79.4)	<0.001
Vasopressor use*	197 (21.4)	164 (48.2)	<0.001
Renal replacement therapy*	2 (0.2)	1 (0.3)	1.000
Mechanical ventilation*	416 (45.3)	252 (74.1)	<0.001
SOFA score*	2 (2–3)	2 (2–3)	0.013
SAPS II score*	32 (25–39)	38 (29–46)	<0.001
Hemoglobin (g/dL)*	10.5 (8.9–12.0)	10.0 (8.5–11.5)	0.001
Platelet count ($\times 10^9$/L)*	175.0 (135.0–225.5)	147.0 (107.0–201.0)	<0.001
INR*	1.4 (1.2)	1.4 (0.6)	0.951
Charlson comorbidity index	3 (1–6)	3 (1–6)	0.747
ISSs			
Maximum AIS score (general)	0 (0–0)	0 (0–0)	0.376
Maximum AIS score (head and neck)	2 (0–3)	2 (0–3)	0.501
Maximum AIS score (face)	0 (0–0)	0 (0–0)	0.915
Maximum AIS score (extremities)	0 (0–2)	0 (0–2)	0.004
Maximum AIS score (chest)	0 (0–1)	0 (0–3)	<0.001
Maximum AIS score (abdomen)	0 (0–0)	0 (0–1)	0.001
ISS	9 (4–16)	13 (9–22)	<0.001
Pre-ICU hospital length of stay	75.0 (40.0–135.5)	63.0 (22.5–120.3)	0.006
Platelet transfusions*	0 (0–0)	0(0–0)	0.004
Fresh frozen plasma transfusions*	0 (0–0)	0 (0–0)	<0.001
Cryoprecipitate transfusions*	0 (0–0)	0 (0–0)	0.001
Red blood cell transfusions*	0 (0–0)	0 (0–350)	<0.001
Outcomes			
Duration of mechanical ventilation (days)	3 (2–4)	7 (5–11)	<0.001
Duration of renal replacement therapy (days)	0 (0–0)	0 (0–0)	<0.001
ICU length of stay (days)	2.9 (1.8–5.3)	8.8 (5.7–13.9)	<0.001
Hospital length of stay (days)	8.2 (5.2–13.7)	15.7 (10.4–24.8)	<0.001
ICU mortality	48 (5.2)	43 (12.6)	<0.001
In-hospital mortality	71 (7.7)	67 (19.7)	<0.001
28-day mortality	103 (11.2)	81 (23.8)	<0.001
60-day mortality	123 (13.4)	97 (28.5)	<0.001
90-day mortality	141 (15.3)	101 (29.7)	<0.001

Data expressed as n (%), median (interquartile range), and mean ± SD.
AIS: Abbreviated injury scale; ICU: Intensive care unit; INR: International normalized ratio; ISS: Injury severity score; MODS: Multiple organ dysfunction syndrome; SAPS II: Simplified acute physiology score II; SD: Standard deviation; SOFA: Sequential organ failure assessment.
* The data were collected within the first 24 h after admission.

Nomogram model

A multivariate logistic regression model was developed using the selected features to predict MODS in trauma-induced sepsis patients. Table 2 presents the odds ratios and *P*-values for each variable. Based on the logistic regression model, a nomogram was constructed to provide a visual representation of MODS probability (Figure 2). Additionally, an interactive web-based nomogram was developed to support clinical application (Figure 3A, available at https://ccmaigroup.shinyapps.io/TraumaSepsis_Mods_predict_nomogram).

Table 2
Multivariable logistic regression for MODS prediction in septic trauma patients using screened predictors.

Variable	Odds ratio (95% CI)	*P*-value
SAPS II score	1.032 (1.016 to 1.050)	<0.001
Max chest AIS	1.261 (1.131 to 1.406)	<0.001
Age (years) (<60 as reference)	1.390 (0.925 to 2.094)	0.113
Weight (kg) (40–74 as reference)		
<40	4.118 (0.194 to 35.949)	0.239
75–109	1.339 (0.954 to 1.882)	0.092
≥110	2.757 (1.562 to 4.855)	<0.001
Invasive line (no as reference)	1.349 (0.895 to 2.043)	0.155
Vasopressor used (no as reference)	1.800 (1.243 to 2.606)	0.002
Mechanical ventilation (no as reference)	2.283 (1.542 to 3.403)	<0.001
Platelet count (×10^9/L) (≥150 as reference)		
<50	3.123 (0.948 to 9.958)	0.054
50–99	2.656 (1.615 to 4.356)	<0.001
100–149	1.462 (1.011 to 2.107)	0.042

AIS: Abbreviated injury scale; CL; MODS: Multiple organ dysfunction syndrome; SAPS II: Simplified acute physiology score II.

Machine learning models

Five machine learning models were developed: KNN, SVC, RF, XGB, and MLP, along with a stacked ensemble model combining these algorithms. The RF model was additionally deployed as a web-based application (available at https://traumasepsismods-8awxzcemtbnphxpydcg8ec.streamlit.app), to enable interactive testing and exploration (Figure 3B). Feature importance analysis was performed using the RF, XGB, and MLP models. Figure 4A presents the feature importance scores displayed as a radar chart. Although the ranking of individual features varied, key predictors consistently emerged. The SAPS II score was the most influential feature across all models, emphasizing its critical role in predicting MODS. Additional important variables included platelet count, vasopressor use, and mechanical ventilation, though their relative importance differed among models. SHAP analysis revealed that invasive procedures, elevated SAPS II scores, severe chest injuries, and lower platelet counts were associated with a higher likelihood of developing MODS (Figure 4B).

Model performance evaluation

In the temporal validation cohort, the proposed models demonstrated superior discriminative performance compared to conventional scoring systems (all $P < 0.05$) (Figure 5). The RF model achieved the highest area under the curve (AUC) with 0.769 (95% confidence interval [CI]: 0.712 to 0.826), while the nomogram model yielded an AUC of 0.757 (95% CI: 0.700 to 0.814). Among the new models, KNN had the

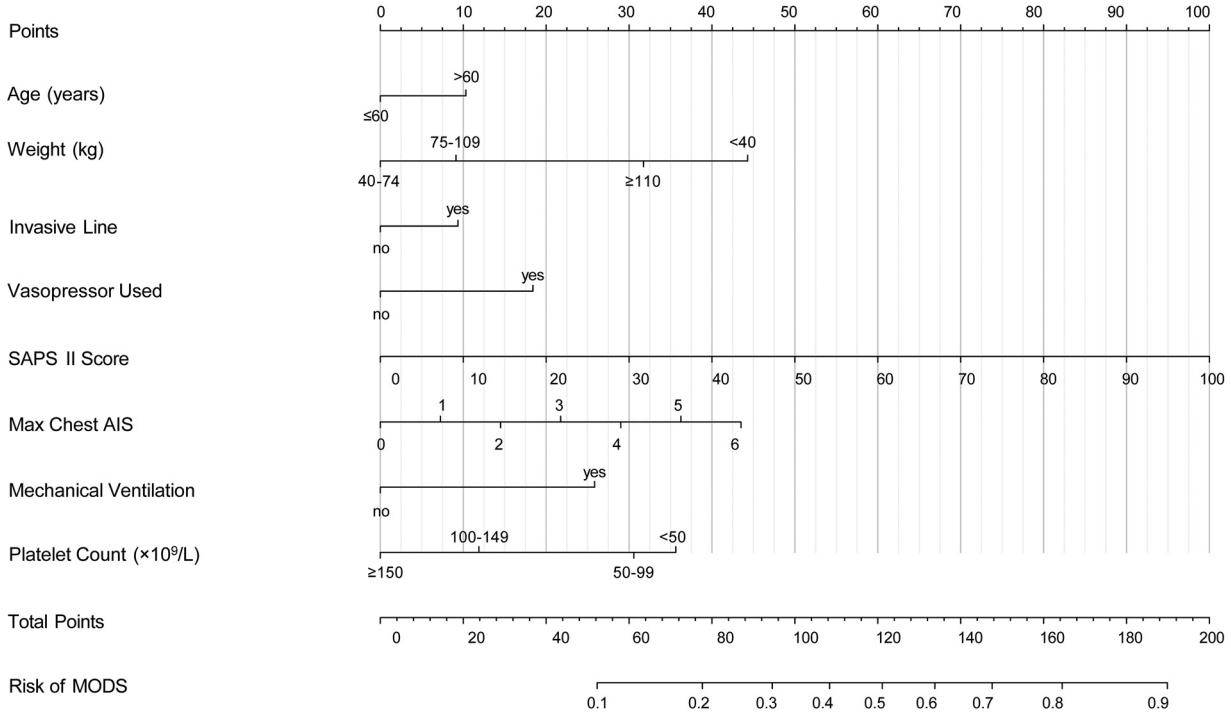

Figure 2. Nomogram for predicting the probability of MODS in patients with trauma-induced sepsis.
AIS: Abbreviated injury scale; MODS: Multiple organ dysfunction syndrome; SAPS II: Simplified acute physiology score II.

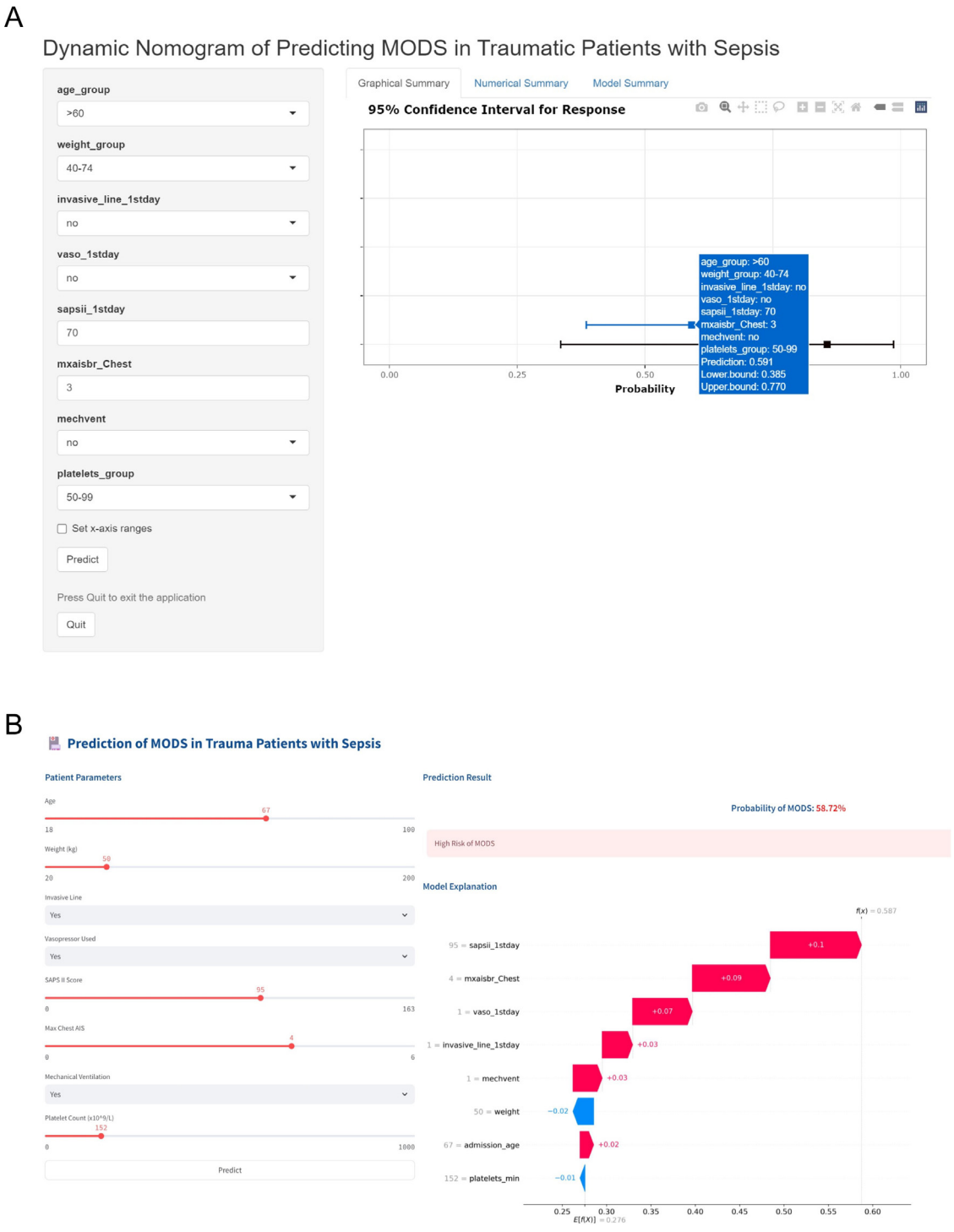

Figure 3. Web-based applications for predicting MODS in patients with trauma-induced sepsis. A: Dynamic nomogram: this interactive web application allows users to input patient-specific variables and obtain a personalized prediction of MODS risk. Users can input values for age, weight, presence of invasive lines, vasopressor use, SAPS II score, maximum AIS score for the chest region, mechanical ventilation, and platelet count. It provides a graphical and numerical summary of the prediction, including a 95% confidence interval. B: RF model predictor: this application utilizes an RF algorithm to predict MODS risk. Users input patient parameters, and the model generates a probability of MODS occurrence. Uniquely, this tool provides a SHAP plot, offering a visual interpretation of how each input variable contributes to the final prediction, enhancing model transparency and interpretability. Both applications feature a disclaimer emphasizing their intended use for research purposes only and the necessity of professional medical judgment in clinical decision-making. These tools aim to assist healthcare providers in risk assessment but should not replace clinical expertise.

AIS: Abbreviated injury scale; MODS: Multiple organ dysfunction syndrome; RF: Random forest; SAPS II: Simplified acute physiology score II; SHAP: SHapley Additive exPlanations.

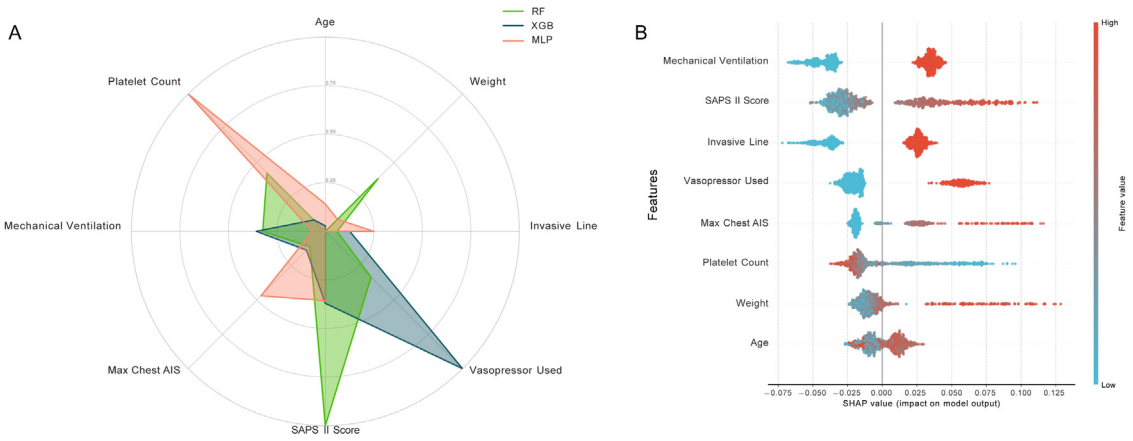

Figure 4. Feature importance and SHAP analysis for predicting MODS in patients with trauma-induced sepsis. A: Feature importance radar chart compares the relative importance of predictive features across three machine learning models: RF, XGB, and MLP. The chart illustrates how each model weighs the importance of different features in predicting MODS. B: SHAP summary plot illustrates the impact of each feature on the model's output using SHAP values. Features are ranked by their overall importance from top to bottom. Each point represents a patient, with color indicating whether the feature value is high (red) or low (blue) for that patient. Points to the right indicate a positive impact on MODS prediction, while points to the left indicate a negative impact.
AIS: Abbreviated injury scale; MLP: Multilayer perceptron; MODS: Multiple organ dysfunction syndrome; RF: Random forest; SAPS II: Simplified acute physiology score II; SHAP: SHapley Additive exPlanations; XGB: XGBoost.

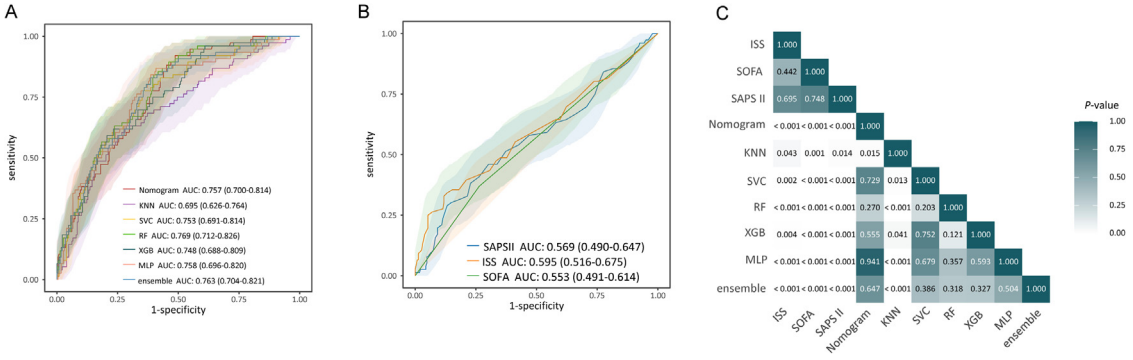

Figure 5. Comparison of model performance for predicting MODS in patients with trauma-induced sepsis. A: ROC curves for newly developed models. B: ROC curves for conventional scoring systems. C: Heat map of statistical differences between model AUCs. The color intensity represents the *P*-value, with darker green indicating higher *P*-values (less significant differences) and lighter colors indicating lower *P*-values (more significant differences).
AUC: Area under the curve; ISS: Injury severity score; KNN: K-nearest neighbor; MLP: Multilayer perceptron; MODS: Multiple organ dysfunction syndrome; RF: Random forest; ROC: Receiver operating characteristic; SAPS II: Simplified acute physiology score II; SOFA: Sequential organ failure assessment; SVC: Support vector classifier; XGB: XGBoost.

Table 3
Performance metrics of RF model for predicting MODS at various cut-off values.

Cut-off value*	Count†	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0.154	259	38.49	97.37	18.86	28.57	95.56
0.191	204	55.26	94.74	42.11	35.29	96
0.222	172	62.5	88.16	53.95	38.95	93.18
0.266	129	66.12	67.11	65.79	39.53	85.71
0.330	84	75	55.26	81.58	50	84.55
0.422	43	75.33	28.95	90.79	51.16	79.31

MODS: Multiple organ dysfunction syndrome; NPV: Negative predictive value; PPV: Positive predictive value; RF: Random forest.
* The threshold probability above which a prediction is classified as positive.
† The number of cases predicted as positive at each cut-off value.

lowest AUC but still outperformed traditional scores. The RF model's performance varied across different cut-off thresholds (Table 3).

Calibration plots for the nomogram and RF models are shown in Figure 6. The nomogram had a calibration slope of 0.900 (95% CI: 0.618 to 1.182) and an intercept of −0.269 (95% CI: −0.557 to −0.020). The RF model demonstrated a calibration slope of 1.634 (95% CI: 1.141 to 2.128) and an intercept of −0.137 (95% CI: −0.407 to 0.133). Brier scores were 0.164 (95% CI: 0.142 to 0.187) for the nomogram and 0.160 (95% CI: 0.142 to 0.187) for the RF model. The decision curve analysis (DCA) illustrated the clinical utility of the nomogram and RF models (Figure 7). Both models consistently provided a higher net benefit compared to conventional scoring systems, including SAPS II, SOFA, and ISS, particularly at threshold probabilities below 0.23.

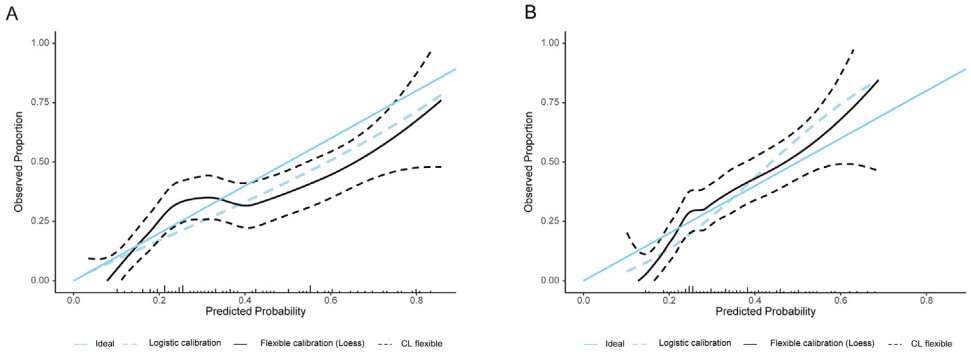

Figure 6. Calibration plots for the nomogram (A) and RF (B) models in predicting MODS in patients with trauma-induced sepsis.
CL: Confidence limit; MODS: Multiple organ dysfunction syndrome; RF: Random Forest.

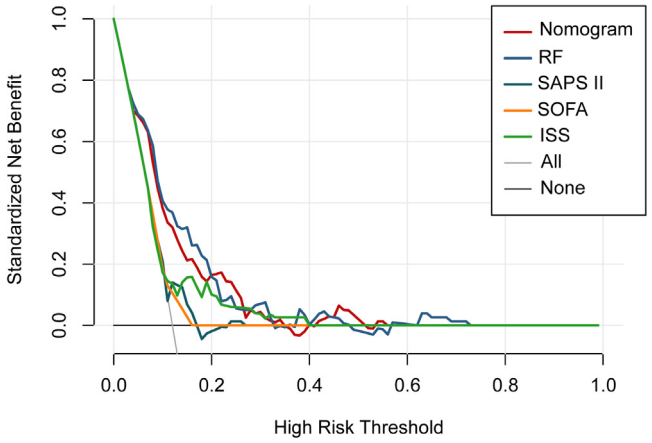

Figure 7. DCA comparing the clinical utility of different models for predicting MODS in patients with trauma-induced sepsis.
DCA: Decision curve analysis; ISS: Injury severity score; MODS: Multiple organ dysfunction syndrome; RF: Random forest; SAPS II: Simplified acute physiology score II; SOFA: Sequential organ failure assessment.

Discussion

This study developed and validated prediction models for MODS in patients with trauma-induced sepsis using the MIMIC-IV version 2.2 database. We employed both traditional statistical methods (nomogram) and advanced machine learning techniques, including KNN, SVC, RF, XGB, MLP, and a stacked ensemble of these models. All newly constructed models significantly outperformed conventional scoring systems such as SAPS II, SOFA, and ISS (all P <0.05). Additionally, we developed web-based applications for the nomogram and RF model, enabling real-time risk assessment and facilitating clinical implementation by expediting the identification of high-risk patients and supporting timely interventions.

Eight key predictors for MODS development in trauma-induced sepsis patients were identified: age, weight, presence of invasive lines on the first ICU day, vasopressor use on the first ICU day, SAPS II score, maximum AIS score for the chest region, mechanical ventilation, and platelet count. These predictors provide valuable insights into the pathophysiological mechanisms underlying MODS. The SAPS II score consistently emerged as the top predictor across models, underscoring the critical role of overall illness severity in MODS development. This finding aligns with prior studies highlighting the relationship between initial illness severity and subsequent organ dysfunction.[20,21] Similarly, mechanical ventilation and vasopressor use were significant predictors, likely reflecting the severity of the patient's condition and the necessity for aggressive life-support measures. While these interventions are essential, they may also contribute to MODS through mechanisms such as ventilator-associated lung injury and systemic inflammation,[22] as well as vasopressor-induced microcirculatory compromise and tissue hypoxia.[23] However, our study design does not allow differentiation between whether these factors directly cause MODS or serve as markers of severe illness. The presence of invasive lines on the first ICU day may signal greater initial trauma severity or hemodynamic instability, further emphasizing its predictive value.[24] Similarly, the maximum AIS score for the chest region underscores the impact of thoracic trauma on MODS development due to its direct effects on critical organs such as the heart and lungs.[25,26] Patient-specific characteristics such as age and weight also proved significant. Advanced age is a well-recognized risk factor for adverse outcomes in critically ill patients,[27] while weight may affect pharmacokinetics, dosing strategies, and overall physiological reserves.[28] Finally, platelet count emerged as a crucial predictor, likely reflecting the interplay between coagulopathy and systemic inflammation in MODS pathogenesis.[29] Thrombocytopenia may indicate consumption coagulopathy, bone marrow suppression, or sepsis-induced thrombocytopenia, all of which are associated with an increased risk of organ dysfunction.

In our study, the RF model (AUC=0.769, 95% CI: 0.712 to 0.826) and the nomogram (AUC=0.757, 95% CI: 0.700 to 0.814) demonstrated comparable performance, with no statistically significant difference observed between them. This result was somewhat unexpected, as we initially anticipated that machine learning models would outperform the nomogram due to their capacity to capture complex, non-linear relationships. Several factors may explain this outcome. First, the relatively small sample size (955 patients in the derivation cohort and 304 in the validation cohort) may have constrained the machine learning models' ability to realize their full potential. Prior research has indicated that large datasets are often necessary for machine learning algorithms to surpass traditional statistical methods.[30,31] Second, the limited number of selected features (eight predictors) may have reduced the complexity of the dataset, limiting the advantages typically associated with

machine learning models in high-dimensional settings.[32,33] We also implemented an ensemble model through stacking multiple algorithms, a common technique for boosting predictive performance. However, the ensemble model (AUC=0.763, 95% CI: 0.704 to 0.821) did not surpass the single RF model (AUC=0.769, 95% CI: 0.712 to 0.826). This could be due to the inherent ensemble nature of the RF algorithm, which may have already captured the optimal predictive potential of our dataset. Future studies could explore alternative ensemble strategies, such as testing different combinations of base models or meta-learners, to better leverage complementary predictive features.[34] Despite the comparable performance of the RF and nomogram models, both significantly outperformed conventional scoring systems, highlighting the importance of selecting relevant, trauma-specific predictors. This finding suggests that predictor selection, rather than the modeling technique itself, maybe the critical factor in enhancing predictive accuracy in this context.

To facilitate clinical implementation, we developed web-based applications for both the nomogram and RF models. These user-friendly platforms enable clinicians to input patient data and receive real-time risk assessments, potentially supporting early identification of high-risk patients and enabling timely interventions. However, these models should be interpreted as tools that augment clinical judgment rather than replace it. Integration into clinical workflows should proceed cautiously, with continuous monitoring and validation across diverse clinical settings to ensure reliability and effectiveness.

However, this study has certain limitations. First, model development was exclusively based on the MIMIC-IV database, which represents a single-center experience from a tertiary care hospital in the United States. While this database provides extensive clinical information, its limited geographic and institutional scope may restrict the generalizability of the models to other healthcare settings with different patient demographics, clinical practices, and treatment protocols. This limitation is particularly relevant when considering the potential application of these models in international contexts, where healthcare systems, resource availability, and patient characteristics may differ substantially. Additionally, although our models demonstrated superior performance compared to conventional scoring systems, this comparison has inherent limitations. Our models were validated within a single-center dataset, whereas scoring systems like SOFA and acute physiology and chronic health evaluation (APACHE) have undergone multi-center validation. This discrepancy underscores the importance of further external validation to strengthen the models' reliability. Second, temporal validation using data from 2017 to 2019 assessed the models' performance over time but did not address the broader need for external validation across geographically and demographically diverse populations. Future studies should conduct multicenter, international validation to evaluate the models' generalizability and identify potential recalibration needs based on specific healthcare environments. Such validation would be essential before recommending these predictive tools for widespread clinical use. Third, the models were constructed using variables available in the MIMIC-IV database, potentially omitting other critical biomarkers or clinical interventions that could influence MODS risk. Incorporating additional clinical, laboratory, and molecular data from diverse sources could enhance predictive accuracy and provide a more comprehensive risk assessment. Lastly, while we developed web-based applications to facilitate clinical implementation, these tools have not yet been tested in real-world clinical settings. Their impact on clinical decision-making, workflow integration, and patient outcomes remains to be evaluated.

Conclusions

In this study, we developed and validated prediction models for MODS in patients with trauma-induced sepsis using the MIMIC-IV database. Both the nomogram and machine learning models outperformed conventional scoring systems in predicting MODS. However, further research is needed to perform external validation in diverse clinical settings and to assess the impact of model implementation on patient outcomes, ensuring their effectiveness and clinical utility across different healthcare environments.

CRediT Authorship Contribution Statement

Jinyu Peng: Writing – original draft, Software, Investigation, Formal analysis. **Yun Li:** Visualization, Data curation. **Chao Liu:** Writing – review & editing. **Zhi Mao:** Writing – review & editing. **Hongjun Kang:** Conceptualization. **Feihu Zhou:** Conceptualization.

Acknowledgments

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Statement

Not applicable.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The author Feihu Zhou is an Editorial Board Member for this journal and was not involved in the editorial review or the decision to publish this article.

Data Availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

[1] Eriksson J, Eriksson M, Brattström O, Hellgren E, Friman O, Gidlöf A, et al. Comparison of the sepsis-2 and sepsis-3 definitions in severely injured trauma patients. J Crit Care 2019;54:125–9. doi:10.1016/j.jcrc.2019.08.019.

[2] Haug VF, Tapking C, Panayi AC, Thiele P, Wang AT, Obed D, et al. Long-term sequelae of critical illness in sepsis, trauma and burns: a systematic review and meta-analysis. J Trauma Acute Care Surg 2021;91(4):736–47. doi:10.1097/TA.0000000000003349.
[3] Eguia E, Bunn C, Kulshrestha S, Markossian T, Durazo-Arvizu R, Baker MS, 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(9):1296–303. doi:10.1097/CCM.0000000000004451.
[4] Guirgis FW, Brakenridge S, Sutchu S, Khadpe JD, Robinson T, Westenbarger R, et al. The long-term burden of severe sepsis and septic shock: sepsis recidivism and organ dysfunction. J Trauma Acute Care Surg 2016;81(3):525–32. doi:10.1097/TA.0000000000001135.
[5] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for sepsis and septic shock (Sepsis-3). JAMA 2016;315(8):801–10. doi:10.1001/jama.2016.0287.
[6] Lord JM, Midwinter MJ, Chen YF, Belli A, Brohi K, Kovacs EJ, et al. The systemic immune response to trauma: an overview of pathophysiology and treatment. Lancet 2014;384(9952):1455–65. doi:10.1016/S0140-6736(14)60687-5.
[7] Bion J, Strunin L. Multiple organ failure: from basic science to prevention. Br J Anaesth 1996;77(1):1–2. doi:10.1093/bja/77.1.1.
[8] Livingston DH, Deitch EA. Multiple organ failure: a common problem in surgical intensive care unit patients. Ann Med 1995;27(1):13–20. doi:10.3109/07853899509031931.
[9] Gourd NM, Nikitas N. Multiple organ dysfunction syndrome. J Intensive Care Med 2020;35(12):1564–75. doi:10.1177/0885066619871452.
[10] Fröhlich M, Lefering R, Probst C, Paffrath T, Schneider MM, Maegele M, et al. Epidemiology and risk factors of multiple-organ failure after multiple trauma: an analysis of 31,154 patients from the TraumaRegister DGU. J Trauma Acute Care Surg 2014;76(4):921–7 discussion 927-8. doi:10.1097/TA.0000000000000199.
[11] Lelubre C, Vincent JL. Mechanisms and treatment of organ failure in sepsis. Nat Rev Nephrol 2018;14(7):417–27. doi:10.1038/s41581-018-0005-7.
[12] Johnson A, 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. doi:10.1038/s41597-022-01899-x.
[13] Clark DE, Black AW. ICDPIC-R Version 1.0.0. Injury 2021;52(11):3545. doi:10.1016/j.injury.2021.02.079.
[14] Liu X, DuMontier C, Hu P, Liu C, Yeung W, Mao Z, et al. Clinically interpretable machine learning models for early prediction of mortality in older patients with multiple organ dysfunction syndrome: an International Multicenter Retrospective Study. J Gerontol A Biol Sci Med Sci 2023;78(4):718–26. doi:10.1093/gerona/glac107.
[15] Kursa MB, Rudnicki WR. Feature selection with the Boruta Package. J Stat Softw 2010;36(11):1–13. doi:10.18637/jss.v036.i11.
[16] Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22(7):707–10. doi:10.1007/BF01709751.
[17] Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, et al. The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. JAMA 1996;276(10):802–10. doi:10.1001/jama.276.10.802.
[18] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594. doi:10.1136/bmj.g7594.
[19] Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc; 2017. p. 4768–77.
[20] Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993;270(24):2957–63. doi:10.1001/jama.270.24.2957.
[21] Peres Bota D, Melot C, Lopes Ferreira F, Nguyen Ba V, Vincent JL. The Multiple Organ Dysfunction Score (MODS) versus the Sequential Organ Failure Assessment (SOFA) score in outcome prediction. Intensive Care Med 2002;28(11):1619–24. doi:10.1007/s00134-002-1491-3.
[22] Dhanireddy S, Altemeier WA, Matute-Bello G, O'Mahony DS, Glenny RW, Martin TR, et al. Mechanical ventilation induces inflammation, lung injury, and extra-pulmonary organ dysfunction in experimental pneumonia. Lab Invest 2006;86(8):790–9. doi:10.1038/labinvest.3700440.
[23] Meng L. Heterogeneous impact of hypotension on organ perfusion and outcomes: a narrative review. Br J Anaesth 2021;127(6):845–61. doi:10.1016/j.bja.2021.06.048.
[24] Pape HC, Halvachizadeh S, Leenen L, Velmahos GD, Buckley R, Giannoudis PV. Timing of major fracture care in polytrauma patients - An update on principles, parameters and strategies for 2020. Injury 2019;50(10):1656–70. doi:10.1016/j.injury.2019.09.021.
[25] Battle C, Hutchings H, Lovett S, Bouamra O, Jones S, Sen A, et al. Predicting outcomes after blunt chest wall trauma: development and external validation of a new prognostic model. Crit Care 2014;18(3):R98. doi:10.1186/cc13873.
[26] Asim M, Amin F, El-Menyar A. Multiple organ dysfunction syndrome: contemporary insights on the clinicopathological spectrum. Qatar Med J 2020;2020(1):22. doi:10.5339/qmj.2020.22.
[27] Flaatten H, De Lange DW, Morandi A, Andersen FH, Artigas A, Bertolini G, et al. The impact of frailty on ICU and 30-day mortality and the level of care in very elderly patients (≥ 80 years). Intensive Care Med 2017;43(12):1820–8. doi:10.1007/s00134-017-4940-8.
[28] Hanley MJ, Abernethy DR, Greenblatt DJ. Effect of obesity on the pharmacokinetics of drugs in humans. Clin Pharmacokinet 2010;49(2):71–87. doi:10.2165/11318100-000000000-00000.
[29] Levi M, Opal SM. Coagulation abnormalities in critically ill patients. Crit Care 2006;10(4):222. doi:10.1186/cc4975.
[30] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22. doi:10.1016/j.jclinepi.2019.02.004.
[31] van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14:137. doi:10.1186/1471-2288-14-137.
[32] Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018;15(141):20170387. doi:10.1098/rsif.2017.0387.
[33] Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol 2020;122:95–107. doi:10.1016/j.jclinepi.2020.03.005.
[34] Zhang Z, Chen L, Xu P, Hong Y. Predictive analytics with ensemble modeling in laparoscopic surgery: a technical note. Laparosc Endosc Robotic Surg 2022;5(1):25–34. doi:10.1016/j.lers.2021.12.003.