BMC Emergency Medicine

RESEARCH

Open Access

Early prediction of renal replacement therapy within 24 hours after septic shock recognition in the emergency department using machine learning: a retrospective analysis of a prospectively collected multicenter registry

Sangun Nah[1], Tae Ho Lim[2], Sung Phil Chung[3], Gil Joon Suh[4], Sung-Hyuk Choi[5], Woon Yong Kwon[4], Won Young Kim[6], Kyuseok Kim[7], Sangchun Choi[1], Je Sung You[3], Han Sung Choi[8], Tae Gun Shin[9*†] and Sangsoo Han[1*†]

Abstract

Background Early identification of patients with septic shock who may soon require renal replacement therapy (RRT) is clinically important but challenging in the emergency department (ED), where definitive indications for RRT often have not yet developed at the time of presentation. Recognizing these patients in advance is important for timely planning of RRT initiation, including coordination of equipment and personnel at the hospital level. This study aimed to develop and validate machine learning (ML) models that predict the need for RRT within 24 h of septic shock recognition in the ED.

Methods We analyzed data from the Korean Shock Society septic shock registry collected from October 2015 to December 2023. Feature selection was performed using least absolute shrinkage and selection operator regression, and five ML models were trained. The best-performing model was selected based on the area under the receiver operating characteristic curve (AUROC). Shapley additive explanations were used to interpret the contribution of each feature.

Results In total, 5361 patients were included in the analysis, of whom 728 (13.6%) required RRT within 24 h. Among the evaluated models, categorical boosting (CatBoost) demonstrated the best discrimination with an AUROC of 0.86 (95% CI, 0.833–0.887), outperforming conventional severity scores such as the Sequential Organ Failure Assessment (AUROC, 0.673 [95% CI, 0.628–0.717]) and the Acute Physiology and Chronic Health Evaluation (AUROC, 0.672 [95% CI, 0.623–0.719]).

†Tae Gun Shin and Sangsoo Han contributed equally to this work as the corresponding author.

*Correspondence:
Tae Gun Shin
drshin88@gmail.com
Sangsoo Han
brayden0819@daum.net

Full list of author information is available at the end of the article

Conclusions The CatBoost model demonstrated moderate discriminative performance for predicting early RRT requirement within 24 h of ED septic shock recognition.

Keywords Renal replacement therapy, Septic shock, Machine learning

Introduction

Septic shock is a life-threatening condition commonly encountered in the emergency department (ED), and patients may present with severe hemodynamic instability, acid–base and electrolyte imbalances, and rapidly progressing multi-organ failure [1–3]. These abnormalities can worsen quickly, and approximately 8–13% of patients with septic shock require renal replacement therapy (RRT) [4]. According to the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines, RRT is recommended in cases of refractory metabolic acidosis, persistent hyperkalemia, unresponsive fluid overload, or uremic complications [5].

However, the RRT indications proposed in the KDIGO guidelines apply to patients who have already developed overt signs of kidney failure [4]. Thus, when septic shock is first recognized in the ED, these criteria offer little direct guidance in predicting which patients will need RRT in the short term. Early in septic shock, physiological abnormalities can appear in heterogeneous ways, making it difficult to determine the need for RRT based solely on simple laboratory tests or a single clinical parameter [6]. Moreover, commonly used renal injury indicators such as creatinine trajectory, urine output, or KDIGO staging are primarily intended to characterize established kidney injury rather than to predict the short-term need for RRT at the time of initial ED recognition. Therefore, clinical judgment often requires synthesis of a broad range of presenting information.

Importantly, the optimal timing of RRT initiation remains controversial. Although ELAIN trial suggested potential benefits of early initiation, large multicenter randomized controlled trials including AKIKI, IDEAL-ICU, and STARRT-AKI have not demonstrated a consistent survival advantage of accelerated strategies compared with delayed approaches [7–11]. These findings indicate that defining a universally optimal biological timing for RRT remains challenging.

Regardless of the superiority of any particular timing strategy, hospitals must still anticipate which patients are likely to require RRT in the immediate future in order to allocate equipment, dialysis personnel, and critical care resources appropriately. Because RRT is a resource-intensive treatment that requires specialized equipment and trained dialysis personnel, and its immediate initiation may be limited in clinical practice because of equipment availability, staffing constraints, and other operational factors [12, 13]. Therefore, predicting the likelihood of RRT initiation within a clinically relevant short-term window may provide operational value by facilitating timely preparation, rather than by mandating premature intervention.

Recently, various machine learning (ML)–based predictive models have shown promise not only in assessing the prognosis of patients with septic shock but also in the early prediction of major interventional decisions such as tracheal intubation, and research continues to explore their clinical applicability [14–16]. Despite these advances, most studies on RRT have focused on risk factor analysis in intensive care unit patients or those who have already developed acute kidney injury, and no ML study has yet predicted the need for short-term RRT initiation based on early information available when septic shock is first recognized in the ED [17, 18]. If such an early prediction model were available, it could provide practical support for preparing RRT equipment and coordinating medical resources. Therefore, we conducted a prediction model development and internal validation study using supervised machine learning algorithms to estimate the probability of RRT initiation within 24 h after septic shock recognition in the ED. Candidate predictors were restricted to variables available at the time of septic shock recognition to ensure clinical applicability. The intended use of the model is to assist emergency physicians at the time of septic shock recognition in identifying patients at high short-term risk of requiring RRT within the subsequent 24 h, thereby facilitating early preparation of dialysis resources, rather than mandating immediate initiation of RRT.

Methods

Data source and study population

This study is a retrospective analysis of prospectively collected data from the Korean Shock Society (KoSS) septic shock registry collected between October 2015 and December 2023. The KoSS is a multicenter research consortium dedicated to improving the diagnosis and management of sepsis. Since October 2015, the KoSS septic shock registry has prospectively collected predefined clinical data from patients who presented to the EDs of 20 tertiary teaching hospitals across South Korea [15]. The registry systematically compiles comprehensive information on patients with sepsis, including demographic characteristics, vital signs, past medical history, laboratory results, treatments administered, and clinical outcomes. Adult patients (aged ≥ 19 years) with either suspected or confirmed infection who exhibited signs of tissue hypoperfusion or refractory hypotension were

included [19]. Because the KoSS registry was established prior to the release of the Sepsis-3 definitions, the inclusion criteria followed the 2012 Surviving Sepsis Campaign standards regarding tissue hypoperfusion and refractory hypotension [20]. To preserve data consistency throughout the research, this original framework was utilized for the entire duration of the study. Hypoperfusion was defined as a serum lactate concentration of ≥ 4 mmol/L. Hypotension was defined as systolic blood pressure (SBP) of ≤ 90 mmHg, mean arterial pressure of ≤ 70 mmHg, or a decrease in SBP of > 40 mmHg from baseline [20]. Refractory hypotension was defined as persistent hypotension despite adequate fluid resuscitation (20–30 mL/kg or at least 1 L of crystalloid solution administered over 30 min) or the requirement for vasopressors to sustain SBP of ≥ 90 mmHg or mean arterial pressure of ≥ 70 mmHg [20, 21]. Initial management, including fluid resuscitation, vasopressor use and timely antibiotic administration, was performed in accordance with the Surviving Sepsis Campaign guidelines [22].

For this study, eligible patients were selected according to the Sepsis-3 criteria. Sepsis was defined as suspected or confirmed infection accompanied by an increase in the Sequential Organ Failure Assessment (SOFA) score of ≥ 2 points from baseline. Septic shock was defined as sepsis with persistent hypotension requiring vasopressors to maintain a mean arterial pressure ≥ 65 mmHg and a serum lactate level > 2 mmol/L [1]. The time of septic shock recognition was retrospectively determined using prospectively recorded registry variables and was defined as the earliest time during the ED encounter at which both vasopressor initiation and a serum lactate level > 2 mmol/L were documented. When pre-existing organ dysfunction data were unavailable, the baseline SOFA score was assumed to be 0 in accordance with the registry protocol. The exclusion criteria were as follows: patients receiving chronic maintenance dialysis, those who underwent RRT more than 24 h after septic shock recognition, those diagnosed with septic shock more than 6 h after ED arrival, and those with do-not-resuscitate (DNR) orders. Ethical approval was obtained from the institutional review boards of all participating centers, and the requirement for informed consent was waived in accordance with the relevant ethical regulations. This study adhered to the TRIPOD-AI guidelines [23].

Primary outcome

The primary outcome was defined as the initiation of RRT within 24 h of septic shock recognition in the ED. RRT initiation was determined according to the clinical judgment of the emergency physicians at each participating center.

Feature selection

We collected a comprehensive set of clinical variables available prior to or at the time of septic shock recognition in the ED. In total, 60 candidate variables were initially included. These encompassed demographic characteristics (age and sex), vital signs at septic shock recognition (SBP, diastolic blood pressure, heart rate, respiratory rate, and body temperature), altered mental status (Glasgow Coma Scale score < 15), comorbidities (hypertension, diabetes mellitus, cardiovascular disease, cerebrovascular accident, chronic lung disease, hematologic cancer, metastatic cancer, chronic kidney disease, chronic liver disease, solid organ transplant, and dementia), suspected infection sources (pulmonary, gastrointestinal, urinary tract, hepatobiliary, central nervous system, bone or soft tissue, catheter-related infection, endocarditis, bloodstream infection, other, and multifocal infection), and initial laboratory findings (white blood cell count, hemoglobin, platelet count, sodium, potassium, chloride, blood urea nitrogen, creatinine, C-reactive protein, procalcitonin, D-dimer, prothrombin time–international normalized ratio, activated partial thromboplastin time, B-type natriuretic peptide, creatine kinase–myocardial band, albumin, aspartate aminotransferase, alanine aminotransferase, glucose, pH, pCO_2, pO_2, bicarbonate, SaO_2, and lactate). In addition, several clinical factors were included: ED-to-septic shock recognition time, residence in a long-term care facility, antibiotic administration before septic shock recognition, steroid administration before septic shock recognition, receipt of unbalanced crystalloid only (defined as administration of 0.9% normal saline exclusively), and MV (mechanical ventilation) before septic shock recognition. All candidate predictors were restricted to variables available prior to or at the time of septic shock recognition in the ED. The SOFA and Acute Physiology and Chronic Health Evaluation (APACHE) II scores were calculated using clinical and laboratory variables measured at the time of septic shock recognition in the ED. The SOFA and APACHE II scores were collected but excluded from model development because they were used as comparators for evaluating the ML models.

Variables with more than 30% missingness were excluded from the analysis; accordingly, B-type natriuretic peptide was not considered for model development (Supplementary Fig. 1) [24].

ML model development and analysis

The study cohort was randomly divided into train (80%) and test (20%) datasets for model development and performance assessment. After the data split, missing values were handled using an iterative chained-equation imputation approach. The imputation model was fitted using the training dataset and subsequently applied to the test

dataset to avoid data leakage. Continuous variables were imputed iteratively, whereas categorical variables were imputed using the most frequent category. Before model construction, all continuous variables underwent preprocessing, which included outlier handling and normalization. Outlier influence was mitigated using a median absolute deviation–based robust transformation to reduce the impact of extreme values on model training. After this step, continuous predictors were standardized using z-score normalization to achieve a mean of 0 and a standard deviation of 1. All preprocessing parameters, including those used for outlier handling and normalization, were derived from the training dataset and subsequently applied to the test dataset. Given the relatively large number of candidate predictors, least absolute shrinkage and selection operator (LASSO) regression was applied within the training dataset as a dimensionality reduction step to limit the number of variables and reduce the risk of overfitting. The selected predictors were then used consistently across all models to facilitate fair comparison of model performance.

We developed supervised binary classification models, including logistic regression, support vector machine, random forest, extreme gradient boosting (XGBoost), and categorical boosting (CatBoost), to predict the primary outcome [25–28]. Hyperparameter optimization was performed using grid search combined with five-fold cross-validation within the train dataset, which supported stable parameter estimation and reduced the risk of overfitting. The tuned hyperparameters for each model are presented in Supplementary Table 1.

Model performance was assessed in the independent test dataset using several metrics, including the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F1-score, and the area under the precision–recall curve (AUPRC). For threshold-dependent metrics, model-specific probability thresholds were determined in the training dataset by maximizing the F1-score using five-fold out-of-fold predictions and subsequently applied to the independent test dataset. The model demonstrating the highest discriminative ability, primarily based on AUROC, was designated as the final model. The AUROCs of the ML models were compared using DeLong's test, using the final model as the reference model for pairwise comparisons. To provide clinical context, the predictive performance of the final ML model was compared with conventional illness-severity scoring systems, specifically the SOFA and APACHE II scores, as well as the SACrA score, a recently proposed clinical score for predicting the need for RRT in patients admitted to the intensive care unit, using the same test dataset [29, 30]. Model calibration was evaluated using calibration curves, and clinical utility was assessed using decision curve analysis (DCA) across a range of threshold probabilities.

To enhance interpretability, Shapley additive explanations (SHAP) were applied to the final model. SHAP summary plots and mean absolute SHAP values were used to identify and visualize the predictors that contributed most substantially to model output, thereby providing insight into the clinical factors associated with prediction of RRT.

Sensitivity analysis

To assess the potential impact of excluding patients who underwent RRT more than 24 h after septic shock recognition ($n = 349$), a sensitivity analysis was performed in which these patients were reclassified into the non-RRT group. Model performance was subsequently re-evaluated using the same modeling framework as in the primary analysis.

Statistical analysis

For continuous variables, normality was assessed using the Shapiro–Wilk test. Because all variables deviated from a normal distribution, they were compared using the Mann–Whitney U test and are presented as medians with interquartile ranges. Categorical variables were compared using Fisher's exact test or the chi-square test and are expressed as numbers and percentages. Statistical significance was defined as a two-tailed p-value of < 0.05. All analyses were conducted using R version 4.2.2 (The R Foundation for Statistical Computing, Vienna, Austria) and Python version 3.12 (Python Software Foundation, Wilmington, DE, USA). The following Python libraries were used: NumPy and Pandas for data processing and manipulation; Matplotlib and Seaborn for data visualization; and scikit-learn for data imputation using an iterative chained-equation imputation approach, feature scaling, LASSO-based feature selection, model construction, and performance evaluation.

Results

Participants

In total, 6124 patients who met the Sepsis-3 criteria for septic shock were screened. Among them, 216 patients receiving chronic maintenance dialysis, 349 who underwent RRT more than 24 h after septic shock recognition, 93 diagnosed with septic shock more than 6 h after ED arrival, and 105 with a DNR order were excluded. Ultimately, 5361 patients met the eligibility criteria (Supplementary Fig. 2). The dataset was randomly split into train and test sets at an 8:2 ratio (train set: 4288; test set: 1073).

The baseline characteristics of the eligible participants are summarized in Table 1. Among the 5361 included patients, 728 (13.6%) underwent RRT within 24 h after the diagnosis of septic shock. The median ages of the

Table 1 Comparison of baseline characteristics between RRT and non-RRT groups

	Non-RRT group (*n* = 4633)	RRT group (*n* = 728)	*p*-value
Age, years	71 [62–79]	70.5 [61–78]	0.326
Female sex	1865 (40.3)	278 (38.2)	0.309
Vital signs at septic shock recognition			
SBP, mmHg	91 [76–112]	90 [73–113]	0.32
DBP, mmHg	56 [47–67]	53 [44–68]	0.015
HR, beats/min	112 [96–129]	108 [91–126]	< 0.001
RR, breaths/min	20 [18–24]	23 [20–28]	< 0.001
BT, ℃	37.6 [36.6–38.6]	36.9 [36.1–38.1]	< 0.001
Altered mental status at septic shock recognition	1133 (24.5)	276 (37.9)	< 0.001
Past medical history			
Hypertension	1962 (42.4)	362 (49.7)	< 0.001
Diabetes mellitus	1598 (34.5)	300 (41.2)	< 0.001
Cardiovascular disease	595 (12.8)	115 (15.8)	0.033
Cerebrovascular accident	578 (12.5)	103 (14.2)	0.23
Chronic lung disease	333 (7.2)	45 (6.2)	0.364
Hematologic cancer	356 (7.7)	68 (9.3)	0.143
Metastatic cancer	1530 (33.0)	157 (21.6)	< 0.001
Chronic kidney disease	271 (5.9)	96 (13.2)	< 0.001
Chronic liver disease	505 (10.9)	102 (14.0)	0.016
Transplant	108 (2.3)	18 (2.5)	0.918
Dementia	313 (6.8)	39 (5.4)	0.182
Suspected infection source			
Pulmonary infection	1430 (30.9)	235 (32.3)	0.469
Gastrointestinal infection	889 (19.2)	166 (22.8)	0.026
Urinary tract infection	1107 (23.9)	188 (25.8)	0.278
Hepatobiliary infection	1106 (23.9)	130 (17.9)	< 0.001
Central nervous system infection	17 (0.4)	3 (0.4)	0.746
Bone or soft tissue infection	140 (3.0)	34 (4.7)	0.026
Catheter-related infection	42 (0.9)	9 (1.2)	0.518
Endocarditis	20 (0.4)	5 (0.7)	0.373
Bloodstream infection	165 (3.6)	30 (4.1)	0.52
Other	123 (2.7)	21 (2.9)	0.816
Multifocal infection	677 (14.6)	130 (17.9)	0.026
Laboratory findings			
White blood cells, $\times 10^3/\mu L$	9.5 [3.8–16.8]	9.8 [3.2–17.9]	0.727
Hemoglobin, g/dL	10.7 [9–12.5]	10.4 [8.7–12.6]	0.227
Platelets, $\times 10^3/\mu L$	133 [71–212.5]	108 [49–193]	< 0.001
Sodium, mmol/L	135 [131–139]	135 [131–139]	0.394
Potassium, mmol/L	4.1 [3.6–4.7]	4.5 [3.8–5.3]	< 0.001
Chloride, mmol/L	100 [96–105]	100 [95–105]	0.038
Blood urea nitrogen, mg/dL	29 [20–42.9]	45.3 [30–65.9]	< 0.001
Creatinine, mg/dL	1.4 [0.99–2.1]	2.8 [1.8–4]	< 0.001
CRP, mg/dL	13.5 [5.8–23.1]	17.1 [7.7–28]	< 0.001
Procalcitonin, ng/mL	11.9 [2.2–40.4]	27.8 [7–79.9]	< 0.001
D-dimer, μg/mL	4.6 [2.2–10.4]	6.3 [2.9–13.7]	< 0.001
PT-INR	1.3 [1.2–1.5]	1.5 [1.2–1.8]	< 0.001
aPTT, seconds	35 [29.7–42.4]	39.6 [32.5–49.9]	< 0.001
CK-MB, ng/mL	1.8 [0.8–4.1]	3.9 [1.6–9.2]	< 0.001
Albumin, g/dL	2.9 [2.4–3.4]	2.7 [2.3–3.2]	< 0.001
AST, U/L	44 [26–100]	65 [33–182]	< 0.001
ALT, U/L	28 [15–62]	32 [18–92]	< 0.001
Glucose, mg/dL	137 [104–196]	133 [92–202]	0.005
pH	7.42 [7.36–7.47]	7.32 [7.21–7.4]	< 0.001

Table 1 (continued)

	Non-RRT group (*n* = 4633)	RRT group (*n* = 728)	*p*-value
pCO_2, mmHg	27.9 [23.6–32.6]	26 [20.7–32.5]	< 0.001
pO_2, mmHg	82 [67–100.4]	83.9 [67–115]	0.004
HCO_3^-, mmol/L	18.2 [15–21.3]	13.7 [10.5–17.3]	< 0.001
SaO_2, %	96.1 [93.2–97.9]	95.9 [92–98]	0.04
Lactate, mmol/L	4.4 [2.9–6.3]	6.4 [4.4–10.3]	< 0.001
SOFA score	6 [4–9]	9 [7–11]	< 0.001
APACHE II score	21 [16–27]	28.5 [21.8–37]	< 0.001
ED-to-septic shock recognition time, hours	1 [0.3–2.1]	0.5 [0.2–1.2]	< 0.001
From long-term care facility	475 (10.3)	62 (8.5)	0.166
Antibiotic administration before septic shock recognition	1152 (24.9)	111 (15.3)	< 0.001
Steroid administration before septic shock recognition	50 (1.1)	5 (0.7)	0.436
Unbalanced crystalloid only	1991 (43.0)	396 (54.4)	< 0.001
MV before septic shock recognition	3549 (76.6)	243 (33.4)	< 0.001

Note: Values are presented as median [interquartile range] or number (percentage)

Abbreviations: RRT, renal replacement therapy; SBP, systolic blood pressure; DBP, diastolic blood pressure; HR, heart rate; RR, respiratory rate; BT, body temperature; CRP, C-reactive protein; PT-INR, prothrombin time–international normalized ratio; aPTT, activated partial thromboplastin time; CK-MB, creatine kinase–myocardial band; AST, aspartate transaminase; ALT, alanine transaminase; SOFA, Sequential Organ Failure Assessment; APACHE, Acute Physiology and Chronic Health Evaluation; ED, emergency department; MV, mechanical ventilation

Table 2 Comparison of diagnostic performance of ML models in test dataset

Models	Accuracy	Sensitivity	Specificity	F1-score	AUPRC	AUROC	*p*-value
CatBoost	0.823 [0.801–0.846]	0.637 [0.562–0.719]	0.852 [0.828–0.875]	0.505 [0.442–0.569]	0.49 [0.427–0.581]	0.86 [0.833–0.887]	-
XGBoost	0.853 [0.832–0.873]	0.575 [0.493–0.658]	0.896 [0.875–0.916]	0.515 [0.453–0.576]	0.484 [0.419–0.572]	0.857 [0.83–0.885]	0.084
Logistic regression	0.827 [0.805–0.848]	0.575 [0.5–0.658]	0.866 [0.844–0.888]	0.475 [0.417–0.531]	0.456 [0.393–0.54]	0.852 [0.823–0.879]	0.83
Random forest	0.869 [0.851–0.887]	0.507 [0.432–0.589]	0.926 [0.908–0.943]	0.512 [0.447–0.578]	0.487 [0.42–0.573]	0.848 [0.818–0.877]	0.341
Support vector machine	0.806 [0.783–0.829]	0.671 [0.596–0.747]	0.827 [0.801–0.852]	0.485 [0.436–0.534]	0.48 [0.413–0.563]	0.845 [0.815–0.874]	0.633

Note: Values are presented as mean [95% confidence interval]

Abbreviations: ML, machine learning; AUPRC, area under the precision–recall curve; AUROC, area under the receiver operating characteristic curve; XGBoost, extreme gradient boosting; CatBoost, categorical boosting

p-values were computed by comparing AUROCs the final model (CatBoost) using DeLong's test

non-RRT and RRT groups were 71 and 70.5 years, respectively ($p = 0.326$). The median SOFA and APACHE II scores were significantly higher in the RRT group than in the non-RRT group (9 vs. 6, $p < 0.001$; 28.5 vs. 21, $p < 0.001$, respectively).

Model performance evaluation

Feature selection was performed using LASSO regression with five-fold cross-validation, resulting in 32 selected variables (Supplementary Fig. 3 and Supplementary Table 2). Performance comparisons were conducted across logistic regression, support vector machine, random forest, XGBoost, and CatBoost models. Model performance was evaluated using AUROC, AUPRC, accuracy, precision, sensitivity, specificity, and F1-score (Table 2). Also, positive and negative predictive values and the optimal thresholds are presented in Supplementary Table 3.

Among the five models, CatBoost achieved the highest AUROC (0.86 [95% CI, 0.833–0.887]), followed by XGBoost (0.857 [95% CI, 0.83–0.885]) and logistic regression (0.852 [95% CI, 0.823–0.879]). These three models demonstrated comparable discriminative performance, with CatBoost showing the most balanced trade-off between sensitivity (0.637 [95% CI, 0.562–0.719]) and specificity (0.852 [95% CI, 0.828–0.875]). Tree-based gradient-boosting models (CatBoost and XGBoost) showed slightly better discrimination than conventional approaches, although the ROC curves of all five models largely overlapped (Fig. 1A), suggesting broadly comparable performance. The CatBoost model demonstrated a clearly wider ROC curve than the conventional severity scoring systems and the SACrA (Fig. 1B). The results of the sensitivity analysis are presented in Supplementary Table 4.

DCA showed that the model provided greater net benefit than the default strategies of treating all or none of the patients when the threshold probability ranged from approximately 0.05 to 0.6 (Fig. 2). The calibration curves

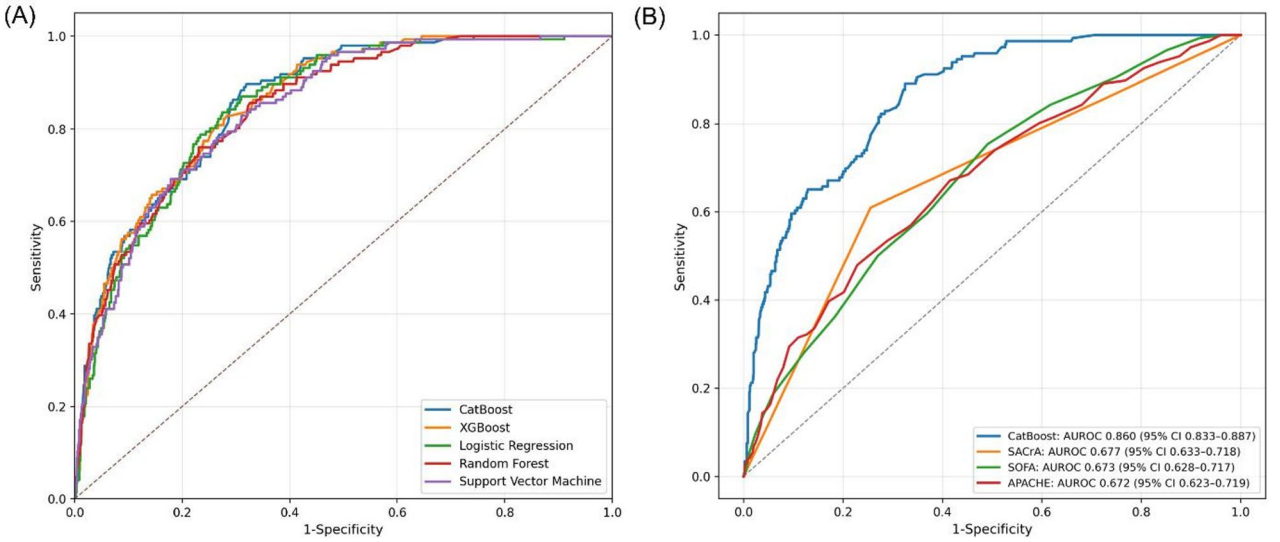

Fig. 1 ROC curves for the ML models and the conventional severity scores (SOFA and APACHE) in predicting RRT (test set). (**A**) ROC curves for the five machine learning models predicting RRT. (**B**) ROC curves for the SOFA, APACHE and SACrA scores predicting RRT. Abbreviations: ROC, receiver operating characteristic; SOFA, Sequential Organ Failure Assessment; APACHE, Acute Physiology and Chronic Health Evaluation; RRT, renal replacement therapy

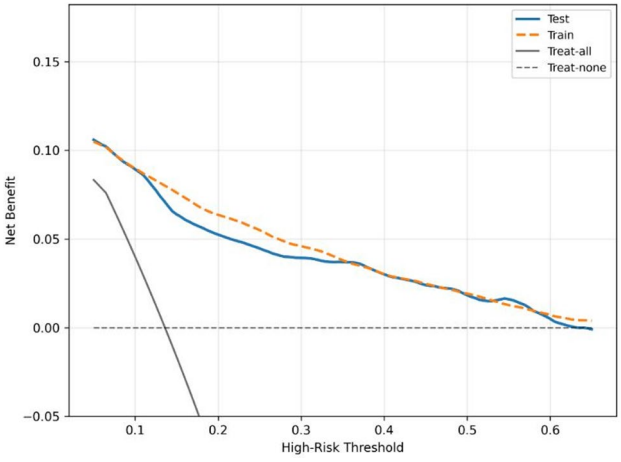

Fig. 2 DCA for the CatBoost model predicting RRT. DCA illustrating the net clinical benefit of the CatBoost model across a range of threshold probabilities in the train (orange dashed line) and test (blue solid line) datasets. The gray solid and dashed lines indicate the default strategies of treating all patients (Treat-all) and treating none (Treat-none), respectively. The model demonstrated a higher net benefit than either default strategy when the threshold probability was approximately between 0.05 and 0.6, suggesting its potential clinical utility in guiding RRT initiation

for the CatBoost model, illustrating the relationship between predicted and actual probabilities of RRT initiation in the test dataset, are presented in Supplementary Fig. 4.

Model explanation

Feature importance was analyzed using SHAP values derived from the CatBoost model. Figure 3A presents the SHAP summary plot, illustrating the direction and magnitude of each variable's contribution to RRT prediction, while Fig. 3B ranks the top 15 features based on their mean absolute SHAP values. Among the top predictors, higher creatinine levels were associated with an increased predicted probability of RRT initiation, whereas higher pH values and the presence of MV before septic shock recognition were associated with a decreased predicted probability.

Discussion

In this study, we developed ML models to predict the need for RRT within 24 h using information available at the time septic shock was first recognized in the ED. Among the evaluated models, CatBoost demonstrated the best performance. However, both sensitivity and the F1-score were suboptimal, with the F1-score showing only moderate performance, underscoring the inherent challenge of achieving fully accurate classification using ED information alone [31]. Even so, the CatBoost model (AUROC, 0.86 [95% CI, 0.833–0.887]) outperformed conventional severity scores, including SOFA (AUROC, 0.673 [95% CI, 0.628–0.717]), APACHE II (AUROC, 0.672 [95% CI, 0.623–0.719]) and SACrA (AUROC, 0.677 [95% CI, 0.633–0.718]), and was able to meaningfully predict RRT initiation using only limited early ED variables. Previous studies evaluating RRT prediction have primarily focused on intensive care unit patients or individuals with established acute kidney injury. Our findings therefore provide clinically relevant insight into the prediction of RRT initiation within 24 h after septic shock recognition in the ED. However, this does not suggest that the model simply reflects overall illness severity. Unlike conventional severity scores that summarize overall illness burden, the present model integrates variables

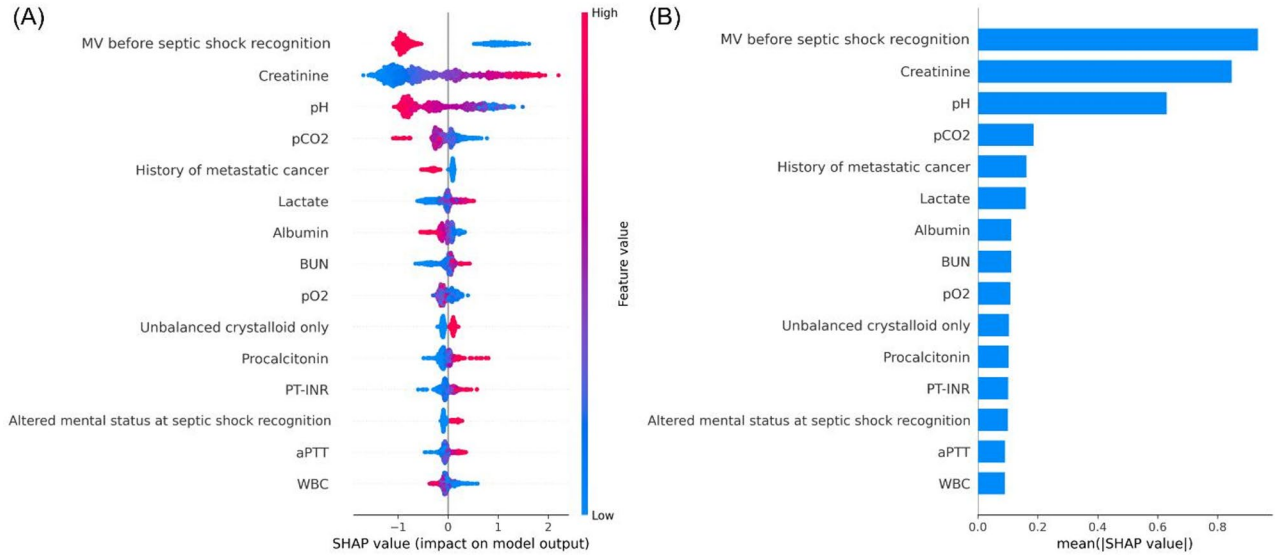

Fig. 3 SHAP summary plot for the top 15 clinical features contributing to the CatBoost model. (**A**) Each point represents a feature's contribution to the model output. The y-axis lists the individual features, while the x-axis shows their corresponding Shapley values. The color gradient indicates whether the feature values are relatively low or high. (**B**) The bar chart illustrates overall feature importance, summarized by the average of the absolute Shapley values. This visualization highlights which covariates had the greatest influence on the final predictive model

more directly related to short-term RRT requirement, including early renal dysfunction and acid–base derangement, together with overall physiological status of the patient. This suggests that the model may capture aspects of RRT risk that are not fully represented by general severity scores.

This study compared the performance of several models, including CatBoost, XGBoost, random forest, support vector machine, and logistic regression. Boosting-based ML models can capture nonlinear relationships and complex patterns among variables, resulting in the most stable and consistent performance [27, 28]. However, although the gradient boosting model performed best overall, the AUROC difference compared with logistic regression was modest, and this difference was not statistically significant, consistent with previous studies showing that ML models are not always markedly superior to traditional statistical approaches [32, 33]. Still, the gradient boosting–based models showed high discriminative performance, with CatBoost achieving the highest AUROC and broadly comparable results across the other evaluation metrics. This may be relevant in the present setting, where the need for RRT in septic shock is likely to reflect the combined pattern of multi-organ physiological derangement rather than a single abnormal variable. In this context, boosting-based models may flexibly capture nonlinear associations and complex interactions among these variables without requiring explicit prespecification of interaction terms [34, 35]. Accordingly, CatBoost was interpreted as a complementary risk-stratification model rather than as a replacement for simpler and more transparent regression-based approaches.

In the DCA, the model yielded greater net benefit than the treat-all or treat-none strategies across the 0.05–0.6 threshold range. This suggests that the model may provide meaningful support for clinical decision-making within a specific risk window, particularly by enabling early identification of high-risk patients during the initial phase when decisions regarding RRT initiation are often difficult to make promptly [36]. In general, lower threshold probabilities reflect situations in which clinicians are willing to act at relatively low levels of predicted risk, whereas higher thresholds indicate that action would only be considered when the predicted risk is sufficiently high. In the emergency setting, this may translate into different levels of clinical attention and early logistical preparation for potential RRT in patients with septic shock. In this sense, the model is better suited as a practical decision-support tool for early risk assessment and resource allocation (e.g., preparing RRT equipment or staffing), rather than as a stand-alone determinant for initiating RRT.

In practice, the model could be integrated into the ED workflow as an automated risk estimation tool embedded within the electronic medical record system and triggered at the time of septic shock recognition using routinely available clinical variables [37]. Once septic shock is identified, the model could automatically generate a predicted probability of RRT requirement within 24 h and present this information to the emergency physician in real time. Patients identified as having a high predicted probability of RRT requirement could prompt earlier nephrology consultation, closer monitoring, and advance logistical preparation for potential RRT

initiation, including coordination of dialysis equipment, trained personnel, and communication with intensive care teams [12, 13]. In this context, the model is best interpreted as a clinical decision-support tool that augments physician judgment and supports early resource planning rather than as a stand-alone determinant for initiating RRT.

The SHAP analysis suggests that the model reflects both renal dysfunction and overall illness severity in septic shock. This finding is clinically plausible because decisions regarding RRT initiation in septic shock are typically based on the combined assessment of kidney injury and the overall physiological status of the patient rather than on renal-specific indicators alone [18]. The negative SHAP contribution of MV suggests that ventilated patients were predicted to have a lower likelihood of requiring RRT within 24 h, which may reflect differences in early organ dysfunction patterns and treatment pathways rather than a direct renal protective effect. Prior sepsis studies have described distinct organ dysfunction phenotypes, such as respiratory-dominant, circulatory-dominant, and renal-dominant patterns [38, 39]. In our study, patients who required MV before septic shock recognition likely belonged to respiratory-dominant phenotypes, in which early renal function may have been relatively preserved. This phenotype difference may partly explain the inverse association observed between MV and short-term RRT requirement. Alternatively, patients who were intubated and mechanically ventilated before septic shock recognition may have received earlier airway stabilization and controlled ventilation at the referring hospital or during prehospital care. Such early stabilization may reduce respiratory muscle oxygen demand and improve systemic oxygen delivery in patients with shock, which could transiently favor renal perfusion in some situations; however, this remains speculative rather than evidence of a causal effect [40].

In this study, increases in creatinine and decreases in pH were both associated with SHAP values shifting toward a higher predicted need for RRT, indicating that early renal dysfunction and the severity of acidemia were each closely linked to short-term RRT requirement. Creatinine reflects a combination of underlying chronic kidney disease, acute tubular injury, and functional decline due to hypoperfusion [18, 41]. Prior studies have consistently shown that early elevation of creatinine is associated with subsequent RRT requirement and mortality [18, 30]. The fact that our model identified creatinine as the top predictor using only early clinical information supports what clinicians have long recognized: its strong prognostic value in renal dysfunction. In septic shock, acidemia arises not only from lactate-driven metabolic acidosis but also from impaired renal acid excretion and mixed acid–base disturbances related to respiratory dysfunction, and its severity is closely associated with an increased need for RRT [17, 42]. The finding that pH retained a substantial SHAP contribution even when lactate and creatinine were included in the model suggests that pH provides additional information not captured by these variables. Taken together, the identification of creatinine and pH as the top predictors suggests that the requirement for RRT initiation within 24 h after septic shock recognition is closely associated with the combined burden of structural renal injury (reflected by elevated creatinine) and systemic metabolic derangement with acidemia (reflected by low pH).

This study was conducted using a multicenter registry and has the strength of developing a practical prediction model based solely on variables that can be readily obtained in the ED. However, several limitations should be noted. First, although the registry was prospectively collected, the analysis was performed retrospectively, so residual confounding cannot be completely excluded. Second, exclusion of patients who underwent RRT more than 24 h after septic shock recognition may have introduced potential selection bias by separating early and late RRT trajectories. To address this concern, we conducted a sensitivity analysis in which these patients were reclassified into the non-RRT group. The results demonstrated only minor changes, suggesting that model discrimination was not substantially affected by this exclusion. Third, important clinical indicators relevant to RRT decision-making—such as urine output, body weight, and cumulative fluid balance—were not available in the KoSS registry and could not be incorporated into the model [18, 43]. Because the registry captures only a single time point, it was not possible to determine the severity of acute kidney injury (e.g., KDIGO stage) or to account for temporal physiological changes in patient status. Baseline SOFA scores were assumed to be 0 when pre-existing organ dysfunction data were unavailable, some degree of misclassification may have occurred. Treatment-related factors such as source control were also not incorporated because the model was designed to use variables available at the time of septic shock recognition in the ED. Fourth, because RRT initiation reflects clinical decision-making rather than a strictly defined physiological threshold and may vary according to institutional practice patterns, the outcome in this study may not represent a universally standardized physiological trigger. Accordingly, the model predicts real-world RRT utilization within a defined short-term window rather than the biologically optimal timing of RRT initiation. Such variation in clinical practice patterns, including thresholds for MV, may also influence model performance when applied in other healthcare settings. Fifth, external validation was limited, and further studies are needed to determine the generalizability of our findings in other settings. In addition,

although the registry includes multiple participating centers, model development and validation were performed using a random train–test split. Therefore, center-based or temporal validation was not conducted, and future studies incorporating these validation strategies would further strengthen the generalizability of the model.

Conclusion

The CatBoost model developed in this study was able to predict initiation of RRT within 24 h after septic shock recognition with reasonable performance using early clinical information available in the ED and outperformed conventional severity scores such as SOFA and APACHE II. However, the model alone cannot determine whether RRT should be initiated, and further evaluation is needed to confirm whether it maintains similar performance across different clinical settings.

Abbreviations

ED	Emergency department
RRT	Renal replacement therapy
KDIGO	Kidney Disease: Improving Global Outcomes
ML	Machine learning
KoSS	Korean Shock Society
SBP	Systolic blood pressure
DNR	Do-not-resuscitate
SOFA	Sequential Organ Failure Assessment
MV	Mechanical ventilation
APACHE	Acute Physiology and Chronic Health Evaluation
MICE	Multiple imputation by chained equation
LASSO	Least absolute shrinkage and selection operator
XGBoost	Extreme gradient boosting
CatBoost	Categorical boosting
AUROC	Area under the receiver operating characteristic curve
AUPRC	Area under the precision–recall curve
DCA	Decision curve analysis
SHAP	Shapley additive explanations

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12873-026-01558-z.

Supplementary Material 1

Acknowledgements

We would like to acknowledge and thank to the Korean Shock Society (KoSS) Investigators: Sang-Min Kim, Seung Mok Ryoo (Asan Medical Center); Gun Tak Lee, Sung Yeon Hwang (Samsung Medical Center); Byuk Sung Ko (Hanyang University Seoul Hospital); Sung-Joon Park (Korea University Guro Hospital); Yoo Seok Park, Jin Ho Beom (Severance Hospital); Taegyun Kim, Yoon Sun Jung (Seoul National University Hospital); Juhyun Song, Kap Su Han (Korea University Anam Hospital); Taeyoung Kong, Eunah Han (Gangnam Severance Hospital); You Hwan Jo, Ji Eun Hwang (Seoul National University Bundang Hospital); Jonghwan Shin, Hui Jai Lee (Seoul Metropolitan Government–Seoul National University Boramae Medical Center); Gu Hyun Kang (Hallym University Kangnam Sacred Heart Hospital) Hanjin Cho, Sejoong Ahn (Korea University Ansan Hospital); Hong Joon Ahn (Chungnam National University Hospital); Kihwan Choi (CHA Gumi Medical Center); Ki Young Jeong, Seok Hun Ko (Kyung Hee University Hospital); Hyo Jin Bang (Seoul St. Mary's Hospital, College of Medicine); Jinwoo Jeoung, Min Joon Seo (Dong-A University Hospital); Heewon Yang (Ajou University Hospital); Chiwon Ahn (Chung-Ang University Hospital); Changsun Kim, Hyungoo Shin (Hanyang University Guri Hospital). And we also thank Ji Eun Moon (Department of Biostatistics, Clinical Trial Center, Soonchunhyang University Bucheon Hospital) for statistical consultation.

Author contributions

Conceptualization: Tae Gun Shin, Sangsoo Han; Methodology: Sangun Nah, Tae Gun Shin, Sangsoo Han; Data acquisition: Tae Ho Lim, Sung Phil Chung, Gil Joon Suh, Sung-Hyuk Choi, Woon Yong Kwon, Won Young Kim, Kyuseok Kim, Je Sung You, Han Sung Choi, Sangsoo Han; Formal analysis and investigation: Tae Ho Lim, Sung Phil Chung, Gil Joon Suh, Sung-Hyuk Choi, Won Young Kim, Kyuseok Kim, Sangchun Choi, Han Sung Choi; Writing – original draft preparation: Sangun Nah; Writing – review and editing: Tae Gun Shin, Sangsoo Han; Supervision: Tae Gun Shin, Sangsoo Han.

Funding

This research was supported by a grant of the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute (grant number: RS-2023-KH134607) and a grant of the Patient-Centered Clinical Research Coordinating Center (grant number: RS-2024-00398566), both funded by the Ministry of Health and Welfare, Republic of Korea to T.G.S., and by the Soonchunhyang University Research Fund to S.H.

Data availability

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

Declarations

Ethics approval and consent to participate

This study was conducted according to the guidelines of the Declaration of Helsinki and was approved by the IRB of each participating center. (1) Asan Medical Center (S2015-1918-0002). (2) Gangnam Sacred Heart Hospital (2015-11-142). (3) Gangnam Severance Hospital (3-2015-0227). (4) Hanyang University Hospital (HYUH 2015-11-013-022). (5) Korea University Anam Hospital (HRPC2016-184). (6) Korea University Kuro Hospital (KUGH15358-001). (7) Samsung Medical Center (SMC2015-09-057-057). (8) Seoul National University Hospital (J-1408-003-599). (9) Seoul National University Bundang Hospital (B-1409/266-401). (10) Severance Hospital (4-2015-0929), (11) Seoul National University Boramae Hospital (IRB-16-2014-36). (12) Soonchunhyang University Bucheon Hospital (IRB No. 2023-04-019). (13) Kyung Hee University Hospital (IRB No. 2023-06-055). (14) Korea University Ansan Hospital (2020AS0021). (15) CHA Bundang Medical Center (IRB No. 2020-02-001). (16) Chungnam National University Hospital (IRB No. 2021-07-055-001). (17) Seoul St. Mary's Hospital (KC21OISI0832). (18) Ajou University Hospital (AJOUIRB-OB-2023-208). (19) Dong-A University Hospital (DAUHIRB-23-108). (20) Chung-Ang University Hospital (2405-156-056). Informed consent was waived in accordance with the relevant ethical regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Emergency Medicine, Soonchunhyang University Bucheon Hospital, 170 Jomaru-ro, Bucheon 14584, Republic of Korea
[2]Department of Emergency Medicine, College of Medicine, Hanyang University, Seoul, Republic of Korea
[3]Department of Emergency Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea
[4]Department of Emergency Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[5]Department of Emergency Medicine, Korea University Guro Hospital, Seoul, Korea
[6]Department of Emergency Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea
[7]Department of Emergency Medicine, CHA Bundang Medical Center, CHA University, Seongnam, Korea
[8]Department of Emergency Medicine, College of Medicine, Kyung Hee University, Seoul, Korea

[9]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Received: 22 December 2025 / Accepted: 20 March 2026
Published online: 27 March 2026

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche J-D, Coopersmith CM. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315(8):801–10.
2. Pandey S. Advances in metabolomics in critically ill patients with sepsis and septic shock. Clin Exp Emerg Med. 2024;12(1):4.
3. Shin TG. Assessment of organ failure in sepsis patients in the emergency department: clinical evaluation, Sequential Organ Failure Assessment (SOFA) score, and future perspectives. Clin Exp Emerg Med. 2024;11(4):327.
4. Hellman T, Uusalo P, Järvisalo MJ. Renal replacement techniques in septic shock. Int J Mol Sci. 2021;22(19):10238.
5. Group K. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl. 2012;2:1.
6. Zarbock A, Nadim MK, Pickkers P, Gomez H, Bell S, Joannidis M, Kashani K, Koyner JL, Pannu N, Meersch M. Sepsis-associated acute kidney injury: consensus report of the 28th Acute Disease Quality Initiative workgroup. Nat Rev Nephrol. 2023;19(6):401–17.
7. Gaudry S, Hajage D, Schortgen F, Martin-Lefevre L, Pons B, Boulet E, Boyer A, Chevrel G, Lerolle N, Carpentier D. Initiation strategies for renal-replacement therapy in the intensive care unit. N Engl J Med. 2016;375(2):122–33.
8. Zarbock A, Kellum JA, Schmidt C, Van Aken H, Wempe C, Pavenstädt H, Boanta A, Gerß J, Meersch M. Effect of early vs delayed initiation of renal replacement therapy on mortality in critically ill patients with acute kidney injury: the ELAIN randomized clinical trial. JAMA. 2016;315(20):2190–9.
9. Barbar SD, Clere-Jehl R, Bourredjem A, Hernu R, Montini F, Bruyère R, Lebert C, Bohé J, Badie J, Eraldi J-P. Timing of renal-replacement therapy in patients with acute kidney injury and sepsis. N Engl J Med. 2018;379(15):1431–42.
10. Investigators S-A, Group CCCT, Australian, Group NZICSCT, Group UKCCR. Timing of initiation of renal-replacement therapy in acute kidney injury. N Engl J Med. 2020;383(3):240–51.
11. Gaudry S, Hajage D, Martin-Lefevre L, Lebbah S, Louis G, Moschietto S, Titeca-Beauport D, La Combe B, Pons B, de Prost N. Comparison of two delayed strategies for renal replacement therapy initiation for severe acute kidney injury (AKIKI 2): a multicentre, open-label, randomised, controlled trial. Lancet. 2021;397(10281):1293–300.
12. Sodhi K, Phillips A, Mishra RC, Tyagi N, Dixit SB, Chaudhary D, Singla MK, Kowdle PC, Kapoor PM. Renal replacement therapy practices in India: a nationwide survey. Indian J Crit Care Med. 2020;24(9):823.
13. An JN, Kim SG, Song YR. When and why to start continuous renal replacement therapy in critically ill patients with acute kidney injury. Kidney Res Clin Pract. 2021;40(4):566.
14. Park SW, Yeo NY, Kang S, Ha T, Kim T-H, Lee D, Kim D, Choi S, Kim M, Lee D. Early prediction of mortality for septic patients visiting emergency room based on explainable machine learning: a real-world multicenter study. J Korean Med Sci. 2024;39(5).
15. Heo JH, Kim T, Shin TG, Suh GJ, Kwon WY, Kim H, Park H, Kim H, Han S. Using machine learning techniques for early prediction of tracheal intubation in patients with septic shock: a multi-center study in South Korea. Acute Crit Care. 2025;40(2):221.
16. Wang Y, Gao Z, Zhang Y, Lu Z, Sun F. Early sepsis mortality prediction model based on interpretable machine learning approach: development and validation study. Intern Emerg Med. 2025;20(3):909–18.
17. Marques F, Fonseca JA, Gameiro J, Gouveia J, Lopes JA. A Risk Prediction Score for Renal Replacement Therapy in Critically Ill Septic-Acute Kidney Injury Patients. Portuguese J Nephrol Hypertens. 2022;36(4):215–22.
18. Song B, Liu P, Liu C, Fu K, Zheng X, Liu Y. Development and validation of a predictive model for continuous renal replacement therapy in sepsis patients using the MIMIC-IV database. Sci Rep. 2025;15(1):21559.
19. Shin TG, Hwang SY, Kang GH, Kim WY, Ryoo SM, Kim K, Jo YH, Chung SP, Joo YS, Beom JH. Korean Shock Society septic shock registry: a preliminary report. Clin Exp Emerg Med. 2017;4(3):146.
20. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. Crit Care Med. 2013;41(2):580–637.
21. Mouncey PR, Osborn TM, Power GS, Harrison DA, Sadique MZ, Grieve RD, Jahan R, Harvey SE, Bell D, Bion JF. Trial of early, goal-directed resuscitation for septic shock. N Engl J Med. 2015;372(14):1301–11.
22. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, Mcintyre L, Ostermann M, Prescott HC. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Crit Care Med. 2021;49(11):e1063–143.
23. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, Van Smeden M. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385.
24. Zhao X, Wei S, Pan Y, Qu K, Yan G, Wang X, Song Y. Early prognosis prediction for non-variceal upper gastrointestinal bleeding in the intensive care unit: based on interpretable machine learning. Eur J Med Res. 2024;29(1):442.
25. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
26. Mammone A, Turchi M, Cristianini N. Support vector machines. Wiley Interdisciplinary Reviews: Comput Stat. 2009;1(3):283–9.
27. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining: 2016; 2016: 785–794.
28. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst. 2018;31.
29. Hassan A, Hong C, Sah SB, Elsayegh S. SOFA vs APACHE II: Determining the need for renal replacement therapy. Chest. 2024;166(4):A6378.
30. Suzuki G, Nishioka S, Kobori T, Masuyama Y, Yamamoto S, Serizawa H, Nakamichi Y, Honda M. SACrA score to predict the initiation of renal replacement therapy in critically ill patients: a single-center retrospective study. Ren Fail. 2024;46(2):2404237.
31. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE. 2015;10(3):e0118432.
32. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. BMC Med Inf Decis Mak. 2020;20(1):251.
33. Wu H, Liao B, Ji T, Ma K, Luo Y, Zhang S. Comparison between traditional logistic regression and machine learning for predicting mortality in adult sepsis patients. Front Med. 2025;11:1496869.
34. Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Stat Sci. 2001;16(3):199–231.
35. Mou C, Yang J, Wu Q, Qin L, Lu J. Progress in sepsis prediction models: from traditional scoring systems to multimodal intelligence and clinical translation. Front Med. 2026;13:1732164.
36. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.
37. Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, Ratliff W, Balu S. A path for translation of machine learning products into healthcare delivery. EMJ Innov. 2020;10:19–00172.
38. Knox DB, Lanspa MJ, Kuttler KG, Brewer SC, Brown SM. Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. Intensive Care Med. 2015;41(5):814–22.
39. Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA. 2019;321(20):2003–17.
40. Theodoulou D, Alevrakis E, Iliopoulou M, Karampitsakos T, Katsaras M, Tsipilis S, Tzimopoulos K, Koutsoukou A, Rovina N. Invasive mechanical ventilation: when and to whom? indications and complications of invasive mechanical ventilation. Pneumon 2020;33(1).
41. Zarbock A, Gomez H, Kellum JA. Sepsis-induced acute kidney injury revisited: pathophysiology, prevention and future therapies. Curr Opin Crit Care. 2014;20(6):588–95.
42. Kellum J. Metabolic acidosis in patients with sepsis: epiphenomenon or part of the pathophysiology? Crit Care Resusc. 2004;6(3):197–203.

43. Mele A, Cerminara E, Häbel H, Rodriguez-Galvez B, Oldner A, Nelson D, Gårdh J, Thobaben R, Jonmarker S, Cronhjort M. Fluid accumulation and major adverse kidney events in sepsis: a multicenter observational study. Ann Intensive Care. 2022;12(1):62.

Publisher's note