JKMS

Original Article
Humanities & Basic Medical Science

Check for updates

OPEN ACCESS

Address for Correspondence:
Jeongwon Heo, MD, PhD
Department of Internal Medicine, School of Medicine, Kangwon National University, 1 Kangwondaehak-gil, Chuncheon 24341, Republic of Korea.
Email: doctorhjw@naver.com

*Sang Won Park and Na Young Yeo contributed equally to this work.

ORCID iDs
Sang Won Park
https://orcid.org/0000-0001-9796-2799
Na Young Yeo
https://orcid.org/0000-0002-3934-8157
Seonguk Kang
https://orcid.org/0009-0007-2333-9794
Taejun Ha
https://orcid.org/0000-0002-2812-6014
Tae-Hoon Kim
https://orcid.org/0009-0003-6855-8107
DooHee Lee
https://orcid.org/0000-0002-5838-6925
Dowon Kim
https://orcid.org/0009-0002-9598-6214

Early Prediction of Mortality for Septic Patients Visiting Emergency Room Based on Explainable Machine Learning: A Real-World Multicenter Study

Sang Won Park [1,2*] Na Young Yeo [3*] Seonguk Kang [4] Taejun Ha [5] Tae-Hoon Kim [6] DooHee Lee [7] Dowon Kim [7] Seheon Choi [7] Minkyu Kim [7] DongHoon Lee [7] DoHyeon Kim [7] Woo Jin Kim [1,8,9] Seung-Joon Lee [8,9] Yeon-Jeong Heo [8,9] Da Hye Moon [8,9] Seon-Sook Han [8,9] Yoon Kim [6,10] Hyun-Soo Choi [6,11] Dong Kyu Oh [12] Su Yeon Lee [12] MiHyeon Park [12] Chae-Man Lim [12] Jeongwon Heo [8,9] and on behalf of the Korean Sepsis Alliance (KSA) Investigators

[1]Department of Medical Informatics, School of Medicine, Kangwon National University, Chuncheon, Korea
[2]Institute of Medical Science, School of Medicine, Kangwon National University, Chuncheon, Korea
[3]Department of Medical Bigdata Convergence, Kangwon National University, Chuncheon, Korea
[4]Department of Convergence Security, Kangwon National University, Chuncheon, Korea
[5]Department of Biomedical Research Institute, Kangwon National University Hospital, Chuncheon, Korea
[6]University-Industry Cooperation Foundation, Kangwon National University, Chuncheon, Korea
[7]Department of Research and Development, ZIOVISION Co. Ltd., Chuncheon, Korea
[8]Department of Internal Medicine, Kangwon National University Hospital, Chuncheon, Korea
[9]Department of Internal Medicine, School of Medicine, Kangwon National University, Chuncheon, Korea
[10]Department of Computer Science and Engineering, Kangwon National University, Chuncheon, Korea
[11]Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul, Korea
[12]Department of Pulmonary and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

ABSTRACT

Background: Worldwide, sepsis is the leading cause of death in hospitals. If mortality rates in patients with sepsis can be predicted early, medical resources can be allocated efficiently. We constructed machine learning (ML) models to predict the mortality of patients with sepsis in a hospital emergency department.

Methods: This study prospectively collected nationwide data from an ongoing multicenter cohort of patients with sepsis identified in the emergency department. Patients were enrolled from 19 hospitals between September 2019 and December 2020. For acquired data from 3,657 survivors and 1,455 deaths, six ML models (logistic regression, support vector machine, random forest, extreme gradient boosting [XGBoost], light gradient boosting machine, and categorical boosting [CatBoost]) were constructed using fivefold cross-validation to predict mortality. Through these models, 44 clinical variables measured on the day of admission were compared with six sequential organ failure assessment (SOFA) components (PaO$_2$/FIO$_2$ [PF], platelets (PLT), bilirubin, cardiovascular, Glasgow Coma Scale score, and creatinine). The confidence interval (CI) was obtained by performing 10,000 repeated measurements via random sampling of the test dataset. All results were explained and interpreted using Shapley's additive explanations (SHAP).

Results: Of the 5,112 participants, CatBoost exhibited the highest area under the curve (AUC) of 0.800 (95% CI, 0.756–0.840) using clinical variables. Using the SOFA components for

JKMS

Seheon Choi (iD)
https://orcid.org/0009-0000-0506-6971
Minkyu Kim (iD)
https://orcid.org/0000-0003-1122-5363
DongHoon Lee (iD)
https://orcid.org/0000-0003-2101-9627
DoHyeon Kim (iD)
https://orcid.org/0009-0004-2626-0510
Woo Jin Kim (iD)
https://orcid.org/0000-0003-2927-370X
Seung-Joon Lee (iD)
https://orcid.org/0000-0003-4597-0281
Yeon-Jeong Heo (iD)
https://orcid.org/0000-0002-8151-7970
Da Hye Moon (iD)
https://orcid.org/0000-0001-5602-9572
Seon-Sook Han (iD)
https://orcid.org/0000-0002-1320-4163
Yoon Kim (iD)
https://orcid.org/0000-0001-5709-5766
Hyun-Soo Choi (iD)
https://orcid.org/0000-0002-3594-8948
Dong Kyu Oh (iD)
https://orcid.org/0000-0002-7511-9634
Su Yeon Lee (iD)
https://orcid.org/0000-0003-1993-7752
MiHyeon Park (iD)
https://orcid.org/0000-0003-0249-0050
Chae-Man Lim (iD)
https://orcid.org/0000-0001-5400-6588
Jeongwon Heo (iD)
https://orcid.org/0000-0001-7694-6730

the same patient, XGBoost exhibited the highest AUC of 0.678 (95% CI, 0.626–0.730). As interpreted by SHAP, albumin, lactate, blood urea nitrogen, and international normalization ratio were determined to significantly affect the results. Additionally, PF and PLTs in the SOFA component significantly influenced the prediction results.

Conclusion: Newly established ML-based models achieved good prediction of mortality in patients with sepsis. Using several clinical variables acquired at the baseline can provide more accurate results for early predictions than using SOFA components. Additionally, the impact of each variable was identified.

Keywords: Clinical Decision Support System (CDSS); Explainable Artificial Intelligence (XAI); Machine Learning; Mortality Prediction, Sepsis

INTRODUCTION

Sepsis is a severe and potentially life-threatening condition affecting millions of people worldwide every year.[1,2] It is a complex syndrome that arises from the body's response to an infection and can lead to organ dysfunction, septic shock, and death if not treated promptly and adequately.[3,4] If we can accurately predict the prognosis of patients with sepsis, we can distribute the limited medical resources more efficiently and improve sepsis prognosis.[5] In particular, early prediction of the prognosis of patients presented to an emergency room (ER) with sepsis is crucial for making informed decisions about their management and optimizing their chances of survival.[6]

Currently, sequential organ failure assessment (SOFA) component scores are used in clinical risk stratification of patients with sepsis; however, they have limitations of poor specificity and sensitivity.[7-11] The SOFA score is one of the main criteria for sepsis-3 definitions. Several previous studies have suggested predictive results for patients with sepsis using the SOFA score; however, they have achieved varying performances in predicting short-term mortality.[12-14] In other words, the generalizability of the prediction models is limited. The SOFA component score comprises six organ scores, each set in the range 0–4 points. It was calculated by simply adding the scores of the six organ systems and did not consider the interaction with various organ systems.[11] Thus, it may have different effects on the mortality risk of patients with sepsis and reduce the predictive power of the model.

Recently, machine learning (ML) has attracted the attention of and gained recognition from clinicians owing to the evolution of computer technology.[15] ML studies have emerged as a promising approach for predicting sepsis mortality, which can help clinicians identify high-risk patients and provide timely and appropriate treatment.[16-18] In addition, the application of new ML techniques and model implementations can be widely used in predictive models of various diseases and can exhibit better performance than conventional logistic regression (LR) or Cox regression analysis.[19,20] Recent studies have shown that ML models have good predictive ability for identifying patients with sepsis who are at a high risk of mortality.[21-23] ML models can analyze large amounts of patient data, such as vital signs, laboratory results, and medical histories, and can identify patterns who may not be apparent to human clinicians.[24,25] Furthermore, the accuracy of mortality predictions can be significantly improved using large-scale data. These models can provide personalized risk predictions for individual patients, thus helping clinicians make informed decisions regarding treatment and management. They can also provide insights into the factors contributing to a patient's

Kang S, Yeo NY, Heo J. Methodology: Park SW, Yeo NY. Software: Park SW, Yeo NY. Validation: Kim Y, Choi HS. Visualization: Park SW, Yeo NY. Writing - original draft: Park SW, Kang S, Yeo NY. Writing - review & editing: Park SW, Kang S, Yeo NY, Han SS, Heo J, Kim Y, Choi HS.

Kim D,[1] Dowon Kim; Kim D,[2] DoHyun Kim.

risk of death, which can help design a personalized treatment plan that can lead to improved patient outcomes and more efficient allocation of healthcare resources. However, there is a lack of evidence demonstrating the advantage of ML algorithms to predict mortality in sepsis patients by leveraging large amounts of clinical data acquired from early-patient conditions and comparing them to SOFA component scores to demonstrate their superiority.

This study aims to develop an ML model to predict the mortality rate of patients with sepsis admitted to an ER using baseline data. In addition, a new ML framework can be constructed using various clinical variables in a large dataset to achieve better performance than the SOFA score. We use explainable artificial intelligence (XAI) to confirm the effects of individual factors on the mortality prediction performance in patients with sepsis.

METHODS

Data source and participants

This prospective cohort study used data from an ongoing nationwide cohort study conducted by the Korean Sepsis Alliance. The patients were enrolled from 19 participating hospitals between September 2019 and December 2020. The protocols for patient enrollment and data collection have been described previously.[26] Patients were included if they were at least 19 years old and diagnosed with sepsis or septic shock in the emergency department. The diagnoses of sepsis and septic shock were based on the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).[4] We investigated patients who were diagnosed with sepsis visiting an ER, regardless of whether they were admitted to a ward or intensive care unit (ICU). Patients were excluded if they were not prescribed antibiotics, prescribed antibiotics that were not in accordance with the guidelines, prescribed antibiotics to which the cultured organism proved to be resistant.[27] The average time that all patients investigated in this study stayed in ER was 12 hours. Empiric therapy was defined as the initial use of antimicrobial treatment without specific identification of the causative pathogen. The empiric therapy appropriateness is evaluated based on the results of drug susceptibility tests or relevant guideline recommendations.[28] Crystalloid fluid was administered for bolus loading. The volume at the consideration of the treating clinician but at least 30 mL/kg.[29]

Data source

We used multidimensional data acquired from 5,112 patients with sepsis (demographic characteristics, clinical information, pre-existing comorbidities, initial laboratory findings, and initial characteristics of infection) as variables (**Table 1**). A total of 44 variables were selected for ML modeling.[30] The final 44 clinical variables were selected by a respiratory specialist after selecting only variables with an NA value of less than 10% from a total of 213 variables in the cohort (**Fig. 1**). For the SOFA component score, six organ system-categorized variables (respiratory, cardiovascular, liver, renal, coagulation, and neurological) were selected and used (**Table 2**, **Supplementary Table 1**).

Data pre-processing

Raw-data prior information can directly affect the optimized classifier performance. In this regard, data pre-processing is important to obtain efficient mortality classification performance using ML algorithms.[31] Data normalization is essential for scaling and transforming data to minimize bias. We removed the NA values that existed in the dataset used in this study and the values that were incorrectly written under the decision of an expert

Table 1. Subject characteristics for clinical variables

Variables	Alive (n = 3,657)	Death (n = 1,455)	P value
Demographic characteristics			
Age	70.2 ± 13.3	71.7 ± 13.0	< 0.001
Sex			< 0.05
Male	2,102 (57.5)	889 (61.1)	
Female	1,555 (42.5)	566 (38.9)	
Height	162.0 ± 9.1	162.3 ± 8.8	-
Weight	57.7 ± 12.9	57.0 ± 12.6	< 0.05
Residence a nursing home	272 (7.4)	110 (7.6)	-
Septic shock	655 (17.9)	450 (30.9)	< 0.001
Clinical information			
MBP (Initial vital sign)	73.3 ± 21.7	69.9 ± 20.9	< 0.001
HR (Initial vital sign)	106.1 ± 24.7	109.8 ± 25.3	< 0.001
RR (Initial vital sign)	23.8 ± 5.2	25.3 ± 6.3	< 0.001
BT (Initial vital sign)	37.5 ± 1.3	37.1 ± 1.2	< 0.001
Clinical Frailty Scale			< 0.001
Very fit	192 (5.3)	43 (3.0)	
Well	403 (11.0)	70 (4.8)	
Managing well	544 (14.9)	156 (10.7)	
Vulnerable	588 (16.1)	229 (15.7)	
Mildly frail	354 (9.7)	154 (10.6)	
Moderately frail	371 (10.1)	165 (11.3)	
Severely frail	659 (18.0)	305 (21.0)	
Very severely frail	535 (14.6)	311 (21.4)	
Terminally ill	11 (0.3)	22 (1.5)	
Pre-existing comorbidities			
Total CCI	5.6 ± 2.6	6.4 ± 2.7	< 0.001
Initial characteristics of infection			
Site of infection			< 0.001
Pulmonary	1,366 (37.4)	662 (45.5)	
Abdominal	1,006 (27.5)	331 (22.7)	
Urinary	608 (16.6)	126 (8.7)	
Skin/soft tissue	123 (3.4)	36 (2.5)	
Catheter-related	18 (0.5)	8 (0.5)	
Systemic infections without clear primary site of infection	326 (8.9)	187 (12.9)	
Neurologic	21 (0.6)	9 (0.6)	
Multiple infection sites	189 (5.2)	96 (6.6)	
Type of infection			< 0.05
Community-acquired	2,203 (60.2)	803 (55.2)	
Nursing home	199 (5.4)	87 (6.0)	
Nursing hospital	491 (13.4)	210 (14.4)	
Hospital	764 (20.9)	355 (24.4)	
Initial laboratory findings			
Lactate	3.2 ± 2.7	5.1 ± 4.1	< 0.001
WBC	12.8 ± 8.8	12.8 ± 10.5	-
ANC	0.8 ± 0.2	0.8 ± 0.2	< 0.001
Hb	10.9 ± 2.5	10.1 ± 2.5	< 0.001
Platelets	187.0 ± 112.7	165.2 ± 126.2	< 0.001
Na	135.9 ± 7.7	135.8 ± 8.9	-
K	4.2 ± 0.9	4.4 ± 1.0	< 0.001
Cl	101.4 ± 8.3	101.1 ± 9.0	-
BUN	35.1 ± 26.9	43.1 ± 29.6	< 0.001
Cr	1.8 ± 1.7	2.1 ± 1.7	< 0.001
Bilirubin	1.6 ± 2.8	2.5 ± 4.7	< 0.001
Albumin	3.0 ± 0.7	2.6 ± 0.6	< 0.001
PT INR	1.4 ± 0.8	1.6 ± 1.5	< 0.001
CRP	25.4 ± 47.2	32.8 ± 57.1	< 0.001
Glucose	168.2 ± 105.5	166.3 ± 103.2	-
Arterial PH	7.4 ± 0.1	7.4 ± 0.1	< 0.001
PaCO$_2$	33.0 ± 12.1	33.9 ± 15.3	-

Table 1. (Continued) Subject characteristics for clinical variables

Variables	Alive (n = 3,657)	Death (n = 1,455)	P value
PaO₂	89.9 ± 45.5	87.4 ± 47.1	-
Bicarbonate	20.8 ± 6.0	18.8 ± 7.0	< 0.001
Surviving sepsis campaign bundles within 1 hr			
Measure lactate level	3,647 (99.7)	1,450 (99.7)	-
Blood culture	3,632 (99.3)	1,440 (99.0)	-
Use of broad-spectrum antibiotics	3,624 (99.1)	1,436 (98.7)	< 0.05
Bolus injection for hypotension	2,776 (75.9)	1,159 (79.7)	< 0.05
Administration of vasopressors	1,894 (51.8)	972 (66.8)	< 0.001
Re-measure lactate level more than 2 mmol/L	1,750 (47.9)	967 (66.5)	< 0.001
Adjunctive corticosteroid treatment			
Treatment using steroid	552 (15.1)	380 (26.1)	< 0.001
Source control			
Infection source control			< 0.001
Non-surgical	3,157 (86.3)	1,312 (90.2)	
Surgical	500 (13.7)	143 (9.8)	
Surgical infection source control	131 (3.6)	40 (2.7)	-
Antibiotics treatment			
Administration of antibiotics before sepsis diagnosis	836 (22.9)	406 (27.9)	< 0.001
Appropriateness of initial empirical therapy			< 0.05
Appropriate	3,260 (89.1)	1,250 (85.9)	
Inappropriate	372 (10.2)	192 (13.2)	

Continuous variables were presented as mean ± SD, and categorical variables were presented as number and percentage. Site of infection refers to suspected or confirmed infection as defined by the MOSAICS II study. Additionally, systemic infections without a clear primary site of infection cannot be selected simultaneously with others. Surviving sepsis campaign bundles indicate whether it will be implemented or not.

MBP = mean blood pressure, HR = heart rate, RR = respiratory rate, BT = body temperature, CCI = Charlson Comorbidity Index, MOSAICS = management of sepsis in Asia's intensive care units II, WBC = white blood cell count, ANC = absolute neutrophil count, Hb = hemoglobin, Na = natrium, K = kalium, Cl = chlorine, BUN = blood urea nitrogen, Cr = creatinine, PT INR = prothrombin time-international normalized ratio, CRP = C-reactive protein, PH = percentage of hydrogen, PaCO₂ = pressure of arterial carbon dioxide, PaO₂ = pressure of arterial oxygen, NA = not applicable.
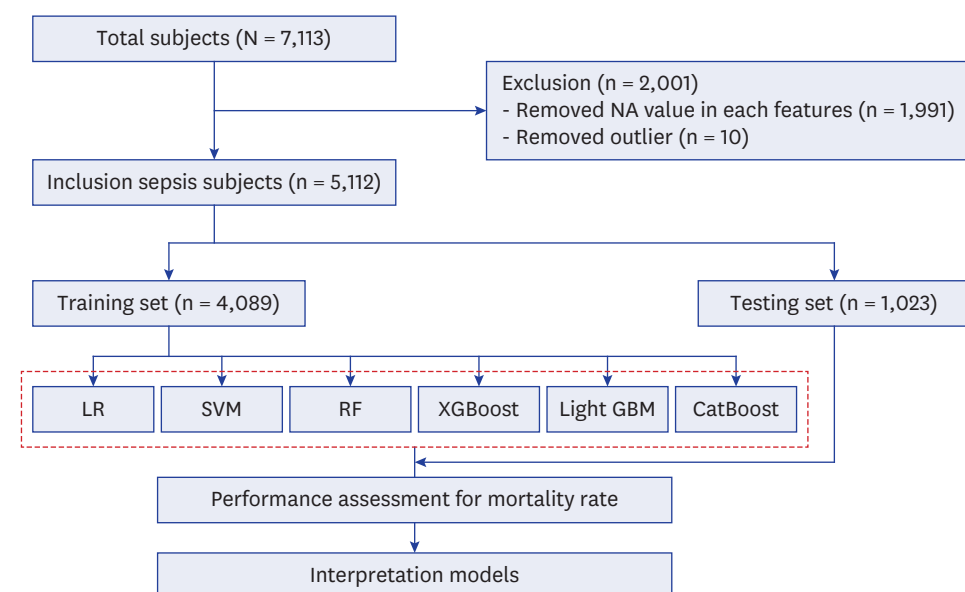
Fig. 1. Flowchart of patient selection and machine-learning-model application for mortality prediction. Of the 7,113 patients diagnosed with sepsis, 5,112 patients were selected after excluding patients with missing values and outliers. Of all selected patients, 80% were used for model training and 20% for model testing.
NA = not available, LR = logistic regression, SVM = support vector machine, RF = random forest, XGBoost = extreme gradient boosting, Light GBM = light gradient boosting machine, CatBoost = categorical boosting.

Table 2. Subject characteristics for SOFA score

Variables	Alive (n = 3,657)	Death (n = 1,455)	*P* value
SOFA_PF			< 0.001
0	**764 (20.9)**	**239 (16.4)**	
1	1,040 (28.4)	251 (17.3)	
2	1,491 (40.8)	720 (49.5)	
3	261 (7.1)	128 (8.8)	
4	101 (2.8)	117 (8.0)	
SOFA_PLT			< 0.001
0	**2,106 (57.6)**	**693 (47.6)**	
1	685 (18.7)	220 (15.1)	
2	552 (15.1)	279 (19.2)	
3	215 (5.9)	162 (11.1)	
4	99 (2.7)	101 (6.9)	
SOFA_BIL			< 0.001
0	**2,398 (65.6)**	**828 (56.9)**	
1	590 (16.1)	233 (16.0)	
2	524 (14.3)	260 (17.9)	
3	94 (2.6)	70 (4.8)	
4	51 (1.4)	64 (4.4)	
SOFA_Cardio			< 0.001
0	**1,346 (36.8)**	**498 (34.2)**	
1	1,670 (45.7)	592 (40.7)	
2	56 (1.5)	21 (1.4)	
3	417 (11.4)	174 (12.0)	
4	168 (4.6)	170 (11.7)	
SOFA_GCS			< 0.001
0	**1,701 (46.5)**	**513 (35.3)**	
1	705 (19.3)	291 (20.0)	
2	588 (16.1)	248 (17.0)	
3	494 (13.5)	233 (16.0)	
4	169 (4.6)	170 (11.7)	
SOFA_Renal			< 0.001
0	**1,699 (46.5)**	**540 (37.1)**	
1	956 (26.1)	394 (27.1)	
2	614 (16.8)	301 (20.7)	
3	561 (15.3)	124 (8.5)	
4	201 (5.5)	96 (6.6)	

Variables were presented as number and percentage.
SOFA = sequential organ failure assessment, SOFA_PF = SOFA score for pressure of oxygen/fraction of inspired oxygen (PaO$_2$/FIO$_2$), SOFA_PLT = SOFA score for platelets, SOFA_BIL = SOFA score for bilirubin, SOFA_Cardio = SOFA score for cardiovascular, SOFA_GCS = SOFA score for Glasgow Coma Scale score, SOFA_Renal = SOFA score for renal.
*Bold values indicate that the function of each organ is normal.

clinician and judged to be outliers. In addition, the skewness method was applied to biased data variables to determine the outliers that exist within quantitative values and transform them into a normal distribution. Finally, max–min normalization pre-processing was performed.

Data split and validation

Of the total data, 80% were used for training, and the remaining data were used as test data. Validation data were used for 20% of the total training set. Of the selected participant data (i.e., all participants), 20% were removed entirely from the cross-validation-based estimation of hyperparameter values for each of the four classification methods. In addition, stratified k-fold cross-validation (k = 5) was performed to avoid label distortions that might occur during model generation and maintain model stability (**Fig. 2**). The stratified k-fold cross-validation technique was similar to the regular k-fold cross-validation, except that stratified sampling was used instead of random sampling. To demonstrate the reliability
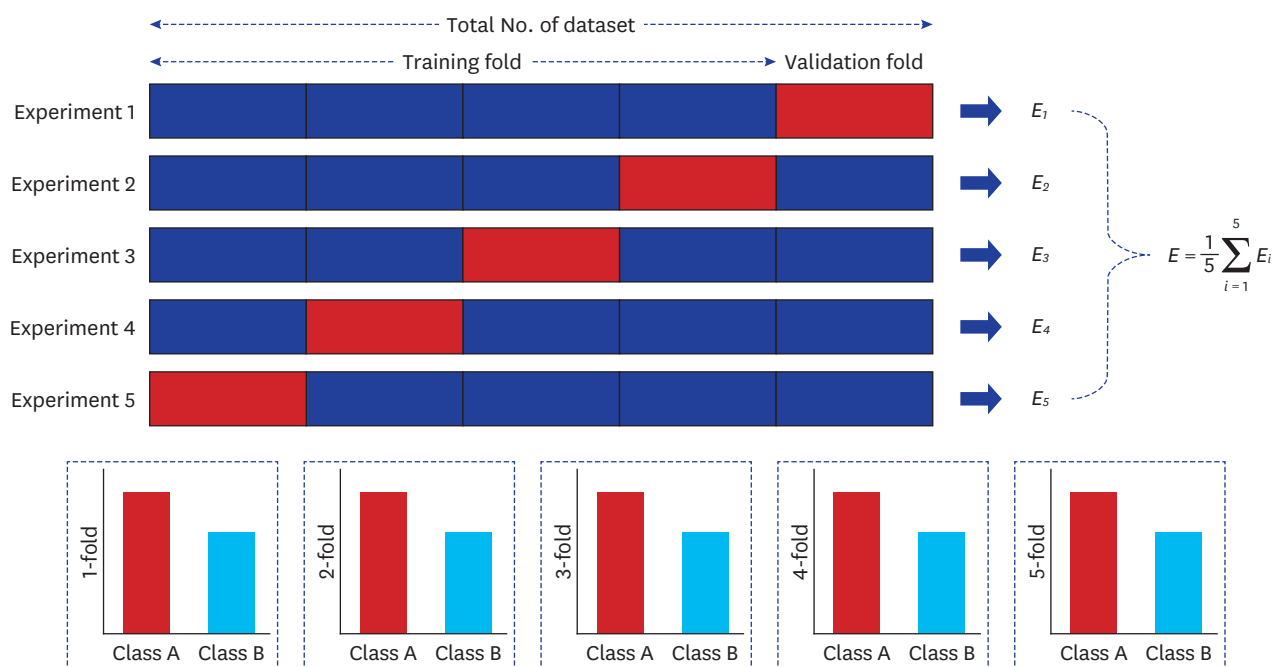
Fig. 2. Stratified k-fold cross-validation. Each result is averaged to obtain the validated performance of the model.

of the individual model results, half of the entire test dataset was randomly sampled, and repeated tests were performed 10,000 times to present the confidence intervals (CIs) for the test results.

Statistical analysis

Continuous variables are presented as mean with standard deviation. The Student's t-tests were used to confirm the difference in mean ratio between alive and death groups. Categorial variables are expressed as number and percentage, and were compared using the χ^2 tests.

ML

LR

As a form of supervised learning in ML, LR is a conventional probabilistic statistical model for classification that has been broadly used across disciplines in medical sciences.[32] The LR models the relation between a continuous independent and categorical dependent variables. It can predict and classify a sample to a group as a probability value between 0 and 1, which learns the relationship between the independent variables $x1$, $x2$, ⋯, xn, and the dependent variable y as a specific function. In other words: $y = (w1x1 + ⋯ + wnxn)$, where $w1$, ⋯, wn are trainable parameters and σ denotes the sigmoid function, such that $σ(t) = 1/(1 + e{-}t)$. In linear regression, the predicted dependent variable falls within the range [-∞, ∞]. The LR to classify binary tasks becomes possible by application of the sigmoid function, which always returns a probability in the range [0, 1].

Support vector machine (SVM)

The linear discriminant function of the SVM is an algorithm that allows classification by defining a decision hyperplane in two or more dimensions.[33] The SVM is a commonly used algorithm in ML for classification tasks and it uses the optimal hyperplane, which maximizes the gap between two groups. The points at which each datum is distributed and the distance

between the classes of parallel hyperplanes passing the support vector through the optimal separating hyperplane are used for classification using margins, such as hard or soft margins, which can be determined by maximizing. The hyperplane is not unique and can be estimated by maximizing the classifier's performance, that is, the classifier's ability to operate satisfactorily with any data.

Random forest (RF)
RF is an ensemble learning method that creates a multitude of decision trees at training time and outputs the class, which is the mode of the classes (classification) or mean prediction (regression) of the individual trees.[34] Each tree in the RF is trained on a randomly selected subset of the data and features, which helps to reduce overfitting and improve the generalizability of the model. During the prediction, each tree in the RF produces a class prediction, and the class with the most votes across all trees is selected as the final prediction.

Gradient boosting
Gradient boosting is an ensemble method that has the advantage of constructing a strong classification by combining several common weak classifiers. In this study, extreme gradient boosting (XGBoost), light gradient boosting machine (Light GBM), and categorical boosting (CatBoost) were widely used for analyzing structural tabular data.[35-37]

XGBoost is a tree-based ensemble-learning algorithm that combines multiple decision trees and uses classification and regression based on gradient descent. It expands the decision trees horizontally (i.e., levelwise) to reduce their depth. It works well on imbalanced datasets and has excellent accuracy and speed. Light GBM is an algorithm similar to XGBoost, but can learn faster for large datasets. It uses a leaf-centered histogram-based segmentation algorithm and applies a leafwise division method to make decisions. The CatBoost algorithm specializes in handling categorical variables. It applies its own categorical data processing method, and can process both numerical and categorical variables. It uses the symmetric tree method and ordered target encoding technique and has a built-in overfitting prevention function. We reflected on the characteristics of the high-dimensional and large-capacity datasets used in this study and confirmed the adequacy of the models. Because all three methods have relative strengths and weaknesses, we compared their performance in predicting mortality in patients with sepsis.

Interpretation model based on Shapley additive explanations (SHAP)

We explained the classification prediction of our model with SHAP,[38] which is a method to explain models with feature importance and inspired by the concept of coalition game theory by replacing "player" in the theory with "data feature" in tabular data. In the ML context, agents correspond to the features of the data, and the goal is to explain the model's prediction. This was deduced to be a linear function of this feature. Original model f is explained using surrogate model g. Surrogate model g is defined as follows:

$$g(z) = \phi_0 + \sum_{j=1}^{M} \phi_j z_j,$$

where z' denotes the coalition vector, the attribution of feature j, and M denotes the maximum number of features.

Evaluation performance

Five model performance metrics (accuracy, recall, precision, F1 score, and specificity) were used to evaluate the proposed model performance. As this study focused on the accurate classification of survival for patients with sepsis, the true positive metric was established for the overall performance evaluation of the classification model.

Tools

All programming work for ML in this study was performed in Python programming language (version 3.9.16) and statistical analysis was conducted in R (version 4.2.2). All ML modeling was performed using Scikit-learn version 1.2.0 and hyperparameter tuning was performed using Optuna, a hyperparameter optimization framework.[39,40]

Ethics statement

The study was approved by the Institutional Review Board of each participating hospital, and the requirement for informed consent was waived because of the non-interventional observational nature of the study. This study was approved by the Institutional Research Board of Kangwon National University Hospital (KNUH-2022-11-010-001), and written informed consent was obtained from all patients.

RESULTS

Subject characteristics

Tables 1 and **2** summarize the subjects used in this study with descriptive statistics of survivors and non-survivors of sepsis from the cohort. **Table 1** lists the participants' information regarding the selected clinical variables, and **Table 2** lists the SOFA scores. The characteristics of subjects with sepsis were stratified by training and test datasets for selected clinical variables and SOFA scores (**Supplementary Tables 2** and **3**).

Model performance

In selected clinical variables, of all six ML methods, area under the curve (AUC) of the CatBoost model (0.800; 95% CI, 0.756–0.840) exhibited the highest discriminative ability (**Table 3**), followed by the Light GBM model (0.795; 95% CI, 0.750–0.836) and XGBoost model (0.797; 95% CI, 0.754–0.838). The accuracy of XGBoost (0.773; 95% CI, 0.744–0.802) also exhibited the highest accuracy among all six ML methods, followed by the CatBoost model (0.769; 95% CI, 0.742–0.798) and Light GBM model (0.763; 95% CI, 0.726–0.797). The RF model (0.736; 95% CI, 0.688–0.783) had the lowest AUC among all ML models. Among the ML models, the CatBoost model exhibited the highest specificity (0.937; 95% CI, 0.910–0.962), followed by the Light GBM (0.937; 95% CI, 0.910–0.962) and XGBoost (0.929; 95% CI, 0.902–0.954) models. The XGBoost model also exhibited the highest AUPRC (0.626, 95% CI 0.558–0.697), similar to the CatBoost (0.621; 95% CI, 0.551–0.694) and Light GBM (0.617; 95% CI, 0.546–0.689) models. All metrics of the different models for predicting sepsis-related mortality are shown in **Fig. 3** and **Table 3**. Most performance metrics of the Boost model were superior to those of the other three models.

For the SOFA score, of all six ML methods, the AUC of the XGBoost model (0.678; 95% CI, 0.626–0.730) exhibited the highest discriminative ability, similar to that of the CatBoost (0.676; 95% CI, 0.623–0.727), Light GBM (0.667; 95% CI, 0.615–0.720), and RF (0.672; 95% CI, 0.620–0.723) models (**Table 3**). The accuracy of CatBoost model (0.732; 95% CI,

Table 3. Model performance results

Variables	ACC	Precision	Recall	F1 score	Specificity	AUC	AUPRC
Clinical variables							
LR	0.731 (0.693–0.769)	0.522 (0.469–0.576)	0.667 (0.586–0.738)	0.585 (0.528–0.64)	0.757 (0.713–0.801)	0.772 (0.726–0.815)	0.597 (0.528–0.67)
SVM	0.729 (0.691–0.767)	0.518 (0.468–0.571)	0.694 (0.621–0.766)	0.593 (0.539–0.646)	0.743 (0.697–0.787)	0.771 (0.724–0.815)	0.596 (0.527–0.67)
RF	0.701 (0.661–0.74)	0.479 (0.423–0.536)	0.584 (0.503–0.662)	0.526 (0.465–0.585)	0.747 (0.702–0.792)	0.736 (0.688–0.783)	0.511 (0.447–0.594)
XGB	0.773 (0.744–0.802)	0.681 (0.586–0.773)	0.381 (0.303–0.462)	0.489 (0.407–0.565)	0.929 (0.902–0.954)	0.797 (0.754–0.838)	0.626 (0.558–0.697)
LGB	0.763 (0.726–0.797)	0.595 (0.521–0.667)	0.516 (0.435–0.6)	0.553 (0.481–0.619)	0.861 (0.825–0.896)	0.795 (0.75–0.836)	0.617 (0.546–0.689)
CatBoost	0.769 (0.742–0.798)	0.687 (0.588–0.786)	0.347 (0.269–0.428)	0.461 (0.374–0.542)	0.937 (0.91–0.962)	0.800 (0.756–0.84)	0.621 (0.551–0.694)
SOFA score							
LR	0.629 (0.587–0.669)	0.395 (0.348–0.442)	0.577 (0.497–0.655)	0.469 (0.412–0.524)	0.649 (0.601–0.697)	0.664 (0.611–0.716)	0.453 (0.392–0.529)
SVM	0.675 (0.636–0.712)	0.4281 (0.363–0.493)	0.430 (0.352–0.510)	0.429 (0.360–0.495)	0.772 (0.730–0.814)	0.655 (0.600–0.707)	0.455 (0.392–0.529)
RF	0.661 (0.620–0.703)	0.426 (0.372–0.480)	0.550 (0.469–0.628)	0.480 (0.420–0.536)	0.705 (0.659–0.751)	0.672 (0.620–0.723)	0.455 (0.393–0.530)
XGB	0.725 (0.703–0.750)	0.5581 (0.415–0.703)	0.165 (0.110–0.228)	0.255 (0.172–0.335)	0.948 (0.924–0.970)	0.678 (0.626–0.730)	0.462 (0.400–0.537)
LGB	0.664 (0.622–0.703)	0.424 (0.367–0.480)	0.505 (0.421–0.586)	0.461 (0.398–0.522)	0.727 (0.680–0.771)	0.667 (0.615–0.720)	0.461 (0.399–0.538)
CatBoost	0.732 (0.708–0.755)	0.605 (0.455–0.755)	0.168 (0.110–0.228)	0.263 (0.179–0.344)	0.956 (0.934–0.975)	0.676 (0.623–0.727)	0.474 (0.408–0.551)

ACC = accuracy, AUC = area under the curve, AUPRC = area under the precision-recall curve, LR = logistic regression, SVM = support vector machine, RF = random forest, XGB = extreme gradient boosting, LGB = light gradient boosting machine, CAT = category boosting, SOFA = sequential organ failure assessment.
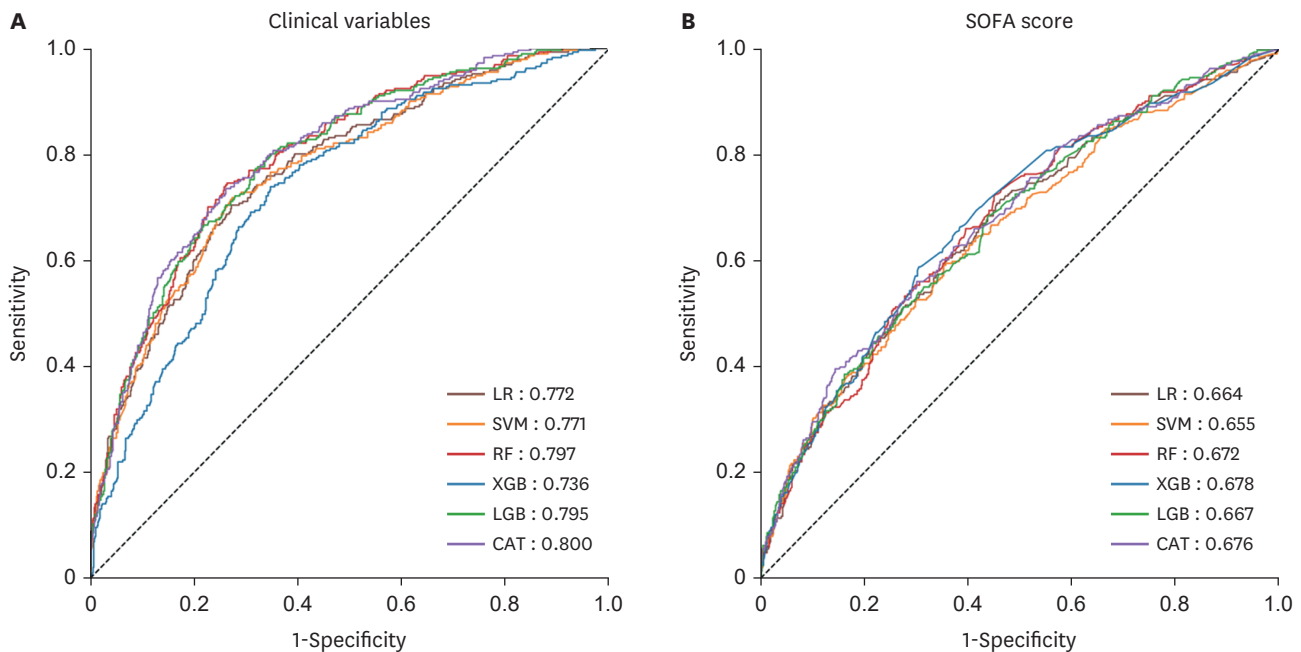
Fig. 3. Receiver operating-characteristic curve of six machine learning models for two datasets. (**A**) Clinical variables. (**B**) SOFA component scores.
LR = logistic regression, SVM = support vector machine, RF = random forest, XGB = extreme gradient boosting, LGB = light gradient boosting machine, CAT = categorical boosting.

0.708–0.755) was the highest among all six models, followed by XGBoost (0.725; 95% CI, 0.703–0.750) and SVM (0.675; 95% CI, 0.636–0.712). LR model (0.629; 95% CI, 0.587–0.669) had the lowest AUC among the ML models.

Among the ML models, CatBoost demonstrated the highest specificity (0.956; 95% CI, 0.934–0.975) followed by XGBoost (0.948; 95% CI, 0.924–0.970). The F1 score of the RF model (0.480; 95% CI, 0.420–0.536) was higher than that of the other models, followed by the Light GBM model (0.461; 95% CI, 0.398–0.522), similar to the results of the LR model (0.469; 95% CI, 0.412–0.524). All metrics of the different models are shown in **Fig. 3** and **Table 3**.

Interpretation of models by SHAP

To interpret the contribution of each variable to each model result, the Shapley value was used to measure the influence of all variables in each model. The SHAP global bar plot explains the global feature importance, where the global importance of each feature is taken as the mean absolute value for that feature over all the given test datasets. These visualizations show the impact on the overall prediction of all models in the order of the relative importance of each variable. **Figs. 4A** and **5A** show the explainable results in the CatBoost and Light GBM models for the clinical variables and SOFA score, respectively. For a more comprehensive understanding of the variable importance in each model, we present all SHAP results in the **Supplementary Figs. 1** and **2**. In the clinical variable dataset, lactate and albumin had greater effects than all other models among the 44 clinical variables, and blood urea nitrogen had a high effect in all models.

The SHAP summary plot explains the information-dense summary of the effect of the top features in a dataset on the model results. For each instance, the explanation is represented by a single dot for each feature flow. **Figs. 4B** and **5B** show the association between feature values and CatBoost and Light GBM in the clinical variable dataset through Shapley values. The SOFA-PaO_2/FiO_2 variable exhibited the highest influence in all models except the SVM model (**Fig. 5**, **Supplementary Figs. 3** and **4**). Platelets have a high influence, which contributes significantly to the performance of all models.

DISCUSSION

In summary, in this study, we performed ML-based mortality-prediction classification using both real-world clinical variables obtained through multicenter and the widely known SOFA score. As a result, we presented the excellence of optimized model performance through clinical information for each patient and presented an explainable ML method to help the clinical decision support system (CDSS). A CDSS can offer a systematic application of health-related knowledge and analysis of available data. It also assists clinical decision to understand the effect of variables for septic patients, and identify opportunities for quality improvement.[41] The effectiveness of several ML algorithms for mortality prediction of sepsis patients based on the 44 clinical-factor combinations utilized in this study was evaluated. Overall, the variables used in this study were easily obtained clinically from patients with sepsis who visited the ER, and the reliability of the model was further demonstrated by the possibility of efficient interpretation and consistency with clinical experience.

Sepsis with high morbidity and mortality is a life-threatening disease requiring prompt diagnosis and treatment to improve patient outcomes. One of the most critical factors in
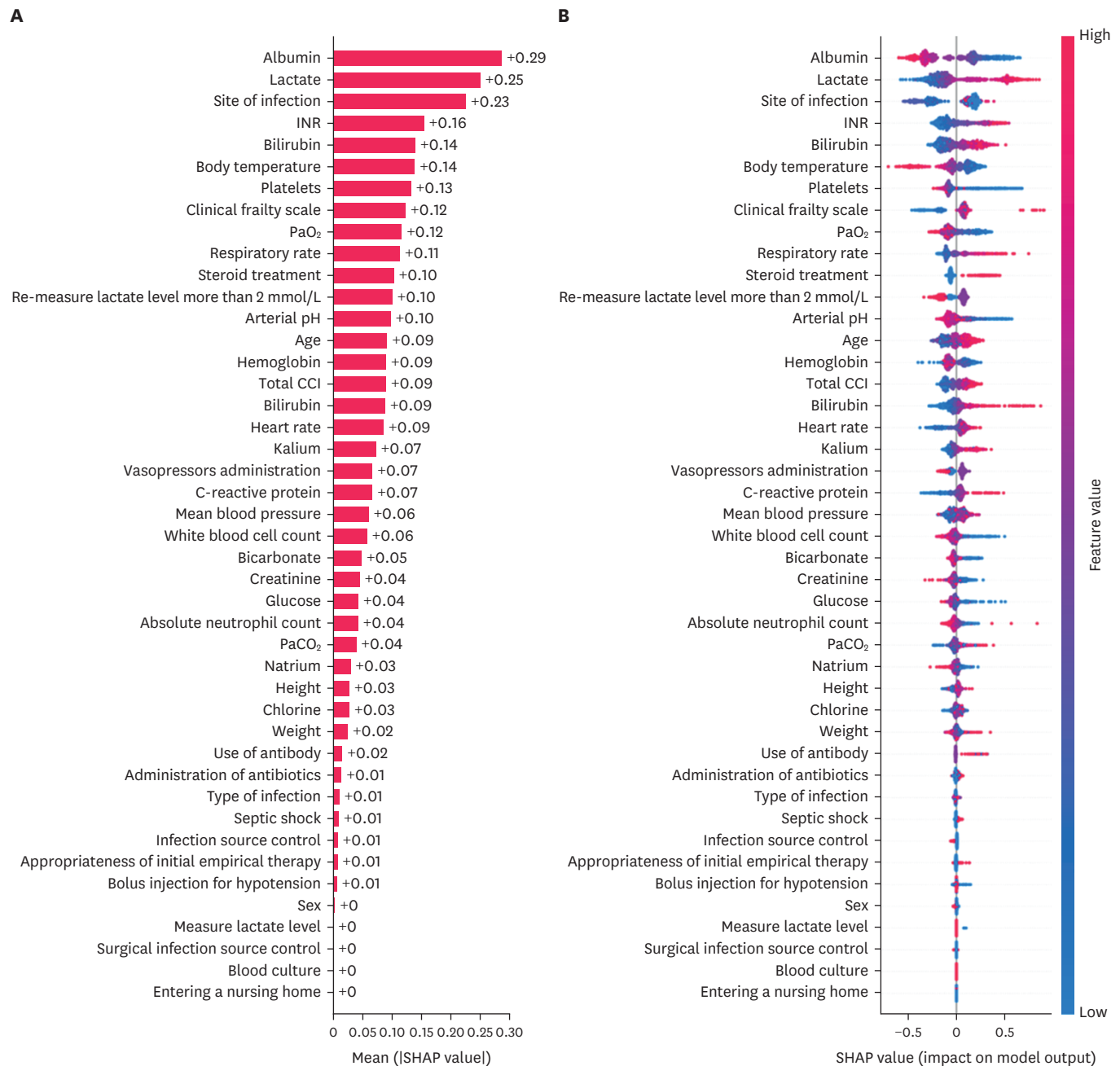
Fig. 4. Feature importance for clinical variables based on interpretation using Shapley's additive explanations of categorical boosting. (**A**) Global importance of each feature as presented mean absolute value. (**B**) Information-dense summary of how the top features in a dataset impact the model output.
PaO$_2$ = pressure of arterial oxygen, PaCO$_2$ = pressure of arterial carbon dioxide, INR = prothrombin time-international normalized ratio.

the management of sepsis is the accurate prediction of patient mortality. The SOFA score is widely known as a tool for assessing organ function and predicting mortality in patients with sepsis. However, recent studies have shown that the accuracy of SOFA score in predicting mortality may be limited.[42-44]

In this study, we suggest that ML based on clinical variables obtained at the baseline in patients with sepsis presented to the emergency department can be used to predict mortality risk in patients with sepsis and has a higher predictive value than SOFA components.
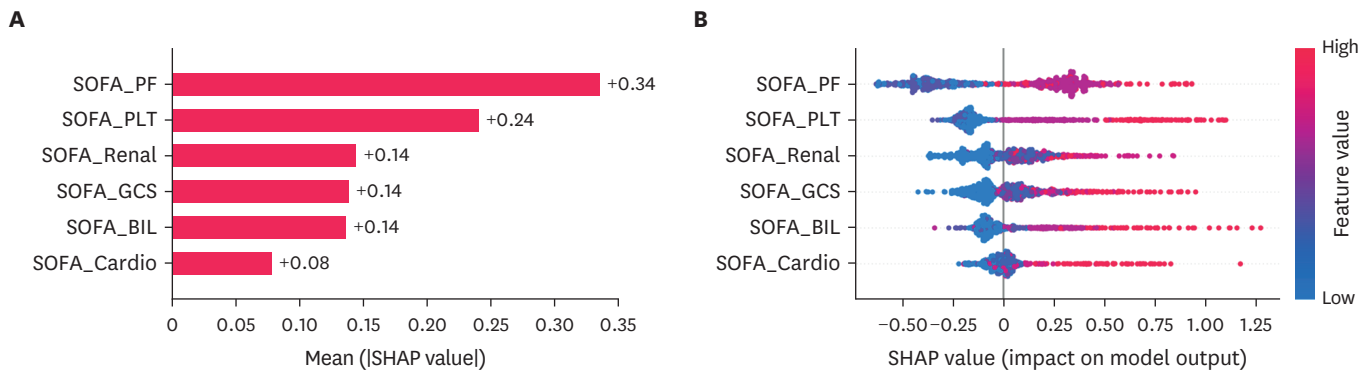
Fig. 5. Feature importance for SOFA component scores based on interpretation using Shapley's additive explanations of the light gradient boosting machine. (**A**) Global importance of each feature presented as the mean absolute value. (**B**) Information-dense summary of how the top features in a dataset impact the model output.

SOFA = sequential organ failure assessment, PF = PaO$_2$/FIO$_2$, PLT = platelets, GCS = Glasgow Coma Scale score, BIL = bilirubin, Cardio = cardiovascular.

Early detection of septic patients at risk of death can help healthcare providers prioritize interventions and effectively allocate resources.[27] Unlike early prediction diagnosis of sepsis, the advantage of early prediction of mortality for septic patients have not been fully explored through research. However, ML-based mortality prediction can calculate and present the risk of death using patient's vital signs, laboratory results, clinical records, etc. Additionally, early mortality prediction may help allocate resources more efficiently and does not mean discontinuing treatment in high-risk patients. Patients identified as having a high risk of mortality may undergo more intensive monitoring and receive more intensive interventions, whereas those assessed to be at lower risk may adhere to standard treatment protocols.[45] Consequently, this risk stratification enables medical doctors to personalize treatment plans for individual patients more effectively. This personalized approach ensures optimal deployment and allocation of resources where they are needed most. Furthermore, healthcare institutions can allocate critical resources like ICU beds, ventilators, and specialized medical staff more efficiently by prioritizing patients at higher risk of mortality.[46] This targeted allocation enhances the likelihood of survival by delivering crucial interventions to those most in need and suggests that constructing predictive models can help a CDSS.[44]

This study, which developed an ML model for predicting sepsis-related mortality in an ER, is a crucial step toward improving sepsis management. The constructed ML models have the potential to analyze vast amounts of data and identify complex patterns that may not be apparent using traditional statistical methods.[47-49] The performances of CatBoost (AUC = 0.800), Light GBM (AUC = 0.795), and XGBoost (AUC = 0.797) were significantly better than those of the other three models for the patient record variables (**Table 3**). The performances of RF (AUC = 0.672), XGBoost (AUC = 0.678), and CatBoost (AUC = 0.676) were significantly better than those of the other three models for the SOFA score components variables (**Table 3**). All models were validated using stratified k-fold (k = 5) cross-validation during training and demonstrated good generalization performance by iterating the model with a random selection of test data and presenting CIs. The implemented prediction was more accurate in classifying mortality risk in patients with sepsis than a previous prediction study whose research objects were also from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database (AUC = 0.61) and other previous studies using SOFA scores, including the results of AUC 0.612–0.752.[12-14,50]

In this context, the development of an ML model using real-world clinical data that can predict sepsis mortality with higher accuracy than the existing SOFA score is a significant achievement. In addition, compared with existing studies, many existing studies have been conducted using MIMIC data to predict the mortality rate of sepsis patients. However, most of them investigated for the patients after admission to the ICU. These studies utilize LR, RF, and SVM, or even though they utilize GBM, they do not differentiate the models in detail. The models have a disadvantage in that the number of variables used is different, and it does not provide superior performance compared to this study even though it is a patient admitted to the ICU. In addition, although they presented the influence of individual variables, it was just done using conventional statistical analysis.[51-53] Thus, the ML models constructed in this study provide more accurate predictions and improve patient outcomes. The development of an ML model in this study aimed to predict the early mortality rate of patients with sepsis who visit an ER, which has significant implications for improving patient outcomes and reducing healthcare costs. The model was found to have higher accuracy in predicting sepsis mortality when using various clinical variables than when using SOFA score based on ML application. It could demonstrate the potential of ML in enhancing sepsis management.

Generally, an artificial intelligence (AI) model, being a black box, cannot have its results, predictions, or decision-making interpreted and explained. Several studies have been conducted to interpret AI models to explain AI behavior. Compared with generally derived feature importance, feature importance produced by SHAP values provide a detailed explanation by considering offsets among all variables for result calculations by carrying out permutation calculations.[54,55] First, we developed a surrogate model, aiming to mirror the original model's behavior and calculate the SHAP values. A surrogate model represents a more straightforward and interpretable version, designed for enhanced ease of explanation. The SHAP values for each feature were determined by assessing how the surrogate model's predictions changed when that feature was altered to its baseline value (usually the average or median of the training dataset). Traditional feature importance might indicate varying levels of significance based on the impact of parameters as the model undergoes successive iterations. Hence, in this research, we utilized the SHAP to quantify the importance of all variables in patient records and characteristics of the SOFA components, which influenced the performance of the ML classifier. Both lactate and albumin are laboratory markers that have been shown to be strongly associated with sepsis and its severity. From the SHAP values in **Fig. 4**, it was confirmed that lactate and albumin were important factors in predicting the mortality rate of patients with sepsis. Lactate, a by-product of anaerobic metabolism, is commonly used as a marker of tissue hypoxia and can result from sepsis-induced hypo-perfusion.[55] Elevated lactate levels have been associated with increased mortality in patients with sepsis. In contrast, albumin is a protein synthesized by the liver and is commonly used as a marker of nutritional status and inflammation. In sepsis, albumin levels often decrease owing to increased capillary permeability and a fluid shift from the intravascular space to the interstitial space. Low albumin levels have been shown to be associated with increased mortality in septic patients. Our studies have shown that patients with sepsis and higher lactate and lower albumin levels are at an increased risk of mortality. Therefore, monitoring these laboratory markers in patients with sepsis may be useful for predicting the prognosis and guiding their management.

However, this study has several limitations. First, we used prospectively collected multi-center data. However, despite using a large amount of data and clinical variables, nonstructural data, such as bio-signals and images, are not included. In the future, it

will be necessary to conduct a multimodal study by combining structural tablet data with nonstructural data. Second, no longitudinal studies were conducted. We performed mortality prediction for patients with sepsis; however, this was a cross-sectional study. The research results can be used as part of a CDSS; longitudinal studies are needed to design more precise and personalized treatments. Third, although this study aims for early mortality prediction in septic patients using mainly ER-obtained clinical variables and a large, continuously updated cohort, it also includes hard-to-detect early variables like infectious source control and initial empirical therapy. This is to secure a large number of investigated patients for this study who can be excluded through exclusion of the that variables, and we considered as important variables in clinical practice. However, as a result of comparison among various ML models, and XAI interpreted through interactions between variables by constructed in this study, the majority of variables that can be quickly obtained in the ER had a significant effect and influence on early mortality prediction. In further study, it is believed that a more accurate early mortality prediction model can be developed for more patients through future prospective study data collection. Lastly, interpretation through SHAP has potential issues, because the attribution of feature importance is normally based on random permutations. This approach has limitations, such as a high correlation among predictors. The presence of high correlation among features can lead to SHAP values that may not precisely represent the actual importance of each feature. In future studies, we plan to implement Kernel SHAP, a method that employs a weighted sampling strategy to calculate each SHAP value, taking into account the inter-correlations among features.

In conclusion, a new ML model for the prediction of sepsis mortality was developed. The developed ML model was demonstrated to predict sepsis mortality with a higher accuracy than the SOFA score; this is a significant achievement in sepsis management. A new ML model was constructed using various clinical variables in a large dataset, resulting in improved performance compared to SOFA. This model has the potential to improve patient outcomes by enabling more accurate predictions and facilitating timely interventions. However, further studies are necessary to validate this model and ensure its suitability for clinical practice.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

Supplementary Table 1
SOFA score criteria

Supplementary Table 2
Data split of clinical variables

Supplementary Table 3
Data split of clinical variables

Supplementary Fig. 1
SHAP figure of clinical variables. Feature importance for clinical variables based on interpretation using Shapley's additive explanations of machine learning models for information dense summary of how the top features in a dataset impact the model output. (**A**) Logistic regression. (**B**) Support vector machine. (**C**) Random forest. (**D**) Extreme gradient boosting. (**E**) Light gradient boosting machine.

Supplementary Fig. 2
Confusion matrix results for clinical variables in all six models. (**A**) LR. (**B**) SVM. (**C**) RF. (**D**) XGB. (**E**) LGB. (**F**) CatBoost.

Supplementary Fig. 3
SHAP figure of SOFA score. Feature importance for SOFA component scores based on interpretation using Shapley's additive explanations for information-dense summary of how the top features in a dataset impact the model output. (**A**) Logistic regression. (**B**) Support vector machine. (**C**) Random forest. (**D**) Light gradient boosting machine. (**E**) Categorical boosting.

Supplementary Fig. 4
Confusion matrix results for SOFA component scores in all six models. (**A**) LR. (**B**) SVM. (**C**) RF. (**D**) XGB. (**E**) LGB. (**F**) CatBoost.

REFERENCES

1. Rudd KE, Kissoon N, Limmathurotsakul D, Bory S, Mutahunga B, Seymour CW, et al. The global burden of sepsis: barriers and potential solutions. *Crit Care* 2018;22(1):232. **PUBMED** | **CROSSREF**

2. Park DW, Chun BC, Kim JM, Sohn JW, Peck KR, Kim YS, et al. Epidemiological and clinical characteristics of community-acquired severe sepsis and septic shock: a prospective observational study in 12 university hospitals in Korea. *J Korean Med Sci* 2012;27(11):1308-14. **PUBMED** | **CROSSREF**

3. Reaven MS, Rozario NL, McCarter MS, Heffner AC. Incidence and risk factors associated with early death in patients with emergency department septic shock. *Acute Crit Care* 2022;37(2):193-201. **PUBMED** | **CROSSREF**

4. 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. **PUBMED** | **CROSSREF**

5. Kim HI, Park S. Sepsis: early recognition and optimized treatment. *Tuberc Respir Dis (Seoul)* 2019;82(1):6-14. **PUBMED** | **CROSSREF**

6. Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* 2019;321(20):2003-17. **PUBMED** | **CROSSREF**

7. Liu H, Zhang L, Xu F, Li S, Wang Z, Han D, et al. Establishment of a prognostic model for patients with sepsis based on SOFA: a retrospective cohort study. *J Int Med Res* 2021;49(9):3000605211044892. **PUBMED | CROSSREF**

8. Thakur R, Naga Rohith V, Arora JK. Mean SOFA score in comparison with APACHE II score in predicting mortality in surgical patients with sepsis. *Cureus* 2023;15(3):e36653. **PUBMED | CROSSREF**

9. Koozi H, Lidestam A, Lengquist M, Johnsson P, Frigyesi A. A simple mortality prediction model for sepsis patients in intensive care. *J Intensive Care Soc* 2023;24(4):372-8. **PUBMED | CROSSREF**

10. Li W, Wang M, Zhu B, Zhu Y, Xi X. Prediction of median survival time in sepsis patients by the SOFA score combined with different predictors. *Burns Trauma* 2020;8:tkz006. **PUBMED | CROSSREF**

11. Pan X, Xie J, Zhang L, Wang X, Zhang S, Zhuang Y, et al. Evaluate prognostic accuracy of SOFA component score for mortality among adults with sepsis by machine learning method. *BMC Infect Dis* 2023;23(1):76. **PUBMED | CROSSREF**

12. Yang J, Liao Y, Dai Y, Hu L, Cai Y. Prediction of prognosis in sepsis patients by the SOFA score combined with miR-150. *Adv Clin Exp Med* 2022;31(1):9-15. **PUBMED | CROSSREF**

13. Liu Z, Meng Z, Li Y, Zhao J, Wu S, Gou S, et al. Prognostic accuracy of the serum lactate level, the SOFA score and the qSOFA score for mortality among adults with Sepsis. *Scand J Trauma Resusc Emerg Med* 2019;27(1):51. **PUBMED | CROSSREF**

14. Li Y, Yan C, Gan Z, Xi X, Tan Z, Li J, et al. Prognostic values of SOFA score, qSOFA score, and LODS score for patients with sepsis. *Ann Palliat Med* 2020;9(3):1037-44. **PUBMED | CROSSREF**

15. Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. *J Transl Med* 2022;20(1):215. **PUBMED | CROSSREF**

16. Kijpaisalratana N, Sanglertsinlapachai D, Techaratsami S, Musikatavorn K, Saoraya J. Machine learning algorithms for early sepsis detection in the emergency department: a retrospective study. *Int J Med Inform* 2022;160:104689. **PUBMED | CROSSREF**

17. Yao RQ, Jin X, Wang GW, Yu Y, Wu GS, Zhu YB, et al. A machine learning-based prediction of hospital mortality in patients with postoperative sepsis. *Front Med (Lausanne)* 2020;7:445. **PUBMED | CROSSREF**

18. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. *Front Med (Lausanne)* 2021;8:607952. **PUBMED | CROSSREF**

19. Du M, Haag DG, Lynch JW, Mittinty MN. Comparison of the tree-based machine learning algorithms to cox regression in predicting the survival of oral and pharyngeal cancers: ANALYSES based on SEER database. *Cancers (Basel)* 2020;12(10):2802. **PUBMED | CROSSREF**

20. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med* 2020;18(1):462. **PUBMED | CROSSREF**

21. Hu CA, Chen CM, Fang YC, Liang SJ, Wang HC, Fang WF, et al. Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. *BMJ Open* 2020;10(2):e033898. **PUBMED | CROSSREF**

22. Rodríguez A, Mendoza D, Ascuntar J, Jaimes F. Supervised classification techniques for prediction of mortality in adult patients with sepsis. *Am J Emerg Med* 2021;45:392-7. **PUBMED | CROSSREF**

23. Greco M, Caruso PF, Spano S, Citterio G, Desai A, Molteni A, et al. Machine learning for early outcome prediction in septic patients in the emergency department. *Algorithms* 2023;16(2):76. **CROSSREF**

24. van Doorn WP, Stassen PM, Borggreve HF, Schalkwijk MJ, Stoffers J, Bekers O, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. *PLoS One* 2021;16(1):e0245157. **PUBMED | CROSSREF**

25. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med Inform Decis Mak* 2020;20(1):251. **PUBMED | CROSSREF**

26. Park S, Jeon K, Oh DK, Choi EY, Seong GM, Heo J, et al. Normothermia in patients with sepsis who present to emergency departments is associated with low compliance with sepsis bundles and increased in-hospital mortality rate. *Crit Care Med* 2020;48(10):1462-70. **PUBMED | CROSSREF**

27. Valera Durán M. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock: 2016. *Rev Electron AnestesiaR* 2017;9:2.

28. Na SJ, Oh DK, Park S, Lee YJ, Hong SB, Park MH, et al. Clinical characteristics and outcomes of neutropenic sepsis: a multicenter cohort study. *Shock* 2022;57(5):659-65. **PUBMED | CROSSREF**

29. Yeo HJ, Lee YS, Kim TH, Jang JH, Lee HB, Oh DK, et al. Vasopressor initiation within 1 hour of fluid loading is associated with increased mortality in septic shock patients: analysis of national registry data. *Crit Care Med* 2022;50(4):e351-60. **PUBMED | CROSSREF**

30. Jeon K, Na SJ, Oh DK, Park S, Choi EY, Kim SC, et al. Characteristics, management and clinical outcomes of patients with sepsis: a multicenter cohort study in Korea. *Acute Crit Care* 2019;34(3):179-91. **PUBMED** | **CROSSREF**

31. Singh D, Singh B. Investigating the impact of data normalization on classification performance. *Appl Soft Comput* 2020;97:105524. **CROSSREF**

32. Feng J, Xu H, Mannor S, Yan S. Robust logistic regression and classification. NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems; December 8-13, 2014; Montreal, Canada. Cambridge, MA, USA: MIT Press; 2014, 253-61.

33. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20(3):273-97. **CROSSREF**

34. Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32. **CROSSREF**

35. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'16; August 13-17, 2016; San Francisco, CA, USA. New York, NY, USA: Association for Computing Machinery; 2016, 785-94.

36. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 2017;30:3149-57.

37. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* 2018;31:6638-48.

38. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;30:4765-74.

39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12(85):2825-30.

40. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; August 4-8, 2019; Anchorage, AK, USA. New York, NY, USA: Association for Computing Machinery; 2019, 2623-31.

41. Amland RC, Hahn-Cover KE. Clinical decision support for early recognition of sepsis. *Am J Med Qual* 2019;34(5):494-501. **PUBMED** | **CROSSREF**

42. Do SN, Dao CX, Nguyen TA, Nguyen MH, Pham DT, Nguyen NT, et al. Sequential Organ Failure Assessment (SOFA) Score for predicting mortality in patients with sepsis in Vietnamese intensive care units: a multicentre, cross-sectional study. *BMJ Open* 2023;13(3):e064870. **PUBMED** | **CROSSREF**

43. Moreno R, Rhodes A, Piquilloud L, Hernandez G, Takala J, Gershengorn HB, et al. The Sequential Organ Failure Assessment (SOFA) Score: has the time come for an update? *Crit Care* 2023;27(1):15. **PUBMED** | **CROSSREF**

44. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016;315(8):762-74. **PUBMED** | **CROSSREF**

45. Schinkel M, Nanayakkara PW, Wiersinga WJ. Sepsis performance improvement programs: from evidence toward clinical implementation. *Crit Care* 2022;26(1):77. **PUBMED** | **CROSSREF**

46. Burney M, Underwood J, McEvoy S, Nelson G, Dzierba A, Kauari V, et al. Early detection and treatment of severe sepsis in the emergency department: identifying barriers to implementation of a protocol-based approach. *J Emerg Nurs* 2012;38(6):512-7. **PUBMED** | **CROSSREF**

47. Rajula HS, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina (Kaunas)* 2020;56(9):455. **PUBMED** | **CROSSREF**

48. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med* 2016;44(2):368-74. **PUBMED** | **CROSSREF**

49. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed* 2022;214:106584. **PUBMED** | **CROSSREF**

50. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electron Mark* 2021;31(3):685-95. **CROSSREF**

51. Ribas Ripoll VJ, Vellido A, Romero E, Ruiz-Rodríguez JC. Sepsis mortality prediction with the Quotient Basis Kernel. *Artif Intell Med* 2014;61(1):45-52. **PUBMED** | **CROSSREF**

52. Zhang Z, Hong Y. Development of a novel score for the prediction of hospital mortality in patients with severe sepsis: the use of electronic healthcare records with LASSO regression. *Oncotarget* 2017;8(30):49637-45. **PUBMED** | **CROSSREF**

53. Li X, Zhou Y, Dvornek NC, Gu Y, Ventola P, Duncan JS. Efficient Shapley explanation for features importance estimation under uncertainty. *Med Image Comput Comput Assist Interv* 2020;12261:792-801.
   PUBMED | **CROSSREF**

54. Bakker J, Nijsten MW, Jansen TC. Clinical use of lactate monitoring in critically ill patients. *Ann Intensive Care* 2013;3(1):12. **PUBMED** | **CROSSREF**

55. Quinlan GJ, Martin GS, Evans TW. Albumin: biochemical properties and therapeutic potential. *Hepatology* 2005;41(6):1211-9. **PUBMED** | **CROSSREF**