



# Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission

Chang Hu · Lu Li · Yiming Li · Fengyun Wang · Bo Hu ·  
Zhiyong Peng

Received: May 6, 2022 / Accepted: June 23, 2022 / Published online: July 14, 2022  
© The Author(s) 2022

## ABSTRACT

**Introduction:** Septic patients requiring intensive care unit (ICU) readmission are at high risk of mortality, but research focusing on the association of ICU readmission due to sepsis and mortality is limited. The aim of this study was to develop and validate a machine learning (ML) model for predicting in-hospital mortality in septic patients readmitted to the ICU using routinely available clinical data.

---

Chang Hu, Lu Li and Yiming Li contributed equally to this work and share first authorship.

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40121-022-00671-3>.

---

C. Hu · L. Li · Y. Li · F. Wang · B. Hu (✉) · Z. Peng (✉)  
Department of Critical Care Medicine, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei, China  
e-mail: [hobbier1979@163.com](mailto:hobbier1979@163.com)

Z. Peng  
e-mail: [zn001590@whu.edu.cn](mailto:zn001590@whu.edu.cn)

C. Hu · L. Li · Y. Li · F. Wang · B. Hu · Z. Peng  
Clinical Research Center of Hubei Critical Care Medicine, Wuhan, Hubei, China

C. Hu · B. Hu  
Jiangsu Provincial Key Laboratory of Critical Care Medicine, Southeast University, Nanjing, Jiangsu, China

**Methods:** The data used in this study were obtained from the Medical Information Mart for Intensive Care (MIMIC-IV, v1.0) database, between 2008 and 2019. The study cohort included patients with sepsis requiring ICU readmission. The data were randomly split into a training (75%) data set and a validation (25%) data set. Nine popular ML models were developed to predict mortality in septic patients readmitted to the ICU. The model with the best accuracy and area under the curve (A.C.) in the validation cohort was defined as the optimal model and was selected for further prediction studies. The SHAPLY Additive explanations (SHAP) values and Local Interpretable Model-Agnostic Explanation (LIME) methods were used to improve the interpretability of the optimal model.

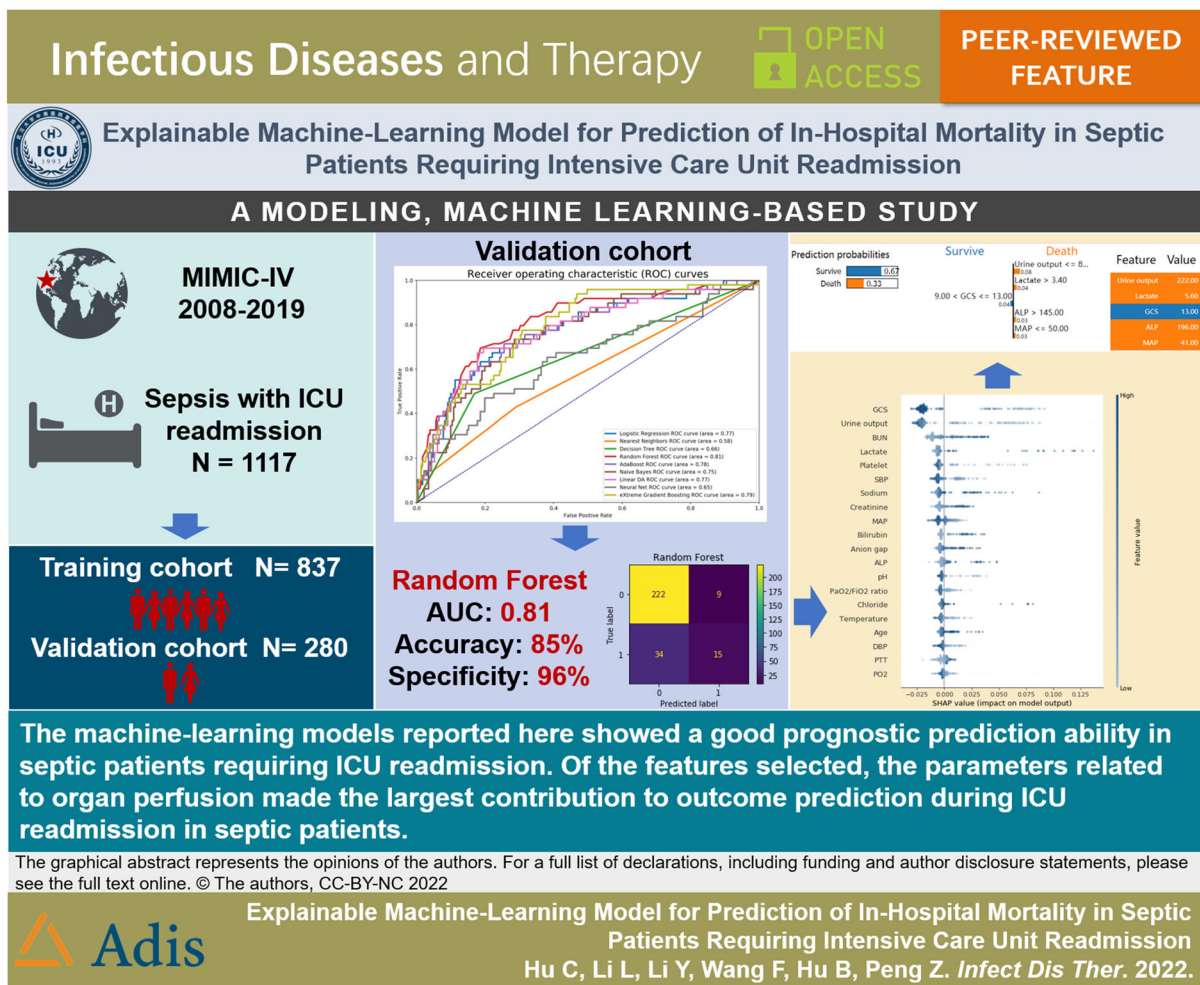
**Results:** A total of 1117 septic patients who had required ICU readmission during the study period were enrolled in the study. Of these participants, 434 (38.9%) were female, and the median (interquartile range [IQR]) age was 68.6 (58.4–79.2) years. The median (IQR) ICU interval duration was 2.60 (0.64–5.78) days. After feature selection, 31 of 47 clinical factors were ultimately chosen for use in model construction. Of the nine ML models tested, the best performance was achieved with the random forest (RF) model, with an A.C. of 0.81, an accuracy of 85% and a precision of 62% in the validation cohort. The SHAP summary analysis revealed that Glasgow Coma Scale score, urine

output, blood urea nitrogen, lactate, platelet count and systolic blood pressure were the top six most important factors contributing to the RF model. Additionally, the LIME method demonstrated how the RF model works in terms of explaining risk of death prediction in septic patients requiring ICU readmission. **Conclusion:** The ML models reported here showed a good prognostic prediction ability in septic patients requiring ICU readmission. Of

the features selected, the parameters related to organ perfusion made the largest contribution to outcome prediction during ICU readmission in septic patients.

**Keywords:** Explainable artificial intelligence; Machine learning; Random Forest; Sepsis; Mortality

Infographic



### Key Summary Points

#### *Why carry out this study?*

Sepsis is among the major causes of readmission to the intensive care unit (ICU), with a considerable negative impact on critically ill patients readmitted to ICU.

Early prediction of deterioration could provide a critical window of time for clinical interventions that might reduce severity.

The objective of this investigation was to develop and validate an explainable machine-learning (ML) model based on clinical variables to estimate the in-hospital mortality in septic patients requiring ICU readmission.

#### *What was learned from the study?*

ML-based algorithms for accurate prediction of mortality in septic patients requiring ICU readmission is possible.

Parameters related to organ perfusion contributed the most to outcome prediction during ICU readmission in septic patients.

SHAP values and the LIME method could help to improve the model's interpretation.

## DIGITAL FEATURES

This article is published with digital features, including a graphical abstract, to facilitate understanding of the article. To view digital features for this article go to <https://doi.org/10.6084/m9.figshare.20260935>.

## INTRODUCTION

The transition of patients from the intensive care unit (ICU) to a hospital ward is one of the highest-risk transitions of care [1, 2]. About 5–7% of patients admitted to the ICU will be transferred back to ICU, and the frequency of ICU readmission appears to have increased over the last 20 years [3]. ICU readmission is associated with higher mortality and longer ICU and hospital lengths of stay [4, 5]. Sepsis, defined as life-threatening organ dysfunction caused by a dysregulated host response to infection, is among the most important reasons for ICU readmission [6] and has a considerable negative impact on critically ill patients [7]. Sepsis can be caused by virtually any infecting organism and is usually accompanied by the production of proinflammatory cytokines with immunosuppressive activity, which increase patient mortality and morbidity [8]. Recent epidemiological evidence suggests that sepsis occurs in more than 1.7 million persons annually in the USA alone, with an estimated mortality of 10–40% [9, 10], and it is increasingly recognized as a serious, worldwide public health priority [11]. Despite constant updates to the Surviving Sepsis Campaign (SSC) guidelines between 2004 and 2021 [12–16], the incidence of sepsis-related deaths remained unacceptably high [9].

Delays in identifying clinical deterioration in critically ill patients usually result in increased morbidity and mortality [17] while, conversely, early prediction of deterioration could provide a critical window of time for clinical interventions that might reduce severity [18]. Recently, a considerable body of literature has grown up around the theme of early detection of clinical outcomes in sepsis by utilizing advanced technologies [19–22], such as machine learning (ML) and artificial intelligence. However, most studies only included the first qualifying ICU admission during a single hospitalization, and excluded multiple qualifying ICU admissions [20–22]. Therefore, to our knowledge, no study has yet developed a specific ML model for predicting in-hospital mortality in septic patients requiring ICU readmission.

Explainability of ML is of great importance to help enhance the trust of medical professionals, because it shows why predictions are made and how parameters contribute to the model [23]. However, almost all of the ML models are black boxes, and the decision-making procedure is complex and hard to interpret and explain in detail [24, 25]. For this reason, Lundberg and colleagues first identified an efficient and trustworthy method to significantly enhance the interpretability of ML models by incorporating desirable properties from game theory [26]. This powerful method has also been successfully used in prediction of the prevention of hypoxemia in patients during surgery [27], prediction of the development of acute kidney injury in patients following cardiac surgery [28] and prediction of sepsis in patients with COVID-19 [29].

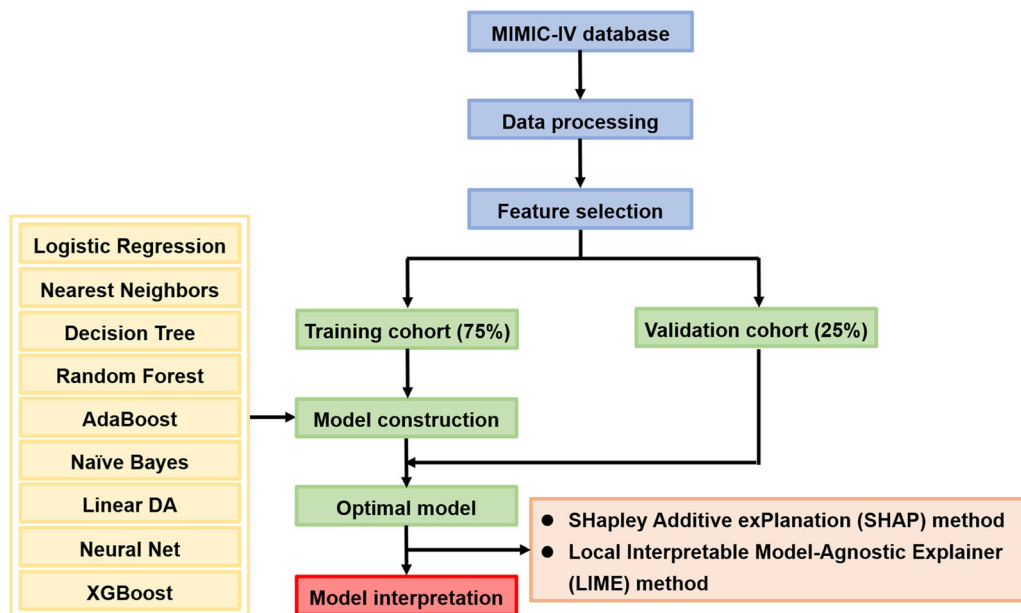
Accordingly, the objective of this investigation was to develop and validate an explainable ML model based on clinical variables to estimate the in-hospital mortality in septic patients requiring ICU readmission.

## METHODS

### Study Design and Data set

This was a modeling study that used a clinical data set extracted from the Medical Information Mart for Intensive Care (MIMIC)-IV database (v.1.0) [30]. The MIMIC-IV is a publicly available single-center critical care database that contains information on > 70,000 critically ill patients admitted to the ICU at the Beth Israel Deaconess Medical Center from 2008 to 2019. After completing a training course called Protecting Human Research Participants that includes Health Insurance Portability and Accountability Act (HIPAA) requirements, we were granted access (author Chang Hu; certification number: 47460147).

The principal procedures of this study were conducted in three main steps. First, we developed nine popular ML models using clinical variables collected on the first 24 h after ICU readmission. Second, we compared the performance of nine ML models in the validation set, and the optimized model with the best accuracy and greatest area under the curve (AUC) was



**Fig. 1** Flow diagram of this study. *MIMIC-IV* Medical Information Mart for Intensive Care IV, *XGBoost* eXtreme Gradient Boosting, *AdaBoost* Adaptive Boosting

chosen as the optimal model. Third, we performed two last interpretation methods to explain the optimal model. The overall workflow of this study is schematically presented in Fig. 1.

The establishment of this database was approved by the Massachusetts Institute of Technology (Cambridge, MA, USA) and Beth Israel Deaconess Medical Center (Boston, MA, USA), and consent had been obtained for the original data collection. Therefore, the ethical approval statement and the need for informed consent from each individual patient were waived for this study because the project did not impact clinical care and all protected health information was deidentified [30]. This study was done in accordance with the Declaration of Helsinki [31] and complied with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [32].

### Patients and Outcomes

We enrolled all adult individuals (> 18 years of age) diagnosed with sepsis in the last 24 h before ICU readmission or in the first 24 h after ICU readmission. The ICU readmission was defined as any return to ICU after a first ICU discharge during the first hospitalization within 1 year. The ICU interval duration was defined as the number of hours between the first ICU discharge and the second ICU admission (Electronic Supplementary Material [ESM] Fig. S1). Patients with third or more ICU admission were not included in our study. Sepsis was defined according to the Sepsis 3.0 criteria [7]. In brief, a confirmed or suspected infection combined with a Sequential Organ Failure Assessment (SOFA) score  $\geq 2$  (ESM Fig. S1). Patients with ICU length of stay < 24 h were excluded. The flow diagram of the inclusion criteria and exclusion criteria are presented in ESM Fig. S2.

The primary endpoint of our study was the area under the receiver operating characteristic curve (AUROC) of the model's prediction. The secondary endpoints of our study were accuracy, precision, recall and F1 score of the model's prediction.

### Data Collection

Clinical data in the MIMIC-IV database were extracted for the first 24 h following patient ICU readmission. These variables included patients' demographics, medical history, vital signs, laboratory parameters, severity of illness scoring systems and outcomes. Demographic data extracted included: age, sex, body weight and height. Then, we collected data on comorbidities, including hypertension, diabetes, congestive heart failure, cerebrovascular disease, chronic pulmonary disease, liver disease, renal disease and tumor, and we calculated the Charlson Comorbidity Index (CCI) [33]. Vital signs, such as heart rate, systolic blood pressure (SBP), diastolic blood pressure, mean artery pressure, respiratory rate, body temperature and SpO<sub>2</sub> were also abstracted. Furthermore, we also collected the laboratory parameters, such as white blood cell counts, platelets, hematocrit, hemoglobin, international normalized ratio, prothrombin time, partial thromboplastin time, alanine aminotransferase, alkaline phosphatase, aspartate aminotransferase, amylase, total bilirubin, creatine phosphokinase, creatine kinase-MB, lactate dehydrogenase, albumin, blood urea nitrogen (BUN), serum creatinine, lactate, pH, PO<sub>2</sub>, PCO<sub>2</sub>, PaO<sub>2</sub>/FiO<sub>2</sub> ratio, base excess, anion gap, bicarbonate, serum calcium, serum chloride, serum sodium, serum potassium, fibrinogen and blood glucose. ICU interval duration, taken to be the number of hours between the first ICU discharge and the second ICU admission, and the cumulative urine output on day 1 in the ICU were also extracted. Finally, we calculated the Glasgow Coma Scale (GCS), SOFA score [34], Oxford Acute Severity of Illness Score (OASIS) [35] and Logistic Organ Dysfunction System (LODS) score [36]. The maximum value and minimum value in the first 24 h of each variable were considered to be different variables in the final data set. A list of each variable is detailed in ESM Table S1.

### Feature Selection and Data Preprocessing

We selected all of the features listed above (except SOFA, OASIS and LODS; for medical



history, only the CCI was included for analysis) for the development of ML models. Feature selection is considered to be a crucial step of data preprocessing. For this study, we proposed a five-stage feature selection method: (1) find columns with a missing fraction > 50%, and features with a missing percentage of not > 50% were retained and filled in with the method of multivariate imputation by chained equations in the R statistical environment [37] (ESM Table S2); (2) find features with only a single unique value; (3) find collinear features as identified by a correlation coefficient > 0.7; (4) find features with zero importance from a gradient boosting machine; and (5) find features that do not contribute to a 95% cumulative feature importance from the gradient boosting machine. The details of feature selection are shown in ESM Fig. S3. Ultimately, we chose 31 clinical features to be included for model construction (ESM Table S3).

### Model Development and Validation

The data set was randomly split into two data sets: a training (75%) data set, which was used to develop the models, and an internal validation (25%) data set, which was used to validate the constructed models. We utilized the following nine representative ML classifier algorithms for clinical feature screening and model construction in the training data set: Logistic Regression, Nearest Neighbors, Decision Tree, Random Forest (RF), adaptive boosting (AdaBoost), Naive Bayes, Linear DA, Neural Net and extreme gradient boost (XGBoost). To ensure maximum use of data, we did not use a cross-validation method. The A.C., as well as accuracy, precision (also called positive predictive value), recall and F1-measure (F1) were calculated for each ML model to be evaluated and compared in the validation cohort. Through comprehensive evaluation of multiple evaluation indicators, the best performing model among the nine ML models was defined as the optimal model and selected for further prediction analysis. Furthermore, we compared the performance of the optimal model-based ML versus the traditional model-based severity of

illness scores (OASIS and SOFA) to predict the mortality in septic patients requiring ICU readmission. Finally, we performed calibration curve to evaluate the consistency of the optimal model.

### Model Explainability

Opening the black box of ML is of great importance to engender trust with the health-care workforce and provide transparency into the ML-based decision-making process. Thus, we implemented SHAPLY Additive explanations (SHAP) values and the Local Interpretable Model-Agnostic Explanation (LIME) method to improve the interpretability for the top-performing model. SHAP values could help to quantify the contribution of each feature and explain how the features affect the output of the optimal prediction model, whereas the LIME method could help to determine ML model explainability for each patient's prediction.

### Statistical Analysis

Differences in patient characteristics that were measured on a continuous scale were compared using the *t*-test or Mann–Whitney *U*-test as appropriate, and categorical variables were evaluated using the Chi-square or Fisher exact test, as appropriate. After feature selection and data preprocessing, we developed nine popular ML-based models to predict mortality in septic patients requiring ICU readmission. Overall performance of each model was assessed via the A.C., accuracy, precision, recall and F1-measure. Moreover, a two-by-two confusion matrix with the number of true positive, false positive, false negative and true negative values was generated for each ML model. Then, the best performing model was applied to the further interpretation. Finally, SHAP summary analysis, SHAP dependence analysis and the LIME method were utilized for model explainability.

SPSS statistical software version 24.0 (IBM Corp., Armonk, NY, USA), R statistical software version 3.6.1 ® Project for Statistical Computing, Vienna, Austria) and Python software

**Table 1** Baseline characteristics of the study cohort

Characteristics	Survivors ( <i>N</i> = 896)	Non-survivors ( <i>N</i> = 221)	<i>P</i> value
<i>Demographic</i>			
Age, year	67.9 (57.7–78.2)	71.9 (62.4–83.1)	< 0.001
Sex			0.118
- Male, <i>n</i> (%)	558 (62.3)	125 (56.6)	
- Female, <i>n</i> (%)	338 (37.7)	96 (43.4)	
Weight, kg	80 (68–94)	78 (67–93)	0.433
Height, cm	170 (163–178)	169 (163–175)	0.232
<i>Comorbidities</i>			
Charlson Comorbidity Index	6 (4–8)	7 (6–9)	< 0.001
Hypertension, <i>n</i> (%)	355 (39.6)	79 (35.7)	0.290
Diabetes, <i>n</i> (%)	304 (33.9)	67 (30.3)	0.307
Congestive heart failure, <i>n</i> (%)	342 (38.2)	88 (39.8)	0.652
Cerebrovascular disease, <i>n</i> (%)	146 (16.3)	47 (21.3)	0.080
Chronic pulmonary disease, <i>n</i> (%)	264(29.5)	56 (25.3)	0.224
Liver disease, <i>n</i> (%)	171 (19.1)	68 (30.8)	< 0.001
Renal disease, <i>n</i> (%)	215 (24.0)	77 (34.8)	0.001
Tumor, <i>n</i> (%)	153 (17.1)	62 (28.1)	< 0.001
<i>Vital signs on day 1</i>			
Heart rate, bpm	88 (79–99)	93 (83–103)	0.001
Systolic blood pressure, mmHg	114 (105–124)	109 (100–120)	< 0.001
Diastolic blood pressure, mmHg	60 (54–67)	58 (51–64)	0.001
Mean arterial pressure, mmHg	75 (70–82)	71 (67–78)	< 0.001
Respiratory rate	19 (17–23)	20 (18–24)	0.003
Body temperature, °C	36.9 (36.7–37.2)	36.7 (36.4–37.1)	< 0.001
SpO <sub>2</sub> , %	93 (90–95)	91 (87–94)	< 0.001
<i>Laboratory findings on day 1</i>			
White blood cell, × 10 <sup>3</sup> /μL	14.3 (10.3–20.1)	15.0 (10.6–20.7)	0.519
Platelets, × 10 <sup>3</sup> /μL	193 (118–275)	142 (68–222)	< 0.001
Hematocrit, %	26.9 (23.8–30.5)	25.6 (22.7–30.1)	0.033
Hemoglobin, g/dL	8.8 (7.8–10.1)	8.2 (7.3–9.9)	0.001
International normalized ratio	1.4 (1.2–1.7)	1.6 (1.4–2.4)	< 0.001
Prothrombin time, s	15.5 (13.5–18.6)	18.1 (14.8–25.3)	< 0.001
Partial thromboplastin time, s	35 (29–54)	44 (32–75)	< 0.001

**Table 1** continued

Characteristics	Survivors ( <i>N</i> = 896)	Non-survivors ( <i>N</i> = 221)	<i>P</i> value
Alanine aminotransferase, U/L	31 (17–68)	33 (18–76)	0.262
Alkaline phosphatase, U/L	90 (63–144)	107 (75–178)	0.002
Aspartate aminotransferase, U/L	42 (25–86)	61 (33–184)	< 0.001
Amylase, U/L	56 (37–95)	61 (28–129)	0.761
Total bilirubin, mg/dL	1.0 (0.5–2.1)	1.3 (0.6–5.0)	0.004
Creatine phosphokinase, U/L	100 (43–284)	130 (60–527)	0.045
Creatine kinase MB, U/L	4 (2–7)	5 (3–14)	0.005
Lactate dehydrogenase, U/L	311 (227–412)	438 (246–805)	< 0.001
Albumin, g/dL	2.8 (2.4–3.2)	2.4 (2.1–3.2)	0.005
Blood urea nitrogen, mg/dL	24 (16–39)	38 (25–60)	< 0.001
Serum creatinine, mg/dL	1.1 (0.8–1.7)	1.6 (1.0–2.7)	< 0.001
Lactate, mmol/L	2.1 (1.4–3.5)	3.0 (1.7–6.4)	< 0.001
pH	7.45 (7.40–7.49)	7.42 (7.36–7.47)	< 0.001
pO <sub>2</sub> , mmHg	73 (49–102)	61 (41–81)	< 0.001
pCO <sub>2</sub> , mmHg	46 (40–55)	47 (39–61)	0.589
PaO <sub>2</sub> /FiO <sub>2</sub> ratio	163 (95–240)	109 (72–178)	< 0.001
Base excess	2 (0–5)	0 (– 3 to 3)	< 0.001
Anion gap	15 (13–18)	18 (14–22)	< 0.001
Bicarbonate, mmol/L	25 (23–28)	24 (20–28)	< 0.001
Serum calcium, mmol/L	8.5 (8.1–9.0)	8.6 (8.0–9.3)	0.077
Serum chloride, mmol/L	105 (101–109)	106 (101–112)	0.020
Serum sodium, mmol/L	139 (137–142)	142 (139–146)	< 0.001
Serum potassium, mmol/L	4.4 (4.1–4.9)	4.5 (4.1–5.1)	0.120
Fibrinogen, mg/dL	266 (203–390)	230 (179–350)	0.042
Blood glucose, mg/dL	172 (138–217)	174 (136–236)	0.441
ICU interval duration <sup>a</sup> , day	2.5 (0.6–5.6)	2.9 (1.0–7.0)	0.073
Urine output on day 1, mL	1535 (1040–2350)	861 (445–1645)	< 0.001
<i>Severity of illness scores</i>			
GCS	13 (10–14)	8 (4–13)	< 0.001
SOFA	6 (4–9)	10 (7–14)	< 0.001
OASIS	34 (28–40)	41 (34–49)	< 0.001



**Table 1** continued

Characteristics	Survivors ( <i>N</i> = 896)	Non-survivors ( <i>N</i> = 221)	<i>P</i> value
LODS	5 (3–8)	9 (7–12)	< 0.001

Data are presented as a number with the percentage in parentheses, or as the median with the interquartile range (IQR) in parentheses

ICU Intensive care unit, GCS Glasgow Coma Scale, SOFA Sequential Organ Failure Assessment, OASIS Oxford Acute Severity of Illness Score, LODS Logistic Organ Dysfunction System

<sup>a</sup>ICU interval duration was defined as the time gap between the first ICU discharge and the second ICU admission

version 3.6.6 (Python Software Foundation, Wilmington, DE, USA) were used for all analyses. All statistical tests were 2-sided, and *P* values < 0.05 were considered to be statistically significant.

## RESULTS

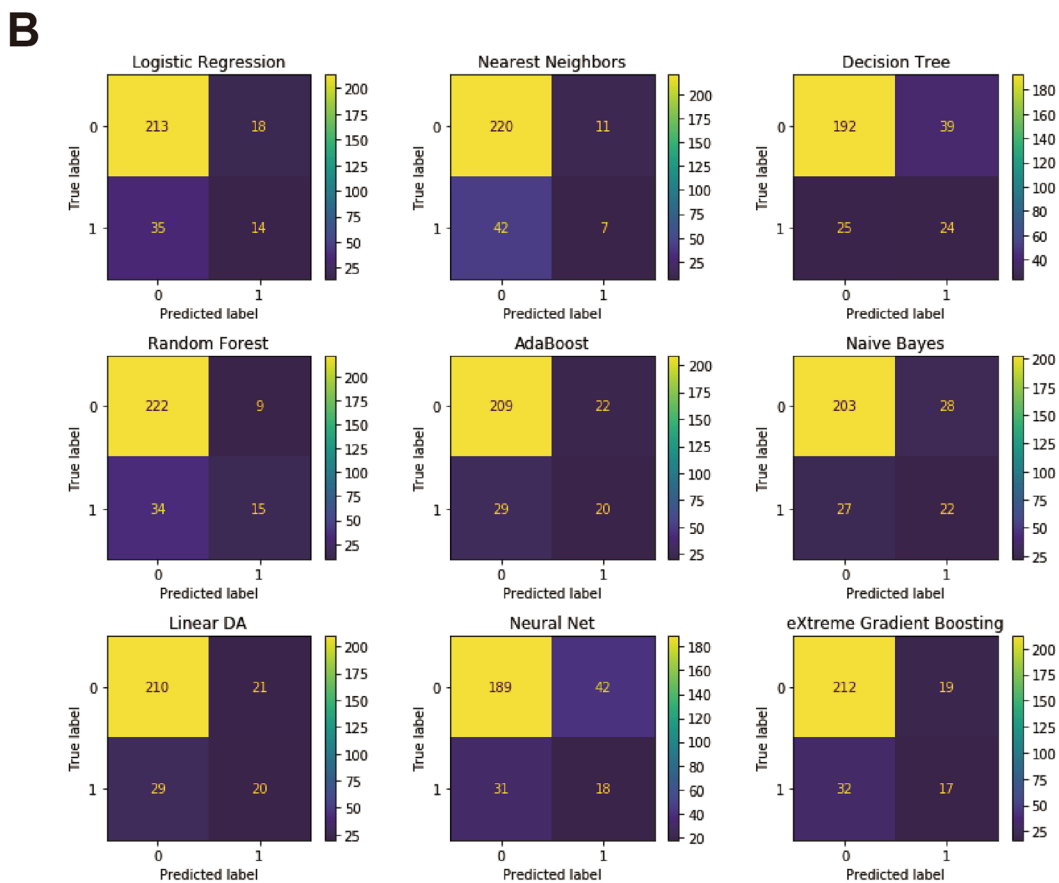
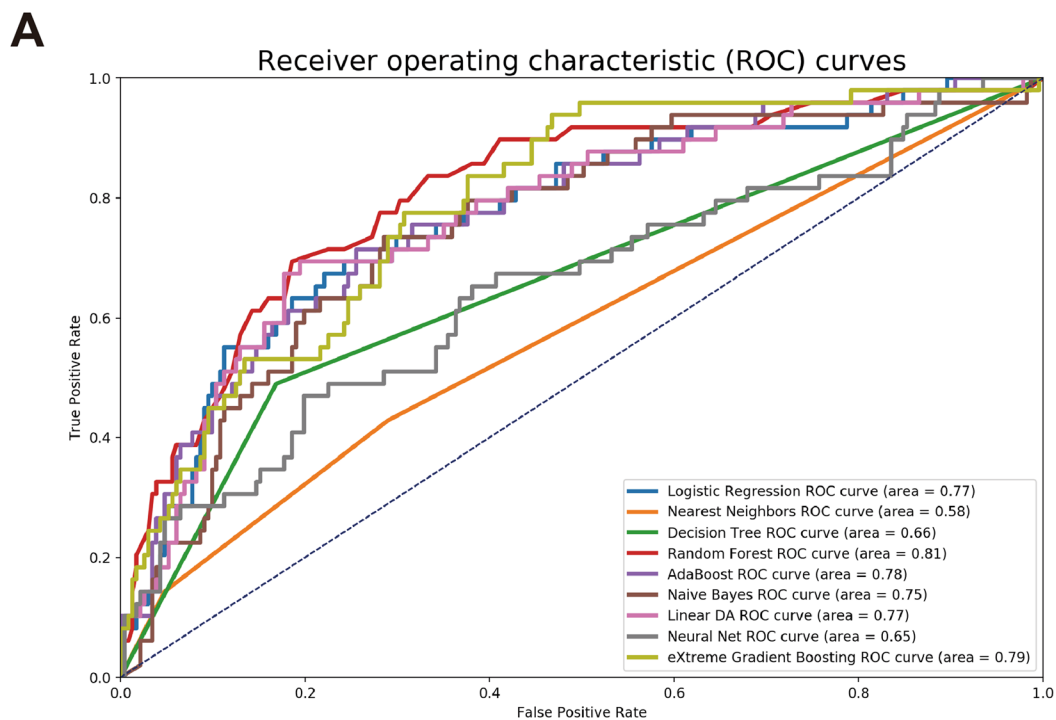
### Cohort Characteristics

Among 76,540 ICU stays registered in the MIMIC-IV database, we identified 1117 patients as having sepsis who were readmitted to the ICU, including 837 in a training cohort and 280 in a validation cohort. The flow chart of this study is presented in ESM Fig. S2. Of these 1117 participants, 434 (38.9%) were female, and the median (interquartile range [IQR]) age was 68.6 (58.4–79.2) years. The median (IQR) ICU interval duration was 2.60 (0.64–5.78) days. The mortality rate of this cohort was 19.8% (221/1117). Table 1 shows the baseline characteristics of the overall study cohort stratified by the survivor status. Compared with those in the survivor group, non-survivors were older (71.9 [IQR 62.4–83.1] vs. 67.9 [57.7–78.2] years; *P* < 0.001); had higher CCI scores (7 [IQR 6–9] vs. 6 [4–8]; *P* < 0.001); and had more comorbidities of liver disease (30.8% vs. 19.1%; *P* < 0.001), renal disease (34.8% vs. 24.0%; *P* = 0.001) and tumor (28.1% vs. 17.1%; *P* < 0.001). In addition, non-survivors were more likely to have lower urine output on day 1 after ICU readmission (861 [IQR 445–1645] vs. 1535 [1040–2350] mL; *P* < 0.001), with higher

organ injury scores, such as SOFA (10 [IQR 7–14] vs. 6 [4–9]; *P* < 0.001), OASIS (41 [IQR 34–49] vs. 34 [28–40]; *P* < 0.001) and LODS (9 [IQR 7–12] vs. 5 [3–8]; *P* < 0.001).

### Model Development and Validation

The overview of study design is shown in Fig. 1. The detailed processes for determining each variable are described in ESM Fig. S3. Missingness is considered to be missing at random, and we performed multiple imputation of chained equations for missing data with missingness < 50%. ESM Fig. S4 presents a good agreement between the imputed values and actual values. After feature selection, we used 31 candidate factors for model construction (ESM Table S3). The performance of each of the nine ML models as assessed by ROC analysis and confusion matrix are presented in Fig. 2. In the validation cohort, the RF model had a promising discriminatory capability with an A.C. of 0.81 for estimating in-hospital mortality among septic patients requiring ICU readmission, compared with the XGBoost (A.C. 0.79), AdaBoost (A.C. 0.78), Logistic Regression (A.C. 0.77), Linear DA (A.C. 0.77), Naive Bayes (A.C. 0.75), Decision Tree (A.C. 0.66), Neural Net (A.C. 0.65) and Nearest Neighbors (A.C. 0.58). Table 2 details the A.C., accuracy, precision, recall and F1 of the nine ML classifiers. The RF approach achieved the best performance with an accuracy of 85% and a precision of 62%, compared with other eight ML models. Additionally, the RF model had a significantly greater A.C. than OASIS and SOFA (RF 0.81; OASIS 0.76; SOFA



◀**Fig. 2** Area under the receiver operator curve (*ROC*) and confusion matrix for each machine learning–based model to estimate mortality in the validation set. **a** ROC curves for the nine machine-learning models to predict in-hospital mortality, **b** confusion matrix for each machine learning–based model to estimate in-hospital mortality. *AdaBoost* adaptive boosting

0.74) (ESM Fig. S5). The calibration curve also showed a good agreement between predictions and actual observations, which demonstrated the consistency of the RF model (ESM Fig. S6). Thus, the RF model was applied to the further interpretation.

**Model Explainability**

Figure 3 shows the SHAP summary plot that orders features based on their importance to predict mortality in the validation cohort. For the RF approach, GCS score, urine output, BUN, lactate, platelet count and SBP were the top six most important features. Figure 4 shows the relationship between the top six features and the prediction of patient outcomes. The SHAP values for these six features exceed zero, representing an increased risk of mortality. The elevated GCS score, cumulative urine output on day 1, platelet count and SBP showed a negative

correlation with mortality due to sepsis, whereas elevated BUN and lactate levels showed a positive correlation.

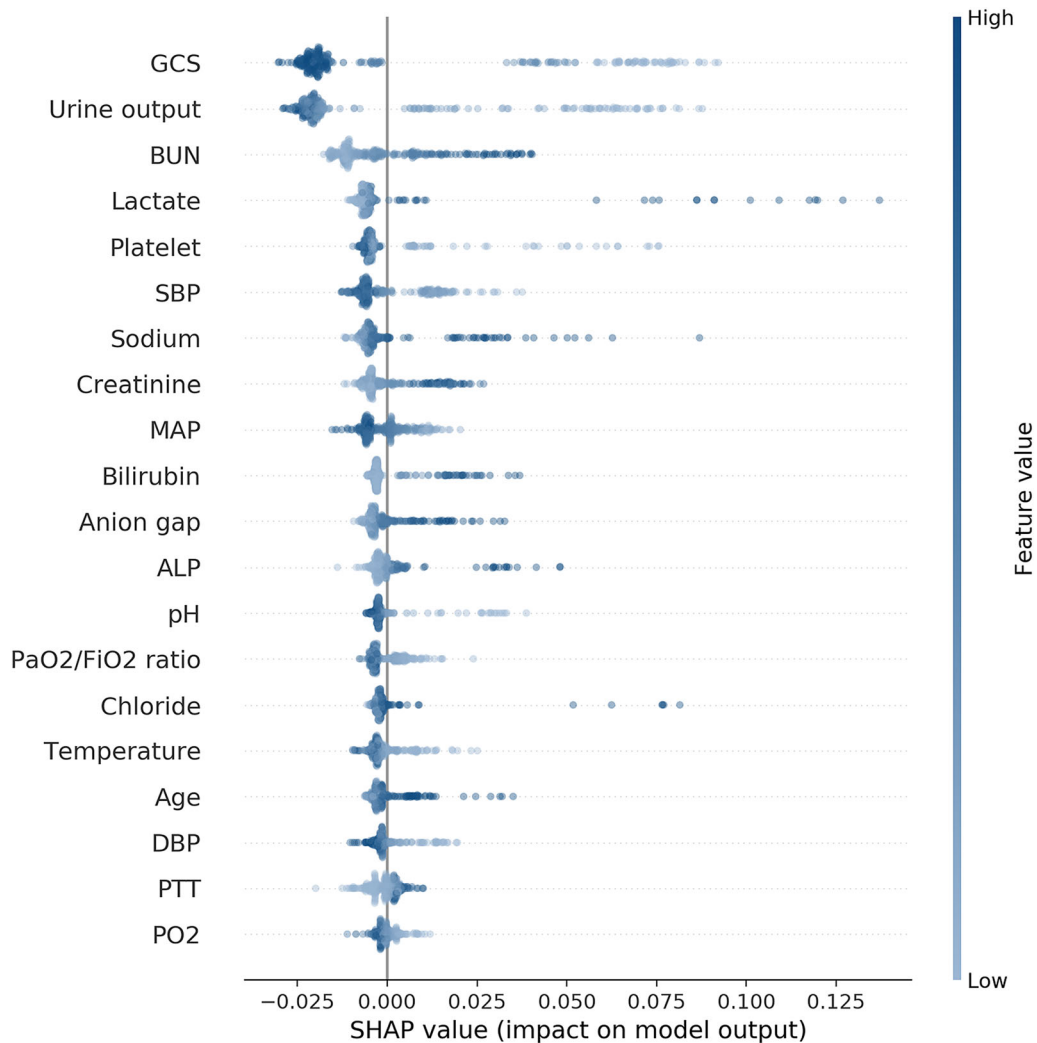
**Model Application**

Figure 5 shows the four prediction patients using the RF model in the validation set. For patient A (Fig. 5a), the predicted probability for in-hospital mortality by the RF model is 2%. The factors detected for predicting a lower mortality in this patient include a high GCS score of 13, a normal urine output of 1845 mL, a normal kidney function with BUN of 7.0 mg/dL, a normal serum sodium level of 137 mmol/L and a normal total bilirubin value of 0.6 mg/dL. For patient B (Fig. 5b), the predicted probability for mortality is 6%. The patient’s risk of mortality is only a decreased SpO<sub>2</sub> value of 87%, whereas the normal urine output of 1465 mL, high GCS score of 13, normal BUN of 14.0 mg/dL and normal partial thromboplastin time of 27.3 s contribute to a negative impact for mortality. For patient C (Fig. 5c), the predicted probability for mortality is 33%. The patient’s oliguria (the cumulative urine output of 222 mL on day 1), elevated lactate value of 5.6 mmol/L, high alkaline phosphatase level of 196 U/L and low mean arterial pressure level of 41 mmHg contribute to increasing the mortality, whereas

**Table 2** Performance of the nine machine-learning classifiers for predicting in-hospital mortality in the validation set

Classifiers	A.C.	Accuracy	Precision	Recall	F1
Random Forest	0.81	84.64%	62%	0.31	0.41
eXtreme Gradient Boosting	0.79	81.79%	47%	0.35	0.40
AdaBoost	0.78	81.79%	48%	0.41	0.44
Logistic Regression	0.77	81.07%	44%	0.29	0.35
Linear DA	0.77	82.14%	49%	0.41	0.44
Naive Bayes	0.75	80.36%	44%	0.45	0.44
Decision Tree	0.66	77.14%	38%	0.49	0.43
Neural Net	0.65	73.93%	30%	0.37	0.33
Nearest Neighbors	0.58	81.07%	39%	0.14	0.21

*A.C.* Area under curve, *AdaBoost* Adaptive Boosting



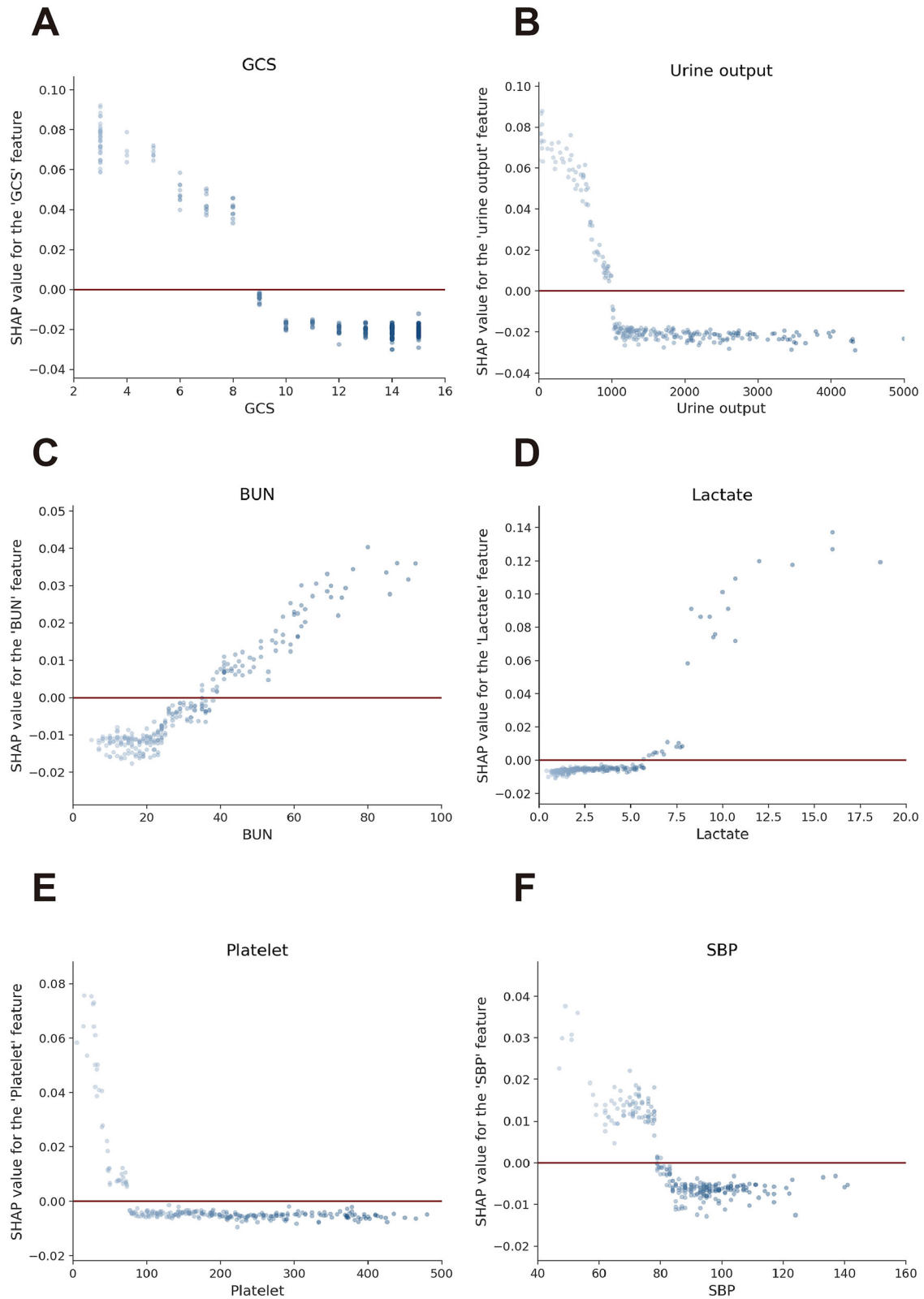
**Fig. 3** Feature importance assessment for RF classifier using SHAP values in the validation set. Each dot represents 1 patient and accumulates vertically to depict the density. The color reflects the high and low values of each feature, with deep color indicating a higher value and light color indicating a lower value. The X-axis of the graph represents the SHAP value, and a positive SHAP value indicates that it contributes positively to predicting the model and that the probability of mortality occurring

is high, and vice versa. *SHAP* SHAPLEY Additive explanation, *RF* Random Forest, *GCS* Glasgow Coma Scale, *BUN* blood urea nitrogen, *SBP* systolic blood pressure, *MAP* mean arterial pressure, *ALP* alkaline phosphatase, *DBP* diastolic blood pressure, *PTT* partial thromboplastin time, *PO2* partial pressure of oxygen, *PaO2/FiO2* ratio of arterial oxygen partial pressure to fractional inspired oxygen

a high GCS score of 13 helps offset this risk. For patient D (Fig. 5d), the predicted probability for mortality is 65%. The patient's risk for a such high mortality because of a low GCS score of 3, a lower urine output of 222 mL and abnormal laboratory findings (high level of glucose, serum sodium and BUN).

## DISCUSSION

Our aim in performing this modeling, ML-based study was to establish an effective, stable and explainable model for predicting mortality in septic patients requiring ICU readmission. The results demonstrated that the RF model was the





◀**Fig. 4** SHAP dependence plot of the top 6 clinical features in the RF classifier. **a** GCS, **b** urine output, **c** BUN, **d** lactate, **e** platelet, **f** SBP. Values are plotted with a scatter plot. A feature has an impact on mortality when its SHAP value is  $> 0$ , and it has an impact on the patient's survival when its SHAP value is  $<$  than 0. *GCS* Glasgow Coma Scale, *BUN* blood urea nitrogen, *SBP* systolic blood pressure, *SHAP* SHapley Additive explanation, *RF* Random Forest

most reliable and accurate of all models tested for predicting outcomes. We also found that GCS score, urine output, BUN, lactate, platelet count and SBP were the top six most important features contributing to the RF model. Moreover, we illustrated the specific impacts of key features on the RF model in four individual patients from the validation cohort. Overall, our study demonstrated that accurate prediction of mortality for septic patients requiring ICU readmission using routinely collected clinical data is possible.

Sepsis is a life-threatening condition with limited therapeutic options. Additionally, survivors of critical illness with sepsis usually have a high ICU readmission rate [38]. However, the vast majority of previously described models were created for septic patients with the first ICU admission, and no one model has yet been constructed to predict outcome for such patients with ICU readmission [20, 39–41]. Therefore, it is necessary to develop and validate a risk as well as an outcome prediction model to assess septic patients requiring ICU readmission.

In the present study, we found that the mortality rate in septic patients requiring ICU readmission was 19.8%, which was approximately 1.6-fold higher than that associated with the first ICU admission for sepsis (12.6%) [22]. In accordance with the previous results, our study also found that readmission of patients to the ICU was associated with worse outcomes, such as increased mortality and increased lengths of stay [3].

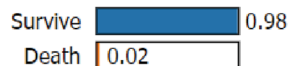
In this study, we extracted the clinical data from the MIMIC-IV database and randomly divided the data set into a training cohort (75%) and a validation cohort (25%). We then

constructed nine popular ML models to predict the mortality risk for septic patients requiring ICU readmission. The results from the validation set showed that the best performing model among the nine ML models tested was the RF model, having the highest A.C., accuracy and precision and classical severity of illness scores (OASIS, SOFA). Given the many data elements included in our models, including patient demographics, medical history and laboratory test results, it is not surprising that the RF model performed well statistically. These findings are in keeping with previous observational studies that confirmed the important role of applying artificial intelligence to predict mortality in patients with sepsis [20, 40].

ML models are usually considered to be a black box: the data go in and the decisions come out, but the processes between input and output are opaque [42–44]. This opaqueness on how decisions are arrived at accounts for why healthcare providers are unable to explain why the ML model makes a such particular prediction or how it works. In this study, we utilized two novel methods, the SHAP value and LIME, to open the black box in our RF model. The SHAP value is a game theoretic approach that is used to explain the output of any ML model with a good performance, whereas the LIME can explain the predictions of the ML model by learning an interpretable model locally around the prediction. The SHAP summary analysis showed that GCS score, urine output, BUN, lactate, platelet count and SBP were the top six most important factors contributing to the RF model. This result is partly in agreement with our earlier observations, which showed that GCS score, BUN and cumulative urine output on day 1 were the key factors contributing to the XGBoost model in predicting mortality among patients with sepsis [22]. In addition, these results demonstrated that the elevated lactate levels, decreased platelet counts and SBP may be even more relevant to predicting increasing risk of mortality in septic patients upon readmission to the ICU, compared with the primary ICU admission for sepsis. Taken further, these results indicated that the parameters related to organ perfusion contribute highly to outcome prediction during ICU

### A

Prediction probabilities



Survive

9.00 < GCS <= 13.00  
0.03  
1375.00 < Urine ou...  
0.03  
BUN <= 18.00  
0.03  
Sodium <= 137.00  
0.03  
0.50 < Bilirubin <=...

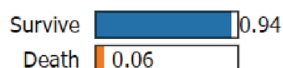
Death

Feature Value

GCS	13.00
Urine output	1845.00
BUN	7.00
Sodium	137.00
Bilirubin	0.60

### B

Prediction probabilities



Survive

1375.00 < Urine ou...  
0.04  
9.00 < GCS <= 13.00  
0.03  
BUN <= 18.00  
0.03  
SPO2 <= 89.00  
0.03  
PTT <= 29.80  
0.02

Death

Feature Value

Urine output	1465.00
GCS	13.00
BUN	14.00
SPO2	87.00
PTT	27.30

### C

Prediction probabilities



Survive

9.00 < GCS <= 13.00  
0.04

Death

Urine output <= 8...  
0.08  
Lactate > 3.40  
0.04  
ALP > 145.00  
0.03  
MAP <= 50.00  
0.03

Feature Value

Urine output	222.00
Lactate	5.60
GCS	13.00
ALP	196.00
MAP	41.00

### D

Prediction probabilities



Survive

GCS <= 9.00  
0.10  
Urine output <= 8...  
0.08  
Glucose > 217.00  
0.05  
Sodium > 143.00  
0.05  
BUN > 44.00  
0.05

Death

Feature Value

GCS	3.00
Urine output	435.00
Glucose	230.00
Sodium	146.00
BUN	55.00

◀**Fig. 5** Results of local interpretable model-agnostic explanations (LIME) for top 5 clinical features using RF classifier applied to the 4 patients in the validation set. Visualization of 4 individual patients for the explanation risk of death prediction, with blue denoting the negative case and orange denoting the positive case. *GCS* Glasgow Coma Scale, *BUN* blood urea nitrogen, *PTT* partial thromboplastin time, *ALP* alkaline phosphatase, *MAP* mean arterial pressure, *RF* Random Forest, *LIME* local interpretable model-agnostic explanations

readmission in septic patients. Therefore, much more attention should be paid to organ perfusion in patients with sepsis readmitted to the ICU. On the other hand, we used the LIME method to clarify the outcomes produced by the model on the four individuals from the validation set, and to explain how the RF model works in mortality prediction for septic patients. These will greatly increase a healthcare provider's trust in the ML model behavior and provide insights into the possibility for clinical use.

Recently, the combination of higher-frequency physiological data streams and artificial intelligence offer promising applications for predicting the onset of sepsis, which can allow for early identification of at-risk patients [45–47]. This promising tool help healthcare providers in making better treatment decisions.

However, the present study still has several limitations. First, the modeling, single-center and retrospective design of this study does not allow causal inferences about the associations between variables and mortality to be drawn. Additionally, the prediction efficiency of the current model may be affected by racial and ethnic differences. Second, the populations in our study were heterogeneous due to the broad time window for sepsis diagnosis. We have attempted to partly mitigate this effect by applying a new definition of sepsis (Sepsis-3) and excluding treatment measures in our final model. Third, our predictive model lacks proper external validation, and this will affect the credibility of the RF model. Fourth, our study could not capture some potential features known to be highly predictive of mortality risk, such as medications, and we could only collect

the features based on available data in the MIMIC-IV. Fifth, we only extracted the clinical data within the first 24 h after ICU readmission, but we failed to assess changes during the ICU stay, which may ignore the effect of feature fluctuations on the trends. Thus, further multicenter prospective studies are needed to validate our findings.

## CONCLUSION

The ML models have shown a good prognostic prediction ability in septic patients requiring ICU readmission. Parameters related to organ perfusion made the major contributions to outcome prediction in septic patients readmitted to the ICU. The SHAP values and LIME method could help to improve interpretation of ML model outcomes.

## ACKNOWLEDGEMENTS

The authors would like to thank the MIMIC-IV program for access to the database.

**Funding.** This work and the journal's rapid service fee were supported by the National Natural Science Foundation of China (Zhiyong Peng; No. 81772046 and 81971816), the Special Project for Significant New Drug Research and Development in the Major National Science and Technology Projects of China (Zhiyong Peng, No. 2020ZX09201007), the Translational Medicine and Interdisciplinary Research Joint Fund of Zhongnan Hospital of Wuhan University (Bo Hu, No. ZNJC202011), and the Subject Cultivation Project of Zhongnan Hospital of Wuhan University (Zhiyong Peng, No. ZNXKPY2021001; Bo Hu, No. ZNXKPY2021002), and the Open Project Program of the Jiangsu Provincial Key Laboratory of Critical Care Medicine (Bo Hu, No. JSKLCCM-2021-02-003).

**Authorship.** All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of

the work as a whole, and have given their approval for this version to be published.

**Author Contributions.** CH, LL, BH and ZP designed this study; CH and LL were responsible for the data collection; CH, YL and FW were responsible for data analysis; CH and BH conducted the manuscript writing; BH and ZP critically revising the manuscript. All authors read and approved the final manuscript.

**Disclosures.** Chang Hu, Lu Li, Yiming Li, Fengyun Wang, Bo Hu and Zhiyong Peng have nothing to disclose.

**Compliance with Ethics Guidelines.** The establishment of this database was approved by the Massachusetts Institute of Technology (Cambridge, MA, USA) and Beth Israel Deaconess Medical Center (Boston, MA, USA), and consent had been obtained for the original data collection. Therefore, the ethical approval statement and the need for informed consent from each individual patient were waived for this study because the project did not impact clinical care and all protected health information was deidentified [30]. This study was done in accordance with the Declaration of Helsinki [31] and complied with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [32].

**Data Availability.** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Open Access.** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit

line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

1. Kripalani S, Jackson AT, Schnipper JL, Coleman EA. Promoting effective transitions of care at hospital discharge: a review of key issues for hospitalists. *J Hosp Med.* 2007;2(5):314–23.
2. Hosein FS, Roberts DJ, Turin TC, Zygun D, Ghali WA, Stelfox HT. A meta-analysis to derive literature-based benchmarks for readmission and hospital mortality after patient discharge from intensive care. *Crit Care.* 2014;18(6):715.
3. Kramer AA, Higgins TL, Zimmerman JE. The association between ICU readmission rate and patient outcomes. *Crit Care Med.* 2013;41(1):24–33.
4. Rosenberg AL, Watts C. Patients readmitted to ICUs\*: a systematic review of risk factors and outcomes. *Chest.* 2000;118(2):492–502.
5. Renton J, Pilcher DV, Santamaria JD, et al. Factors associated with increased risk of readmission to intensive care in Australia. *Intensive Care Med.* 2011;37(11):1800–8.
6. Brown SE, Ratcliffe SJ, Kahn JM, Halpern SD. The epidemiology of intensive care unit readmissions in the United States. *Am J Respir Crit Care Med.* 2012;185(9):955–64.
7. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA.* 2016;315(8):801–10.
8. Cecconi M, Evans L, Levy M, Rhodes A. Sepsis and septic shock. *Lancet.* 2018;392(10141):75–87.
9. Rhee C, Dantes R, Epstein L, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009–2014. *JAMA.* 2017;318(13):1241–9.
10. Peltan ID, McLean SR, Murnin E, et al. Prevalence, characteristics, and outcomes of emergency department discharge among patients with sepsis. *JAMA Netw Open.* 2022;5(2): e2147882.

11. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S. Recognizing sepsis as a global health priority—a WHO resolution. *N Engl J Med*. 2017;377(5):414–7.
12. Dellinger RP, Carlet JM, Masur H, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Crit Care Med*. 2004;32(3):858–73.
13. Dellinger RP, Levy MM, Carlet JM, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med*. 2008;36(1):296–327.
14. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med*. 2013;39(2):165–228.
15. Rhodes A, Evans LE, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Crit Care Med*. 2017;45(3):486–552.
16. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*. 2021;47(11):1181–247.
17. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. *N Engl J Med*. 2020;383(20):1951–60.
18. Dong J, Feng T, Thapa-Chhetry B, et al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. *Crit Care*. 2021;25(1):288.
19. Agor JK, Paramita N, Ozaltn OY. Prediction of sepsis related mortality: an optimization approach. *IEEE J Biomed Health Inform*. 2021;25(11):4207–16.
20. Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med*. 2020;18(1):462.
21. Wernly B, Mamandipoor B, Baldia P, Jung C, Osmani V. Machine learning predicts mortality in septic patients using only routinely available ABG variables: a multi-centre evaluation. *Int J Med Inform*. 2021;145: 104312.
22. Hu C, Li L, Huang W, et al. Interpretable machine learning for early prediction of prognosis in sepsis: a discovery and validation study. *Infect Dis Ther*. 2022;11(3):1117–32.
23. Ploug T, Holm S. The four dimensions of contestable AI diagnostics—a patient-centric approach to explainable AI. *Artif Intell Med*. 2020;107: 101901.
24. Reddy S. Explainability and artificial intelligence in medicine. *Lancet Digit Health*. 2022;4(4):e214–5.
25. Watson DS, Krutzinna J, Bruce IN, et al. Clinical applications of machine learning algorithms: beyond the black box. *BMJ*. 2019;364: 1886.
26. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1): 56–67.
27. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng*. 2018;2(10):749–60.
28. Tseng PY, Chen YT, Wang CH, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. 2020;24(1):478.
29. Tang G, Luo Y, Lu F, et al. Prediction of sepsis in COVID-19 using laboratory indicators. *Front Cell Infect Microbiol*. 2020;10: 586054.
30. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3: 160035.
31. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191–4.
32. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63.
33. Charlson ME, Carrozzino D, Guidi J, Patierno C. Charlson Comorbidity Index: a critical review of clinimetric properties. *Psychother Psychosom*. 2022;91(1):8–35.
34. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315(8):762–74.
35. Johnson AE, Kramer AA, Clifford GD. A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy. *Crit Care Med*. 2013;41(7):1711–8.
36. Le Gall JR, Klar J, Lemeshow S, et al. The Logistic Organ Dysfunction system. A new way to assess



- organ dysfunction in the intensive care unit. ICU Scoring Group. *JAMA*. 1996;276(10):802–10.
37. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338: b2393.
38. Maharaj R, Terblanche M, Vlachos S. The utility of ICU readmission as a quality indicator and the effect of selection. *Crit Care Med*. 2018;46(5): 749–56.
39. Banerjee S, Mohammed A, Wong HR, Palaniyar N, Kamaleswaran R. Machine learning identifies complicated sepsis course and subsequent mortality based on 20 genes in peripheral blood immune cells at 24 H post-ICU admission. *Front Immunol*. 2021;12: 592303.
40. Yao RQ, Jin X, Wang GW, et al. A machine learning-based prediction of hospital mortality in patients with postoperative sepsis. *Front Med (Lausanne)*. 2020;7:445.
41. van Doorn W, Stassen PM, Borggreve HF, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. *PLoS ONE*. 2021;16(1): e0245157.
42. Price WN. Big data and black-box medical algorithms. *Sci Transl Med*. 2018;10(471):eao5333. <https://doi.org/10.1126/scitranslmed.aao5333>.
43. The Lancet Respiratory Medicine. Opening the black box of machine learning. *Lancet Respir Med*. 2018;6(11):801.
44. Savage N. Breaking into the black box of artificial intelligence. *Nature*. 2022.<https://doi.org/10.1038/d41586-022-00858-1>.
45. Mohammed A, Van Wyk F, Chinthala LK, et al. Temporal differential expression of physiometers predicts sepsis in critically ill adults. *Shock*. 2021;56(1):58–64.
46. Kamaleswaran R, Akbilgic O, Hallman MA, West AN, Davis RL, Shah SH. Applying artificial intelligence to identify physiometers predicting severe sepsis in the PICU. *Pediatr Crit Care Med*. 2018;19(10):e495–503.
47. Deng HF, Sun MW, Wang Y, et al. Evaluating machine learning models for sepsis prediction: a systematic review of methodologies. *iScience*. 2022;25(1):103651.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.