

A Time-Incorporated SOFA-Based Explainable Machine Learning Model for Mortality Prediction in Critically Ill Patients

Yang Liu

Nanjing University Medical School

Kun Gao

Nanjing Medical University

Hongbin Deng

Nanjing Medical University

Tong Ling

Nanjing University

Jiajia Lin

Nanjing University Medical School

Xianqiang Yu

Southeast University Medical College

Xiangwei Bo

Southeast University Medical College

Jing Zhou

Southeast University Medical College

Lin Gao

Nanjing University Medical School

Peng Wang

Nanjing Medical University

Zhihui Tong

Nanjing University Medical School

Yuxiu Liu

Nanjing Medical University

Yinghuan Shi

Nanjing University

Lu Ke (✉ ctgkelu@nju.edu.cn)

Nanjing University <https://orcid.org/0000-0001-8093-5073>

Yang Gao

Nanjing University

Weiqin Li

Research

Keywords: Organ dysfunction, SOFA, Time dimension, Machine learning, Mortality prediction

Posted Date: November 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1026029/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Organ dysfunction (OD) assessment is essential in intensive care units (ICUs). However, no OD scoring system has so far considered the duration of OD, which is clinically relevant. This study aimed to develop and validate an ICU mortality prediction model based on the Sequential Organ Failure Assessment (SOFA) score, incorporating the time dimension with machine learning methods.

Methods: Data from the eICU Collaborative Research Database and Medical Information Mart for Intensive Care (MIMIC) -III were mixed for model development, and the MIMIC-IV dataset and Nanjing Jinling Hospital Surgery ICU (SICU-JL) dataset were used for external testing. Adult patients in the ICUs for more than 72 hours were deemed eligible. The total SOFA score and individual scores were calculated every 12 hours for the first three days of ICU admission. Time-dimensional variables were derived from the consecutively recorded SOFA scores and individual scores for each organ. A modified SOFA model incorporating the time dimension (T-SOFA) was stepwise constructed to predict ICU mortality using multiple machine learning algorithms. The predictive performance was assessed with the area under the receiver operating characteristic curves (AUROC). Also, we utilized the SHapley Additive exPlanations (SHAP) algorithm for data visualization and model explainability.

Results: We extracted a total of 66,709 ICU patients from the mixed datasets for model development and 15,423 patients for validation. The T-SOFA M3 that incorporated the time dimension features and age, using the XGBoost algorithm, significantly outperformed the original SOFA scores (AUROC 0.800 95% CI [0.787-0.813] vs. 0.693 95% CI [0.678-0.709], $p < 0.01$) in the validation set. Good prediction performance was maintained for the T-SOFA M3 in test Set A and test Set B, with AUROC of 0.803, 95% CI [0.791-0.815], and 0.830, 95% CI [0.789-0.870], respectively. Significant contributors demonstrated by the SHAP analysis included total SOFA score, Respiration-score, CNS-score, age, Cardiovascular-score, and SOFA Organ dysfunction Unalleviated Time Index.

Conclusions: A SOFA-based, time-incorporated prediction model was developed and validated by machine learning algorithms, showing satisfactory predictability and medical interpretability.

Background

Organ injury or dysfunction caused by a primary illness is a common reason for intensive care unit (ICU) admission[1]. Although the original diseases of critically ill patients vary, organ dysfunction(OD) is associated with substantial morbidity and mortality[2]. Therefore, OD assessment is a routine in ICUs to determine the impact of the underlying disease and the effects of therapies.

OD assessment system has evolved from the Multiple Organ Dysfunction Score (MODS) to the sequential organ failure assessment (SOFA) score, which was extended from sepsis to all critically ill patients after the establishment of the Sepsis-3 criteria [3–5]. Theoretically, OD evaluation should be conducted along three dimensions: the severity of each OD, the number of organ systems involved, and the duration of each OD. OD is a dynamic process rather than a static event [4], so a temporal assessment can assist

clinicians in understanding the progression of the disease and potentially facilitating decision-making. Unfortunately, to date, none of the existing OD scoring systems have taken the time dimension into account.

The assessment of time dimension is complicated by the fact that the duration of OD could be different from one system to another, and changes in OD during ICU stay are multidirectional. Previous studies proposed concepts such as Δ SOFA to reflect the direction and magnitude of variation, but information on the duration of each OD remains unaccountable [6]. Artificial intelligence algorithms have recently had a wide range of applications in healthcare practice [7], where machine learning or deep learning algorithms could process time-series data to improve prediction accuracy by extracting potentially valuable time features[8]. However, the cost is that the prediction process became a "black box" (unexplainable),sacrificing medical interpretability.

This study aimed to develop and validate a novel ICU mortality prediction model based on the SOFA score by incorporating time dimension through a machine learning algorithm. Furthermore, the impact of time dimension on ICU mortality was further clarified by interpretable algorithms.

Methods

Data sources

The study data was extracted from three real-world, open-source databases: Medical Information Mart for Intensive Care (MIMIC-III) version 1.4, MIMIC-IV version 1.0, and eICU-Collaborative Research Database (eICU-CRD) version 2.0 and one local dataset: Nanjing Jinling Hospital Surgery ICU (SICU-JL) database. The MIMIC-III database retrospectively included clinical data from all critically ill patients at Beth Israel Deaconess Medical Center from 2001 to 2012, including demographic data, bedside vital sign measurements (approximately once per hour), laboratory test results, records of therapeutic interventions and medical care and mortality (including in-hospital and after-discharge)[9]. MIMIC-IV is an updated version of the MIMIC database, containing data collected between 2008 and 2019[10]. We selected patients admitted between 2013 and 2019 from the MIMIC-IV database to distinguish it from the MIMIC-III database, thereby making it an external dataset. The eICU-CRD is another public database consisting of more than 200,000 admissions to ICUs from 208 hospitals across the US between 2014 and 2015, which provided a degree of generalizability to the training model due to its comprehensive geographic coverage and a more broad range of critical illnesses[11]. The SICU-JL database contained 2,155 electronic health records from the Surgical ICU of Jinling Hospital, Nanjing University, between 2018 to 2020. The following inclusion criteria were used for the present study: 1. adult critically ill patients aged \geq 18 years; 2. ICU stay at least 72 hours. The criterion-compliant data from the MIMIC-III and the eICU-CRD were mixed and randomly divided in a ratio of 8:2, where 80% of the selected admissions were used as a training cohort. The remaining 20% were served as an internal validation cohort to verify the predictive performance and optimize hyperparameter settings to prevent over-fitting. The data from the MIMIC-IV database was used as an external test set A, and data from the SICU-JL dataset served as external test

set B to ensure the generalizability of the model. We had obtained proof of completion of the course on human research and signed a data use agreement on PhysioNet, and approved by MIT's Institutional Review Board. Data collection from the SICU-JL dataset was approved by the institutional review board of the Jinling hospital. The study was conducted in accordance with the tenets of the Declaration of Helsinki. Informed consent was waived by the IRB-JL in light of the retrospective nature.

Data extraction and preprocessing

Demographic information (e.g., age, gender, ethnicity, etc.), clinical data related to SOFA score (e.g., Oxygen partial pressure, mean arterial pressure, Glasgow Coma Scale (GCS) score, etc.), and the outcomes of patients were extracted from the dataset. The raw clinical data required to calculate the SOFA score were collected every 12 hours after ICU admission, with T_1 for the first 12 hours at admission and T_6 at 72 hours after ICU entry. The total SOFA score and the sub-scores were calculated simultaneously, thereby resulting in six consecutive time points.

Definition of time dimension variables

We defined several novel variables to reflect the temporary change of OD as follows (Fig. 1): 1) The time point when the total SOFA score or a sub-score ≥ 2 was counted as an Organ dysfunction unit Time (OFuT) out of the six time points. We defined the sum of OFuT multiplied by the mean value of the total SOFA score or the corresponding sub-scores as the OFT Index (OFTI). 2) The difference in the total SOFA score or a sub-score between two adjacent time points ($\Delta\text{SOFA} = \text{SOFA}_{T_{x+1}} - \text{SOFA}_{T_x}, x \in [1, 2, 3, 4, 5]$) ≥ 0 was recorded as an Organ dysfunction Unalleviated unit Time (OUuT), and the summed OUuT multiplied by the mean value of the total SOFA score or the corresponding sub-scores was called the OUT Index (OUTI). 3) The time point when $\Delta\text{SOFA} > 0$ was defined as an Organ dysfunction Aggravated unit Time (OAuT), and the summed OAuT multiplied by the mean value of the total SOFA score or the corresponding sub-scores was defined as an OAT Index (OATI). 4) With the sub-score value (0-4 points) or the total score value (0-24 points) as the y-axis and time (0-3 days) as the x-axis, the study time points were connected, constituting the SOFA curves. The area under the SOFA curves was described as the SOFA-Time Area Under the Curve (STAUC), which could reflect the cumulative changes of OD over the 72 hours.

Conventional time series features of total SOFA score and sub-scores (e.g., Mean, maximum, minimum, variance) were also included for analysis [12].

Prediction model development, testing, and explanation

We used a supervised classification learning algorithm to construct a prediction model for ICU mortality based on SOFA score. The maximum SOFA score during the first 72 hours was included as the original SOFA as a baseline reference [13]. Then we progressively incorporate related variables into the prediction model. First, the total SOFA score and sub-scores at each time point were included. The model was trained using four machine learning algorithms, eXtreme Gradient Boosting (XGBoost) [14], Support Vector Machines (SVM) [15], Random Forest (RF) [16], and Logistic regression to obtain the T-SOFA M1. In

the second step, the time dimension was taken into account by incorporating the abovementioned temporary variables to train a prediction model called T-SOFA M2. In the third step, given that age is a well-proven risk factor for all-cause mortality in ICUs[17], we added age to M2, thereby creating T-SOFA M3. All training models would be performed in the internal validation set to prevent over-fitting. The performance of T-SOFA were further compared with the original SOFA and tested using datasets with public database and local database.

We used a unified framework, SHAP (SHapley Additive exPlanations), to obtain the impact of each variable on the designated outcome (ICU mortality), which could interpret the contribution of different organ systems and time features to the prediction model. Variables with an impact value ≥ 0.1 , as indicated by the SHAP algorithm, were considered significant for ICU mortality.

Missing data preprocessing

Most of the variables involved in this study were time-series data. Missing data were imputed by the last observation carried forward (LOCF)[18]. If an original variable required for SOFA score calculation is missing and cannot be imputed by proximity, it is filled using the mean value of the variables in the data set at the corresponding time point, when the missing data are normally distributed; otherwise, the median is used.

Statistical analysis

Differences in clinical characteristics between the training set, internal validation set, and external test sets were compared using the ANOVA or Kruskal-Wallis test, and categorical variables were compared using Pearson's chi-squared test. Comparisons of differences in temporary variables between survivors and non-survivors were made using a two independent sample t-test. The primary outcome was all-cause mortality in the ICUs, and the predictive performance was evaluated by the area under the receiver operating characteristic curves (AUROC). Differences in predictive performance between models were compared with a non-parametric DeLong test with a two-sided $p < 0.05$ criterion for statistical significance[19]. Statistical analyses were performed using Stata 15.0 software, and the machine learning algorithm modeling was implemented in Python 3.7 and R 3.6.3

Results

A total of 66,709 patients from the MIMIC-III database and the eICU-CRD were used to develop the mortality prediction model. Then 80% of the patients were randomly allocated for training and 20% for internal validation. 13,929 patients from the MIMIC-IV database and 1,494 patients from the SICU-JL database served as the external test set A and test set B, respectively (**Additional file 1. Figure S1**). The modeling cohort consisting of both the training and internal validation sets showed significant differences from the external test cohort in terms of baseline clinical characteristics of admissions and outcomes (**Table 1**). The patients from both test sets had ICU mortality around 9.7% (test set A:1358/13929, test set B:145/1494), which were higher than those in the training and internal validation

sets (9% [4785/53367] and 9% [1206/13342], respectively). Almost all the SOFA-related temporary variables and age differed statistically between survivors and non-survivors (**Additional file 1. Table S1**).

Prediction performance assessment

In the validation set, the original SOFA score showed an AUROC of 0.693 (95%CI 0.678-0.709) in predicting ICU mortality. The XGBoost algorithm outperformed the others (**Table 2.** and **Additional file 1. Table S2**) among the four machine learning algorithms and was therefore adopted as the modeling algorithm. T-SOFA M1 incorporated individual organ systems and multiple time points had an AUROC of 0.759 (95% CI [0.744-0.773]). T-SOFA M2, which further included temporary variables, had an AUROC of 0.787(95% CI [0.774-0.801]), higher than T-SOFA-M1 ($p < 0.01$). The AUROC for T-SOFA M3 was elevated to 0.800 (95% CI [0.787-0.813]) when adding age, which is significantly better than T-SOFA M2 ($p < 0.01$). To further confirm these findings, the prediction performance of each model was then tested in external test sets. The T-SOFA M3 outperformed the original SOFA and all the other models($p < 0.01$). The AUROC of T-SOFA M3 was 0.803 (95% CI [0.791-0.815]) in test set A, and the AUROC of original SOFA, T-SOFA M1 and T-SOFA M2 were 0.739 (95% CI [0.726-0.755]), 0.776 (95% CI [0.763-0.789]) and 0.797 (95% CI [0.784-0.809]), respectively (Fig. 2C). In the local test set B, T-SOFA M3 yielded an AUROC of 0.830 (95% CI [0.789-0.870]) and remained significantly better than the original SOFA score (AUROC=0.773 95% CI [0.728-0.818], $p < 0.01$).

Clinical interpretability of prediction models

Among all the variables, the significant contributors to ICU mortality appeared to be the total SOFA score, the respiration score, the CNS score, the age, the cardiovascular score, and the SOFA OUTI, with variable important for the projection(VIP) value of 0.2861, 0.2836, 0.2594, 0.2548, 0.1437 and 0.125, respectively (Fig. 3A). The SHAP values for each feature were visualized using a summary plot (Fig. 3B). Additionally, the prediction model provided individualized mortality risk factor analysis of instances through the SHAP algorithm (**Additional file 1. Figure S2**).

Discussion

In this study, a SOFA-based, time-incorporated mortality risk prediction model(T-SOFA M3) for critically ill patients was developed and validated. We defined several novel temporary variables to describe the dynamic change of OD quantitatively from different perspectives. The results showed that a prediction model incorporating time dimension and age using machine learning algorithms had a significantly better predictive performance for ICU mortality than the original SOFA.

The SHAP algorithm was applied to evaluate the contribution of the time dimension. The results showed that various OD types have different impacts with the respiration system, the CNS system, and the cardiovascular system ranking the top three, consistent with what are used in the popular qSOFA score[5, 20]. Among all the temporary variables, the impacts of SOFA OUTI, which reflects the duration of

persistent(unalleviated) OD, ranks highest, even higher than that of renal dysfunction, indicating the importance of the time dimension. The better interpretability provided by the SHAP model can potentially further guide treatment and facilitate clinical decision-making. Furthermore, our model derived from the impacts of each variable could be used for individual prediction, which would be of clinical use for risk stratification in ICU settings, which was impossible with previously proposed modified SOFA models[6, 13, 21–23].

Although efforts to develop new predictive models based on the SOFA score were repeatedly presented in the literature [6, 13, 20, 24–27], the T-SOFA M3 has several unique advantages. Firstly, the training set was obtained from the mixed datasets (MIMIC-III and eICU-CRD) with a large sample size, which may compensate for the remarkable heterogeneity among ICUs in the US. Secondly, the model was externally tested using an open-source multi-center dataset and a local single-center dataset, respectively, demonstrating excellent predictive performance and geographical generalizability. More, we collected total SOFA scores and individual scores every 12 hours within 72 hours of ICU admission, which was relatively infrequent and therefore doable in real practice, implicating its potential for future clinical use.

In recent years, the application of the SOFA score has in a broader range of diseases. Although the SOFA score was not initially developed to predict mortality but to evaluate comorbidities[4], it has become a well-established tool for prognosis prediction[28]. As a consequence, a lot of modified SOFA models have emerged to overcome the limitations of the SOFA score [29]. The modified SOFA systems incorporating the SOFA score derivatives like Δ SOFA, MAX SOFA, Mean SOFA, etc., could significantly outperform SOFA at admission, suggesting that characteristics representing changes in OD over time could help improve the performance.

Lilian Minne et al. systematically reviewed the performance of SOFA-based models for predicting mortality in critically ill patients and concluded that MAX SOFA had the best predictive performance (AUROC range = 0.792 to 0.922), while Δ SOFA showed variable performance across studies (AUROC range = 0.51 to 0.828) according to the definitions [13] [29]. Another alternative option was to optimize the SOFA by replacing individual components with more clinically reliable ones. Previous studies have attempted to use the Richmond Agitation-Sedation Scale (RASS) score as a substitution as the Glasgow Coma Scale (GCS) score tends to be inaccurately recorded. However, they failed to find the superiority of RASS score modified SOFA over the original version[30]. Adding other systems or variables to the original SOFA, such as the gastrointestinal system, was also considered and tested. Yehudit Apherstein et al. proposed the Novel Gastrointestinal Dysfunction Assessment Tool (Resting energy expenditure daily balance, Gastric residual volume, vomiting, and Bowel movements) as the seventh organ dysfunction assessment criterion of the modified SOFA and demonstrated improved prediction accuracy[21]. However, their modified SOFA score did not gain much popularity due to the subjective and inaccurate nature of assessing gastrointestinal dysfunction [31]. In the T-SOFA M3, age was the only additional variable apart from the original SOFA-required data since it was readily available and there was a strong correlation between age and mortality [32, 33].

There are some limitations to our study. First of all, both the training and validation sets were extracted from retrospective data, which inevitably contained missing data and recording errors. Although multiple efforts were made to minimize the potential bias, there was still some data distortion. Therefore, prospective studies are necessary to further validate the discriminative ability of the model. Moreover, our model requires consecutive 72-hour SOFA data, which denies its use in the "very early" (≤ 24 hours) stage of ICU admission.

Conclusions

We developed and validated a SOFA-based time-incorporated mortality prediction model by machine learning algorithms, with satisfactory discrimination and medical interpretability. Further prospective clinical studies are required to demonstrate the value of time dimension in clinical practice.

Abbreviations

ICU
Intensive Care Unit
OD
Organ Dysfunction
MODS
Multiple Organ Dysfunction Syndrome
SOFA
Sequence Organ Failure Assessment
MIMIC
Medical Information Mart for Intensive Care
eICU-CRD
eICU Collaborative Research Database
OFTI
Organ dysFunction Time Index
OFuT
Organ dysFunction unit Time
OUTI
Organ dysfunction Unalleviated Time Index
OUuT
Organ dysfunction Unalleviated unit Time
OATI
Organ dysfunction Aggravated Time Index
OAuT
Organ dysfunction Aggravated unit Time
STAUC

SOFA-time Area Under Curve
LR
Logistic Regression
XGBoost
eXtreme Gradient Boosting
SVM
Support Vector Machine
RF
Random Forest
AUROC
Area Under the receiver operating characteristic Curve
ROC
Receiver Operating characteristic Curve
SHAP
SHapley Additive exPlanations
LOCF
Last Observation Carried Forward
BMI
Body Mass Index
GCS
Glasgow Coma Scale
Std
Standard deviation
ASC
Absolute Sum of changes
var
variance
skew
Skewness
RASS
Richmond Agitation-Sedation Scale

Declarations

Acknowledgements: The authors thank MedicalSystem Company of Suzhou to provide technical support for SIUC-JL dataset development and clinical data collection. The authors also thank the eICU Collaborative Research Database and the Medical Information Mart for Intensive Care-III/IV database for the data provided.

Funding: This study was supported by the National Natural Science Foundation of China (82070665).

Availability of data: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions: Ke L and Li WQ contributed to the conception of the research; Liu Y and Gao K contributed to the design of the research; Deng HB contributed to the acquisition of the open-source data; Shi YH and Gao Y contributed to the construction of machine learning algorithms; Ling T and Gao K contributed to the analysis and interpretation of the data; and Ke L, Liu Y and Lin JJ drafted the manuscript. Li WQ, Gao Y, Liu YX, Tong ZH, YuXQ, Bo XW, ZhouJ, GaoL, WangP, critically revised the manuscript. All authors agree to be fully accountable for ensuring the integrity and accuracy of the work, and read and approved the final manuscript.

Ethics approval and consent to participate

This study was an analysis of the open-source anonymized databases and local databases(2019NZKY-003-03) with preexisting IRB approval.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Supplementary Information

Supplemental digital content is available for this article.

References

1. Nates JL, Nunnally M, Kleinpell R, Blosser S, Goldner J, Birriel B, Fowler CS, Byrum D, Miles WS, Bailey H *et al*: **ICU Admission, Discharge, and Triage Guidelines: A Framework to Enhance Clinical Operations, Development of Institutional Policies, and Further Research.** *Crit Care Med* 2016, **44**(8):1553–1602.
2. El-Menyar A, Thani HA, Zakaria ER, Zarour A, Tuma M, AbdulRahman H, Parchani A, Peralta R, Latifi R: **Multiple Organ Dysfunction Syndrome (MODS): Is It Preventable or Inevitable?** *International Journal of Clinical Medicine* 2012, **03**(07):722–730.
3. Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ: **Multiple Organ Dysfunction Score: A reliable descriptor of a complex clinical outcome.** *Critical Care Medicine* 1995, **23**(10):1638–1652.
4. Vincent J-L, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart C, Suter P, Thijs LG: **The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure.** In.: Springer-Verlag; 1996.

5. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M *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):762-774.
6. Karakike E, Kyriazopoulou E, Tsangaris I, Routsis C, Vincent J-L, Giamarellos-Bourboulis EJ: **The early change of SOFA score as a prognostic marker of 28-day sepsis mortality: analysis through a derivation and a validation cohort**. *Critical Care* 2019, **23**(1).
7. Schwalbe N, Wahl B: **Artificial intelligence and the future of global health**. *The Lancet* 2020, **395**(10236):1579–1586.
8. Alistair, E., W., Johnson, Mohammad, M., Ghassemi, Shamim, Nemati, Katherine: **Machine Learning and Decision Support in Critical Care**. *Proceedings of the IEEE* 2016.
9. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG: **MIMIC-III, a freely accessible critical care database**. *Sci Data* 2016, **3**:160035.
10. Johnson A B, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R.: **MIMIC-IV (version 0.4)**. In. *PhysioNet*. 2020.
11. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O: **The eICU Collaborative Research Database, a freely available multi-center database for critical care research**. *Sci Data* 2018, **5**:180178.
12. Christ M BN, Neuffer J, Kempa-Liehr AW.: **Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)**. *Neurocomputing*. 2018:307:372–377.
13. Minne L, Abu-Hanna A, de Jonge E: **Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review**. *Crit Care* 2008, **12**(6):R161.
14. Chen T, Guestrin C: **XGBoost: A Scalable Tree Boosting System**. In: *the 22nd ACM SIGKDD International Conference*. 2016; 2016.
15. Zhang H GCC: **Support Vector Machines versus Boosting**. 2016.
16. Bishop CM: **Pattern recognition and machine learning**: springer; 2006.
17. Thorsen-Meyer H-C, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, Strøm T, Chmura PJ, Heimann M, Dybdahl L *et al*: **Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records**. *The Lancet Digital Health* 2020, **2**(4):e179-e191.
18. Lachin JM: **Fallacies of last observation carried forward analyses**. *Clinical trials* 2016, **13**(2):161–168.
19. Demler OV, Pencina MJ, D'Agostino RB, Sr.: **Misuse of DeLong test to compare AUCs for nested models**. *Stat Med* 2012, **31**(23):2577–2587.
20. Kwon YS, Baek MS: **Development and Validation of a Quick Sepsis-Related Organ Failure Assessment-Based Machine-Learning Model for Mortality Prediction in Patients with Suspected Infection in the Emergency Department**. *J Clin Med* 2020, **9**(3).
21. Aperstein Y, Cohen L, Bendavid I, Cohen J, Grozovsky E, Rotem T, Singer P: **Improved ICU mortality prediction based on SOFA scores and gastrointestinal parameters**. *PLoS One* 2019, **14**(9):e0222599.

22. Aşuroğlu T, Oğul H: **A deep learning approach for sepsis monitoring via severity score estimation.** *Computer Methods and Programs in Biomedicine* 2021, **198**.
23. Grissom CK, Brown SM, Kuttler KG, Boltax JP, Jones J, Jephson AR, Orme JF: **A modified sequential organ failure assessment score for critical care triage.** *Disaster medicine and public health preparedness* 2010, **4(4):277–284**.
24. Elias A, Agbarieh R, Saliba W, Khoury J, Bahouth F, Nashashibi J, Azzam ZS: **SOFA score and short-term mortality in acute decompensated heart failure.** *Scientific Reports* 2020, **10(1)**.
25. Chen FC, Kung CT, Cheng HH, Cheng CY, Tsai TC, Hsiao SY, Su CM: **Quick Sepsis-related Organ Failure Assessment predicts 72-h mortality in patients with suspected infection.** *Eur J Emerg Med* 2019, **26(5):323–328**.
26. Cardoso FS, Abrales JG, Sy E, Ronco JJ, Bagulho L, McPhail MJ, Karvellas CJ: **Lactate and number of organ failures predict intensive care unit mortality in patients with acute-on-chronic liver failure.** *Liver Int* 2019, **39(7):1271–1280**.
27. Jentzer JC, Bennett C, Wiley BM, Murphree DH, Keegan MT, Barsness GW: **Predictive value of individual Sequential Organ Failure Assessment sub-scores for mortality in the cardiac intensive care unit.** *PLoS One* 2019, **14(5):e0216177**.
28. Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, Pilcher DV, Australian, New Zealand Intensive Care Society Centre for O, Resource E: **Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit.** *JAMA* 2017, **317(3):290–300**.
29. Lambden S, Laterre PF, Levy MM, Francois B: **The SOFA score-development, utility and challenges of accurate assessment in clinical trials.** *Crit Care* 2019, **23(1):374**.
30. Vasilevskis EE, Pandharipande PP, Graves AJ, Shintani A, Tsuruta R, Ely EW, Girard TD: **Validity of a Modified Sequential Organ Failure Assessment Score Using the Richmond Agitation-Sedation Scale.** *Crit Care Med* 2016, **44(1):138–146**.
31. **Gastrointestinal symptoms during the first week of intensive care are associated with poor outcome: a prospective multicentre study.** *Intensive care medicine* 2013, **39(5):899-909**.
32. Schlapbach LJ, Straney L, Bellomo R, MacLaren G, Pilcher D: **Prognostic accuracy of age-adapted SOFA, SIRS, PELOD-2, and qSOFA for in-hospital mortality among children with suspected infection admitted to the intensive care unit.** *Intensive care medicine* 2018, **44(2):179–188**.
33. Matics TJ, Sanchez-Pinto LN: **Adaptation and validation of a pediatric sequential organ failure assessment score and evaluation of the sepsis-3 definitions in critically ill children.** *JAMA pediatrics* 2017, **171(10):e172352-e172352**.

Figures

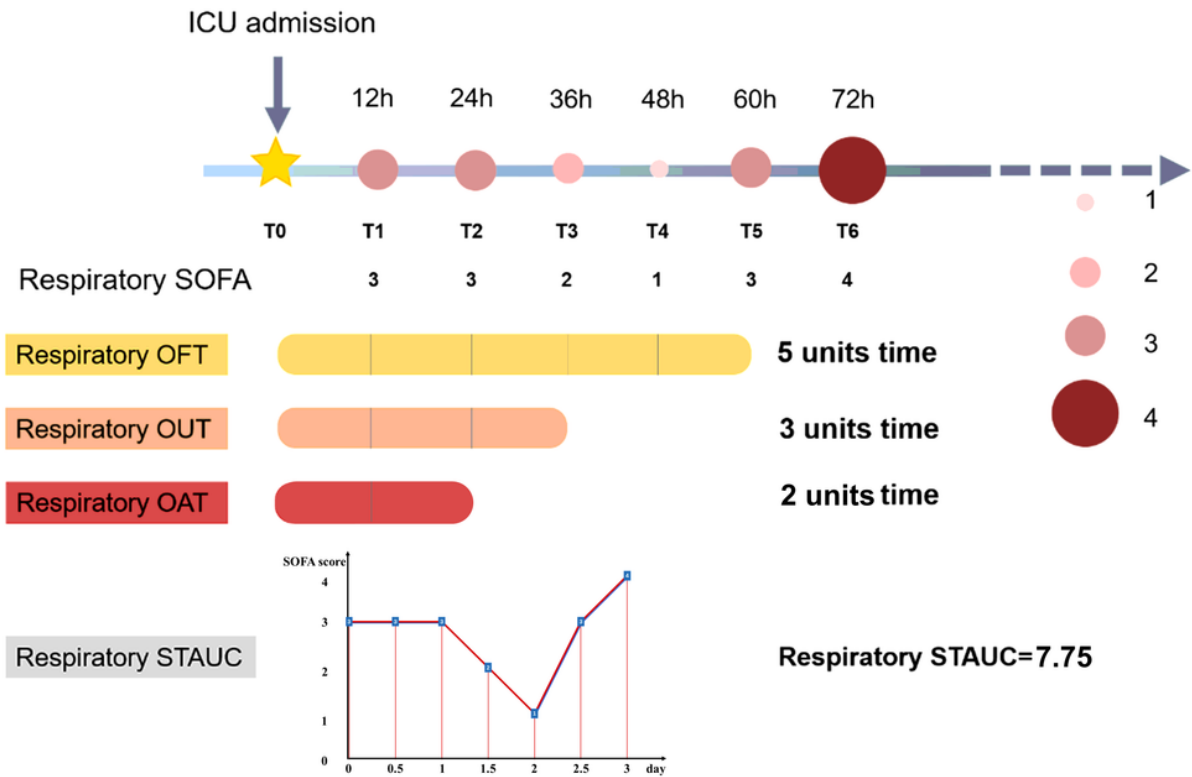


Figure 1

Diagram of the time dimensional characteristics associated with calculating the respiratory score in the instance.

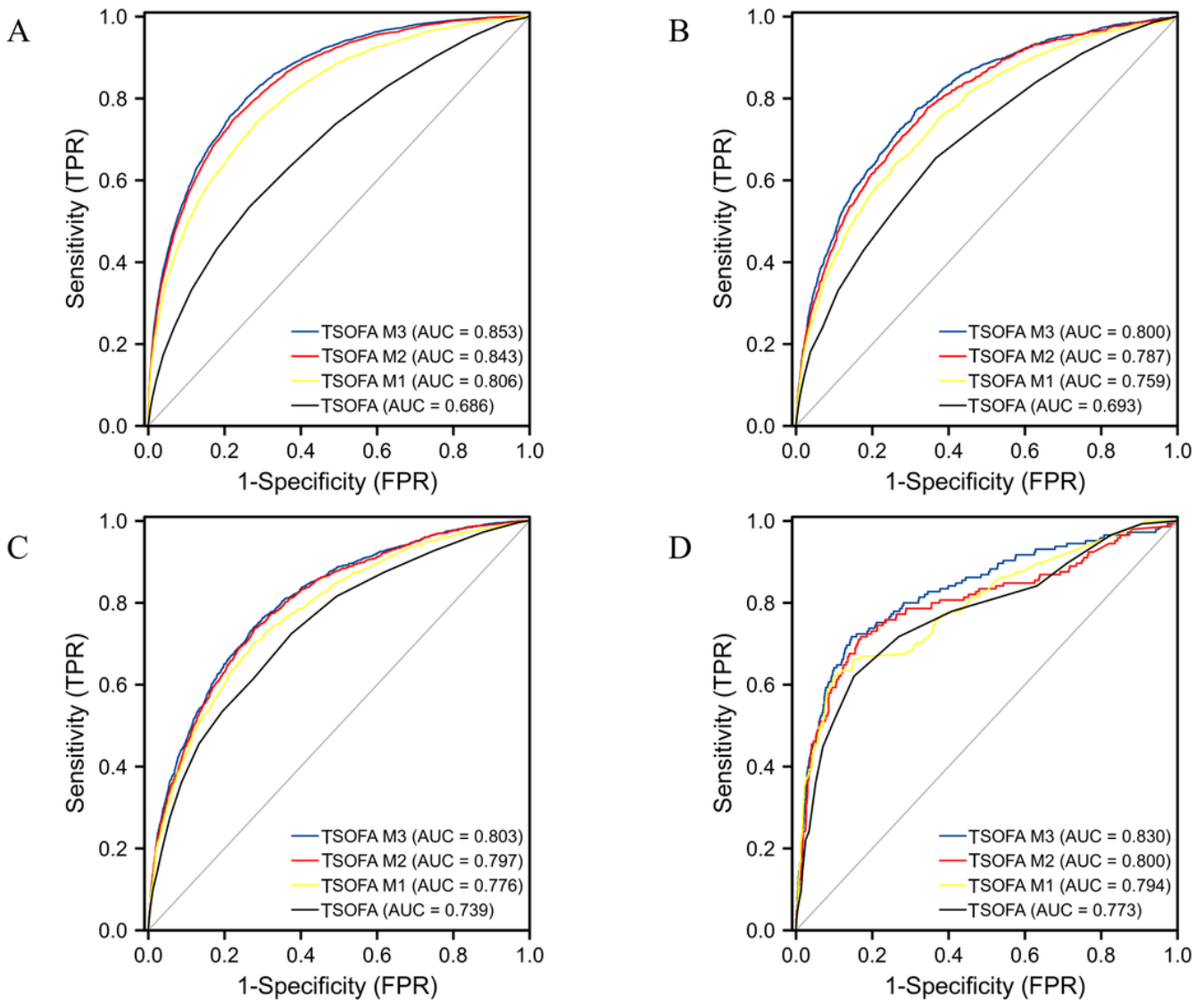


Figure 2

Receiver operating characteristic (ROC) curves for the machine learning models and the original SOFA score in predicting ICU mortality in different datasets: A. Training set; B. Validation set; C. Test set A; D. Test set B

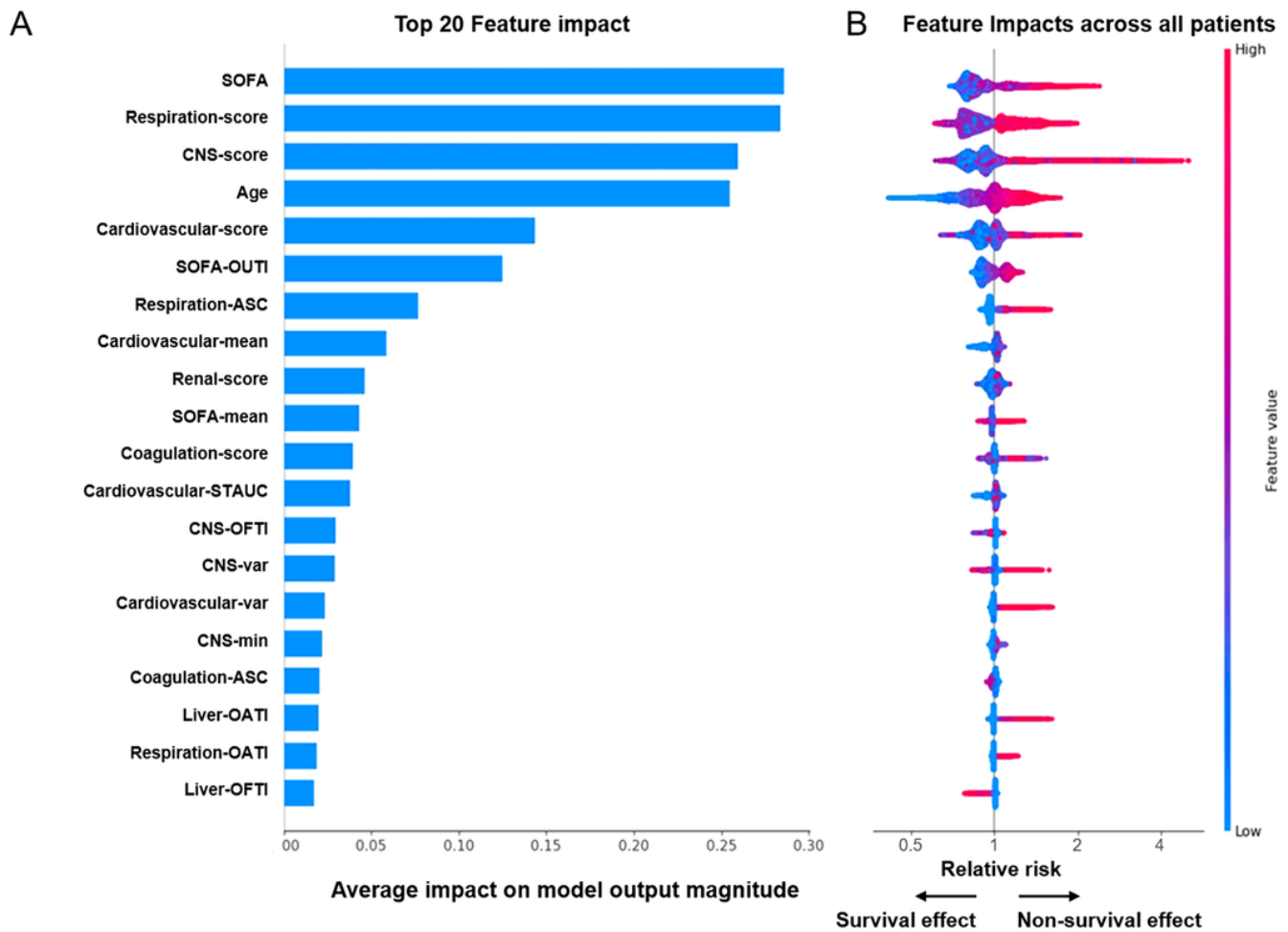


Figure 3

The impact of the input features on predictions. (A) Top 20 input features with impact on mortality prediction. (B) The relationship between each feature and the ICU mortality, with the colors reflecting the magnitude of each feature value (high in red, low in blue), with the points to the left of the central axis representing features that move the prediction towards survival and the points to the right representing features that drive the prediction in a non-survival direction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)