



Enhancing Sepsis prognosis: Integrating social determinants and demographic variables into a comprehensive model for critically ill patients

Elie Sarraf^a, Alireza Vafaei Sadr^b, Vida Abedi^b, Anthony S. Bonavia^{c,*}

^a Department of Anesthesiology and Perioperative Medicine, Penn State Milton S Hershey Medical Center, Hershey, PA 17033, USA

^b Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA 17033, USA

^c Division of Critical Care Medicine, Department of Anesthesiology and Perioperative Medicine, Penn State Milton S Hershey Medical Center, Hershey, PA 17033, USA

ARTICLE INFO

Keywords:

Sepsis
Machine learning
Sequential organ failure assessment score
MIMIC database
Mortality

ABSTRACT

Background: The Sequential Organ Failure Assessment (SOFA) score monitors organ failure and defines sepsis but may not fully capture factors influencing sepsis mortality. Socioeconomic and demographic impacts on sepsis outcomes have been highlighted recently.

Objective: To evaluate the prognostic value of SOFA scores against demographic and social health determinants for predicting sepsis mortality in critically ill patients, and to assess if a combined model increases predictive accuracy.

Methods: The study utilized retrospective data from the MIMIC-IV database and prospective external validation from the Penn State Health cohort. A Random Forest model incorporating SOFA scores, demographic/social data, and the Charlson Comorbidity Index was trained and validated.

Findings: In the MIMIC-IV dataset of 32,970 sepsis patients, 6,824 (20.7%) died within 30 days. A model including demographic, socioeconomic, and comorbidity data with SOFA scores improved predictive accuracy beyond SOFA scores alone. Day 2 SOFA, age, weight, and comorbidities were significant predictors. External validation showed consistent performance, highlighting the importance of delta SOFA between days 1 and 3.

Conclusion: Adding patient-specific demographic and socioeconomic information to clinical metrics significantly improves sepsis mortality prediction. This suggests a more comprehensive, multidimensional prognostic approach is needed for accurate sepsis outcome predictions.

List of abbreviations

AUROC	area under the receiver operating characteristic curve
CCI	Charlson comorbidity index
ICU	intensive care unit
MIMIC	Medical Information Mart for Intensive Care
NPV	negative predictive value
PPV	positive predictive value
PSH	Penn State Health
SOFA	sequential (or sepsis-associated) organ failure assessment

1. Introduction

The Sepsis-related Organ Failure Assessment (SOFA) score, initially devised by Vincent et al. [1], was designed to understand the progression and impact of organ dysfunction in sepsis, complementing other

severity scores. Later known as the ‘Sequential’ Organ Failure Assessment score, it was incorporated into the diagnostic criteria for sepsis (Sepsis-3) in 2016 [2]. The SOFA score provides an objective measure of sequential organ dysfunction or recovery during a patient’s hospitalization, contrasting with other scoring systems like the APACHE II score [3]. It relies heavily on laboratory data, which may not always reflect the current clinical condition of critically ill patients. Additionally, among the six SOFA domains, four hinge on laboratory results, potentially delaying the identification of critical changes in a patient’s condition. Although other severity-of-illness scores exist, no scoring system exclusively predicts mortality in sepsis [4,5]. Recognizing its limitations, a review is currently underway to update and improve the SOFA score for more accurate, sepsis-specific applications [6].

Recent literature suggests that socioeconomic status and demographic background may play a more definitive role in the outcomes of sepsis than previously recognized [7–10]. In fact, artificial intelligence

* Corresponding author.

E-mail address: abonavia@pennstatehealth.psu.edu (A.S. Bonavia).

<https://doi.org/10.1016/j.jcrc.2024.154857>

Available online 11 July 2024

0883-9441/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

has facilitated data-driven approaches to uncovering existing healthcare disparities [11]. These social determinants of health (SDoH) may provide immediate, specific insight into patient risk without the need for expensive tests. Thus, our study sought to assess the predictive power of SOFA scores relative to demographic and socioeconomic factors for sepsis prognosis. We hypothesized that a combined approach using SOFA scores measured on the days following sepsis onset, together with social health determinants, could offer a more accurate prediction of mortality.

2. Methods

We accessed anonymized patient data from the Medical Information Mart for Intensive Care (MIMIC)-IV database (version 2.2, Jan 6 2023), which includes critical care data from Beth Deaconess Medical Center [12]. This resource was chosen for its comprehensive data and patient mortality information. Using existing code [13], we calculated SOFA scores for patients fitting Sepsis-3 criteria [2].

From the onset of suspected infection, we tracked the highest daily SOFA score and its change between each day of critical illness. Additionally, we computed daily changes in SOFA (delta SOFA) and cumulative daily SOFA scores (sum SOFA). This method ensured data

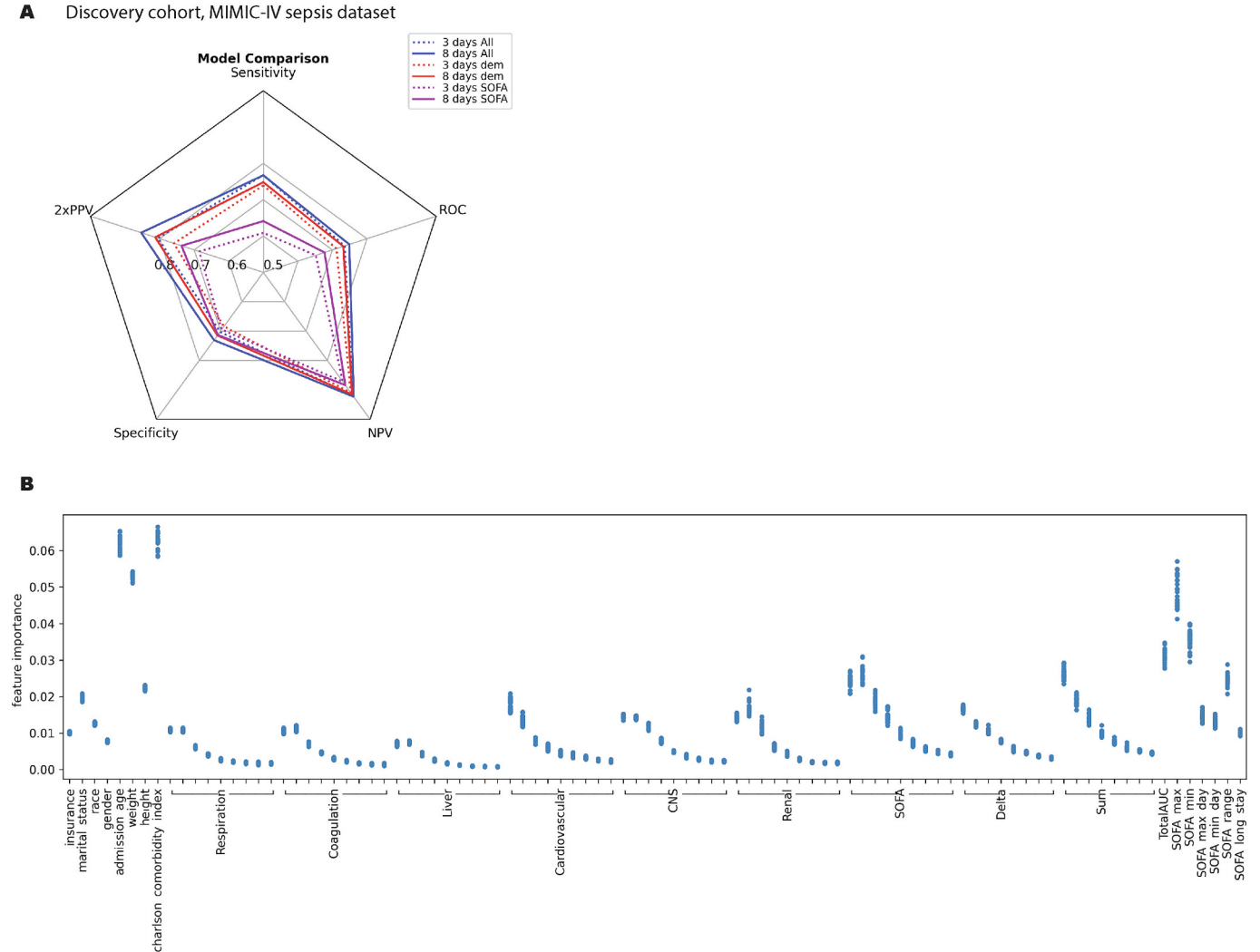


Fig. 1. Comparative metrics and feature importance for 30-day mortality prediction based on parameters derived from Medical Information Mart for Intensive Care (MIMIC)-IV dataset. (A) This spider plot illustrates the efficacy of different predictive models over a span of 3 and 8 days following a sepsis diagnosis. ‘SOFA’ represents models utilizing exclusively SOFA-related parameters, including total SOFA score, its daily change (delta), cumulative sum, and AUROC values. ‘Dem’ extends the ‘SOFA’ model by incorporating demographic and socioeconomic factors such as insurance status, marital status, ethnicity, age, gender, body weight, height, and the Charlson comorbidity index. ‘All’ encompasses all ‘Dem’ variables plus daily, organ-specific scores (e.g., day-to-day SOFA respiratory scores). Metrics of model performance include positive predictive value (PPV), negative predictive value (NPV), and the receiver operating characteristic (ROC) curve area. (B) This figure displays the relative feature importance of various parameters in predicting 30-day mortality following sepsis. The x-axis is segmented by square brackets indicating an 8-day timeline post-sepsis onset, with day 1 at the bracket’s left end progressing to day 8 at the right. The parameters include daily SOFA component scores for respiration, coagulation, liver function, cardiovascular stability, central nervous system (CNS) activity, and renal performance. ‘Delta SOFA’ quantifies the day-over-day variation in SOFA scores, while ‘sum SOFA’ aggregates SOFA scores over two consecutive days. ‘Total AUC’ represents the cumulative SOFA score up to the current day. ‘SOFA max’ and ‘SOFA max day’ denote the peak SOFA score recorded for a patient and the specific day it was registered. Conversely, ‘SOFA min’ and ‘SOFA min day’ indicate the lowest SOFA score and its corresponding day. ‘SOFA range’ is the time span between the highest and lowest SOFA scores, providing a measure of fluctuation in organ function. ‘SOFA long stay’ tallies the total days a patient’s SOFA scores are monitored within the ICU, reflecting the duration of critical care received.

completeness, with no need for imputation except when patients were discharged from the ICU. We also gathered demographic and sociological data, specifically gender, age, weight, height, insurance status (Medicare, Medicaid, private insurance), marital status and race as well as the Charlson Comorbidity Index (CCI, [14]), focusing on 30-day mortality as the primary outcome.

Using Python’s *scikit-learn* library (version 1.3.2) [15], we trained a Random Forest model, evaluating it through 5-fold cross-validation. We tested six model types, including (1) models that were trained on 3-day and 8-day retrospective data, (2) models with and without daily SOFA organ component measures, and (3) models with and without patient socioeconomic, demographic and comorbidity data. Metrics such as feature importance, PPV, NPV, sensitivity, specificity, and AUROC were assessed during each iteration.

We then externally validated our best predictive model with real-world, observational data from critically ill patients having sepsis and forming part of a prospective research cohort at Penn State Health (PSH) from Aug 2020 – Feb 2024. Eligible participants were adults over 18 years, identified within 48 h of critical illness onset. Sepsis was defined according to Sepsis-3 criteria, and critical illness required continuous intravenous vasopressors or ongoing respiratory support. Patients with active hematologic malignancies or those on immune-altering therapies

were excluded.

Given that the validation data set included SOFA scores measured every other day, the original model was re-tuned to only use odd-numbered days. We also compared feature importance between a streamlined dataset for a Random Forest model with both the MIMIC and PSH data. Our analysis code is available on Github, and the study adhered to TRIPOD guidelines for predictive model reporting [16].

3. Results

Of 299,712 unique patients included in the MIMIC-IV database, 32,970 had sepsis. Among these, 57.8% were male and the mean age was 66.7 ± 16 years. Supplemental Table 1 details the SDoH factors included in the study. A total of 6,824 patients (20.7%) died within 30 days of the sepsis onset.

For each of the six scenarios described above, 25 different machine learning models were trained. Fig. 1A describes five metrics pertaining to the model having the highest comparative AUROC (75%, with 95% CI 73–77%). This model demonstrated 77% sensitivity (95% CI 75–77%), 74% accuracy (95% CI 72–76%) and a precision of 43% (95% CI 39–47%). The most effective model combined demographic, socioeconomic, and comorbidity data with total SOFA score and SOFA organ

A Validation cohort, Penn State Health sepsis dataset

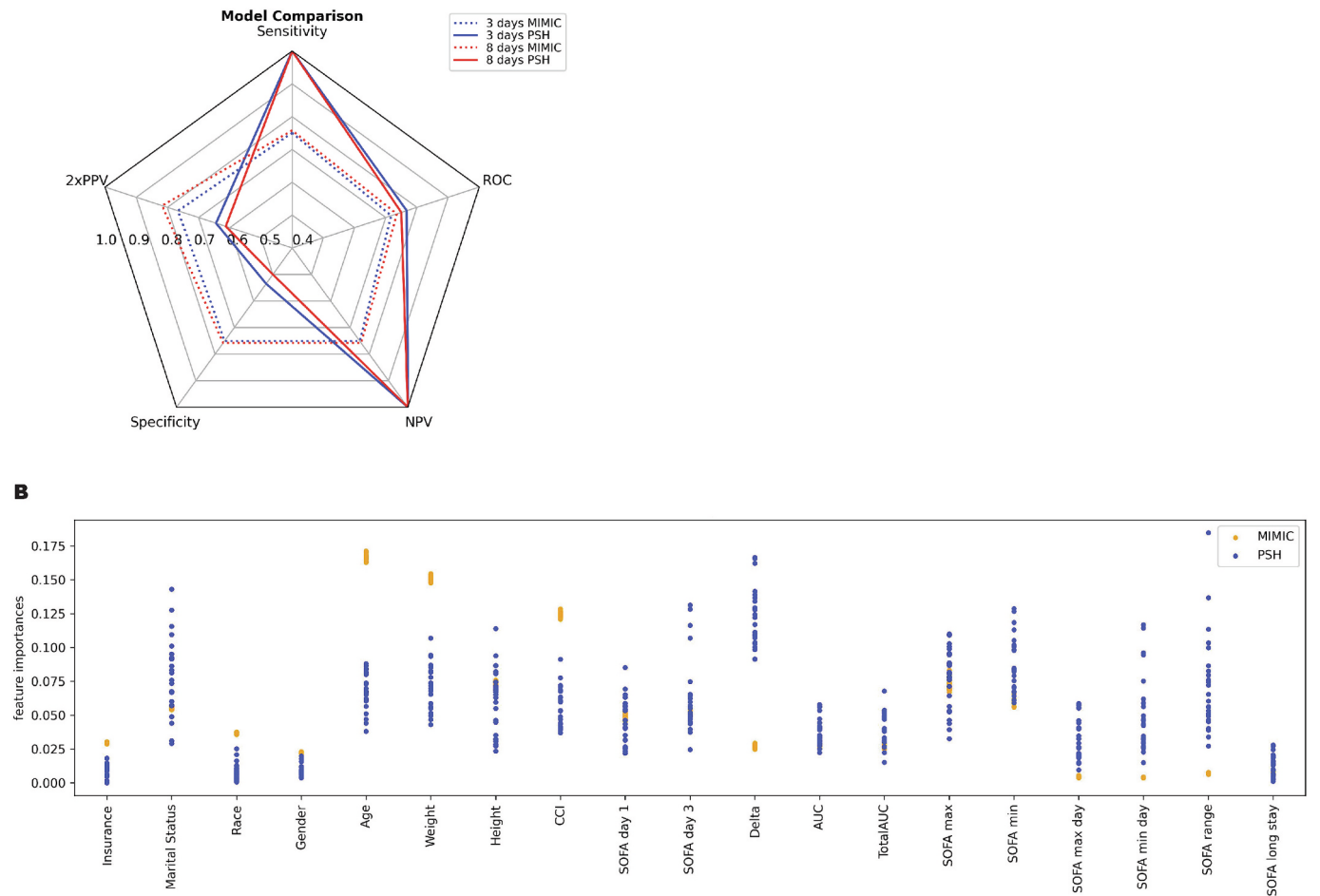


Fig. 2. External validation of 30-day mortality prediction model, utilizing data from septic patients treated at Penn State Health. (A) This spider plot compares the performance of the predictive model trained on the MIMIC dataset on a validation subset of MIMIC (‘MIMIC’) versus the external validation data obtained from Penn State Electronic Medical Record (‘PSH’). All abbreviations are defined in the legend for Fig. 1. (B) This figure compares the relative feature importance in predicting 30-day mortality following sepsis for MIMIC (blue) and PSH (orange). We extracted patient characteristics and clinical variables from the first three days of illness onset alone. To minimize the risk of model overfitting, we streamlined the number of features following the initial discovery phase.

component measures over an 8-day period post-sepsis. However, adding just demographic, socioeconomic, and comorbidity data to the total SOFA score (i.e., without individual organ dysfunction measures) significantly enhanced 30-day mortality prediction as compared with total SOFA score alone. This improvement was evident in analyses conducted at both 3 and 8 days after sepsis onset.

With respect to feature importance, SOFA organ component scores measured on day 2 were most predictive of 30-day mortality (Fig. 1B). Among organ systems affected by sepsis, cardiovascular dysfunction, renal dysfunction, and central nervous system dysfunction were most predictive of 30-day mortality. Equally important was the relative feature importance of age, weight, height, marital status and CCI as compared with organ-specific or overall SOFA measures (Fig. 1B).

Figure 2A illustrates the results of external validation using prospective data from 105 septic, critically ill patients undergoing care at PSH. In this data set, 52.4% of patients were male, with a mean age was 66.5 ± 15 years. The microbial sources of sepsis in this cohort are enumerated in Supplementary Table 2. Nineteen patients (18.1%) died within 30 days. When individual metrics were compared between discovery and validation cohorts, there was a noted increase in sensitivity and negative predictive value, and to a minor degree AUROC, at the expense of a lower specificity and positive predictive value. The negative predictive value of the model derived from MIMIC data appeared to decrease when using data derived from every other day of critical illness. Otherwise, the model's performance remained consistent.

Interestingly, in the streamlined model, age, weight, height and CCI continued to comprise the most important features (Fig. 2B). The importance of marital status was variable as compared with results derived from the MIMIC model, while 'delta SOFA' between days 1–3 of critical illness proved to be far better correlated with mortality in the PSH data set.

4. Discussion

Historically, a higher SOFA score, signaling severe organ failure, has been closely linked with poorer clinical outcomes [17,18]. However, our analysis proposes a paradigm shift. Our pivotal finding is the pronounced impact of patient-specific and social risk factors on 30-day mortality, overshadowing the predictive relevance of organ dysfunction severity [10].

Our data indicate that, by day 2 of sepsis, certain organ dysfunction measures can predict mortality with some degree of reliability. This finding is consistent with previous literature that indicates enhanced prognostic value of SOFA score when measured later in a patient's hospitalization [19,20]. Yet, it is the integration of a patient's age, weight, height, marital status and comorbidity profile (as encapsulated by the Charlson Comorbidity Index) that significantly amplifies prognostic precision as compared with the SOFA score alone or any of its individual organ components. By Day 8, models enriched with these variables not only sustained but also enhanced the predictive accuracy of mortality, with model performance metrics surpassing those based solely on clinical measures.

The inclusion of individual organ dysfunction parameters only modestly improved the model's performance, reinforcing the premise that, while clinical measures of organ dysfunction are not to be overlooked, they are evidently less predictive of patient outcomes compared to SDoH. This is a critical observation, as it underscores the limitations of current clinical-only prognostic models and highlights the potential for improved risk stratification through the incorporation of other, readily available, patient-specific factors.

We externally validated the best model generated from publicly available data on a subset of patients receiving health care at our institution. We found that, while the predictive model remained stable, delta SOFA between days 1 and 3 of critical illness was a far stronger predictor of mortality in our patient population. The prognostic utility of delta SOFA has been previously reported [21], although it is not

currently in widespread clinical use.

Past efforts to improve on the SOFA score have produced mixed results [22–24]. However, our research supports these endeavors, suggesting that an enhanced SOFA score incorporating SDoH could sharpen the predictive accuracy for mortality in sepsis patients. This augmented model may be particularly beneficial for patients who, due to cognitive impairments caused by their illness, are unable to provide a detailed medical history. Given the large regional variability in socioeconomic status and ethnic diversity, it may also be necessary to adjust the model to reflect this local variability. Methods such as federated learning can be utilized for this purpose [25].

Although our study offers valuable insights, it is important to acknowledge its limitations. A significant constraint was that our validation dataset recorded SOFA scores every 48 h instead of every 24 h. This less frequent recording could have affected the accuracy of validation, given that peak SOFA scores usually manifest within the first 24 h of sepsis onset. Furthermore, our exclusive use of the Sepsis-3 diagnostic criteria might not encompass the full range of clinical presentations and outcomes. This is particularly relevant since other frameworks, like Sepsis-2 [26], are still widely used in clinical practice. Finally, as the validation cohort was small, with minimal numbers for many subgroups, one must be cautious in interpreting the data in a general manner; rather, it should be approached as a preliminary glance at the model performance. Further external validation, using diverse populations, a community healthcare setting and temporal separation of development/validation datasets are underway to address this limitation.

Our study's findings have implications beyond the immediate scope of sepsis and can be applied to various models across different medical domains [27,28]. Integrating SDoH and demographic data with clinical scoring systems can enhance predictive accuracy and provide a more holistic understanding of patient outcomes. This approach is particularly relevant in chronic diseases such as diabetes and heart failure, where socioeconomic factors profoundly impact disease progression and mortality. Moreover, incorporating SDoH into predictive models for infectious diseases like COVID-19 can improve resource allocation and patient management in pandemics, highlighting the broader applicability of our findings across diverse medical fields.

In conclusion, our investigation underscores the imperative to revisit the prognostic frameworks for sepsis. It is evident that a multidimensional approach, encompassing both clinical and non-clinical factors, is crucial for a more accurate prediction of outcomes. Our work contributes to the growing body of evidence that supports the integration of broader patient information, extending beyond the confines of physiological and laboratory measures, into prognostic models for sepsis.

Ethics approval and consent to participate

De-identified and publicly-available human data was accessed via Physionet [29] and utilized in accordance with PhysioNet Credentialed Health Data Use Agreement 1.5.0. Prospective validation data was collected in accordance with the Penn State Human Subjects Protection Office (IRB study# 15328, approved 7/27/2020).

Consent for publication

Not applicable.

Funding

Funding was provided by the National Institute of General Medical Sciences R35GM150695 (ASB), endowment funds from the Department of Anesthesiology at Penn State Health (ES), and start-up funds from the Department of Public Health Sciences (VA).

Authors' contributions

ES contributed to data curation and preprocessing. ES and AVS were responsible for the development and implementation of the machine learning algorithms and performed the data analysis. VA provided critical feedback, and helped shape the research, analysis, and manuscript. ASB conceptualized the study, assisted in the interpretation of the computational data analysis, supervised the project, and led manuscript writing. All authors read and approved the final manuscript.

CRediT authorship contribution statement

Elie Sarraf: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Alireza Vafaei Sadr:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis. **Vida Abedi:** Writing – review & editing, Writing – original draft, Supervision, Methodology. **Anthony S. Bonavia:** Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests.

Availability of data and materials

The datasets analyzed during the current study are available in the MIMIC-IV version 2.2 database repository, available at <https://physionet.org/content/mimiciv/2.2>. Validation data from PSH will be made available from the investigator on reasonable request. The code used to analyze MIMIC-IV data is publicly available and can be accessed here: <https://github.com/rasman/SOFA>.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jcrc.2024.154857>.

References

- [1] Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the working group on Sepsis-related problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22(7):707–10.
- [2] 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.
- [3] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13(10):818–29.
- [4] 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.
- [5] Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3 – from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* 2005;31(10):1345–55.
- [6] Moreno R, Singer M, Rhodes A. Why the sequential organ failure assessment score needs updating? *Crit Care Sci* 2024;36. e20240296en.
- [7] Minejima E, Wong-Beringer A. Impact of socioeconomic status and race on Sepsis epidemiology and outcomes. *J Appl Lab Med* 2021;6(1):194–209.
- [8] Galiatsatos P, Brigham EP, Pietri J, Littleton K, Hwang S, Grant MC, et al. The effect of community socioeconomic status on sepsis-attributable mortality. *J Crit Care* 2018;46:129–33.
- [9] Galiatsatos P, Follin A, Alghanim F, Sherry M, Sylvester C, Daniel Y, et al. The association between neighborhood socioeconomic disadvantage and readmissions for patients hospitalized with Sepsis. *Crit Care Med* 2020;48(6):808–14.
- [10] Galiatsatos P, Kachalia A, Belcher HME, Hughes MT, Kahn J, Rushton CH, et al. Health equity and distributive justice considerations in critical care resource allocation. *Lancet Respir Med* 2020;8(8):758–60.
- [11] Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? *AMA J Ethics* 2019;21(2):E167–79.
- [12] Johnson AE, Stone DJ, Celi LA, Pollard TJ. The MIMIC code repository: enabling reproducibility in critical care research. *J Am Med Inform Assoc* 2018;25(1):32–9.
- [13] Johnson A, Pollard T, Blundell J, Chahin A, Erinhong Gow B, Schubert M, et al. MIT-LCP/mimic-code: MIMIC Code v2.4.0. In: v2.4.0 edn; Zenodo; 2023.
- [14] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373–83.
- [15] Pedregosa F, Varoquaux G, Gramfort A, Vincent M, Bertrand T, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [16] Patzer RE, Kaji AH, Fong Y. TRIPOD reporting guidelines for diagnostic and prognostic studies. *JAMA Surg* 2021;156(7):675–6.
- [17] Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. 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.
- [18] Falcao ALE, Barros AGA, Bezerra AAM, Ferreira NL, Logato CM, Silva FP, et al. The prognostic accuracy evaluation of SAPS 3, SOFA and APACHE II scores for mortality prediction in the surgical ICU: an external validation study and decision-making analysis. *Ann Intensive Care* 2019;9(1):18.
- [19] Moreno R, Vincent JL, Matos R, Mendonca A, Cantraine F, Thijs L, et al. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Working group on Sepsis related problems of the ESICM. *Intensive Care Med* 1999;25(7):686–96.
- [20] Vincent JL, de Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. *Crit Care Med* 1998;26(11):1793–800.
- [21] de Grooth HJ, Geenen IL, Girbes AR, Vincent JL, Parienti JJ, Oudemans-van Straaten HM. SOFA and mortality endpoints in randomized controlled trials: a systematic review and meta-regression analysis. *Crit Care* 2017;21(1):38.
- [22] Arakawa M, Levy JH, Fujimori K, Kondo K, Iba T. A new SOFA score calculation to improve the predictive performance for mortality in sepsis-associated disseminated intravascular coagulopathy patients. *J Crit Care* 2021;64:108–13.
- [23] Lee HJ, Ko BS, Ryoo SM, Han E, Suh GJ, Choi SH, et al. Modified cardiovascular SOFA score in sepsis: development and internal and external validation. *BMC Med* 2022;20(1):263.
- [24] Liu Y, Gao K, Deng H, Ling T, Lin J, Yu X, et al. A time-incorporated SOFA score-based machine learning model for predicting mortality in critically ill patients: a multicenter, real-world study. *Int J Med Inform* 2022;163:104776.
- [25] Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020;10(1):12598.
- [26] Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS international Sepsis definitions conference. *Crit Care Med* 2003;31(4):1250–6.
- [27] Pappas G, Queen S, Hadden W, Fisher G. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl J Med* 1993;329(2):103–9.
- [28] Stringhini S, Carmeli C, Jokela M, Avendaño M, Muennig P, Guida F, et al. Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women. *Lancet* 2017;389(10075):1229–37.
- [29] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000;101(23):E215–20.