PREDICTING MORTALITY IN INTENSIVE CARE UNITS FOR SEPSIS PATIENTS USING MACHINE LEARNING

CAMILO ANDRES SANTOS ORTIZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FÍSICO-MECÁNICA
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
MAESTRÍA EN INGENIERÍA DE TELECONUMICACIONES
BUCARAMANGA
2025

PREDICTING MORTALITY IN INTENSIVE CARE UNITS FOR SEPSIS PATIENTS USING MACHINE LEARNING

CAMILO ANDRES SANTOS ORTIZ

Thesis submitted in partial fulfillment of the requirements for the degree of Master in Telecommunications Engineering.

Advisor
Carlos A. Fajardo, Ph.D.

Co-advisor
Said Pertuz, Ph.D

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FÍSICO-MECÁNICA
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
MAESTRÍA EN INGENIERÍA DE TELECONUMICACIONES
BUCARAMANGA
2025

CONTENT

LIST OF FIGURES

LIST OF TABLES

ANNEXES

RESUMEN

TÍTULO PREDICCIÓN DE LA MORTALIDAD EN UNIDADES DE CUIDADOS INTENSIVOS PARA PACIENTES CON SEPSIS MEDIANTE APRENDIZAJE AUTOMÁTICO [*]

AUTOR: CAMILO ANDRES SANTOS ORTIZ [**]

PALABRAS CLAVE: Factores clínicos, eICU, UCI, MIMIC-IV, Predicción de mortalidad, Sepsis

DESCRIPCIÓN: La sepsis es una de las principales causas de mortalidad en la UCI en todo el mundo, y la identificación precoz de los pacientes de alto riesgo es fundamental. Sin embargo, la complejidad de la sepsis dificulta la predicción de la mortalidad. El objetivo de este estudio es identificar factores clínicos significativos con el fin de desarrollar modelos procesables para la predicción de la mortalidad en pacientes con sepsis en diversos entornos clínicos. Este estudio retrospectivo y anidado utilizó datos de 15.100 pacientes de UCI del conjunto de datos MIMIC-IV v3.0 para el entrenamiento del modelo y de 8.201 pacientes del conjunto de datos MIMIC-IV v3.0 para el entrenamiento del modelo. y 8.201 pacientes del conjunto de datos eICU v2.0 para la validación externa. Se identificó a los pacientes con sepsis según los criterios criterios de Sepsis-3. Se entrenaron ocho modelos de aprendizaje automático y se seleccionó el modelo de mejor rendimiento en función del AUC más alto. AUC. Los factores clínicos significativos se identificaron mediante odds ratios y odds ratios ajustados para evaluar su relevancia predictiva. El modelo de mejor rendimiento se obtuvo a partir del valor AUC más alto, alcanzando un AUC de 0,84 en el conjunto de datos MIMIC-IV y 0,75 en el conjunto de datos eICU. La selección secuencial de características redujo el modelo a 25 factores clínicos sin comprometer de DeLong (AUC de 0,84, valor p = 0,507). Además, el análisis de odds ratio y los parámetros de laboratorio como redictores significativos de mortalidad. El modelo de aprendizaje automático, aprovechando datos clínicos significativos como los signos vitales y los resultados de laboratorio, demuestra una gran capacidad de predicción de la mortalidad en la UCI en pacientes con sepsis.

---

[*] Tesis

[**] Facultad de Ingeniería Físico-Mecánica. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunucaciones. Director: Carlos A. Fajardo, Ph.D. Codirector: Said Pertuz, Ph.D.

ABSTRACT

TITLE: PREDICTING MORTALITY IN INTENSIVE CARE UNITS FOR SEPSIS PATIENTS USING MACHINE LEARNING [*]

AUTOR: CAMILO ANDRES SANTOS ORTIZ [**]

Keywords: Clinical factors, eICU, ICU, MIMIC-IV, Mortality prediction, Sepsis

Description: Sepsis is a leading cause of ICU mortality worldwide, and early identification of high-risk patients is critical. However, the complexity of sepsis makes mortality prediction challenging. This study aims to identify significant linical factors to develop actionable models for robust mortality prediction in sepsis patients across diverse clinical settings. This retrospective, nested study utilized data from 15,100 ICU patients from the MIMIC-IV v3.0 dataset for model training and 8,201 patients from the eICU v2.0 dataset for external validation. atients were identified as having sepsis according to the Sepsis-3 criteria. Eight machine learning models were trained, and the top-performing model was selected based on the highest AUC. Significant clinical factors were identified using odds ratios and adjusted odds ratios to assess their predictive relevance. The best performing model was obtained from the highest AUC value, achieving an AUC of 0.84 on the MIMIC-IV dataset and 0.75 on the eICU dataset. Sequential Feature Selection reduced the model to 25 clinical factors without compromising predictive performance, as confirmed by the DeLong's test (AUC of 0.84, p-value=0.507). Furthermore, odds ratio analysis highlighted vital signs and laboratory parameters as significant predictors of mortality. The machine learning model, leveraging significant clinical data such as vital signs and laboratory results, demonstrates strong predictive capability for ICU mortality in sepsis patients.

---

[*] Thesis

[**] Facultad de Ingeniería Físico-Mecánica. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunucaciones. Advisor: Carlos A. Fajardo, Ph.D. Electronic Engineer. Co-advisor: Said Pertuz, Ph.D.

INTRODUCTION

Sepsis is defined as a potentially life-threatening systemic response of the body to infection. This response triggers wides-pread inflammation and causes tissue damage throughout the body, leading to *multi-organ dysfunction* and *septic shock* [1] [2]. Sepsis causes alterations in multiple systems and the activation of inflammatory markers, underscores its complexity and variability in Intensive Care Unit (ICU) environment.

0.1. The importance of mortality prediction in ICU sepsis patients

Sepsis represents a global health challenge, ranking among the leading causes of mortality worldwide [3] [4]. According to the World Health Organization, sepsis accounts for approximately 20 % of all global deaths [5]. Its rapid progression, often leading to fatal outcomes in hours, highlights the need for accurate mortality prediction to facilitate timely and effective treatment, thus improving patient outcomes.
However, predicting mortality in sepsis patients remains challenging. Clinical diagnosis is often hampered by atypical presentations, while ICU clinicians face the simultaneous management of multiple patients, especially in resource-constrained settings

---

[1] Mervyn Singer et al. «The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)». En: *JAMA* 315.8 (2016), págs. 801-810. DOI: 10.1001/jama.2016.0287.

[2] Cé Caraballo y Fabián Jaimes. «Organ Dysfunction in Sepsis: An Ominous Trajectory From Infection To Death». English. En: *The Yale journal of biology and medicine* 92.4 (2019), págs. 629-640.

[3] Konrad Reinhart et al. «Recognizing Sepsis as a Global Health Priority — A WHO Resolution». En: *New England Journal of Medicine* 377.5 (2017), págs. 414-417. DOI: 10.1056/NEJMp1707170.

[4] L Evans et al. «Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021., 2021, 47». En: *DOI: https://doi. org/10.1007/s00134-021-06506-y* (2021), págs. 1181-1247.

[5] World Health Organization (WHO). *Sepsis*. 2024.

[6]. This underscores the urgency for innovative approaches, such as leveraging machine learning models, to enhance diagnostic accuracy and clinical decision-making.

0.2. Scoring systems and sepsis diagnostic criteria

In routine clinical practice, structured scoring systems are employed to assess the severity of patients in ICUs and predict their prognosis, including mortality risk. Notably, many of these tools have been specifically tailored for sepsis by focusing on the early identification of organ dysfunction and the inflammatory cascade that characterizes septic patients. These tools are designed for rapid bedside application, prioritizing simplicity through models that utilize a limited number of factors with small integer coefficients, enhancing interpretability. Such scoring systems facilitate risk classification, anticipation of health outcomes, and optimization of clinical activities [7].

One of the most widely used scoring systems is the Sequential Organ Failure Assessment (SOFA) score, which was developed based on the guidelines of the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [1]. This score evaluates organ dysfunction and its progression in patients with sepsis, with the aim of predicting patient mortality. Its design specifically targets the early detection of multi-organ failure that is the hallmark of sepsis. The SOFA score is calculated considering the performance of various organ systems, such as respiratory, cardiovascular, hematological, hepatic, neurological and renal. A higher SOFA score

---

[6] Mahanazuddin Syed et al. «Application of Machine Learning in Intensive Care Unit (ICU) Settings Using MIMIC Dataset: Systematic Review». En: *INFORMATICS-BASEL* 8.1 (2021). DOI: 10.3390/informatics8010016.

[7] H. Mumtaz et al. «APACHE scoring as an indicator of mortality rate in ICU patients: a cohort study». En: *Annals of Medicine and Surgery* 85.3 (2023). Consultado el [fecha en que accediste al artículo], págs. 416-421. DOI: 10.1097/MS9.0000000000000264.

indicates a higher probability of mortality [8].

On the other hand, the Systemic Inflammatory Response Syndrome (SIRS) criteria was designed to identify the systemic inflammatory response. Objectively, it is defined by the presence of at least two of the following criteria: a) body temperature greater than 38 ℃ or less than 36 ℃, b) heart rate greater than 90 bpm, c) respiratory rate greater than 20 breaths per minute or partial CO_2 pressure less than 32 mmHg, d) leukocyte count greater than 12,000 or less than 4,000 cells/μL, or more than 10 % immature forms (bands) [9]. These criteria were originally established to capture the early inflammatory surge observed in sepsis, enabling prompt clinical intervention.

Similarly, there are other scoring systems for assessing critical patients, such as the Acute Physiology and Chronic Health Evaluation (APACHE), which is designed to estimate the risk of death in the ICU. It is routinely measured within the first 24 hours of admission and includes several clinical parameters, such as patient demographics, clinical characteristics, and laboratory results. Although APACHE is a general ICU scoring system, several of its variables, such as organ function and vital signs are directly applicable to the assessment of sepsis severity. The APACHE score ranges from 0 to 71, where lower scores (0–10) indicate a low risk of mortality, moderate scores (11–20) suggest an increased risk, and higher scores correspond to a significantly elevated probability of death. [10].

---

[8] E. E. Christensen et al. «Mortality and Sequential Organ Failure Assessment Score in Patients With Suspected Sepsis: The Impact of Acute and Preexisting Organ Failures and Infection Likelihood». En: *Critical Care Explorations* 5.2 (2023), e0865. DOI: 10.1097/CCE.0000000000000865.

[9] RK Chakraborty y B Burns. *Systemic Inflammatory Response Syndrome*. [Updated 2023 May 29]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK547669/. Treasure Island (FL): StatPearls Publishing, 2025.

[10] R Venkataraman et al. «Mortality Prediction Using Acute Physiology and Chronic Health Evaluation II and Acute Physiology and Chronic Health Evaluation IV Scoring Systems: Is There a Difference?» En: *Indian Journal of Critical Care Medicine* 22.5 (2018), págs. 332-335. DOI: 10.4103/ijccm.IJCCM_422_17.

All of these scoring systems are based on a number of clinical factors, including vital signs, laboratory results and organ function, and provide an objective assessment of the patient's critical condition. Notably, many of these scoring systems have been validated in specific cohorts of sepsis patients, reinforcing their utility in detecting and managing the unique condition of sepsis. However, the intentional simplicity of these scoring systems, designed to enable real-time calculation in ICU with limited resources, often results in a sparse scoring framework that fails to capture the full complexity of sepsis. This limitation may overlook subtle, non-linear physiological interactions that are essential for understanding the dynamic progression of organ dysfunction. Recent computational advances, such as integer-constrained machine learning [11], suggest ways to improve precision while maintaining clinical interpretability, although their integration into sepsis frameworks remains exploratory.

0.3. Advances in machine learning for mortality prediction in sepsis patients

In recent years, machine learning has gained significant attention due to the increase in data generation in healthcare, with approximately 86 % of hospitals using machine learning or predictive models to improve patient care [12] [13] [14]. With the increasing availability of large health data repositories, such as electronic health records, and advances in computing power, traditional statistical methods are being complemen-

---

[11] Jiachang Liu et al. «FasterRisk: Fast and Accurate Interpretable Risk Scores». En: *Advances in Neural Information Processing Systems* 35 (2022).

[12] Bill Siwicki. «86 % of healthcare companies use some form of AI». En: *Healthcare IT News* (2017).

[13] Jessica Kent. «Over 80 % of Health Execs Have Artificial Intelligence Plans in Place». En: *HealthTech Analytics* (2020).

[14] Amir Masoud Rahmani et al. «Machine Learning (ML) in Medicine: Review, Applications, and Challenges». En: *MATHEMATICS* 9.22 (2021). DOI: 10.3390/math9222970.

ted or replaced by ML approaches to classify and predict health outcomes [15] [16].
A particularly active research area focuses on predicting in-hospital mortality among ICU patients. Studies indicate that in-hospital mortality is influenced by multiple factors and should be evaluated comprehensively, considering a broad range of patient-specific and hospital-related risk factors [17]. This outcome is primarily driven by conditions such as respiratory diseases, cardiovascular events, sepsis, and stroke, among others. Additionally, disease severity and quality of care significantly impact patient survival outcomes [18] [19] [20]. Several investigations have aimed to mitigate this critical health issue by analyzing extensive sets of clinical factors collected from hospitalized

---

[15] Jason E Black, Jacqueline K Kueper y Tyler S Williamson. «An introduction to machine learning for classification and prediction». En: *Family Practice* 40.1 (2023), págs. 200-204. DOI: 10.1093/fampra/cmac104.

[16] Massimo Bertolini et al. «Machine Learning for industrial applications: A comprehensive literature review». En: *EXPERT SYSTEMS WITH APPLICATIONS* 175 (2021). DOI: 10.1016/j.eswa.2021.114820.

[17] Ana M. Porcel-Gálvez et al. «Factors Associated with In-Hospital Mortality in Acute Care Hospital Settings: A Prospective Observational Study». En: *International Journal of Environmental Research and Public Health* 17.21 (2020), pág. 7951. DOI: 10.3390/ijerph17217951.

[18] H. Li et al. «A machine learning-based prediction of hospital mortality in mechanically ventilated ICU patients». En: *PLoS ONE* 19.9 (2024), e0309383. DOI: 10.1371/journal.pone.0309383.

[19] Negin Ashrafi et al. «Deep learning model utilization for mortality prediction in mechanically ventilated ICU patients». En: *Informatics in Medicine Unlocked* 49 (2024), pág. 101562. DOI: https://doi.org/10.1016/j.imu.2024.101562.

[20] Hans-Christian Thorsen-Meyer 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». En: *The Lancet Digital Health* 2.4 (2020), e179-e191. DOI: 10.1016/S2589-7500(20)30018-2.

patients [21] [22] [23] [24] [25] [26].

A critical area of research in ICU clinical practice is the mortality risk among sepsis patients. Despite the life threatening nature of sepsis, only a limited number of studies have specifically focused on predicting mortality in this population. Recent advances in machine learning have demonstrated significant promise in forecasting

---

[21] Whitney D. Gannon et al. «Outcomes and Mortality Prediction Model of Critically Ill Adults With Acute Respiratory Failure and Interstitial Lung Disease». En: *Chest* 153.6 (2018), págs. 1387-1395. DOI: https://doi.org/10.1016/j.chest.2018.01.006.

[22] Nicholas Brajer et al. «Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission». En: *JAMA Network Open* 3.2 (2020), e1920733. DOI: 10.1001/jamanetworkopen.2019.20733.

[23] Sofia Fika et al. «A novel mortality prediction model for the current population in an adult intensive care unit». En: *Heart & Lung* 47.1 (2018), págs. 10-15. DOI: https://doi.org/10.1016/j.hrtlng.2017.10.009.

[24] S. R. Moonesinghe et al. «Risk Stratification Tools for Predicting Morbidity and Mortality in Adult Patients Undergoing Major Surgery: Qualitative Systematic Review». En: *Anesthesiology* 119.4 (2013), págs. 959-981. DOI: 10.1097/ALN.0b013e3182a4e94d.

[25] Eitan Kerem et al. «Prediction of Mortality in Patients with Cystic Fibrosis». En: *New England Journal of Medicine* 326.18 (1992), págs. 1187-1191. DOI: 10.1056/NEJM199204303261804.

[26] T. B. Enger et al. «Prediction of Mortality in Adult Patients with Severe Acute Lung Failure Receiving Veno-Venous Extracorporeal Membrane Oxygenation: A Prospective Observational Study». En: *Critical Care* 18 (2014), R67. DOI: 10.1186/cc13824.

both the onset of sepsis and subsequent patient mortality [27] [28] [29] [30] [31] [32] [33] using techniques such as logistic regression, random forest, and XGBoost, among others. These machine learning models take advantage of a wide range of clinical factors, including patient demographics, comorbidities, vital signs, and laboratory values, to develop robust predictive tools that facilitate proactive interventions and comprehensive monitoring, in many cases outperforming even traditional scoring systems such

---

[27] A. Alanazi et al. «Machine Learning for Early Prediction of Sepsis in Intensive Care Unit (ICU) Patients». En: *Medicina (Lithuania)* 59 (7 2023). DOI: 10.3390/medicina59071276.

[28] H.M. Giannini et al. «A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice». En: *Critical Care Medicine* 47 (11 2019), págs. 1485-1492. DOI: 10.1097/CCM.0000000000003891.

[29] Supreeth P. Shashikumar et al. «Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics». En: *Journal of Electrocardiology* 50.6 (2017), págs. 739-743. DOI: https://doi.org/10.1016/j.jelectrocard.2017.08.013.

[30] Khandaker Reajul Islam et al. «Machine Learning-Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review». En: *Journal of Clinical Medicine* 12.17 (2023). DOI: 10.3390/jcm12175658.

[31] R. Z. Wang et al. «Predictive Models of Sepsis in Adult ICU Patients». En: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. 2018, págs. 390-391. DOI: 10.1109/ICHI.2018.00068.

[32] Joseph Guillén et al. «Predictive models for severe sepsis in adult ICU patients». En: *2015 Systems and Information Engineering Design Symposium*. 2015, págs. 182-187. DOI: 10.1109/SIEDS.2015.7116970.

[33] Dong Wang et al. «A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients». En: *Frontiers in Public Health* 9 (2021). DOI: 10.3389/fpubh.2021.754348.

as SAPS-II and SOFA [34] [35] [36] [37]. However, the models presented in the reviewed studies do not prioritize the selection of the most significant clinical factors for ICU sepsis patients, but tend to include a broader range of clinical factors compared to traditional clinical scores.

0.4. Challenges in the selection and validation of the clinical factors

One of the current challenges in this field is to develop more actionable models with consistent predictive performance across different ICU settings, ensuring their robustness. To address this, some studies have focused on creating interpretable models without compromising predictive performance by implementing machine learning approaches such as Random Forests, Support Vector Machines, and Decision Trees. These models aim to facilitate adoption by healthcare professionals in clinical

---

[34] JY Park et al. «Predicting Sepsis Mortality in a Population-Based National Database: Machine Learning Approach». En: *J Med Internet Res* 24.4 (2022), e29982. DOI: 10.2196/29982.

[35] A. Boussina et al. «Impact of a deep learning sepsis prediction model on quality of care and survival». En: *NPJ Digital Medicine* 7.1 (2024). Erratum in: NPJ Digit Med. 2024 Jun 12;7(1):153. doi: 10.1038/s41746-024-01149-x, pág. 14. DOI: 10.1038/s41746-023-00986-6.

[36] Mehtap Selcuk, Oguz Koc y A. Sevtap Kestel. «The prediction power of machine learning on estimating the sepsis mortality in the intensive care unit». En: *Informatics in Medicine Unlocked* 28 (2022), pág. 100861. DOI: https://doi.org/10.1016/j.imu.2022.100861.

[37] Jie Weng et al. «Development and validation of a score to predict mortality in ICU patients with sepsis: a multicenter retrospective study». En: *Journal of Translational Medicine* 19.1 (2021), págs. 1-12.

practice [38] [39] [40], while maintaining strong predictive performance for clinically relevant applications in ICU environments.

In addition, there has been a growing emphasis on reducing the number of clinical factors required by these algorithms. Techniques like SHapley Additive exPlanations (SHAP) and algorithm-driven feature selection have been used to identify significant clinical factors [41] [42], to simplify model complexity and enhancing their actionability. However, despite their promise, many of these approaches lack comprehensive validation, particularly with respect to the robustness of the methods used to select significant clinical factors and their applicability to diverse datasets. Most validations have been performed on a single dataset, raising questions about the performance of the models when applied to data from different ICU settings. Addressing these gaps is essential to ensure the robustness and clinical utility of machine learning models for sepsis mortality prediction.

For these models to be integrated into clinical practice, they must be validated for both robustness and actionability. Robustness refers to the performance of the mo-

---

[38] G. Kong, K. Lin e Y. Hu. «Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU». En: *BMC Medical Informatics and Decision Making* 20.1 (2020), pág. 251. DOI: 10.1186/s12911-020-01271-2.

[39] H. Koozi et al. «A simple mortality prediction model for sepsis patients in intensive care». En: *Journal of the Intensive Care Society* 24.4 (2023). Epub 2023 Feb 1, págs. 372-378. DOI: 10.1177/17511437221149572.

[40] Zhixuan Zeng et al. «Development and validation of a novel blending machine learning model for hospital mortality prediction in ICU patients with Sepsis». En: *BioData Mining* 14.1 (2021), pág. 40. DOI: 10.1186/s13040-021-00276-5.

[41] Zhengyu Jiang et al. «An explainable machine learning algorithm for risk factor analysis of in-hospital mortality in sepsis survivors with ICU readmission». En: *Computer Methods and Programs in Biomedicine* 204 (2021), pág. 106040. DOI: https://doi.org/10.1016/j.cmpb.2021.106040.

[42] Jiayi Gao et al. «Prediction of sepsis mortality in ICU patients using machine learning methods». En: *BMC Medical Informatics and Decision Making* 24.1 (2024), pág. 228. DOI: 10.1186/s12911-024-02630-z.

dels across different ICU settings, ensuring error generalization and optimal results on new data. Actionability means that the models are easy for healthcare professionals to use in their clinical routine, particularly when they utilize a limited number of clinical factors.

This study aims to identify significant clinical factors to build actionable models for the robust prediction of mortality in sepsis patients in ICUs across different clinical settings. By leveraging two large datasets and employing rigorous feature selection and statistical analysis, we seek to address existing limitations and propose actionable solutions.

1. OBJECTIVES

1.1. General objective

Developing a predictive machine learning algorithm for mortality in Intensive Care Units of sepsis patients using electronic health records.

1.2. Specific objectives

- To curate the Medical Information Mart for Intensive Care (MIMIC) and the eICU Collaborative Research databases for utilization in the feature engineering, training, validation, and testing stages of machine learning models
- To implement a machine learning algorithm for the prediction of mortality in Intensive Care Units of sepsis patients by selecting relevant features.
- To evaluate the performance of machine learning models on the eICU Collaborative Research database.

2. DATA COLLECTION

This retrospective and nested study employed two publicly available critical care datasets: MIMIC-IV v3.0 and eICU v2.0. The MIMIC-IV v3.0, derived from Beth Israel Deaconess Medical Center, provides detailed clinical information from a single institution, ensuring consistency in patient management protocols [43]. In contrast, eICU v2.0 aggregates data from multiple ICUs, capturing a broader spectrum of clinical scenarios and practices [44]. To assess the robustness of our approach, MIMIC-IV was used for training and validation, while the eICU was reserved exclusively for external testing, thus challenging the adaptability of the model to different clinical environments.

The patient inclusion criteria were the following: only ICU-admitted patients were considered. To ensure data robustness, the analysis included only patients with a single ICU admission and a minimum ICU stay of 24 hours. Sepsis cases were identified based on the Sepsis-3 definition [1]. Once patients with sepsis have been identified, only those with complete demographic data are included in the cohort. Among non-survivors, only those whose death occurred within the ICU were considered. The inclusion process is illustrated in Figure 1.

This research analyzed a cohort of 15,100 patients admitted to ICU at Beth Israel Deaconess Medical Center, Boston, extracted from the MIMIC-IV v3.0 dataset for model training, and 8,201 patients from large multicenter ICUs from the eICU v2.0 critical care dataset for testing.

---

[43] Alistair E. W. Johnson et al. «MIMIC-IV, a freely accessible electronic health record dataset». En: *Scientific Data* 10 (1 2023), pág. 1. DOI: 10.1038/s41597-022-01899-x.

[44] Tom J. Pollard et al. «The eICU Collaborative Research Database, a freely available multi-center database for critical care research». En: *Scientific Data* 5.1 (2018), pág. 180178. DOI: 10.1038/sdata.2018.178.

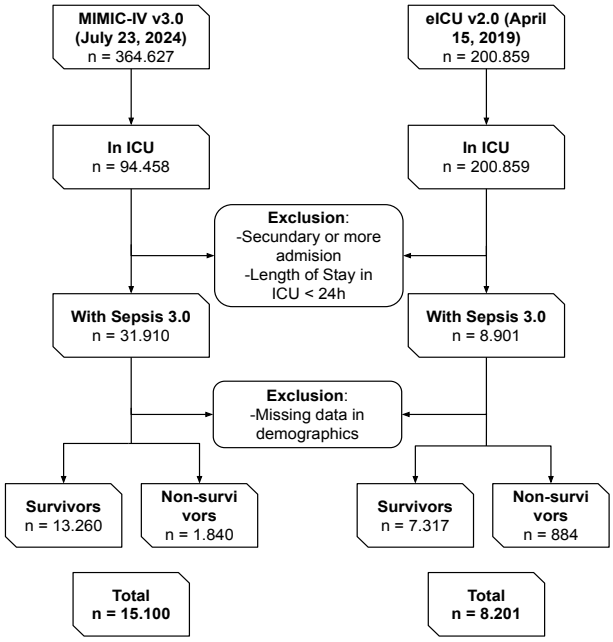

Figure 1. Inclusion and exclusion criteria for the patient cohort from the MIMIC-IV v3.0 and eICU v2.0 databases.

The selection of the initial set of clinical factors was assisted by an expert. To handle missing data, factors with approximately 20 % missing values were imputed using the miceforest method [45]. Miceforest is a multiple imputation method based on Random Forests, which iteratively predicts and fills in missing values while preserving the original data distribution.

As a result, a total of 77 clinical factors were extracted from both databases, including vital signs, laboratory results, pharmacologic interventions, demographic information, Glasgow Coma Scale and SOFA score. A detailed breakdown of all clinical factors can be found in Annex1.

[45] The miceforest Development Team. *miceforest: Fast, Memory Efficient Imputation with LightGBM*. Versión 6.0.3. 2024.

3. STATISTICAL ANALYSIS

For the statistical analysis, we evaluated demographic clinical factors, vital signs, Glasgow Coma Scale, and SOFA score. The Kolmogorov Smirnov test was applied to assess whether each clinical factor followed a normal distribution. The results of this test indicated that none of the analyzed factors met this criterion. Consequently, for continuous clinical factors, the Mann-Whitney U test was used to compare survivors and non-survivors, with a significance level set at 0.05, identifying statistically significant differences between the groups. Similarly, for categorical clinical factors, the Chi-square test was performed, also with a significance of 0.05. Among the analyzed factors, gender was the only one that did not show statistically significant differences between the groups. This analysis was conducted for the clinical factors from both the MIMIC-IV v3.0 dataset and the eICU v2.0 dataset, as shown in Table 1 and Table 2, respectively.

Clinical factors	Survivors	Non-Survivors	p-value
Gender **(n %)**			0.543
Female	44.02 %	44.82 %	
Male	55.98 %	55.18 %	
Median (Q1,Q3)			
LOS ICU	3.0 (1.0, 6.0)	4.0 (2.0, 9.0)	10^{-49}
Age	68.0 (56.0, 79.0)	73.0 (61.7, 82.0)	10^{-29}
Mean Blood Pressure mean	76.0 (70.2, 83.5)	73.8 (67.9, 81.4)	10^{-20}
Diastolic Blood Pressure mean	61.7 (55.5, 68.8)	59.3 (53.2, 66.9)	10^{-18}
Systolic Blood Pressure mean	114.3 (105.2, 126.5)	108.6 (100.8, 121.2)	10^{-41}
Respiration rate mean	19.2 (16.9, 22.1)	21.4 (18.4, 24.6)	10^{-73}
Heart rate mean	85.6 (74.6, 97.8)	90.1 (76.8, 104.1)	10^{-20}
Oxygen Saturation mean	97.0 (95.6, 98.4)	96.6 (94.7, 98.4)	10^{-14}
Temperature mean	36.8 (36.6, 37.2)	36.7 (36.4, 37.1)	10^{-27}
Glasgow Coma Scale min	15.0 (13.0, 15.0)	15.0 (13.0, 15.0)	10^{-35}
Verbal Glasgow Coma Scale	4.0 (1.0, 5.0)	1.0 (0.0, 5.0)	10^{-59}
Motor Glasgow Coma Scale	6.0 (5.0, 6.0)	5.0 (1.0, 6.0)	10^{-122}
Eyes Glasgow Coma Scale	3.0 (2.0, 4.0)	3.0 (1.0, 4.0)	10^{-103}
SOFA score	8.0 (6.0, 10.0)	11.0 (8.0, 14.0)	10^{-198}

Table 1. Comparative analysis of some clinical factors between survivors and non-survivors in the MIMIC-IV v3.0 dataset.

Clinical factors	Survivors	Non-Survivors	p-value
Gender **(n %)**			0.948
Female	48.53 %	48.71 %	
Male	51.47 %	51.29 %	
Median (Q1,Q3)			
LOS ICU	67.0 (43.0, 118.0)	80.0 (41.0, 180.2)	10^{-6}
Age	67.0 (55.0, 77.0)	71.0 (60.0, 81.0)	10^{-13}
Mean Blood Pressure mean	73.9 (68.2, 81.5)	69.5 (64.9, 76.0)	10^{-35}
Diastolic Blood Pressure mean	60.0 (54.0, 66.8)	56.2 (51.0, 62.8)	10^{-27}
Systolic Blood Pressure mean	110.5 (102.2, 121.1)	104.0 (96.8, 113.6)	10^{-35}
Respiration rate mean	20.5 (17.6, 23.9)	23.1 (19.6, 26.9)	10^{-37}
Heart rate mean	90.8 (79.7, 102.4)	99.3 (87.0, 112.4)	10^{-37}
Oxygen Saturation mean	97.1 (95.6, 98.4)	96.2 (94.0, 97.9)	10^{-26}
Temperature mean	36.9 (36.6, 37.3)	36.8 (36.4, 37.3)	10^{-8}
Glasgow Coma Scale min	15.0 (13.0, 15.0)	14.0 (10.0, 15.0)	10^{-38}
Verbal Glasgow Coma Scale	5.0 (4.0, 5.0)	4.0 (1.0, 5.0)	10^{-27}
Motor Glasgow Coma Scale	6.0 (6.0, 6.0)	6.0 (5.0, 6.0)	10^{-25}
Eyes Glasgow Coma Scale	4.0 (4.0, 4.0)	4.0 (3.0, 4.0)	10^{-30}
SOFA score	7.0 (4.0, 9.0)	10.0 (7.0, 13.0)	10^{-148}

Table 2. Comparative analysis of some clinical factors between survivors and non-survivors in the eICU v2.0 dataset.

4. MACHINE LEARNING MODEL DEVELOPMENT

Before training the machine learning models, the data were standardized using Z-score normalization to ensure comparability across factors. Eight machine learning models were evaluated in this study: Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Gradient Boosting, Multilayer Perceptron, XGBoost, and LightGBM. The models were trained using 5-fold cross-validation on the MIMIC-IV v3.0 dataset. Additionally, to assess the robustness of the models, the eICU v2.0 dataset comprising data from diverse ICU settings was used as an independent test set.

The Area Under the Curve (AUC) metric was used to compare the performance of the models. The AUC is particularly suitable for binary classification tasks, as it provides a robust assessment of a model's ability to discriminate between classes. It summarises model performance at all possible decision thresholds, shown in the Receiver Operating Characteristic curve (ROC curve) that illustrates the trade-off between true positive rates and false positive rates, making it an ideal metric for this study.

According to the 5-fold cross-validation on the MIMIC-IV v3.0 dataset, the model with the highest AUC was selected for hyperparameter optimization. This optimization was conducted using Optuna [46], a Bayesian optimization-based library that efficiently explores the hyperparameter space by selecting promising values through probabilistic models.

To enhance the robustness of the results, bootstrapping was applied to estimate

---

[46] Takuya Akiba et al. «Optuna: A Next-Generation Hyperparameter Optimization Framework». En: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM. 2019, págs. 2623-2631.

confidence intervals for the AUC values on both datasets. Furthermore, we aimed to maximize the F1-score to find the model's decision threshold, ensuring a balanced approach between false negatives and false positives. The AUC of the optimized model was compared with the optimal cut-off point, the SOFA score, a widely used clinical reference for assessing mortality in ICU patients with sepsis.

Finally, we categorized the datasets with the assistance of a healthcare professional into five groups: demographic, laboratory, pharmacological, vital signs, and Glasgow Coma Scale. We then evaluated the AUC performance on both the MIMIC-IV v3.0 and eICU v2.0 datasets. Additionally, we explored the combination of different clinical factors, integrating vital signs with each of the other categories, and applied the same approach to laboratory factors to assess their predictive contributions.

5. METHODS

In the following sections, we apply some techniques to identify significant clinical factors. We begin with a traditional approach, employing the Sequential Feature Selector (SFS) in Section 5.1. The model trained using the clinical factors selected by SFS yielded results comparable to those of the model incorporating all clinical factors, as assessed through DeLong's test in Section 5.2.

To assess the actionability and clinical relevance of clinical factors selection, we conducted an Odds Ratio (OR) analysis in Section 5.3 to identify the clinical factors most strongly associated with the outcome. Additionally, we calculated the Adjusted Odds Ratio (AOR) to evaluate the independent contribution of each factor category while controlling for potential confounders. This comprehensive approach enhances interpretability and ensures the development of a clinically actionable model for mortality prediction in sepsis patients.

5.1. Traditional Feature Selection

There are several traditional methods to select the most relevant factors employed in machine learning models. One of them is the Sequential Feature Selector (SFS), which sequentially adds or removes factors based on their contribution to model performance. In this case, we use forward selection, where the algorithm starts without any factors and adds them one by one, checking the impact of each feature on the model. The performance improvement is evaluated by the AUC. At each iteration, the factor that most improves model performance is selected, with a 5-fold cross-validation process. This makes the SFS a robust method for identifying optimal set of clinical factors based on this performance metric.

5.2. Model Performance Comparison

To assess whether the AUCs obtained from the model using the categories of clinical factors and the model derived through SFS were statistically similar, we applied DeLong's test. The DeLong's test is a non-parametric statistical method designed to compare the areas under two correlated ROC curves. This test assesses whether the observed difference in AUC values is statistically significant. The null hypothesis is proposed which assumes that there is no statistically significant difference between the two ROC curves, while the alternative hypothesis suggests the presence of a significant difference. The significance level was set at 0.05, meaning that if the p-value is below this value, the null hypothesis is rejected, indicating a significant difference between the models.

5.3. Identification of Significant Clinical Factors

To identify significant clinical factors, both OR and AOR analyses were performed. The OR provides a measure of the association between a clinical factor and the outcome, comparing the odds of the outcome occurring in the presence versus the absence of the factor. On the other hand, the AOR adjusts for potential confounding factors, allowing for a more accurate estimate of the relationship between the clinical factor and the outcome by controlling for other factors that might distort the observed association. The confounding adjustment helps identify whether the initial observed association is due to the factor itself or if it is influenced by other related factors. The SOFA score was excluded from the OR and AOR analyses to focus on assessing the individual association of clinical factor categories with the outcome.

5.3.1. Association Analysis: Odds Ratio Calculation For the OR analysis, we calculated ORs along with their corresponding confidence intervals for clinical factors

across five categories: demographic, laboratory, pharmacological, vital signs, and Glasgow Coma Scale. We then evaluated the individual association of each clinical factor category with the outcome.

Specifically, we analyzed combinations of vital signs with the other four categories (demographic, laboratory, pharmacological, and Glasgow Coma Scale) and similarly examined laboratory results in combination with the remaining four categories.

5.3.2. Confounding Adjustment: Adjusted Odds Ratio Estimation For the AOR analysis, a clinical factor identified as having a relationship with the outcome in the OR analysis was selected as the exposure factor. Confounding factors were then added to the model to assess their impact. If there was a noticeable change (either an increase or decrease) in the OR upon adjustment, it was considered that the added factors were confounding the relationship between the exposure and the outcome. This adjustment ensures that the estimated relationship between the exposure and the outcome reflects the true effect, accounting for the influence of other factors.

6. RESULTS

6.1. Machine learning models

The training process, employing 5-fold cross-validation on the MIMIC-IV v3.0 database, yielded the results illustrated in Figure 2. Similarly, the performance on the independent test dataset (eICU v2.0) is shown in Figure 3. Tree-based models, including Random Forest, Gradient Boosting, and LightGBM, consistently outperformed other algorithms in both datasets, being LightGBM achieving the highest AUC.

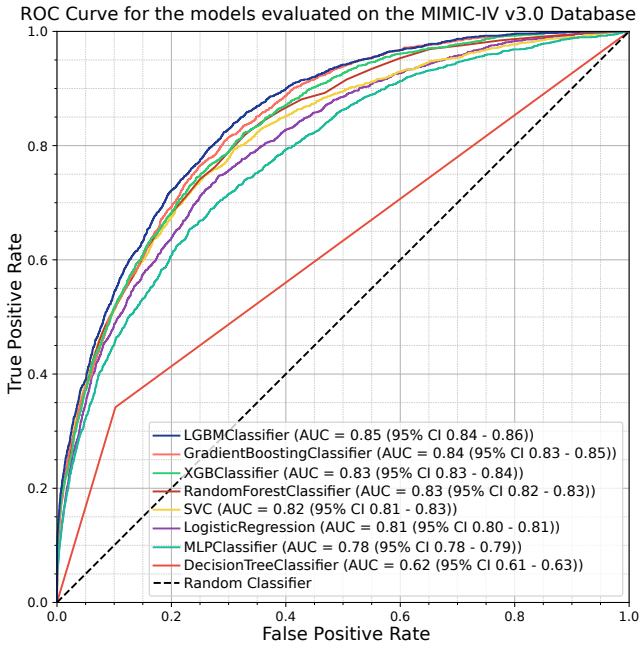

Figure 2. ROC curves for the eight machine learning models by 5-fold cross-validation in the MIMIC-IV v3.0 database.

Given its superior performance in MIMIC-IV v3.0 database, the LightGBM (LGBM) model was selected for hyperparameter optimization using Optuna. After fine-tuning, LightGBM achieved an AUC of 0.84 on the MIMIC-IV v3.0 dataset (95 % CI: 0.83–0.85) and an AUC of 0.76 on the eICU v2.0 dataset (95 % CI: 0.75–0.76).

To balance sensitivity and specificity, the optimal decision threshold was determined by maximizing the F1-score, resulting in a threshold of 0.162. This threshold yielded a sensitivity of 77 % on the test dataset, underscoring the model's ability to reduce false negatives while maintaining overall predictive performance.

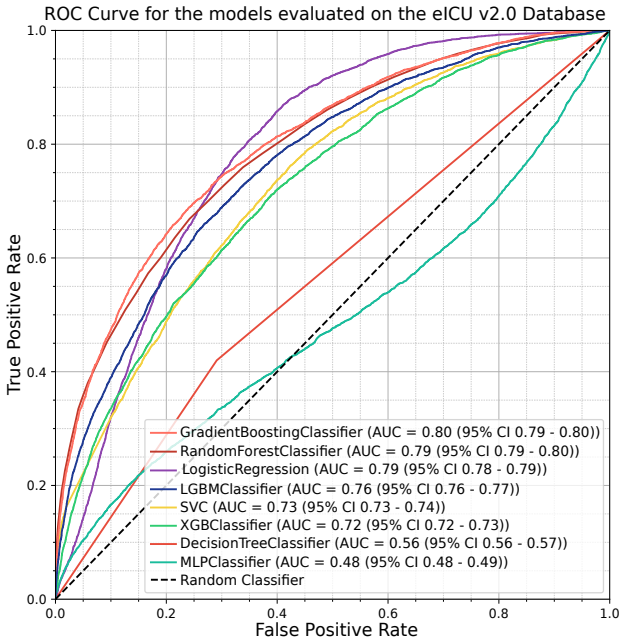

Figure 3. ROC curves for the eight machine learning models in the eICU v2.0 database.

In contrast, traditional scoring methods such as SOFA exhibited lower performance, with an AUC of 0.69 on the MIMIC-IV v3.0 dataset (95 % CI: 0.68–0.70) and an AUC of 0.72 on the eICU v2.0 dataset (95 % CI: 0.70–0.74). These findings underscore the superior predictive capability of LightGBM for mortality prediction in ICU sepsis patients, as illustrated in Annex 2.

The AUC results for the five categories of clinical factors are presented in Table 3 and in Table 4, highlighting the performance of the LGBM model trained exclusively on vital signs data. This model demonstrated strong actionability, achieving comparable performance to the model with all clinical factors in the eICU v2.0 dataset. A

similar behavior was observed with the LGBM model trained on laboratory factors. Notably, the model combining both vital signs and laboratory factors outperformed the model with all clinical factors on the eICU v2.0. This suggests that a simplified and actionable model, relying on a reduced set of significant clinical factors, could offer a valuable advantage in resource-limited ICU settings.

LGBM Models	**MIMIC-IV**			
	AUC (95 % CI)	**Spec.**	**Sens.**	**F1**
77 clinical factors	**0.84 (0.83-0.85)**	**39 %**	**58 %**	**47 %**
25 clinical factors	**0.84 (0.83-0.85)**	**40 %**	**53 %**	**46 %**
Demographic	0.60 (0.58-0.61)	15 %	58 %	24 %
Laboratory	**0.71 (0.70-0.72)**	**34 %**	**35 %**	**35 %**
Pharmacologic	0.65 (0.63-0.66)	24 %	35 %	29 %
Vital signs	**0.75 (0.74-0.76)**	**35 %**	**43 %**	**39 %**
Glasgow Coma Scales	0.63 (0.62-0.65)	21 %	45 %	29 %
Laboratory + Glasgow Coma Scales	0.74 (0.73-0.75)	35 %	41 %	37 %
Laboratory + Pharmacologic	0.72 (0.71-0.74)	32 %	41 %	36 %
Laboratory + Demographic	0.78 (0.77-0.79)	35 %	44 %	39 %
Vital signs + Glasgow Coma Scales	0.78 (0.77-0.79)	34 %	51 %	40 %
Vital signs + Pharmacologic	0.75 (0.74-0.76)	35 %	41 %	38 %
Vital signs + Demographic	0.80 (0.79-0.81)	35 %	49 %	41 %
Vital signs + Laboratory	**0.78 (0.77-0.79)**	**41 %**	**42 %**	**42 %**

Table 3. Performance of each LGBM model by category, evaluated using 5-fold cross-validation on the MIMIC-IV v3.0 dataset. Note: “Spec” indicates specificity, “Sens” indicates sensitivity, and “F1” represents the F1-score.

6.2. Traditional methods to identify significant clinical factors: Sequential Feature Selector

The Sequential Feature Selector with forward selection identified an optimal subset of 25 factors for model performance, as outlined in Annex 3. The selection process, assessed using AUC with 5-fold cross-validation, achieved an AUC of 0.84 (95 % confidence interval: 0.83–0.85) in the MIMIC-IV v3.0 database, which is the same performance obtained using all 77 clinical factors. In the eICU v2.0 dataset, the mo-

LGBM Models	eICU			
	AUC (95 % CI)	Spec.	Sens.	F1
77 clinical factors	**0.76 (0.75-0.76)**	**18 %**	**77 %**	**29 %**
25 clinical factors	**0.73 (0.72-0.74)**	**18 %**	**71 %**	**29 %**
Demographic	0.55 (0.54-0.55)	11 %	91 %	19 %
Laboratory	**0.72 (0.72-0.73)**	**27 %**	**45 %**	**34 %**
Pharmacologic	0.60 (0.59-0.61)	20 %	34 %	25 %
Vital signs	**0.76 (0.75-0.77)**	**20 %**	**69 %**	**32 %**
Glasgow Coma Scales	0.55 (0.54-0.56)	21 %	12 %	16 %
Laboratory + Glasgow Coma Scales	0.74 (0.73-0.75)	32 %	39 %	35 %
Laboratory + Pharmacologic	0.73 (0.72-0.74)	27 %	46 %	34 %
Laboratory + Demographic	0.69 (0.68-0.70)	16 %	72 %	26 %
Vital signs + Glasgow Coma Scales	0.77 (0.77-0.78)	23 %	65 %	34 %
Vital signs + Pharmacologic	0.76 (0.75-0.77)	23 %	64 %	34 %
Vital signs + Demographic	0.73 (0.72-0.74)	16 %	79 %	27 %
Vital signs + Laboratory	**0.79 (0.78-0.80)**	**25 %**	**64 %**	**36 %**

Table 4. Performance of each LGBM model by category, evaluated externally on the eICU v2.0 dataset. Note: “Spec” indicates specificity, “Sens” indicates sensitivity, and “F1” represents the F1-score.

del with 25 factors achieved an AUC of 0.73 (95 % confidence interval: 0.72–0.74), slightly lower than the AUC obtained with 77 clinical factors, which obtained an AUC of 0.76 (95 % confidence interval: 0.75–0.76). These findings underscore the effectiveness of the SFS method in identifying a reduced but highly predictive set of factors. Notably, the use of approximately 25 clinical factors achieves the same level of performance as employing all 77. Among these, the SOFA score emerges as a critical factor for predicting sepsis mortality, consistent with its central role in the clinical definition of sepsis. Other key factors include urine output, Glasgow Coma Scale, specific laboratory values such as lactate, and certain vital signs. These findings align with previous studies and reflect factors routinely monitored in clinical practice within the ICU.

6.3. Comparison of Area Under Curves through DeLong's test

In this analysis, we compared the ROC curve generated using 77 clinical factors from the MIMIC-IV v3.0 dataset to the ROC curve generated using 25 clinical factors from the same dataset. A similar comparison was performed for the eICU v2.0 dataset. As shown in Table 5, the results for the MIMIC-IV dataset yielded a p-value greater than 0.05, supporting the null hypothesis and indicating no statistically significant difference between the two ROC curves. Conversely, for the eICU dataset, the p-value was close to zero, supporting the alternative hypothesis and revealing a statistically significant difference between the ROC curves.

Dataset	**LGBM Models AUC (95 % CI)**		**p-value**
	77 factors	25 factors	
MIMIC-IV	0.84 (0.83-0.85)	0.84 (0.83-0.85)	0.507
eICU	0.75 (0.75-0.76)	0.73 (0.72-0.74)	10^{-25}

Table 5. DeLong's test in the MIMIC-IV v3.0 and eICU v2.0 databases to assess statistical significance between the LightGBM model with 77 clinical factors and the LightGBM model with 25 clinical factors

6.4. Association measurement: Odds Ratio

The OR analysis, presented in Table 6, reveals that the vital signs category has the highest OR at 2.252 (95 % confidence interval: 2.193–2.313), closely followed by the laboratory factors category with an OR of 1.951 (95 % confidence interval: 1.903–2.000). These values indicate that patients exhibiting changes in these categories are approximately twice as likely to experience mortality in ICU sepsis cases compared to other categories. In other words, vital signs and laboratory factors are strongly associated with an increased risk of mortality in this context.

On the other hand, the demographic and Glasgow Coma Scales categories showed lower OR values, suggesting a much weaker or negligible association with the risk of

mortality. This highlights the relatively minor role these factors play in predicting mortality outcomes compared to the substantial influence of vital signs and laboratory factors.

According to these findings, we analyzed combinations of clinical factor categories, as shown in Table 6. The combination yielding the highest OR was vital signs and laboratory data, with an OR of 2.458 (95 % CI: 2.392–2.525), slightly higher than the OR obtained when including all categories without SOFA, which was 2.280 (95 % CI: 2.214–2.348).

Notably, combinations that included vital signs consistently resulted in higher OR values compared to those containing only laboratory data. This outcome aligns with expectations, given that the OR for vital signs alone exceeds that of laboratory factors individually.

These results reinforce the critical role of both vital signs and laboratory data in mortality prediction, as their inclusion consistently yields the highest OR values. Notably, the combination of these two categories achieves an even higher OR than the full set of clinical factors, suggesting that a model based solely on vital signs and laboratory values could serve as an effective and simplified predictive tool for ICU sepsis mortality. The findings from both OR analyses further support this approach, highlighting the strong predictive value of these clinical variables in this context.

6.5. Confounding analysis: Adjusted Odds Ratio

In the confounding analysis, where vital signs were considered as the exposure factor, the results shows that laboratory factors introduced the greatest confounding effect. This is evident from the substantial change observed in the AOR compared to the OR, indicating that laboratory values strongly influence the relationship between vital signs and the mortality prediction. Conversely, the Glasgow Coma Scale had the least confounding impact, as its AOR remained nearly identical to its OR.

The combination of laboratory and pharmacological factors showed the largest confounding effect, suggesting a strong interaction between these factors. In contrast, the combination of laboratory and demographic factors showed the least confounding effect, with a minimal rate of change between OR and AOR. The results of both analyses are reported in Table 6.

Clinical factors category	**OR (95 % CI)**	**AOR (95 % CI)**
Demographic	1.15 (1.12-1.19)	1.18 (1.10-1.26)
Laboratory	1.95 (1.90-2.00)	1.60(1.50-1.70)
Pharmacologic	1.36 (1.33-1.39)	1.26 (1.19-1.34)
Vital signs	2.25 (2.19-2.31)	—
Glasgow Coma Scales	1.22 (1.19-1.25)	1.20 (1.13-1.28)
Laboratory + Glasgow Coma Scales	2.02 (1.97-2.08)	1.59 (1.49-1.69)
Laboratory + Pharmacologic	1.96 (1.91-2.00)	1.60 (1.50-1.70)
Laboratory + Demographic	1.89 (1.83-1.94)	1.86 (1.74-1.99)
Vital signs + Glasgow Coma Scales	2.30 (2.24-2.36)	—
Vital signs + Pharmacologic	2.20 (2.15-2.26)	—
Vital signs + Demographic	2.213 (2.14-2.27)	—
Vital signs + Laboratory	2.45 (2.39-2.52)	—
All categories	2.28 (2.21-2.34)	—

Table 6. Odds Ratio and Adjusted Odds Ratio analysis for each combination of clinical factor categories. The Adjusted Odds Ratio was conducted for vital signs category.

7. DISCUSSION

This study identified which sets of clinical factors are significant to mortality prediction in ICU patients with sepsis. The primary motivation behind this analysis was to simplify the current clinical workflow, which often involves a large number of factors, including traditional severity scores, and an overwhelming volume of patient data. By focusing on a reduced set of significant factors, such as vital signs and laboratory analyses, this research enables the development of actionable and robust predictive models for predicting mortality in sepsis patients in ICU. This approach holds significant potential, particularly in resource-limited ICU settings, where streamlined and efficient decision-making tools are essential.

The identification of significant clinical factors using traditional methods, such as the Sequential Feature Selector, produces performance metrics comparable to those of models incorporating the full set of clinical factors, as confirmed by DeLong's test (AUC = 0.84, p-value = 0.507) in this study. The model with 25 selected factors achieved an AUC of 0.84 (95 % CI: 0.83–0.85) on the MIMIC-IV v3.0 dataset and an AUC of 0.73 (95 % CI: 0.72–0.74) on the eICU v2.0 dataset. However, despite rigorous evaluation using 5-fold cross-validation, there remains a possibility that this set of clinical factors could be replaced by an alternative set, potentially compromising clinical actionability. These performances surpass those of the SOFA score, a widely used clinical scoring system, which achieved an AUC of 0.69 on the MIMIC-IV v3.0 dataset (95 % CI: 0.68–0.70) and 0.72 on the eICU v2.0 dataset (95 % CI: 0.70–0.74).

The OR analysis revealed a weaker association between the pharmacological category and mortality prediction, including critical treatments such as vasopressors and vasoactive drugs, crucial for sepsis mortality in ICUs. Similarly, the Glasgow Coma Scale, a clinician-assessed measurement, showed less pronounced association

than expected.

In addition, both OR and AOR analyses highlighted the strong association of vital signs and laboratory analyses with mortality outcomes, either individually or in combination with other categories. This underscores the critical role of these clinical factors in predicting sepsis mortality in ICU patients. Models trained using only vital signs and laboratory data achieved an AUC of 0.78 on the MIMIC-IV v3.0 dataset (95 % CI: 0.77 – 0.79) and 0.79 on the eICU v2.0 dataset (95 % CI: 0.78 – 0.80). Their predictive performance is comparable to models incorporating all available clinical factors, relying solely on vital signs and laboratory analyses enhances model actionability, which is particularly valuable in resource-limited ICUs where access to comprehensive clinical data may be restricted. Furthermore, these models significantly outperform the SOFA score.

A common aspect of some studies is that the clinical validation of the model is performed exclusively on the MIMIC dataset [47] [48] [49] [50]. This practice limits the robustness and actionability of the model, as its evaluation relies on data with the same distribution, without testing its performance in different ICU settings. One of the strengths of our research is the validation of the model using two independent datasets, MIMIC-IV v3.0 and eICU v2.0. A 5-fold cross-validation is conducted on the MIMIC-IV dataset,

---

[47] F. Mao et al. «Machine Learning Approach for Sepsis Risk Assessment in Ischemic Stroke Patients». En: *Journal of Intensive Care Medicine* (2025). Epub ahead of print. DOI: 10.1177/08850666241308195.

[48] Zhijiang Yu et al. «Prediction of 30-day mortality for ICU patients with Sepsis-3». En: *BMC Medical Informatics and Decision Making* 24.1 (2024), pág. 223. DOI: 10.1186/s12911-024-02629-6.

[49] X. Pan et al. «Evaluate prognostic accuracy of SOFA component score for mortality among adults with sepsis by machine learning method». En: *BMC Infectious Diseases* 23.1 (2023), pág. 76. DOI: 10.1186/s12879-023-08045-x.

[50] C. Hu, L. Li, Y. Li et al. «Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission». En: *Infectious Diseases and Therapy* 11 (2022), págs. 1695-1713. DOI: 10.1007/s40121-022-00671-3.

while a hold-out validation is performed on the eICU dataset. This presents a challenge, as publicly available datasets sharing the same clinical factors for comparison are limited. However, the reliance on vital signs and laboratory analyses, factors commonly available across clinical settings, simplifies this issue and opens the possibility for future validation studies using additional datasets.

Furthermore, the techniques employed to select significant clinical factors in these studies, such as tree-based models and LASSO regression [47] [49], as well as their interpretability analyses using methods such as SHAP [47] [48] and LIME [50], have limitations in terms of clinical actionability. These approaches mainly focus on the relative importance of variables within the model, but do not necessarily consider their relevance from a medical perspective. Our research addresses the selection of significant clinical factors by applying approaches such as odds ratio and adjusted odds ratio analyses, which guarantee not only greater clinical validity but also facilitate interpretation by healthcare professionals. This allows us to identify which clinical categories have a stronger association with ICU mortality in sepsis patients, demonstrating that the proposed model outperforms studies developing models for mortality prediction in ICU sepsis patients [49]. Moreover, our model is actionable and robust to different ICU settings, providing a more accurate tool for clinical decision-making in sepsis management.

8. CONCLUSIONS

This study presents a comprehensive analysis identifying key clinical factors for predicting mortality in ICU patients with sepsis, demonstrating that fewer factors can be leveraged to create robust and actionable predictive models. By focusing on vital signs and laboratory data, the proposed machine learning model outperformed traditional scoring systems, offering significant promise for accurate mortality prediction. Importantly, this streamlined approach not only enhances model performance but also ensures its applicability in resource limited ICU settings, where comprehensive data might not always be available. The findings underscore the potential for developing predictive tools that can support early intervention strategies and ultimately improve patient outcomes in critical care, marking a significant step toward the integration of machine learning in ICU monitoring systems.

BIBLIOGRAPHY.

Akiba, Takuya et al. «Optuna: A Next-Generation Hyperparameter Optimization Framework». En: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM. 2019, págs. 2623-2631 (vid. pág. 26).

Alanazi, A. et al. «Machine Learning for Early Prediction of Sepsis in Intensive Care Unit (ICU) Patients». En: *Medicina (Lithuania)* 59 (7 2023). DOI: 10.3390/medicina59071276 (vid. pág. 16).

Ashrafi, Negin et al. «Deep learning model utilization for mortality prediction in mechanically ventilated ICU patients». En: *Informatics in Medicine Unlocked* 49 (2024), pág. 101562. DOI: https://doi.org/10.1016/j.imu.2024.101562 (vid. pág. 14).

Bertolini, Massimo et al. «Machine Learning for industrial applications: A comprehensive literature review». En: *EXPERT SYSTEMS WITH APPLICATIONS* 175 (2021). DOI: 10.1016/j.eswa.2021.114820 (vid. pág. 14).

Black, Jason E, Jacqueline K Kueper y Tyler S Williamson. «An introduction to machine learning for classification and prediction». En: *Family Practice* 40.1 (2023), págs. 200-204. DOI: 10.1093/fampra/cmac104 (vid. pág. 14).

Boussina, A. et al. «Impact of a deep learning sepsis prediction model on quality of care and survival». En: *NPJ Digital Medicine* 7.1 (2024). Erratum in: NPJ Digit Med. 2024 Jun 12;7(1):153. doi: 10.1038/s41746-024-01149-x, pág. 14. DOI: 10.1038/s41746-023-00986-6 (vid. pág. 17).

Brajer, Nicholas et al. «Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission». En: *JAMA Network Open* 3.2 (2020), e1920733. DOI: 10.1001/jamanetworkopen.2019.20733 (vid. pág. 15).

Caraballo, Cé y Fabián Jaimes. «Organ Dysfunction in Sepsis: An Ominous Trajectory From Infection To Death». English. En: *The Yale journal of biology and medicine* 92.4 (2019), págs. 629-640 (vid. pág. 10).

Chakraborty, RK y B Burns. *Systemic Inflammatory Response Syndrome*. [Updated 2023 May 29]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK547669/. Treasure Island (FL): StatPearls Publishing, 2025 (vid. pág. 12).

Christensen, E. E. et al. «Mortality and Sequential Organ Failure Assessment Score in Patients With Suspected Sepsis: The Impact of Acute and Preexisting Organ Failures and Infection Likelihood». En: *Critical Care Explorations* 5.2 (2023), e0865. DOI: 10.1097/CCE.0000000000000865 (vid. pág. 12).

Development Team, The miceforest. *miceforest: Fast, Memory Efficient Imputation with LightGBM*. Versión 6.0.3. 2024 (vid. pág. 22).

Enger, T. B. et al. «Prediction of Mortality in Adult Patients with Severe Acute Lung Failure Receiving Veno-Venous Extracorporeal Membrane Oxygenation: A Prospective Observational Study». En: *Critical Care* 18 (2014), R67. DOI: 10.1186/cc13824 (vid. pág. 15).

Evans, L et al. «Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021., 2021, 47». En: *DOI: https://doi. org/10.1007/s00134-021-06506-y* (2021), págs. 1181-1247 (vid. pág. 10).

Fika, Sofia et al. «A novel mortality prediction model for the current population in an adult intensive care unit». En: *Heart & Lung* 47.1 (2018), págs. 10-15. DOI: https://doi.org/10.1016/j.hrtlng.2017.10.009 (vid. pág. 15).

Gannon, Whitney D. et al. «Outcomes and Mortality Prediction Model of Critically Ill Adults With Acute Respiratory Failure and Interstitial Lung Disease». En: *Chest* 153.6 (2018), págs. 1387-1395. DOI: https://doi.org/10.1016/j.chest.2018.01.006 (vid. pág. 15).

Gao, Jiayi et al. «Prediction of sepsis mortality in ICU patients using machine learning methods». En: *BMC Medical Informatics and Decision Making* 24.1 (2024), pág. 228. DOI: 10.1186/s12911-024-02630-z (vid. pág. 18).

Giannini, H.M. et al. «A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice». En: *Critical Care Medicine* 47 (11 2019), págs. 1485-1492. DOI: 10.1097/CCM.0000000000003891 (vid. pág. 16).

Guillén, Joseph et al. «Predictive models for severe sepsis in adult ICU patients». En: *2015 Systems and Information Engineering Design Symposium*. 2015, págs. 182-187. DOI: 10.1109/SIEDS.2015.7116970 (vid. pág. 16).

Hu, C., L. Li, Y. Li et al. «Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission». En: *Infectious Diseases and Therapy* 11 (2022), págs. 1695-1713. DOI: 10.1007/s40121-022-00671-3 (vid. págs. 39, 40).

Islam, Khandaker Reajul et al. «Machine Learning-Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review». En: *Journal of Clinical Medicine* 12.17 (2023). DOI: 10.3390/jcm12175658 (vid. pág. 16).

Jiang, Zhengyu et al. «An explainable machine learning algorithm for risk factor analysis of in-hospital mortality in sepsis survivors with ICU readmission». En: *Computer Methods and Programs in Biomedicine* 204 (2021), pág. 106040. DOI: https://doi.org/10.1016/j.cmpb.2021.106040 (vid. pág. 18).

Johnson, Alistair E. W. et al. «MIMIC-IV, a freely accessible electronic health record dataset». En: *Scientific Data* 10 (1 2023), pág. 1. DOI: 10.1038/s41597-022-01899-x (vid. pág. 21).

Kent, Jessica. «Over 80 % of Health Execs Have Artificial Intelligence Plans in Place». En: *HealthTech Analytics* (2020) (vid. pág. 13).

Kerem, Eitan et al. «Prediction of Mortality in Patients with Cystic Fibrosis». En: *New England Journal of Medicine* 326.18 (1992), págs. 1187-1191. DOI: 10.1056/NEJM199204303261804 (vid. pág. 15).

Kong, G., K. Lin e Y. Hu. «Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU». En: *BMC Medical Informatics and Decision Making* 20.1 (2020), pág. 251. DOI: 10.1186/s12911-020-01271-2 (vid. pág. 18).

Koozi, H. et al. «A simple mortality prediction model for sepsis patients in intensive care». En: *Journal of the Intensive Care Society* 24.4 (2023). Epub 2023 Feb 1, págs. 372-378. DOI: 10.1177/17511437221149572 (vid. pág. 18).

Li, H. et al. «A machine learning-based prediction of hospital mortality in mechanically ventilated ICU patients». En: *PLoS ONE* 19.9 (2024), e0309383. DOI: 10.1371/journal.pone.0309383 (vid. pág. 14).

Liu, Jiachang et al. «FasterRisk: Fast and Accurate Interpretable Risk Scores». En: *Advances in Neural Information Processing Systems* 35 (2022) (vid. pág. 13).

Mao, F. et al. «Machine Learning Approach for Sepsis Risk Assessment in Ischemic Stroke Patients». En: *Journal of Intensive Care Medicine* (2025). Epub ahead of print. DOI: 10.1177/08850666241308195 (vid. págs. 39, 40).

Moonesinghe, S. R. et al. «Risk Stratification Tools for Predicting Morbidity and Mortality in Adult Patients Undergoing Major Surgery: Qualitative Systematic Review». En: *Anesthesiology* 119.4 (2013), págs. 959-981. DOI: 10.1097/ALN.0b013e3182a4e94d (vid. pág. 15).

Mumtaz, H. et al. «APACHE scoring as an indicator of mortality rate in ICU patients: a cohort study». En: *Annals of Medicine and Surgery* 85.3 (2023). Consultado el [fecha en que accediste al artículo], págs. 416-421. DOI: 10.1097/MS9.0000000000000264 (vid. pág. 11).

Pan, X. et al. «Evaluate prognostic accuracy of SOFA component score for mortality among adults with sepsis by machine learning method». En: *BMC Infectious Diseases* 23.1 (2023), pág. 76. DOI: 10.1186/s12879-023-08045-x (vid. págs. 39, 40).

Park, JY et al. «Predicting Sepsis Mortality in a Population-Based National Database: Machine Learning Approach». En: *J Med Internet Res* 24.4 (2022), e29982. DOI: 10.2196/29982 (vid. pág. 17).

Pollard, Tom J. et al. «The eICU Collaborative Research Database, a freely available multi-center database for critical care research». En: *Scientific Data* 5.1 (2018), pág. 180178. DOI: 10.1038/sdata.2018.178 (vid. pág. 21).

Porcel-Gálvez, Ana M. et al. «Factors Associated with In-Hospital Mortality in Acute Care Hospital Settings: A Prospective Observational Study». En: *International Journal of Environmental Research and Public Health* 17.21 (2020), pág. 7951. DOI: 10.3390/ijerph17217951 (vid. pág. 14).

Rahmani, Amir Masoud et al. «Machine Learning (ML) in Medicine: Review, Applications, and Challenges». En: *MATHEMATICS* 9.22 (2021). DOI: 10.3390/math9222970 (vid. pág. 13).

Reinhart, Konrad et al. «Recognizing Sepsis as a Global Health Priority — A WHO Resolution». En: *New England Journal of Medicine* 377.5 (2017), págs. 414-417. DOI: 10.1056/NEJMp1707170 (vid. pág. 10).

Selcuk, Mehtap, Oguz Koc y A. Sevtap Kestel. «The prediction power of machine learning on estimating the sepsis mortality in the intensive care unit». En: *Informatics in Medicine Unlocked* 28 (2022), pág. 100861. DOI: https://doi.org/10.1016/j.imu.2022.100861 (vid. pág. 17).

Shashikumar, Supreeth P. et al. «Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics». En: *Journal of Electrocardiology* 50.6 (2017), págs. 739-743. DOI: https://doi.org/10.1016/j.jelectrocard.2017.08.013 (vid. pág. 16).

Singer, Mervyn et al. «The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)». En: *JAMA* 315.8 (2016), págs. 801-810. DOI: 10.1001/jama.2016.0287 (vid. págs. 10, 11, 21).

Siwicki, Bill. «86 % of healthcare companies use some form of AI». En: *Healthcare IT News* (2017) (vid. pág. 13).

Syed, Mahanazuddin et al. «Application of Machine Learning in Intensive Care Unit (ICU) Settings Using MIMIC Dataset: Systematic Review». En: *INFORMATICS-BASEL* 8.1 (2021). DOI: 10.3390/informatics8010016 (vid. pág. 11).

Thorsen-Meyer, Hans-Christian 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». En: *The Lancet Digital Health* 2.4 (2020), e179-e191. DOI: 10.1016/S2589-7500(20)30018-2 (vid. pág. 14).

Venkataraman, R et al. «Mortality Prediction Using Acute Physiology and Chronic Health Evaluation II and Acute Physiology and Chronic Health Evaluation IV Scoring Systems: Is There a Difference?» En: *Indian Journal of Critical Care Medicine* 22.5 (2018), págs. 332-335. DOI: 10.4103/ijccm.IJCCM_422_17 (vid. pág. 12).

Wang, Dong et al. «A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients». En: *Frontiers in Public Health* 9 (2021). DOI: 10.3389/fpubh.2021.754348 (vid. pág. 16).

Wang, R. Z. et al. «Predictive Models of Sepsis in Adult ICU Patients». En: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. 2018, págs. 390-391. DOI: 10.1109/ICHI.2018.00068 (vid. pág. 16).

Weng, Jie et al. «Development and validation of a score to predict mortality in ICU patients with sepsis: a multicenter retrospective study». En: *Journal of Translational Medicine* 19.1 (2021), págs. 1-12 (vid. pág. 17).

(WHO), World Health Organization. *Sepsis*. 2024 (vid. pág. 10).

Yu, Zhijiang et al. «Prediction of 30-day mortality for ICU patients with Sepsis-3». En: *BMC Medical Informatics and Decision Making* 24.1 (2024), pág. 223. DOI: 10.1186/s12911-024-02629-6 (vid. págs. 39, 40).

Zeng, Zhixuan et al. «Development and validation of a novel blending machine learning model for hospital mortality prediction in ICU patients with Sepsis». En: *BioData Mining* 14.1 (2021), pág. 40. DOI: 10.1186/s13040-021-00276-5 (vid. pág. 18).

ANNEXES

Annex A. List of 77 clinical factors

This appendix provides a detailed breakdown of the 77 clinical factors utilized in the study, categorized into five groups: (1) four clinical-demographic factors, as shown in Table 7; (2) 22 vital signs, presented in Table 8; (3) 32 laboratory analyses, detailed in Table 9; (4) 14 pharmacological clinical factors, outlined in Table 10; and (5) five factors related to the Glasgow Coma Scale and SOFA score, summarized in Table 11.

Demographic clinical factors
Gender, Race, Age, Length of stay in ICU

Table 7. List of 4 demographic clinical factors.

Vital signs clinical factors
Heart rate min, Heart rate max Heart rate mean,
Systolic blood pressure min, Systolic blood pressure max,
Systolic blood pressure mean, Diastolic blood pressure min
Diastolic blood pressure max, Diastolic blood pressure mean
Mean blood pressure min, Mean blood pressure max
Mean blood pressure mean, Respiration rate min
Respiration rate max, Respiration rate mean
Temperature min, Temperature max, Temperature mean
SpO2 min, SpO2 max, SpO2 mean, Urine output

Table 8. List of 22 vital signs clinical factors.

Laboratory clinical factors
Lactate min, Lactate max, Hematocrit min Hematocrit max, Hemoglobin min, Hemoglobin max Platelets min, Platelets max, White blood cells min White blood cells max, Albumin min, Albumin max Anion gap min, Anion gap max, Bicarbonate min Bicarbonate max, Blood urea nitrogen min Blood urea nitrogen max, Calcium min Calcium max, Chloride min, Chloride max Creatinine min, Creatinine max, Glucose min Glucose max, Sodium min, Sodium max Potassium min, Potassium max, Bilirubin min Bilirubin max

Table 9. List of 32 laboratory clinical factors.

Pharmacologic clinical factors
Cefalosporine, Macrolide, Meropenem, Metronidazole Penicillin, Quinolone, Vancomycin, Dopamine Epinephrine, Norepinephrine, Phenylephrine Vasopressin, Dobutamine, Milrinone

Table 10. List of 14 pharmacologic clinical factors.

Glasgow Coma Scales and SOFA
Glasgow Coma Scale min, Glasgow Coma Scale motor Glasgow Coma Scale verbal, Glasgow Coma Scale eyes SOFA score

Table 11. List of 5 Glasgow Coma Scales and SOFA.

Annex B. SOFA score performance

In this appendix, Figure 4 displays the ROC curve derived from the SOFA score model, which was trained using 77 clinical factors with 5-fold cross-validation on the MIMIC-IV v3.0 database and subsequently tested on the eICU v2.0 database.

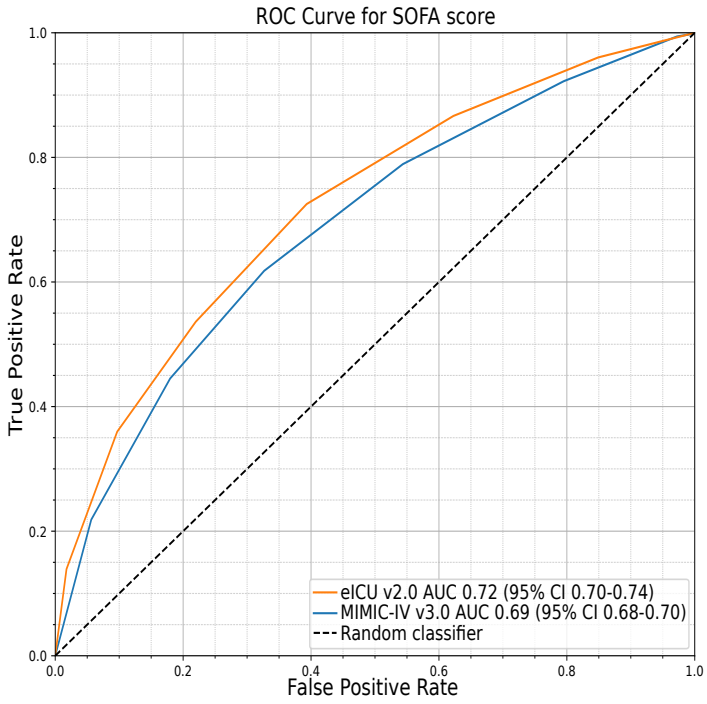

Figure 4. ROC curve of the SOFA score trained with 77 clinical factors with 5-fold cross-validation in the MIMIC-IV v3.0 database and tested in the eICU v2.0 database.

Annex C. List of 25 clinical factors

This appendix details the list of 25 clinical factors, as shown in Table 12, which were generated using the sequential feature selector algorithm.

Clinical Factors
SOFA score
Urine output
Vasopressin
Glasgow Coma Scale eye
Glasgow Coma Scale verbal
Glasgow Coma Scale motor
Vancomycin
Cefalosporine
Blood urea nitrogen max
Sodium max
Platelets max
Platelets min
Lactate max
Glucose min
Chloride max
Bilirubin max
Anion gap max
Systolic blood pressure mean
Respiration rate mean
Heart rate max
SpO2 min
SpO2 max
Temperature mean
Length of stay in ICU
Age

Table 12. List of 25 clinical factors.