



medicina *intensiva*

<http://www.medintensiva.org/en/>



ORIGINAL ARTICLE

Toward transparent clinical decision support in the ICU: A multi-window, model-agnostic explainability approach for 72-h mortality prediction using the eICU collaborative research database

Daniel Gallardo Gómez^{a,b,1}, Sara Díaz^{b,1}, Javier Quintero Sosa^{a,b,1},
Claudia Jiménez Vázquez^b, Álvaro Ritoré-Hidalgo^b, Antonio Gutiérrez-Pizarraya^{a,c},
Miguel Ángel Armengol de la Hoz^{b,*}, María Recuerda Núñez^{d,2}, Ángel Estella^{d,e,f,2}

^a Health Technology Assessment Area (AETSA), Progress and Health Public Foundation (FPS), Seville, Spain

^b Data Science Lab, Progress and Health Public Foundation (FPS), Seville, Spain

^c Research Group IBiS CS-017 'Lung Cancer Research, Seville, Spain

^d Intensive Care Unit, University Hospital of Jerez, Jerez, Spain

^e Department of Medicine, University of Cádiz, Cádiz, INIBICA, Spain

^f Biomedical Research and Innovation Institute of Cádiz (INIBICA) Research Unit, University of Cádiz, Cádiz, Spain

Received 16 July 2025; accepted 12 October 2025

KEYWORDS

Intensive Care Unit;
Mortality;
Database;
Machine learning;
Clinical decision
support system

Abstract

Objective: This study has two main objectives: (1) to develop a multi-model framework for predicting Intensive Care Unit (ICU) mortality within the first 72 h of admission; and (2) to introduce a novel model-agnostic explainability approach classification that enables variable-level interpretation of predicted probabilities.

Design: Retrospective study using a multi-model machine learning approach, analyzing data across multiple time windows and incorporating demographic, clinical, and biochemical variables.

Setting: ICU mortality.

Patients or participants: Patients included in the eICU database over 16 years old who have been admitted to ICUs in 2014 and 2015 with available data within the first 72 h after ICU admission. A total of 106,449 patients were included in the analyses.

Interventions: No clinical interventions were applied; this was a retrospective analysis for predictive model development and evaluation.

Main variables of interest: Demographic, clinical, and biochemical variables collected across multiple time windows.

* Corresponding author.

E-mail address: fpsdata.fps@juntadeandalucia.es (M.Á. Armengol de la Hoz).

¹ Co-first authors, contributed equally.

² Co-senior authors, contributed equally.

<https://doi.org/10.1016/j.medine.2025.502349>

2173-5727/© 2025 Published by Elsevier España, S.L.U.

Please cite this article as: D. Gallardo Gómez, S. Díaz, J. Quintero Sosa et al., Toward transparent clinical decision support in the ICU: A multi-window, model-agnostic explainability approach for 72-h mortality prediction using the eICU collaborative research database, *Medicina Intensiva*, <https://doi.org/10.1016/j.medine.2025.502349>

Results: A total of 106,449 were included (mean age 62.6 years, 46% women), with an overall 72-h mortality of 4.8%. Random Forest models achieved one of the best results in terms of predictive performance metrics, with F1-scores of 0.93, 95% CI 0.93 to 0.94; 0.92, 95% CI 0.92 to 0.93 and 0.83, 95% CI 0.83 to 0.85 across the three temporal data windows. Due to these metrics, the ability to predict deaths, and the biological plausibility of the predictions, Random Forest models were selected from all those studied.

Conclusions: The proposed multi-model approach significantly improves 72-h ICU mortality prediction. Moreover, we outline a model-agnostic strategy for variable-level interpretation of predicted probabilities, which may facilitate transparency and support future applications in clinical decision support.

© 2025 Published by Elsevier España, S.L.U.

PALABRAS CLAVE

Unidad de cuidados intensivos;
Mortalidad;
Base de datos;
Aprendizaje automático;
Sistema de apoyo a la decisión clínica

Hacia un apoyo transparente a la toma de decisiones clínicas en la UCI: un enfoque de explicabilidad multiventana e independiente del modelo para la predicción de la mortalidad a 72 horas utilizando la base de datos de investigación colaborativa eICU

Resumen

Objetivo: Este estudio tiene dos objetivos principales: (1) desarrollar un multimodelo para predecir la mortalidad en la Unidad de Cuidados Intensivos (UCI) dentro de las primeras 72 horas de ingreso; e (2) introducir una novedosa metodología aplicable a cualquier modelo de clasificación que permita la interpretación a nivel variable de las probabilidades predichas.

Diseño: Estudio retrospectivo utilizando un enfoque de aprendizaje automático multimodelo, analizando los datos a través de múltiples ventanas temporales e incorporando variables demográficas, clínicas y bioquímicas.

Ámbito: mortalidad en UCI.

Pacientes o participantes: Pacientes incluidos en la base de datos eICU mayores de 16 años ingresados en UCI en 2014 y 2015 con datos disponibles en las primeras 72 horas tras el ingreso en UCI, lo que significa que se incluyeron en los análisis un total de 106.449 pacientes.

Intervenciones: No se aplicaron intervenciones clínicas; se trató de un análisis retrospectivo para el desarrollo y la evaluación de modelos predictivos.

Variables de interés principales: Variables demográficas, clínicas y bioquímicas recogidas a lo largo de múltiples ventanas temporales.

Resultados: Se incluyeron 106,449 pacientes (edad media 62,6 años; 46% mujeres), con una mortalidad global a 72 horas del 4,8%. Los modelos Random Forest consiguieron uno de los mejores resultados en cuanto a métricas de rendimiento predictivo, con F1-score de 0.93, IC del 95 %: 0.93 a 0.94; 0.92, IC del 95 %: 0.92 a 0.93 y 0.83, IC del 95 %: 0.83 a 0.85 en las tres ventanas de datos temporales. Debido a estas métricas, la capacidad de predecir muertes y la plausibilidad biológica de las predicciones, se seleccionaron los modelos de Random Forest de entre todos los estudiados.

Conclusiones: El enfoque multimodelo propuesto mejora significativamente la predicción de la mortalidad en la UCI a las 72 horas. Asimismo, presentamos una estrategia agnóstica al modelo para la interpretación a nivel de variable de las probabilidades predichas, que puede contribuir a la transparencia y servir de base para futuras aplicaciones en sistemas de apoyo a la decisión clínica.

© 2025 Publicado por Elsevier España, S.L.U.

Introduction

Decisions about Intensive Care Unit (ICU) admission and continued stay are often made under uncertainty, with incomplete data and rapidly evolving physiology.¹⁻³ Conventional severity scores (e.g., APACHE, SAPS) support population-level description but have shown variable reliability and limited individual-level prognostic utility,

especially around the peri-admission period when clinical status can change quickly.⁴⁻⁸

Early ICU deaths represent a clinically distinct, high-severity phenotype in which stabilization fails despite prompt care, and where delays or inadequate early response can have outsized consequences for outcomes and adverse events: personal costs to the patient beyond the risks of invasive treatments, limitations on family visits, increased

risk of healthcare-associated infections, potential need for patient sedation, substantial economic costs, and unnecessary consumption of limited ICU resources.⁹ Evidence underscores how frequent early mortality is: in disease-specific cohorts (e.g., community-acquired septic shock), ~56% of ICU deaths have been reported within the first 72 h of admission¹⁰; a >600-patient cohort of community-acquired septic shock reported 14.4% mortality¹¹; and in a 9-year ICU cohort of >6500 patients, 42.8% died within the first 5 days of ICU stay.¹² In our eICU cohort, 4.82% of all admissions died within 72 h. Together, these data highlight the 72-h window as a critical period for triage and escalation decisions.

Rather than a single static prediction, we examine three progressively extended peri-admission windows: A (–24 h to ICU admission), B (–24 h to +24 h), and C (–24 h to +48 h). These windows mirror real checkpoints in ICU workflow-admission triage (A) and reassessment for continued ICU stay (B, C).^{2,3} Although windows B and C are temporally closer to the outcome, they address different clinical decisions (ongoing ICU allocation and escalation) than window A. This framing mitigates concerns about proximity-to-outcome by aligning each model with a distinct decision point.

Leveraging the growing ecosystem of publicly available healthcare datasets, we develop a multi-model framework that predicts 72-h mortality using routinely collected demographic, clinical, and biochemical variables across peri-admission windows, advancing data-driven clinical knowledge,^{13–16} and we introduce a model-agnostic, probability-based explainability approach that standardizes variable-level interpretation across classifiers.^{17,18} This work is a methodological step toward transparent, temporally aware tools that could support future CDSS in intensive care, not a ready-to-deploy system.

Patients and methods

Retrospective data were obtained from the eICU Collaborative Research Database, a freely available multi-center database for critical care research.¹⁹ The eICU Collaborative Research Database is populated with data from a combination of 335 critical care units at 208 hospitals throughout the continental United States. The data in the collaborative database covers patients who were admitted to critical care units in 2014 and 2015. Data on patients admitted to intensive care have been collected in this database, such as vital sign measurements, care plan documentation, disease severity measures, diagnostic information, treatment information, biomarkers, and blood sample parameters, among others.

To ensure replicability, a deterministic approach was carried out based on seeds' establishment, which ensures that the sequence of random numbers generated by the algorithm is the same every time the code is run. To ensure transparency, we provide open access to all our data extraction, filtering, wrangling, and modeling, and table creation procedures through a publicly available GitHub repository.²⁰

All the analyses were conducted in R (version 4.3.1).²¹ We used the *dplyr* package for data wrangling²²; the *caret* package (version 6.0–94) to train our predictive models²³;

and the *ggplot2* package (version 3.5.0)²⁴ for data plotting and visualization.

Study population

We included patients over 16 years old who have been admitted to ICUs with available data within the first 72 h after ICU admission. After applying these selection criteria, and conducting data cleaning and management procedures, a total of 106,449 patients were included in the analyses. The flowchart of patient selection from the eICU database, including the exclusion criteria, is presented in Fig. 1. Besides, in the Supplementary File the procedure carried out in order to obtain the final datasets used is described.

Outcome and predictors

The outcome of interest was ICU mortality within the first 72 h after admission, which presented imbalance classes (mean prevalence of alive patients=95%; and deceased patients=5%). Data to predict mortality included: (1) personal and demographic data as patient age, gender, and ethnicity; and (2) clinical data and biomarkers such as the time between hospital admission and ICU admission, differences between analyzed time points of bicarbonate, sodium, blood urea nitrogen (BUN), glucose, potassium, hemoglobin, and calcium serum concentrations, differences in respiratory and heart rates, differences in blood pressures parameters (i.e., diastolic, systolic, and mean values), administration of norepinephrine, underlying conditions as diabetes, congestive heart failure, chronic obstructive pulmonary disease (COPD), cancer, or chronic kidney disease. We also recorded Charlson Comorbidity index, and coronary artery bypass grafting past surgery, or past ICU stay due to cardiac arrest, congestive heart failure, cerebrovascular accident, diabetic ketoacidosis, myocardial infarction, rhythm disturbance, kidney or pulmonary sepsis, and whether the hospital was linked to university teachings or not.

Data analysis

We developed a multi-model approach based on three ML models to predict 72-h ICU mortality, each using a different time frame (data windows A, B, and C) as input data: (1) model A uses data from 24 h before patient admission up to the moment of ICU admission; (2) model B from 24 h pre-admission to 24 h post-admission; and (3) model C from 24 h pre-admission to 48 h post-admission (Fig. 2).

The rationale for these windows is not only methodological but also clinical, as they correspond to distinct decision points in ICU workflow. Model A supports decisions regarding ICU admission, integrating pre-admission and admission data. Model B enables early reassessment after 24 h in the ICU, when treatment response becomes observable. Model C, although temporally closer to the outcome and therefore yielding the best predictive performance, is designed for continued ICU stay evaluation at 48 h, reflecting a common practice of intensivists to reassess prognosis and adapt resource allocation.

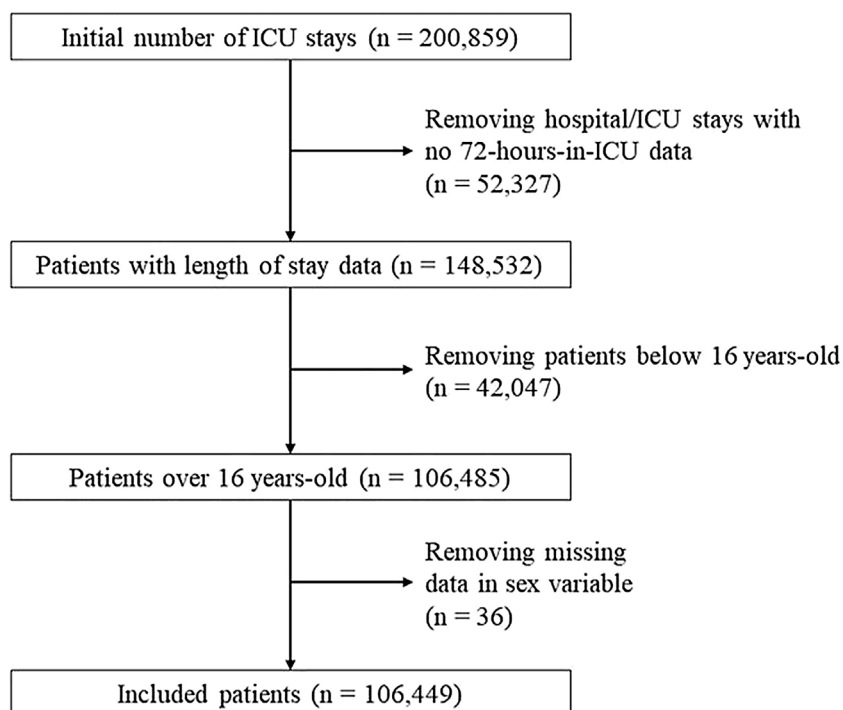


Figure 1 Flowchart of patient selection from eICU dataset.

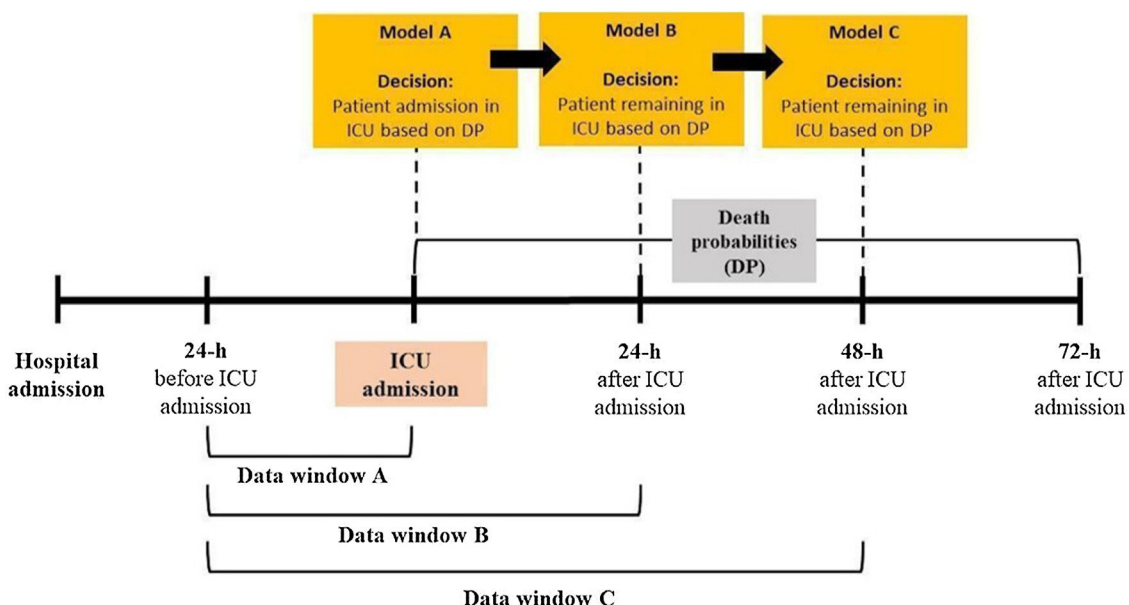


Figure 2 Schematic of the three data windows that are considered in the multi-model approach.

Thus, the three models should not be viewed as alternatives but as a sequential decision-support framework, aligning with the evolving nature of critical care management and capturing different phases of the peri-admission period.

For each data window, several classification ML algorithms have been trained and benchmarked to tackle our research problem. In particular, the following models have been tested: stochastic gradient boosting, Random Forest, eXtreme gradient boosting, boosted logistic regression,

(robust) linear discriminant analysis, partial least squares, and boosted classification trees.²⁵ We used 10-fold repeated cross-validation with random search for parameter tuning, and down-sampling method to tackle class imbalance.

The outcome obtained from the different ML algorithms was the probability of 72-h ICU mortality. Once these probabilities have been obtained, it is necessary to adjust the threshold at which the model predicts patient death or survival. We have set clinically optimal thresholds considering (1) the cost of false positive errors (i.e. people who were

predicted to die, and they actually do not), (2) performance measures, and (3) the prevalence of the target outcome in our population. Additionally, we plotted the precision-recall curve applying 0.02-increases from 0 to 1 for classification thresholds.

Taking into account all the ML algorithms that have been designed, we selected the best models for each data window based on: (1) biological plausibility of the posterior predictions on the test dataset (i.e., whether predicted responses on variables of interest make sense biologically), (2) performance measures such as F1-score and balanced accuracy obtained on the test and train datasets, and (3) number of true positives (i.e., deaths).

Explainability

Next, in order to explore variable contributions in a consistent way across models, we implemented a model-agnostic explainability approach based on marginal predicted probabilities. Unlike previous approaches, this method standardizes the interpretation process across any classification model, allowing consistent and temporally stratified visualization of each variable's marginal contribution to mortality risk.²⁶

For each selected model, we identified the three variables with the highest predictive relevance and visualized their marginal associations with the predicted probability of 72-h ICU mortality.

For continuous variables (e.g., respiratory rate difference), predicted probabilities were estimated across the observed range in the test dataset. We fitted a cubic spline to summarize the relationship between the variable and mortality risk, including a 95% confidence interval (CI) based on bootstrap sampling (10 iterations).

For categorical variables (e.g., norepinephrine administration), we estimated the median predicted probability and associated 95% confidence interval for each category using the same bootstrap procedure.

The formulas proposed for the analysis are presented below. First, for the case of categorical variables:

$$\mathcal{P}_k = \text{median} \left\{ \mathcal{P}_k^{(b)} \right\}_{b=1}^B \quad \text{with} \quad CI = [\mathcal{P}_{k,0.025}, \mathcal{P}_{k,0.975}],$$

Where $\mathcal{P}_k^{(b)}$ is the mean predicted probability for all patients with $X_j = c_k$ in the bootstrap b and c_k is the categorical variable k .

For the case of continuous variables, the equation of the probability of death is:

$$\mathcal{P}(Y = 1 | X_j = x_j) = \frac{1}{B} \sum_{b=1}^B f^{(b)}(x_j, \mathbf{z}_i^{(b)}),$$

Where $f^{(b)}$ is the prediction of the model in the bootstrap sample b , and $\mathbf{z}_i^{(b)}$ represents the remaining fixed covariates for that bootstrap. The range of x_j is traversed completely (e.g. from -40 to +40 for the respiratory rate difference) and fitted by a cubic spline on the (x_j, \mathcal{P}) .

Bias and fairness analysis

We studied if the models perform consistently across different potentially vulnerable subgroups of patients. The analyzed subgroups were categorized by age, according to classifications defined by the World Health Organization (i.e., children, adolescence, youth, adults, and older adults)²⁷; sex (i.e., men and women); ethnicity (i.e., Hispanic, African American, Caucasian, Native American, Asian, or other); and hospital financial status assessed by whether the hospital had linked teaching actions (i.e., yes or no).

Results

As we mentioned in subsection 2.1, we included 106,449 patients admitted to ICU from the eICU database (mean age = 62.65; SD = 17.43), from which 49,148 (46.20%) were women and the majority of patients considered themselves Caucasians (77%). From the study population, 5126 patients died within the first 72 h from ICU admission (4.82%). Information about constant variables (i.e., sociodemographic, and clinical history data), and time-dependent variables (i.e., variables that varied across data windows) is fully detailed in [Tables 1 and 2](#), respectively.

Models' performance

As we mentioned in Section 2, in our study several models have been trained and analyzed for each data window in order to achieve the best one. Details about models' performance measures of all the trained models are shown in [Table 3](#).

As can be seen in [Table 3](#), Random Forest (rf) is one of the models that present better performance metrics for the different temporal windows. Besides, eXtreme Gradient Boosting (xgb) and Boosting Classification Trees (ada) also present good results. Taking into account the (1) biological plausibility of the posterior predictions on the test dataset, (2) performance measures, and (3) number of true positives, an optimal threshold of 0.40 (i.e., if the probability of mortality is 60% or less, the prediction was that the patient will live) was determined for patient status classification (see the Supplementary Figs. 1-3). We optimized our predictive models using the F1-score. Random Forest was the trained model that achieved the best trade-off between F1-scores, the capacity to predict true positives (deaths), and biological plausibility of the predictions within all data windows. Other models such as eXtreme Gradient Boosting, boosted classification trees, and stochastic gradient boosting, also performed well. As can be seen in [Table 3](#), comparing F1-scores between selected models, the model C performed better ($F1_{\text{mean}} = 0.93$; 95% CI 0.93 to 0.94) on average than model A ($F1_{\text{mean}} = 0.83$; 95% CI 0.83 to 0.85), and B ($F1_{\text{mean}} = 0.92$; 95% CI 0.92 to 0.93) predicting death probabilities.

Models' explainability

We applied the above method to the best-performing Random Forest models for each data window. The most important predictors differed by model. For data window

Table 1 Constant study sample characteristics.

	All patients	ICU mortality		SMD (95% CI)
		Alive	Exitus	
<i>N</i>	106,449	101,322	5127	
Age (mean, SD)	62.65 (17.43)	62.31 (17.47)	69.33 (15.12)	0.40 (0.38, 0.43)*
Sex: Men (<i>n</i> , %)	49,148 (46.17)	46,787 (46.18)	2361 (46.05)	0.00 (−0.03, 0.03)
Ethnicity (<i>n</i> , %)				
African American	11,724 (11.00)	11,208 (11.10)	516 (10.10)	
Asian	1518 (1.40)	1440 (1.40)	78 (1.50)	
Caucasian	82,091 (77.10)	78,065 (77.00)	4026 (78.50)	
Hispanic	4297 (4.00)	4127 (4.10)	170 (3.30)	
Native American	793 (0.70)	754 (0.70)	39 (0.80)	
Other/Unknown	6026 (5.70)	5728 (5.70)	298 (5.80)	
Teaching status = University (<i>n</i> , %)	27,433 (25.80)	25,944 (20.60)	1489 (29.00)	0.19 (0.17, 0.22)*
Hours from general admission to ICU admission (mean, SD)	26.13 (99.32)	24.98 (97.13)	48.88 (133.48)	0.24(0.21,0.27)*
Diabetes diagnosis (<i>n</i> , %)	31,300 (29.40)	29,866 (29.50)	1434 (28.00)	−0.03 (−0.06, −0.005)*
Congestive heart failure history (<i>n</i> , %)	14,982 (14.10)	13,950 (13.80)	1032 (20.10)	0.17 (0.14, 0.20)*
COPD history (<i>n</i> , %)	15,352 (14.40)	14,439 (14.30)	913 (17.80)	0.10 (0.07, 0.12)*
Cancer history (<i>n</i> , %)	13,901 (13.10)	12,890 (12.70)	1011 (19.70)	0.19 (0.16, 0.22)*
Renal injury history (<i>n</i> , %)	12,778 (12.00)	11,869 (11.70)	909 (17.70)	0.17 (0.14, 0.20)*
Charlson score (mean, SD)	3.79 (2.69)	3.73 (2.67)	5.01 (2.89)	0.48 (0.45, 0.50)*
CABG (<i>n</i> , %)	2992 (2.80)	2980 (2.90)	12 (0.20)	−0.22 (−0.25, −0.19)*
Cardiac arrest (<i>n</i> , %)	1957 (1.80)	948 (0.90)	1009 (19.70)	0.65 (0.62, 0.69)*
Congestive heart failure diagnosis (<i>n</i> , %)	3567 (3.40)	3404 (3.40)	163 (3.20)	−0.01 (−0.04, 0.02)
Cerebral vascular diagnosis (<i>n</i> , %)	4228 (4.00)	4057 (4.00)	171 (3.30)	−0.04 (−0.07, −0.009)*
Diabetic ketoacidosis (<i>n</i> , %)	3716 (3.50)	3702 (3.70)	14 (0.30)	−0.25 (−0.27, −0.22)*
Myocardial infarction (<i>n</i> , %)	5180 (4.90)	5067 (5.00)	113 (2.20)	−0.15 (−0.18, −0.12)*
Rhythm disturbance (<i>n</i> , %)	3064 (2.90)	2988 (2.90)	76 (1.50)	−0.10 (−0.12, −0.07)*
Renal sepsis (<i>n</i> , %)	3045 (2.90)	2869 (2.80)	176 (3.40)	0.04 (0.007, 0.06)*
Pulmonary sepsis (<i>n</i> , %)	4220 (4.00)	3705 (3.70)	515 (10.00)	0.25 (0.22, 0.28)*

Notes. COPD: chronic obstructive pulmonary disease; CABG: coronary artery bypass graft surgery. SMD: standardized mean difference; these standardized mean differences were computed calculating the difference between both groups (Expired – Alive) divided by the pooled standard deviation difference; these standardized mean differences were computed calculating the difference between both groups (Expired – (greater SMD values indicate a greater magnitude of the variable in the “Expired” group)). A threshold of |SMD|<0.10 was considered adequate balance. * means statistical significance. Confidence intervals (95% CI) were also reported; if the CI did not include zero, the imbalance was considered statistically different from zero.

A, the most relevant variables were: time from general hospital admission to ICU admission, age, and history of cardiac arrest. For data windows B and C, the most relevant variables were norepinephrine administration, respiratory rate difference, and temperature difference (see Supplementary Figs. 4–6).

For each of these variables, we plotted the predicted probability of death as a function of the variable’s value (Fig. 3). This representation allows a better understanding of the association between individual predictors and mortality risk, regardless of the underlying model structure.

The methodology can be applied to any classification model. It aims to provide a reproducible and interpretable summary of variable-level effects without relying on model-specific mechanisms.

For the initial window, previous to ICU admission (model A), the predicted probability of 72-h ICU mortality gradually increased if patients spent more than 10h in general hospital wards, had more than 45–50 years old (e.g., from 45 to 65 years old there was an increase on mortality probabili-

ty of 12%), and suffered a cardiac arrest (the probability of death of patients who suffered a cardiac arrest = 0.92 [95% CI 0.82 to 0.97]; the probability of patients who did not = 0.32 [95% CI 0.23 to 0.43]). Although this association is clinically expected, its consistent emergence as a key predictor confirms the biological plausibility of the framework and strengthens confidence in its application to less obvious variables.

Regarding the second window, patients that required norepinephrine administration had associated higher mortality risk (0.66; 95% CI 0.50–0.77) than those who did not (0.24; 95% CI 0.16–0.32). In case of respiratory rate difference, if patients presented high values, the probability of mortality increased (e.g., −25 chest raises doubled mortality probabilities compared to no difference between before and after ICU admission). Patients whose body temperature varied more than 1 degrees Celsius (°C) had increased mortality probabilities. Similar patterns were found in the last addressed window (model C).

Table 2 Time-dependent study sample characteristics.

	Data window A (–24 h – ICU admission)				Data window B (–24 h –24 h)				Data window C (–24 h –48 h)			
	All patients	ICU mortality		SMD (95% CI)	All patients	ICU mortality		SMD (95% CI)	All patients	ICU mortality		SMD (95% CI)
		Alive	Exitus			Alive	Exitus			Alive	Exitus	
Bicarbonate	–0.84 (1.40)	–0.83 (1.37)	–0.94 (1.83)	–0.08 (–0.11, –0.05) ^a	0.13 (3.40)	0.20 (3.30)	–1.27 (4.89)	–0.43 (–0.46, –0.40) ^a	0.68 (3.96)	0.79 (3.85)	–1.40 (5.31)	–0.55 (–0.58, –0.53) ^a
Sodium	0.14 (1.39)	0.13 (1.35)	0.26 (1.92)	0.09 (0.07, 0.12) ^a	1.01 (3.49)	0.95 (3.39)	2.08 (4.98)	0.32 (0.30, 0.35) ^a	0.97 (4.16)	0.91 (4.05)	2.31 (5.73)	0.34 (0.31, 0.37) ^a
Respiratory rate	–0.16 (4.09)	–0.19 (4.01)	0.40 (5.35)	0.14 (0.12, 0.17) ^a	–0.28 (6.64)	–0.18 (6.33)	–2.11 (10.93)	–0.29 (–0.32, –0.26) ^a	–0.50 (6.65)	–0.33 (6.21)	–3.85 (12.08)	–0.53 (–0.56, –0.50) ^a
BUN	0.67 (2.11)	0.66 (2.07)	0.88 (2.76)	0.10 (0.08, 0.13) ^a	–1.66 (7.89)	–1.79 (7.76)	0.86 (9.75)	0.34 (0.31, 0.36) ^a	–2.62 (11.37)	–2.81 (11.26)	1.19 (12.70)	0.35 (0.32, 0.38) ^a
Diastolic BP	–3.87 (12.12)	–3.85 (11.88)	–4.19 (16.13)	–0.03 (–0.06, 0.00)	–5.46 (18.31)	–5.30 (17.84)	–8.69 (25.62)	–0.19 (–0.21, –0.16) ^a	–3.90 (18.71)	–3.57 (18.13)	–10.30 (26.99)	–0.36 (–0.39, –0.33) ^a
Systolic BP	–5.02 (17.78)	–5.01 (17.49)	–5.18 (22.78)	–0.01 (–0.04, 0.02)	–6.54 (27.22)	–6.27 (26.61)	–11.69 (37.02)	–0.20 (–0.23, –0.17) ^a	–4.50 (28.13)	–3.98 (27.34)	–14.69 (39.33)	–0.38 (–0.41, –0.35) ^a
Mean BP	–2.91 (9.96)	–2.92 (9.79)	–2.76 (12.90)	0.02 (–0.01, 0.04)	–4.44 (19.05)	–4.27 (18.55)	–7.68 (26.88)	–0.18 (–0.21, –0.15) ^a	–3.08 (19.73)	–2.75 (19.10)	–9.60 (28.73)	–0.35 (–0.38, –0.32) ^a
Temperature	–0.01 (0.42)	–0.01 (0.41)	–0.01 (0.61)	0.00 (–0.03, 0.03)	0.10 (0.81)	0.10 (0.75)	0.20 (1.50)	0.12 (0.10, 0.15) ^a	0.07 (0.81)	0.06 (0.75)	0.27 (1.52)	0.26 (0.23, 0.29) ^a
Glucose	–2.01 (56.02)	–2.32 (56.32)	4.27 (49.36)	0.12 (0.09, 0.15) ^a	–23.09 (109.75)	–23.54 (109.36)	–14.11 (116.73)	0.09 (0.06, 0.11) ^a	–24.22 (109.95)	–24.69 (109.41)	–14.82 (119.63)	0.09 (0.06, 0.12) ^a
Potassium	–0.01 (0.27)	–0.01 (0.26)	0.00 (0.36)	0.04 (0.01, 0.07) ^a	–0.07 (0.66)	–0.07 (0.64)	0.02 (0.95)	0.14 (0.11, 0.16) ^a	–0.11 (0.72)	–0.12 (0.70)	0.09 (1.02)	0.29 (0.26, 0.32) ^a
Hemoglobin	–0.63 (0.95)	–0.63 (0.94)	–0.59 (1.05)	0.04 (0.01, 0.07) ^a	–0.69 (1.38)	–0.69 (1.35)	–0.68 (1.86)	0.01 (–0.02, 0.04)	–0.86 (1.54)	–0.86 (1.51)	–0.85 (2.00)	0.01 (–0.02, 0.04)
Calcium	–0.28 (0.43)	–0.28 (0.42)	–0.27 (0.59)	0.02 (–0.01, 0.05)	–0.32 (0.65)	–0.31 (0.63)	–0.44 (1.07)	–0.20 (–0.23, –0.17) ^a	–0.28 (0.71)	–0.27 (0.68)	–0.47 (1.13)	–0.28 (–0.31, –0.25) ^a
Heart rate	–1.51 (11.79)	–1.56 (11.64)	–0.58 (14.46)	0.08 (0.06, 0.11) ^a	–5.44 (20.06)	–5.35 (19.35)	–7.17 (30.81)	–0.09 (–0.12, –0.06) ^a	–5.69 (20.96)	–5.50 (20.09)	–9.46 (33.61)	–0.19 (–0.22, –0.16) ^a
Norepinephrine (n, %)	3620 (3.40)	2984 (2.90)	636 (12.40)	0.36 (0.34, 0.39) ^a	8357 (7.90)	6328 (6.20)	2029 (39.60)	0.87 (0.84, 0.90) ^a	8614 (8.10)	6479 (6.40)	2135 (41.60)	0.91 (0.88, 0.93) ^a

Legend. BUN: blood urea nitrogen; BP: blood pressure. All values are presented as mean (SD) unless other units are specifically stated. SMD: standardized mean difference; these standardized mean differences were computed calculating the difference between both groups (Expired – Alive) divided by the pooled standard deviation (greater SMD values indicate a greater magnitude of the variable in the “Expired” group).

^a Statistical significance (95% confidence interval did not include the zero).

Table 3 Balanced accuracy, F1-score and recall for different data windows and models.

	Data window A (-24h - ICU admission)			Data window B (-24h - +24h)			Data window C (-24h - +48h)		
	Balanced accuracy	F1-score	Recall	Balanced accuracy	F1-score	Recall	Balanced accuracy	F1-score	Recall
gbm	0.68	0.82 (0.81,0.82)	0.77	0.81	0.89 (0.89,0.90)	0.83	0.85	0.92 (0.91,0.92)	0.87
rf	0.69	0.83 (0.83,0.85)	0.86	0.80	0.92 (0.92,0.93)	0.93	0.83	0.93 (0.93,0.94)	0.94
ranger	0.68	0.83 (0.83,0.84)	0.89	0.79	0.91 (0.91,0.92)	0.92	0.81	0.93 (0.92,0.93)	0.93
xgb	0.69	0.85 (0.84,0.85)	0.86	0.82	0.91 (0.90,0.92)	0.89	0.86	0.93 (0.93,0.93)	0.91
lr	0.67	0.85 (0.75,0.91)	0.84	0.72	0.95 (0.90,0.96)	0.94	0.78	0.92 (0.90,0.95)	0.90
Linda	0.67	0.86 (0.85,0.87)	0.90	0.76	0.92 (0.91,0.93)	0.92	0.80	0.93 (0.92,0.93)	0.93
kernelpls	0.63	0.87 (0.87,0.88)	0.96	0.71	0.92 (0.91,0.92)	0.97	0.73	0.93 (0.93,0.93)	0.97
ada	0.68	0.86 (0.85,0.87)	0.92	0.82	0.91 (0.91,0.92)	0.89	0.86	0.93 (0.92,0.93)	0.89
cforest	0.69	0.83 (0.82,0.83)	0.86	0.79	0.92 (0.91,0.92)	0.93	0.67	0.97 (0.97,0.98)	0.99
lda	0.68	0.86 (0.85,0.88)	0.89	0.77	0.92 (0.92,0.92)	0.91	0.79	0.93 (0.92,0.93)	0.92

Notes. The 95% confidence interval of the F1-scores were calculated obtaining the standard deviation (SD) of the F1-scores from the k-folds of the same model, and computing lower and upper interval bounds as the mean F1-score - 1.96*SD, and F1-score + 1.96*SD, respectively. Positive cases are those patients that expired.

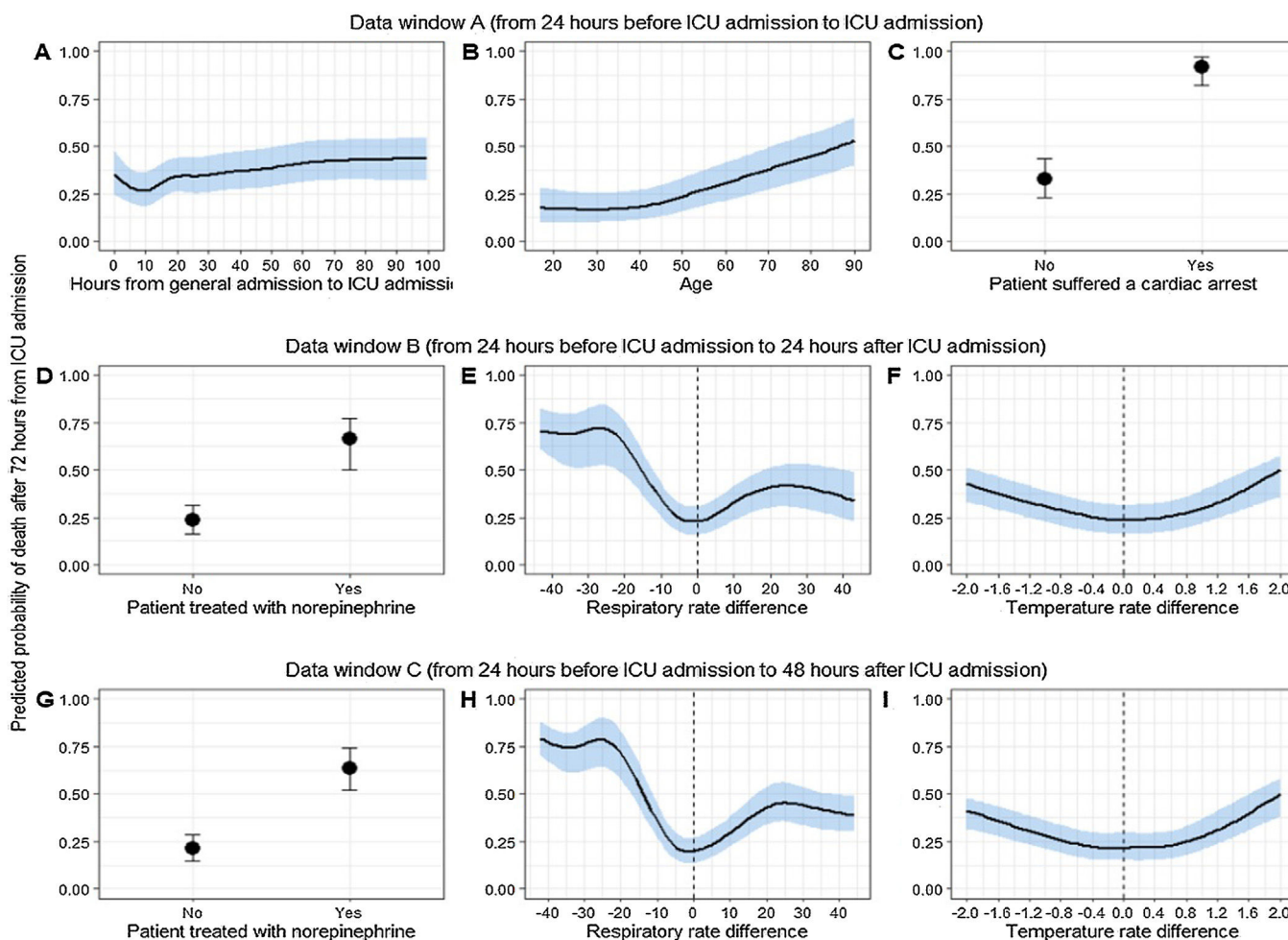


Figure 3 Predicted mortality probabilities for each key predictor within models A, B, and C. The point estimate represents the median probability value, with error bars and shades indicating the 95% CI.

Clinical decision-support system

The main objective of this research was to develop a multi-model approach based on ML to predict the early mortality within the first 72 h of ICU admission taking into account three different temporal configurations. Table 4 illustrates how this system can support clinical decision-making by comparing the 72-h mortality probabilities predicted by our multi-model approach with actual patient outcomes, highlighting potential outcomes for different cases. In particular, the multi-model approach developed would have supported the decision to decline ICU admission of 132 patients who ultimately expired within the first 72 h: 73 of these patients, with a 95% CI upper bound above 60% prior to ICU admission, remained in the ICU for more than 24 h, while 32 patients under the same conditions stayed for more than 48 h.

Table 4 presents four different examples: 1) patient ID #3342213 had a very high mortality probability before ICU admission (using data from window A) and expired within 72 h in the ICU; 2) patient ID #923407 had a medium mortality probability value before ICU admission (model A) but a high one after 24 h in ICU (model B) and expired within 72 h in the ICU; 3) patient ID #386538 is an example of system failure where a low and medium probability of mortality value was predicted using models A and B, but the patient finally died after less than 37 h of ICU admission; 4) patient ID #163973 had a very low mortality probability before ICU admission (model A), after 24 and 48 h (models B and C), and was discharged from the ICU.

Conversely, this multi-model approach underestimated the mortality probability in 120 patients who remained in the ICU for more than 48 h, with predictions yielding 95% CI lower bounds below 60%. For instance, patient ID #386538 stayed in the ICU and had a low mortality risk estimated by our model (data window B), but ultimately expired. When assessing time efficiency, the system potentially saved 283.58 h, compared to 276.01 h spent on patients who did not survive, resulting in a 7.54-h positive trade-off in decision-making efficiency.

This reduction in ICU stay time could potentially translate into lower healthcare costs per patient and increased bed availability for new admissions. However, these implications should be considered exploratory and illustrative, as our study was not designed to directly evaluate economic outcomes. Further validation in prospective and context-specific studies would be required before drawing firm conclusions on resource management.

Bias and fairness analysis

In order to analyse in depth the results obtained, the population has been separated into different groups. In particular, as we have mentioned in Section 2.5, we have decided to analyse four different classifications separating the population according to age, sex, ethnicity and hospital financial status.

In Table 5 we have presented the different F1-scores obtained for the global test dataset and for each group mentioned before for each data window. This can also be seen in Supplementary Figs. 7–9.

As can be seen in Table 5 there are some differences between the classes defined in the different groups, the most significant one is the observed age group between the class ‘‘Older adults’’ and the rest.

Discussion

One of the main goals of this study was to develop a multi-model framework capable of estimating early ICU mortality risk based on different temporal data configurations. While our best-performing model (model C) relies on information very close to the outcome and therefore cannot be considered a stand-alone clinical decision tool, this work represents a methodological contribution. Specifically, it provides a first step toward the development of transparent, temporally-aware approaches that may eventually support clinical decision-making in ICU settings, but further validation and adaptation are required before any practical implementation. This framework may enhance clinical decision-making under uncertainty, particularly for patients who do not clearly fall into high- or low-risk categories. In such cases, the tool offers additional support to guide ICU admission decisions, where uncertainty is often greatest.

Unlike traditional ICU prognostic scores, our approach integrates large-scale clinical databases with experiential knowledge, resulting in a multi-model strategy based on Random Forest classifiers across three peri-admission windows (24 h pre-admission to admission, 24 h pre- to 24 h post-admission, and 24 h pre- to 48 h post-admission). These models achieved F1-scores of 0.92, 0.95, and 0.96, respectively, in predicting 72-h ICU mortality.

Previous ML approaches^{5,8,14,28} relied mainly on severity scales such as APACHE, which have shown limited predictive power. Recent evidence supports that algorithms like XGBoost outperform traditional scores in ICU mortality prediction.²⁹ In addition, a recent study conducted in Spain demonstrated the feasibility of applying ML for early prediction of in-hospital cardiac arrest, further illustrating the potential of these approaches in critical care.³⁰ In line with this, we optimized our models using the F1-score, a metric that balances precision and recall in imbalanced datasets. Unlike AUC-ROC, which may yield overly optimistic estimates, the F1-score provides a more accurate identification of positive cases. These results highlight the advantage of ML approaches that incorporate broader peri-admission data beyond conventional severity scores.

A secondary objective was to enhance model explainability through a classifier-agnostic framework. We summarized variable contributions by plotting predicted mortality probabilities across observed ranges for continuous predictors and by group for categorical variables, with CIs obtained via bootstrapped resampling. Although SHAP values were also explored, our approach offers a simpler and more interpretable alternative that can be generalized to different models. This framework facilitated clearer interpretation of predictors such as norepinephrine use, age, and respiratory rate, providing clinicians with complementary insights without requiring full understanding of internal model mechanics.

As shown in Fig. 3, key predictors of poor prognosis included delay in ICU admission,³¹ older age,³² respiratory

Table 4 Use cases of the clinical decision-support system.

ID ^a	Data window	Death probability (95% CI)	Multi-model approach decision		Actual ICU length of stay ^b	Final status within 72 h of ICU admission	Efficiency ^c
			ICU admission	ICU remaining			
3,342,213	A	0.927 (0.780 to 0.989)	No	-	67.68	Expired	67.68
923,407	A	0.566 (0.439 to 0.663)	Yes	-	28.20	Expired	4.20
	B	0.719 (0.624 to 0.742)	-	No			
386,538	A	0.237 (0.147 to 0.478)	Yes	-	36.23	Expired	-36.23
	B	0.544 (0.429 to 0.626)	-	Yes			
163,973	A	0.210 (0.175 to 0.334)	Yes	-	67.47	Alive (discharged)	0
	B	0.043 (0.022 to 0.077)	-	Yes			
	C	0.056 (0.028 to 0.101)	-	Yes			

Note. ^aEach ID is assigned as an anonymised patient-unit-stay in the dataset used for this study. ^bHours that a patient actually spent in the ICU. ^cHours that the healthcare system would save through the efficient optimization of limited resources. For example, for ID #923407, 24 h after ICU admission our system yielded a high probability of death, recommending not remaining this patient in ICU (Actual ICU length of stay - system prediction = efficiency; 28.20-24 = +4.20 h).

Table 5 F1-scores obtained for the test dataset and for groups for each data window.

	Data window A (-24 h - ICU admission)	Data window B (-24 h - +24 h)	Data window C (-24 h - +48 h)	
Global		0.92	0.95	0.96
Sex	Women	0.92	0.95	0.96
	Men	0.93	0.95	0.96
Age	Adolescents	0.99	0.99	0.99
	Youth	0.99	0.99	0.99
	Adults	0.96	0.97	0.97
	Older adults	0.87	0.92	0.94
Ethnicity	Hispanic	0.93	0.96	0.96
	African American	0.94	0.96	0.96
	Caucasian	0.92	0.95	0.96
	Native American	0.93	0.94	0.95
	Asian	0.92	0.96	0.97
	Other/Unknown	0.93	0.94	0.95
Teaching status	Yes	0.91	0.95	0.96
	Not	0.93	0.95	0.96

rate fluctuations,³³ and vasopressor use.³⁴ These factors are consistent with prior evidence: age is often accompanied by comorbidities, respiratory rate is clinically significant but frequently underdocumented, and norepinephrine use reflects hemodynamic instability rather than being causative per se, that is why interpretation of predictors must be cautious. Additionally, the strong association between cardiac arrest and early mortality, although clinically expected, reinforces the biological plausibility of our framework and

serves as a ‘‘positive control’’ that validates its ability to highlight relevant predictors.

These predictions are threshold-dependent: we classified patients as high risk when the 95% CI lower bound exceeded 60%. However, thresholds may vary according to context. For instance, during the COVID-19 pandemic, intensivists often adjusted criteria due to resource scarcity, highlighting the need for adaptable thresholds in clinical practice.

This multi-model approach may in the future support hospital resource management, for example by informing decisions on allocation of beds, ventilators, and medical staff. However, such implications are exploratory and were not directly evaluated in this study. Further validation in prospective, context-specific settings is needed before translating these findings into economic or resource-related conclusions.

Our models produce probabilistic rather than deterministic outputs and should support, not replace, ICU decision-making. Generalizability may be limited across systems with differing case-mixes and workflows; thus, local, contemporary validation with discrimination/calibration assessment and (re)calibration or retraining is required before CDSS use. The historical eICU cohort (2014–2015), chosen for reproducibility, may not mirror current practice owing to temporal/dataset shift. Ongoing monitoring and periodic updating are advisable. The open-source, model-agnostic pipeline facilitates site-specific validation and maintenance.

This study carries important clinical and ethical implications. ML models trained on large ICU datasets may help overcome the evidence gap caused by excluding critically ill patients from many trials, which often leads to ineffective treatment extrapolation. By providing patient-specific predictions, our approach could reduce futile interventions, alleviate end-of-life suffering and family distress, and lower both economic and opportunity costs. Moreover, the models' performance supports their integration into ICU decision-making as a complement to clinicians' expertise.³⁵ Considering that nearly one in three ICU patients dies within 72 h of admission,^{10,11} a substantial number could benefit from this multi-model strategy. This model-agnostic framework enhances explainability, is adaptable to any supervised classification problem, and can be retrained for different populations, offering a scalable path toward transparent decision-support in critical care.

Conclusion

Beyond estimating a patient's mortality probability, our model-agnostic explainability approach reports the relative contribution and direction of influence of each predictor at the individual level, enhancing transparency and clinical interpretability. These outputs should be viewed as methodological insights rather than a ready-to-use decision aid. Prior to any real-time deployment, temporal external validation, prospective impact evaluation, and integration with clinical expertise are required, with model recalibration or updating as appropriate.

Our multi-model ML framework aims to integrate predictive evidence with clinicians' judgment and patient history to support more personalized decisions, which could help reduce unnecessary ICU admissions and optimize resource allocation. At the same time, responsible implementation must address ethical considerations—including communicating prediction uncertainty, auditing and mitigating bias across subgroups, and preserving medical autonomy—to enable an equitable and trustworthy translation into clinical practice.

Ethical approval

The eICU database is exempt from institutional review board approval due to the retrospective design, lack of direct patient intervention, and the security schema, for which the re-identification risk was certified as meeting safe harbor standards by an independent privacy expert (Privacert, Cambridge, MA) (Health Insurance Portability and Accountability Act Certification no. 1031219-2). The analysis of de-identified, publicly available data, such as that from the eICU database, does not constitute human subjects research as defined by 45 CFR 46.102, and therefore does not require approval or exemption from an ethics committee or institutional review board.

CRedit authorship contribution statement

D.G.G. and S.D. contributed equally to study design, data analysis, and drafting. J.Q.S. and C.J.V. supported model training and interpretation. Á.R. implemented the explainability framework. A.G.P. provided clinical validation. M.A.A.H. led the methodological design, supervised model development, and coordinated the study. M.R.N. and Á.E. supervised clinical aspects and contributed to interpretation.

All authors reviewed and approved the final manuscript.

Funding

No funding.

Declaration of Generative AI and AI-assisted technologies in the writing process

ChatGPT version 4 (OpenAI) was used to support the drafting and refinement of the manuscript, including language editing and structural suggestions, under the supervision of the corresponding author. No part of the scientific analysis or model development was conducted using generative AI.

Declaration of competing interest

The authors declare that there is no Conflict of interest regarding the publication of this article.

Acknowledgements

Thanks to the Laboratory for Computational Physiology of the Massachusetts Institute of Technology and Phillips for contributing to data openness and collaboration through eICU.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at [doi:https://doi.org/10.1016/j.medicine.2025.502349](https://doi.org/10.1016/j.medicine.2025.502349).

References

1. Jackson M, Cairns T. Care of the critically ill patient. *Surgery (Oxf)*. 2021;39:29–36.
2. Reader TW, Reddy G, Brett SJ. Impossible decision? An investigation of risk trade-offs in the intensive care unit. *Ergonomics*. 2018;61:122–33.
3. Bassford C, Griffiths F, Svantesson M, Ryan M, Krucien N, Dale J, et al. Developing an intervention around referral and admissions to intensive care: a mixed-methods study. *Southampton (UK): NIHR Journals Library*; 2019. PMID: 31765112.
4. Sánchez-Casado M, Hostigüela-Martín VA, Raigal-Caño A, Labajo L, Gómez-Tello V, Alonso-Gómez G, et al. Predictive scoring systems in multiorgan failure: a cohort study. *Med Intensiva*. 2016;40(3):145–53, <http://dx.doi.org/10.1016/j.medin.2015.03.005>. English, Spanish. Epub 2015 May 26. PMID: 26022940.
5. Cox EGM, Wiersema R, Eck RJ, Kaufmann T, Granholm A, Vaara ST, et al. External validation of mortality prediction models for critical illness reveals preserved discrimination but poor calibration. *Crit Care Med*. 2023;51:80–90.
6. Prasad PA, Correia J, Fang MC, Fisher A, Correll M, Oreper S, et al. Performance of point-of-care severity scores to predict prognosis in patients admitted through the emergency department with COVID-19. *J Hosp Med*. 2023;18:413–23.
7. Quintairo A, Pilcher D, Salluh JIF. ICU scoring systems. *Intensive Care Med*. 2023;49:223–5.
8. Liu X, Shen M, Lie M, Zhang Z, Liu C, Li D, et al. Evaluating prognostic bias of critical illness severity scores based on age, sex, and primary language in the United States: a retrospective multicenter study. *Crit Care Explor*. 2024;6:e1033, <http://dx.doi.org/10.1097/CCE.0000000000001033>.
9. Rousseau AF, Prescott HC, Brett SJ, Weiss B, Azoulay E, Creteur J, et al. Long-term outcomes after critical illness: recent insights. *Crit Care*. 2021;25:108.
10. Andersen SK, Montgomery CL, Bagshaw SM. Early mortality in critical illness - a descriptive analysis of patients who died within 24 hours of ICU admission. *J Crit Care*. 2020;60:279–84.
11. Vallés J, Diaz E, Carles Oliva J, Martínez M, Navas A, Mesquida J, et al. Clinical risk factors for early mortality in patients with community-acquired septic shock: the importance of adequate source control. *Med Intensiva (Engl Ed)*. 2020;45(9):541–51, <http://dx.doi.org/10.1016/j.medin.2020.05.008>. Epub ahead of print. PMID: 32654923.
12. Mezzaroba AL, Larangeira AS, Morakami FK, Junior JJ, Vieira AA, Costa MM, et al. Evaluation of time to death after admission to an intensive care unit and factors associated with mortality: a retrospective longitudinal study. *Int J Crit Illn Inj Sci*. 2022;12(3):121–6, <http://dx.doi.org/10.4103/ijciis.ijciis.98.21>. Epub 2022 Sep 20. PMID: 36506928; PMCID: PMC9728075.
13. Choi MH, Kim D, Choi EJ, Choi JH, Lee J, Park S. Mortality prediction of patients in intensive care units using machine learning algorithms based on electronic health records. *Sci Rep*. 2022;12:7180, <http://dx.doi.org/10.1038/s41598-022-11226-4>.
14. Yeh YC, Kuo YT, Kuo KC, Cheng YW, Liu DS, Lai F, et al. Early prediction of mortality upon intensive care unit admission. *BMC Med Inform Decis Mak*. 2024;24(1):394.
15. Sauer CM, Dam TA, Celi LA, Faltys M, de la Hoz MAA, Adhikari L, et al. Systematic review and comparison of publicly available ICU data sets—a decision guide for clinicians and data scientists. *Crit Care Med*. 2022;50:e581–8.
16. Bao C, Deng F, Zhao S. Machine-learning models for prediction of sepsis patients' mortality. *Med Intensiva (Engl Ed)*. 2023;47(6):315–25, <http://dx.doi.org/10.1016/j.medine.2022.06.024>. Epub 2022 Nov 4. PMID: 36344339.
17. Li Z. Extracting spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost. *Comput Environ Urban Syst*. 2022;96:101845.
18. Muralidhara CKB. Interpretability of classification & regression ensemble models; 2024.
19. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O, et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data*. 2018;5:180178, <http://dx.doi.org/10.1038/sdata.2018.178>.
20. Data-Science-Laboratory-FPS. *icu_server_72hrs_ethic* [Internet]. 2025. Available from: https://github.com/Data-Science-Laboratory-FPS/icu_server_72hrs_ethic.
21. R: The R Project for Statistical Computing [Internet]. 2024. Available from: <https://www.r-project.org/>.
22. Wickham H, François R, Henry L, others. *dplyr: A grammar of data manipulation*; 2020.
23. Kuhn M. Building predictive models in R using the caret package. *J Stat Soft*. 2008;28:1–26.
24. Wickham H. Getting Started with ggplot2. In: *ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing; 2016. p. 11–31.
25. Japkowicz N, Shah M. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press; 2011.
26. Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transplant*. 2020;55:675–80.
27. World Health Organization, Available from: In: *Universal health coverage*. UHC Compendium. Life course distribution [Internet]; 2024 <https://www.who.int/universal-health-coverage/compendium/interventions-by-life-course>
28. Balkan B, Essay P, Subbian V. Evaluating ICU Clinical Severity Scoring Systems and Machine Learning Applications: APACHE IV/IVa Case Study. In: *Annu Int Conf IEEE Eng Med Biol Soc*; 2018. p. 4073–6.
29. Lim L, Gim U, Cho K, Yoo D, Ryu HG, Lee HC, Available from: Real-time machine learning model to predict short-term mortality in critically ill patients: development and international validation. *Critical Care* [Internet]. 2024;28(76) <https://ccforum.biomedcentral.com/articles/10.1186/s13054-024-04866-7>
30. Socias Crespi L, Gutiérrez Madroñal L, Fiorella Sarubbo M, Borges-Sa M, Serrano García A, López Ramos D, et al. Application of a machine learning model for early prediction of in-hospital cardiac arrests: Retrospective observational cohort study. *Med Intensiva (Engl Ed)*. 2025;49(2):88–95, <http://dx.doi.org/10.1016/j.medine.2024.07.004>. Epub 2024 Aug 29. PMID: 39214814.
31. Kiekkas P, Tzenalis A, Gklava V, Stefanopoulos N, Voyagis G, Aretha D. Delayed admission to the intensive care unit and mortality of critically ill adults: systematic review and meta-analysis. *BioMed Res Int*. 2022;2022:4083494.
32. Vallet H, Guidet B, Boumendil A, De Lange DW, Leaver S, Szczeklik W, et al. The impact of age-related syndromes on ICU process and outcomes in very old patients. *Ann Intensive Care*. 2023;13(1):68.
33. Garrido D, Assioun JJ, Keshishyan A, Sanchez-Gonzalez MA, Goubran B. Respiratory rate variability as a prognostic factor in hospitalized patients transferred to the intensive care unit. *Cureus*. 2018;10(1):e2100.
34. Motiejunaite J, Deniau B, Blet A, Gayat E, Mebazaa A. Inotropes and vasopressors are associated with increased short-term mortality but not long-term survival in critically ill patients. *Anaesth Crit Care Pain Med*. 2022;41(1):101012.
35. Blanch L, Abillama FF, Amin P, Christian M, Joynt GM, Myburgh J, et al. Triage decisions for ICU admission: report from the Task Force of the World Federation of Societies of Intensive and Critical Care Medicine. *J Crit Care*. 2016;36:301–5, <http://dx.doi.org/10.1016/j.jcrr.2016.06.014>.