



# Stratifying Mortality Risk in Intensive Care: A Comprehensive Analysis Using Cluster Analysis and Classification and Regression Tree Algorithms

Antonio Romanelli<sup>1</sup> · Salvatore Palmese<sup>1</sup> · Serena De Vita<sup>1</sup> · Alessandro Calicchio<sup>2</sup> · Renato Gammaldi<sup>1</sup>

Received: 5 February 2024 / Accepted: 21 April 2024 / Published online: 15 May 2024  
© The Author(s) 2024

## Abstract

**Background** Machine learning (ML) can be promising for stratifying patients into homogeneous groups and assessing mortality based on score combination. Using ML, we compared mortality prediction performance for clustered and non-clustered models and tried to develop a simple decision algorithm to predict the patient's cluster membership with classification and regression trees (CART).

**Methods** Retrospective study involving patients requiring ICU admission (1st January 2011–16th September 2022). Clusters were identified by combining Charlson Comorbidity Index (CCI) plus Simplified Acute Physiology Score II (SAPS II) or Sequential Organ Failure Assessment (SOFA). Intercluster and survival analyses were performed. We analyzed the relationship with mortality with multivariate logistic regressions and receiver operating characteristic curves (ROC) for models with and without clusters. Nested models were compared with Likelihood Ratio Tests (LRT). Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were compared for non-nested models. With the best model, we used CART to build a decision tree for patient's membership.

**Results** Our sample consisted of 2605 patients (mortality 59.7%). For both score combinations, we identified two clusters (A and B for CCI+SAPS II,  $\alpha$  and  $\beta$  for CCI+SOFA). Belonging to cluster B/ $\beta$  was associated with shorter survival times (Peto-Peto p-values < 0.0001) and increased mortality (Odds-ratio 4.65 and 5.44, respectively). According to LRT and ROC analysis, clustered models performed better, and CCI+SOFA showed the lowest AIC and BIC values (AIC = 3021.21, BIC = 3132.65). Using CART ( $\beta$  cluster positive case) the accuracy of the decision tree was 94.8%.

**Conclusion** Clustered models significantly improved mortality prediction. The CCI+SOFA clustered model showed the best balance between complexity and data fit and should be preferred. Developing a user-friendly decision-making algorithm for cluster membership with CART showed high accuracy. Further validation studies are needed to confirm these findings.

**Keywords** Charlson Comorbidity Index · Simplified Acute Physiology Score II · Sequential Organ Failure Assessment · Cluster analysis · Classification and regression tree · Mortality

## 1 Introduction

The critical care landscape is continuously evolving, with intensive care unit (ICU) scores being pivotal in stratifying mortality risks among heterogeneous patient populations. Although tools like the Charlson Comorbidity Index (CCI) [1], Simplified Acute Physiology Score II (SAPS II) [2], and Sequential Organ Failure Assessment (SOFA) [3] are routinely used, they are not without limitations. These scores, while informative, often fail to capture the multifaceted nature of an ICU patient's condition due to their focus on isolated variables.

✉ Antonio Romanelli  
antonioromanelli86@gmail.com

<sup>1</sup> Department of Anaesthesia and Intensive Care Unit, AOU "San Giovanni Di Dio E Ruggi D'Aragona", Via San Leonardo, 84125 Salerno, Campania, Italy

<sup>2</sup> Department of Anaesthesia and Intensive Care, "Santa Maria Delle Grazie" Hospital, Via Domiziana, 80078 Pozzuoli, Campania, Italy

From a clinical point of view, it would be more attractive to know which group of patients has a higher mortality risk than the individual patient's risk. In this context, machine learning (ML), applying unsupervised/supervised algorithms, can be promising in stratifying patients into homogeneous groups and accurately assessing mortality based on score combination [4].

Cluster analysis (CA), a group of multivariate mathematical algorithms, is an unsupervised ML algorithm that divides data into groups or clusters based on similar characteristics without a priori hypotheses [5]. CA quantifies the similarity between individuals within a population based on multiple specified variables. In the ICU, CA can be applied to identify subgroups of patients with similar demographic, clinical, and prognostic characteristics, providing a detailed overview and a more in-depth assessment of the cluster-related mortality risk [6, 7]. However, the choice of variables represents the main source of variability and lack of repeatability of the results, limiting generalizability.

A second important aspect for the clinician is to know whether a patient, upon entry into the ICU, belongs to a specific cluster or not. The classification and regression tree (CART) algorithm, a supervised ML algorithm, is a predictive modeling technique that creates a binary decision model predicting the value of a target variable by learning simple decision rules inferred from the data features [8].

In this monocentric retrospective study, we used ML to perform CA based on the alternative combination of CCI plus SAPS II and CCI plus SOFA. We compared mortality prediction performance for clustered and non-clustered models. Then, we tried to develop a simple decision algorithm to predict the patient's membership to a cluster using CART.

## 2 Materials and Methods

### 2.1 Patients' Enrollment

This retrospective study involved all patients admitted in regional hub hospital and requiring ICU admission from 1st January 2011 to 16th September 2022. During the COVID-19 pandemic, we considered patients with negative SARS-CoV-2 nasopharyngeal swabs.

Our ICU is part of the ProSafe project network, supported by the Italian Group for the Evaluation of Interventions in Intensive Care Medicine (GiViTI). The project's purposes are the systematic collection of data and the continuous evaluation and performance comparison of departments.

After collecting written informed consent from conscious patients or from relatives of unconscious patients, demographic data, clinical characteristics on admission and in the first 24 h, the main complications that arose, therapies and procedures performed during the ICU stay, and the outcome

at discharge (including from the hospital) were registered on the national clinical information ProSafe system.

Given the retrospective design of this study, local ethical committee approval was not required. Patient privacy has been rigorously protected in accordance with current national legislation (General Data Protection Regulation of the European Union n° 2016/679 and the Italian Legislative Decrees n° 196/2003 and 101/2018). The study was conducted following the International Conference on Harmonisation Good Clinical Practice guidelines and the 2008 Declaration of Helsinki provisions.

The present study was conducted according to Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [9].

### 2.2 Data Collection

According to the ProSafe project network, all these data were collected on ICU admission and in the first 24 h:

- Demographic: age, gender, body mass index (BMI), comorbidities, patient provenance (same or another hospital), ward before ICU admission (emergency department, surgical or medical ward, other ICU), the reason for ICU admission, and time from in-hospital presentation to ICU admission;
- Function/organ impairments at ICU admission, defined as respiratory, brain, renal, cardiovascular, coagulation, liver, and metabolic failure;
- Sepsis or septic shock on admission;
- Trauma;
- Surgical and non-surgical procedures before ICU admission;
- Development of in-ICU infection;
- Percutaneous tracheostomy and need for vasopressor/inotropes administration in ICU;
- ICU length of stay (LOS) and outcome (death or alive).

Patients' clinical conditions were stratified according to the CCI, the SAPS II, the SOFA, and the Glasgow Coma Scale (GCS) [10]. Supplementary File 1 provided details about data collection.

### 2.3 Records Selection

Only local records were retired and anonymously stored on an electronic sheet (Microsoft Excel).

In the identification phase, we applied the following exclusion criteria:

- Age < 18 years-old;
- Patients with previous LOS > 15 days before ICU admission;

- Transfer from other hospitals, ICU, sub-intensive and long-term care units;
- Admission to ICU for palliative care and post-operative monitoring;
- BMI > 60 kg/m<sup>2</sup>;
- Suffering from AIDS.

Then, in the screening phase, we evaluated missing data. We considered the missing data as missing completely at random, and the observations were deleted.

## 2.4 Statistics

Supplementary File 1 reported details about statistics. A preliminary explorative examination of the records fulfilling our criteria was performed. Descriptive statistics were computed. Categorical data were reported as absolute numbers and percentages. Continuous data were reported as mean  $\pm$  standard deviation or median and first-third quartiles [q<sub>1</sub>–q<sub>3</sub>], according to distribution (tested with the Shapiro–Wilk test).

For clustering, we used the alternative combination of CCI plus SAPS II and CCI plus SOFA. We chose these variables for two reasons: the first is that they are scores widely used in ICUs, and the second is that, as scoring systems, they allow the patient's clinical conditions to be objectively defined. After the normalization of continuous variables [11], the optimal number of clusters (k) was estimated with the package NbClust (version 3.0.1) [12]. For partitioning, we used the Partitioning Around Medoids (PAM) algorithm [13].

Intercluster analysis was performed with proper tests and according to k number. The effect size between clusters and variables was analyzed by computing Pearson (r) or Spearman ( $\rho$ ) coefficient for continuous variables or Cramer's V coefficient for categorical variables [14]. According to the results, coefficients were interpreted as reported in Supplementary File 1. Survival analysis was performed with the Kaplan–Meier method. Median survival time and 28-day survival probabilities with 95% confidence interval (CI<sub>95%</sub>) were computed. The differences in survival rates among clusters were tested with the Peto–Peto test [15].

We performed multivariate logistic regression analysis to find the relationship between cluster membership and ICU mortality. We selected these variables: cluster, gender, BMI, time from in-hospital presentation to ICU admission, ward, clinical condition on ICU admission, and surgical and radiological procedure before ICU admission. The odds ratio (OR) with CI<sub>95%</sub> were computed. Clustered models were compared with the no-clustered model (model 1). For nested models, the difference in deviance was tested with Likelihood Ratio Tests (LRT). If the difference in deviance was not statistically significant, the simplest model was selected

based on the parsimony principle. To evaluate the balance between complexity and data fit for non-nested models we used the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The Receiver Operating Characteristic (ROC) curve was plotted to evaluate the model's predictive performance. The Area Under Curve (AUC) and CI<sub>95%</sub> were computed with the DeLong method [16]. Sensitivity (Se) and specificity (Sp) with CI<sub>95%</sub> were also computed. The differences between AUCs were tested.

With the best predictive model, we developed a simple decision algorithm to predict the patient's cluster membership with the CART algorithm [17] (Supplementary File 1). Since our goal was to provide an easy-to-use algorithm, we chose a split number of 3 nodes in the pruning phase. The authors believe an excessive number of nodes, beyond the probable overfitting, can be interpreted as complex. The effectiveness of the decision model was evaluated with the confusion matrix. Accuracy, with CI<sub>95%</sub>, unweighted Cohen's Kappa, Se, Sp, positive (PPV), and negative (NPV) predictive values were computed.

R-Studio (Posit Software<sup>®</sup> 2009–2023, version 2023.03.1) was used for the analysis. Two-tailed p-values < 0.05 were considered significant for all hypotheses, with  $\alpha = 0.05$ . For multiple comparisons, the p-value was adjusted using the Bonferroni correction. Data are reported in tables and plots.

## 3 Results

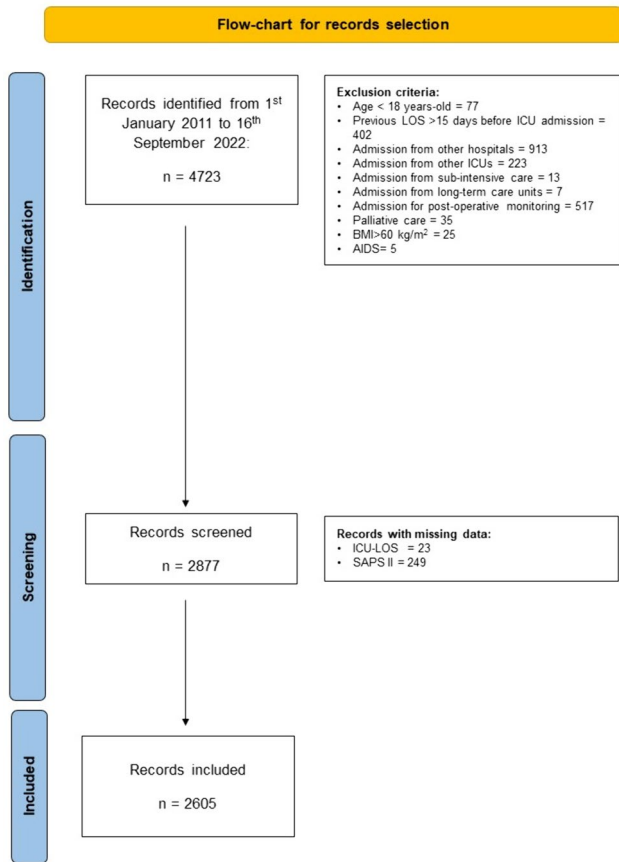
### 3.1 Population Characteristics

From 1st January 2011 to 16th September 2022, 4723 patients were admitted to ICU. Applying exclusion criteria and removing missing data (Fig. 1), the preliminary descriptive analysis was performed on 2605 observations (Table 1).

Briefly, 1543 patients (59.2%) were male, with a median age of 70.0 years [57.0–80.0] and a BMI of 25.1 kg/m<sup>2</sup> [22.1–27.8]. The median time from in-hospital to ICU admission was 0.0 days [0.0–2.0]; 1078 patients (41.4%) came from the emergency department, 915 (35.1%) from surgical, and 612 (23.5%) from medical wards. The medians CCI, SAPS II, SOFA, and GCS were 4.0 [2.0–7.0], 67.0 [51.0–85.0], 10.0 [7.0–13.0], and 5.0 [5.0–9.0], respectively.

The three most common comorbidities were hypertension (1456 patients, 55.9%), moderate/severe COPD (746 patients, 28.6%), and congestive heart failure (727 patients, 27.9%).

The three most common vital/organ function impairments at ICU admission were lung (2382 patients, 91.4%), brain (1582 patients, 60.7%), and cardiocirculatory (1518 patients, 58.2%). Ninety-nine patients (3.8%) showed clinical features



**Fig. 1** Flow chart for records selection. From the 1st of January 2011 to the 16th of September 2022, 4723 observations were retrieved. Once the exclusion criteria were applied, 2877 observations were screened for missing data. Missing data were considered as missing completely at random, and observations were removed. For the analysis, 2605 records were included

suggesting sepsis, while 165 patients (6.3%) had septic shock. Patients with trauma were 437 (16.8%).

Before ICU admission, 995 patients (38.2%) had surgical procedures, while 220 patients (8.4%) had radiological.

Sepsis and septic shock during ICU LOS occurred in 334 (12.8%) and 214 (8.2%) patients, respectively. Percutaneous tracheostomy was performed in 463 patients (17.8%), and 1780 patients (68.3%) required vasopressor/inotrope drugs.

The median ICU LOS was 4.0 days [1.0–11.0], and 1556 patients (59.7%) died during ICU stay.

### 3.2 Clustering and Intercluster Analysis

NbClust found that the best  $k$  number was two for CCI+SAPS II (proposed by eight indexes) and CCI+SOFA (proposed by seven indexes). Figure 2 shows the cluster plots. Clusters were labeled as “A” (1081 observations) and “B” (1524 observations) for CCI+SAPS, and “ $\alpha$ ” (1276 observations) and “ $\beta$ ” (1329 observations) for CCI+SOFA.

Supplementary Table 1 shows the main intercluster differences with the computed statistics.

No significant differences were found for BMI, peptic ulcer, connective tissue disease, and lung and liver failure on admission for CA based on CCI+SAPS II, and for gender, BMI, dementia, peptic ulcer, connective tissue disease, and lung and liver failure on admission for CA based on CCI+SOFA.

According to  $\rho$  and  $V$  coefficients (Table 2), in CA based on CCI+SAPS, gender, liver disease, dementia, hematological malignancy, drug/alcohol addiction, and brain and coagulation failure on admission showed a very weak relationship. In contrast, CCI and SAPS II showed a strong relationship. Age, SOFA, congestive heart failure, chronic kidney disease, and ICU administration of vasopressor/inotrope showed a moderate relationship. All other variables showed a weak relationship.

In CA based on CCI+SOFA, neurodegenerative disease, hematological malignancy, drug/alcohol addiction, and coagulation failure on admission showed a very weak relationship. SAPS II, SOFA, and ICU administration of vasopressor/inotrope showed a strong relationship, while age, CCI, congestive heart failure, and chronic kidney disease showed a moderate relationship. All other variables showed a weak relationship.

### 3.3 Survival Analysis

For CCI+SAPS II clustering, the median ICU LOS was 6.0 days [2.0–15.0] for cluster A and 2.0 days [0.0–8.0] for cluster B, with a statistically significant difference ( $p$ -value < 0.0001). The survival rates were 62.2% (603 patients) and 24.7% (376 patients) for clusters A and B, respectively, with differences statistically significant ( $p$ -value < 0.0001). Kaplan–Meier analysis (Fig. 3A) showed that median survival times were 18.0 days (CI<sub>95%</sub> 16.0–23.0 days) for cluster A and 4.0 days (CI<sub>95%</sub> 4.0–5.0 days,  $p$ -value < 0.0001) for cluster B.

For CCI+SOFA clustering, the median ICU LOS was 6.0 days [2.0–15.0] for the cluster  $\alpha$  and 2.0 days [0.0–7.0] for the cluster  $\beta$ , with a statistically significant difference ( $p$ -value < 0.0001). The survival rates were 61.1% (780 patients) and 20.2% (269 patients) for cluster  $\alpha$  and  $\beta$ , respectively, with differences statistically significant ( $p$ -value < 0.0001). Kaplan–Meier analysis (Fig. 3B) showed that median survival times were 17.0 days (CI<sub>95%</sub> 16.0–20.0 days) for the cluster  $\alpha$  and 3.0 days (CI<sub>95%</sub> 2.0–4.0 days,  $p$ -value < 0.0001) for the cluster  $\beta$ .

### 3.4 Multivariate Logistic Regression and Predictive Performance

Table 3 shows the multivariate logistic regression results. For model 1, time from in-hospital to ICU admission,

**Table 1** The table shows the main population characteristics (2605 observations)

Variable	Dataset (n = 2605)		
	Result	Min–Max	
Gender, male (%)	1543 (59.2%)	-	
Age (years)	70.0 [57.0–80.0]	18.0–99.0	
BMI (kg/m <sup>2</sup> )	25.1 [22.1–27.8]	13.4–59.2	
Time from in-hospital to ICU admission (days)	0.0 [0.0–2.0]	0.0–15.0	
Ward (%)	Emergency	1078 (41.4%)	-
	Surgical	915 (35.1%)	-
	Medical	612 (23.5%)	-
CCI	4.0 [2.0–7.0]	0.0–15.0	
SAPS II	67.0 [51.0–85.0]	7.0–122.0	
SOFA	10.0 [7.0–13.0]	0.0–22.0	
GCS	5.0 [5.0–9.0]	3.0–15.0	
<i>Comorbidities</i>			
Hypertension (%)	1456 (55.9%)	-	
Moderate/severe COPD (%)	746 (28.6%)	-	
Congestive Heart Failure (%)	727 (27.9%)	-	
Diabetes (%)	No insulin	108 (4.1%)	-
	Insulin	422 (16.2%)	-
Cerebral Vascular Disease (%)	379 (14.5%)	-	
Chronic Kidney disease (%)	Moderate/severe	281 (10.8%)	-
	ESRD	81 (3.1%)	-
Arrhythmia (%)	330 (12.7%)	-	
History of Myocardial infarction (%)	312 (12.0%)	-	
Solid tumor (%)	Localised	181 (6.9%)	-
	Metastatic	91 (3.5%)	-
Peripheral Vascular Disease (%)	270 (10.4%)	-	
Hemiplegia (%)	112 (4.3%)	-	
Liver disease (%)	Mild	50 (1.9%)	-
	Moderate/Severe	48 (1.8%)	-
Dementia (%)	91 (3.5%)	-	
Neurodegenerative disease (%)	63 (2.4%)	-	
Haematological malignancy (%)	56 (2.1%)	-	
Drug/alcohol addiction (%)	41 (1.6%)	-	
Peptic ulcer (%)	38 (1.4%)	-	
Connective tissue disease (%)	36 (1.4%)	-	
<i>Clinical condition on admission</i>			
Lung failure (%)	2382 (91.4%)	-	
Brain failure (%)	1582 (60.7%)	-	
Cardiocirculatory failure (%)	1518 (58.2%)	-	
Kidney failure (%)	1514 (58.1%)	-	
Metabolic failure (%)	1207 (46.3%)	-	
Coagulation failure (%)	77 (2.9%)	-	
Liver failure (%)	13 (0.5%)	-	
Infection (%)	Sepsis	99 (3.8%)	-
	Septic Shock	165 (6.3%)	-
Trauma (%)	437 (16.8%)	-	
<i>Procedure before ICU admission</i>			
Surgical (%)	995 (38.2%)	-	
Radiological procedure (%)	220 (8.4%)	-	



**Table 1** (continued)

Variable	Dataset (n = 2605)	
	Result	Min–Max
<i>In ICU infection</i>		
Sepsis (%)	334 (12.8%)	–
Septic Shock (%)	214 (8.2%)	–
<i>In ICU treatment</i>		
Percutaneous tracheostomy (%)	463 (17.8%)	–
Vasopressor/inotrope administration (%)	1780 (68.3%)	–
ICU LOS (days)	4.0 [1.0–11.0]	0.0–172.0
ICU deaths (%)	1556 (59.7%)	–

Categorical data are expressed as absolute numbers and percentages. Continuous data are reported as median and interquartile range [IQR], with minimum and maximum values

previous admission in the medical ward, brain, cardiocirculatory and metabolic failure, and septic shock were risk factors for mortality. Contrarily, lung failure, sepsis, and trauma were factors related to the reduction in mortality risk.

For clustered models, belonging to cluster B or  $\beta$  represented an independent risk factor for the increase in mortality, with an OR of 4.65 (CI<sub>95%</sub> 3.87–5.60) and 5.44 (CI<sub>95%</sub> 4.50–6.59), respectively.

In the CCI+SAPS II clustered model, previous admission in the medical ward, brain and cardiocirculatory failure, and septic shock were factors related to an increase in mortality, while kidney failure was associated with a reduction in mortality. In the CCI+SOFA clustered model, brain and cardiocirculatory failure, and septic shock were factors related to an increase in mortality, while kidney failure was associated with a reduction in mortality.

Deviance differences between model 1 (deviance 3315.2) and models based on CCI+SAPS II (deviance 3032.6) and CCI+SOFA (deviance 2983.2) clustering were statistically significant (p-value < 0.0001). According to the AIC and BIC values, the best non-nested model was the CCI+SOFA clustered model (AIC = 3021.21, BIC = 3132.65).

Figure 4 shows the ROC curves for the three models. Se and Sp were 66.06% (CI<sub>95%</sub> 59.20–79.89%) and 58.03% (CI<sub>95%</sub> 42.99–64.27%) for model 1, 63.39% (CI<sub>95%</sub> 59.58–68.06%) and 77.44% (CI<sub>95%</sub> 72.75–80.33%) for CCI+SAPS II clustered model, and 73.02% (CI<sub>95%</sub> 67.02–79.93%) and 70.76% (CI<sub>95%</sub> 66.71–75.51%) for CCI+SOFA clustered model. The AUCs were 0.6528 (CI<sub>95%</sub> 0.6317–0.6738, poor accuracy) for model 1, 0.7398 (CI<sub>95%</sub> 0.7201–0.7594, moderate accuracy) for CCI+SAPS II clustered model, and 0.7511 (CI<sub>95%</sub> 0.7320–0.7702, moderate accuracy) for CCI+SOFA clustered model. According to the ROC curves analysis, both clustered models showed higher AUC compared to model 1 (adjusted p-value < 0.0001). However, no difference was noted between clustered models AUCs (adjusted p-value 0.2213).

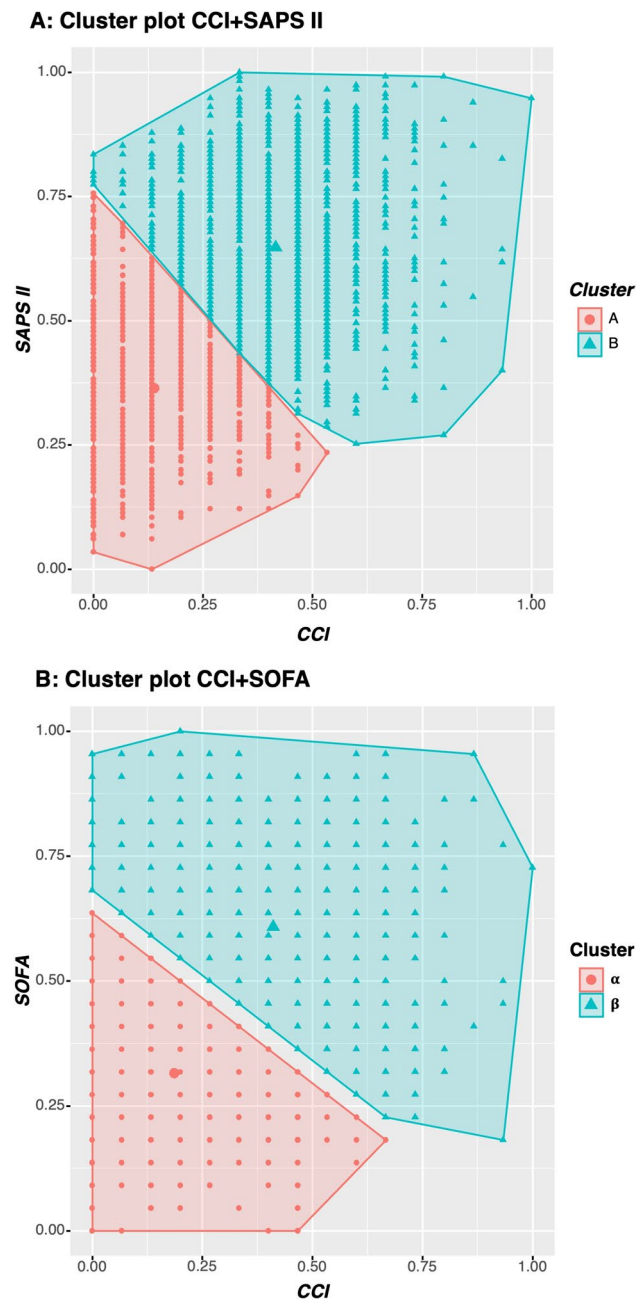
### 3.5 Prediction and Decision Tree

Supplementary Fig. 1 shows the decision tree obtained with the CART algorithm (cp = 0.001, node split = 15). As described in the methods, the tree was pruned with a 3-node split (complexity parameter = 0.034). Figure 5 shows the pruned decision tree. Considering the  $\beta$  cluster as the positive case, the accuracy of the pruned decision tree was 94.8% (CI<sub>95%</sub> 93.9–95.6%), with Kappa 0.896, indicating that the model's predictions are highly reliable. The Se and Sp were 95.3% (CI<sub>95%</sub> 94.0–96.3%) and 93.4% (CI<sub>95%</sub> 93.0–95.5%), respectively, with high PPV (94.6%, CI<sub>95%</sub> 93.3–95.6%) and NPV (95.0%, CI<sub>95%</sub> 93.7–96.0%).

## 4 Discussion

To the best of our knowledge, CCI, SAPS II, and SOFA scores were individually evaluated as prognostic tools within the population requiring ICU admission, and studies have been conducted to establish the best predictive value about complications onset and mortality [18, 19].

However, single scores were validated in defined clinical settings according to different outcomes and observational periods. The CCI (0–33 scale points) was initially developed to predict 1-year mortality in a mixed population of internal medicine patients using comorbidity derived from chart review and consists of 19 selected conditions [1]. The SAPS II (0–163 scale points) is a severity score and has become one of the European ICU's most widely used scoring systems. It is based on the worst values measured on 17 variables during the initial 24 h of ICU stay, with a higher score indicating a more severe illness [2]. SOFA (0–24 scale points) focuses on six organ systems (respiratory, cardiovascular, hepatic, coagulation, renal, and neurological). Each system is scored from 0 to 4 based on the degree of dysfunction or failure, with higher scores indicating more severe organ impairment. This tool is widely used in the ICU to



**Fig. 2** Cluster plots for CCI+SAPS II (**A**) and CCI+SOFA (**B**). **A** Clusters were labeled as “A” (1081 observations) and “B” (1524 observations). The medoids coordinate  $[X_{CCI}, Y_{SAPS II}]$  for clusters were [0.13, 0.36] and [0.40, 0.63]. **B** Clusters were labeled as “ $\alpha$ ” (1276 observations) and “ $\beta$ ” (1329 observations). The medoids coordinate  $[X_{CCI}, Y_{SOFA}]$  for clusters were [0.20, 0.32] and [0.40, 0.59]

predict patient outcomes, guide therapeutic decisions, and evaluate the effectiveness of interventions, especially in the context of sepsis [3].

ML based on CA and CART provides useful clinical information, overcoming the limits of a classic approach based on a single score. The primary study’s finding was that

a combination of CCI+SAPS II and CCI+SOFA allowed the identification of patient clusters characterized by different mortality risks. In detail, belonging to cluster B/ $\beta$  represented an independent risk factor for mortality and shorter survival times when compared to cluster A/ $\alpha$ . Moreover, clustered models improved the mortality prediction ability, such as suggested by moderate accuracy AUC, when compared to the no-clustered model (poor accuracy).

CA can represent a tool to identify patients with peculiar care needs, allowing the redesign of ICU organization based on care processes rather than admission diagnosis [20]. The last aspect is fascinating since the CA, identifying individuals with specific needs, could ensure a better allocation of economic and human resources, reducing waste and improving outcomes. Beyond the organizational aspects, identifying homogenous clusters allows clinicians to adopt the right therapy for the right patients at the right time.

CA has been proposed to personalize treatment in several clinical settings, such as acute respiratory distress syndrome [21], sepsis [22], and out-of-hospital cardiac arrest [23]. However, while most studies focus on categories of patients, our study focuses on a general population, affected by different acute pathologies.

A second aspect that emerges is that clusters do not exist but are the result of complex statistical processing based on the choice of specific variables and justified only by a statistically significant association with a clinical outcome. Identifying a cluster does not necessarily coincide with identifying a patient “prototype”, such as demonstrated by studies showing that the cluster membership is very unstable across different cohorts [24, 25]. In our analysis, this aspect emerges from the different allocation of the observations in a cluster depending on the variables considered for CA and the different degrees of association, as evidenced by the magnitude of the  $\rho$  and  $V$  coefficients between the clusters and the considered variables.

We found a strong association between the use of vasopressors/inotropes in the ICU when clustering was performed with the combination of CCI+SOFA. This result is due to the different items used in SAPS II and SOFA to calculate the final score. While the SAPS II only records the worst systolic pressure value, the SOFA also reports the administration of vasopressor/inotropic drugs. This aspect may also be responsible for the best balance between model complexity and data fit for clustered model based on CCI+SOFA rather than CCI+SAPS II.

Recent literature emphasized the importance of considering vasopressor dosage in the early phase of septic shock and how it correlates with mortality outcomes. Roberts et al. [26] evaluated the association between vasopressor dosing intensity and 30-day in-hospital mortality in patients with septic shock. It was observed that increasing vasopressor dosing intensity during the first 24 h after septic shock was

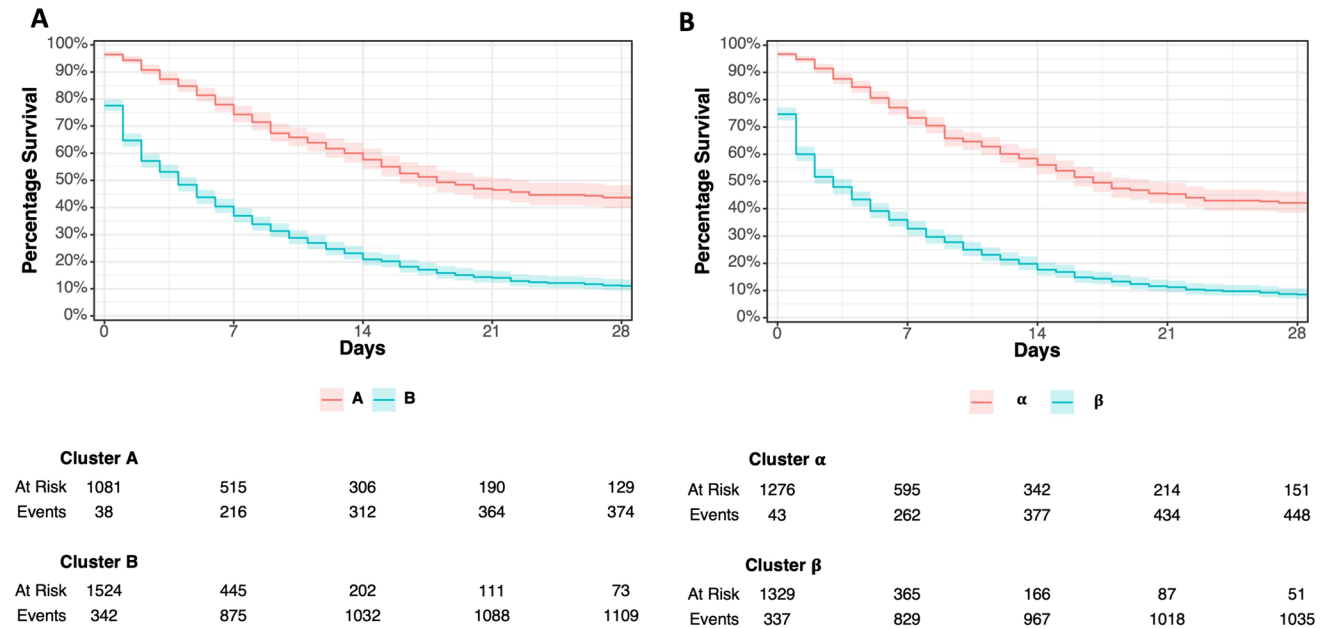
**Table 2** The table shows the effect size between variables and different clustering

Variable	Strenght of relationship									
	CCI+SAPS II clusters					CCI+SOFA clusters				
	VW	W	M	S	VS	VW	W	M	S	VS
Gender	[Red bar]					-				
Age	[Orange bar]					[Orange bar]				
BMI	-					-				
Time from in hospital to ICU admission	[Yellow bar]					[Yellow bar]				
Ward	[Yellow bar]					[Yellow bar]				
CCI	[Green bar]					[Orange bar]				
SAPS II	[Green bar]					[Green bar]				
SOFA	[Orange bar]					[Green bar]				
GCS	[Yellow bar]					[Yellow bar]				
Hypertension	[Yellow bar]					[Yellow bar]				
Moderate/Severe COPD	[Yellow bar]					[Yellow bar]				
Congestive Heart Failure	[Orange bar]					[Orange bar]				
Diabetes	[Yellow bar]					[Yellow bar]				
Cerebral Vascular Disease	[Yellow bar]					[Yellow bar]				
Chronic Kidney disease	[Orange bar]					[Orange bar]				
Arrhythmia	[Yellow bar]					[Yellow bar]				
History of Myocardial infarction	[Yellow bar]					[Yellow bar]				
Solid tumor	[Yellow bar]					[Yellow bar]				
Peripheral Vascular Disease	[Yellow bar]					[Yellow bar]				
Hemiplegia	[Yellow bar]					[Yellow bar]				
Liver disease	[Red bar]					[Yellow bar]				
Dementia	[Red bar]					-				
Neurodegenerative disease	[Yellow bar]					[Red bar]				
Haematological malignancy	[Red bar]					[Red bar]				
Drug/alcohol addiction	[Red bar]					[Red bar]				
Peptic ulcer	-					-				
Connective tissue disease	-					-				
On admission, Lung failure	-					-				
On admission, Brain failure	[Red bar]					[Yellow bar]				
On admission, Cardiocirculatory failure	[Yellow bar]					[Yellow bar]				
On admission, Kidney failure	[Yellow bar]					[Yellow bar]				
On admission, Metabolic failure	[Yellow bar]					[Yellow bar]				
On admission, Coagulation failure	[Red bar]					[Red bar]				
On admission, Liver failure	-					-				
On admission, Infection	[Yellow bar]					[Yellow bar]				
On admission, Trauma	[Yellow bar]					[Yellow bar]				
Before ICU admission, Surgery	[Yellow bar]					[Yellow bar]				
Before ICU admission, Radiological procedure	[Yellow bar]					[Yellow bar]				
In ICU Sepsis or Septic Shock	[Yellow bar]					[Yellow bar]				
In ICU treatment, Percutaneous Tracheostomy	[Yellow bar]					[Yellow bar]				
In ICU treatment, Vasopressor/inotrope use	[Orange bar]					[Green bar]				



**Table 2** (continued)

We use a color code to describe the relationship: red for very weak (VW), yellow for weak (W), orange for moderate (M), green for strong (S), and blue for very strong (VS)



**Fig. 3** The figure shows the 28-day survival curves for CCI+SAPS II (A) and CCI+SOFA (B) clustering with the risk tables. A The predicted 28-day survival probabilities for clusters A and B were 43.7% (CI<sub>95%</sub> 39.7–48.1%) and 11.1% (CI<sub>95%</sub> 9.2–13.4%). Peto-Peto

$\chi^2=448$  on 1 degrees of freedom, p-value < 0.0001. B The predicted 28-day survival probabilities for cluster alpha and beta were 42.2% (CI<sub>95%</sub> 38.5–46.3%) and 8.6% (CI<sub>95%</sub> 6.9–10.8%). Peto-Peto  $\chi^2=589$  on 1 degrees of freedom, p-value < 0.0001

associated with increased mortality. Vallabhajosyula et al. [27] reported that vasopressor burden during early critical illness appears to be a marker for unfavorable outcomes. The elaborated score, incorporating the cumulative vasoactive medication usage within the first 24 h, performed better than SOFA for 28-day mortality prediction.

Although further studies are necessary to evaluate the generalizability of this evidence in different settings, according to our results, a score that evaluates the administration of vasopressor/inotropic drugs among the items may be preferred in CA to predict mortality.

The identification of clusters characterized by a high mortality risk represents important but incomplete information. From a clinical point of view, it would be much more interesting to identify whether a patient with specific features belongs to a cluster at the time of ICU admission. The application of the CART algorithm in this study highlights its potential in the nuanced field of ICU patient data analysis, where rapid and accurate decision-making is crucial. Our findings demonstrate that CART effectively identifies critical thresholds of CCI and SOFA that are simple to compute and evaluate, offering a clear, interpretable model for predicting cluster membership with high accuracy. These aspects should facilitate its application in daily clinical practice, providing targeted interventions for high-risk ICU patients.

The CART algorithm has shown significant potential in the context of ICU. Alghatani et al. [28], in a study on predicting ICU LOS and mortality using patient vital signs, developed decision tree-based methods like CART, demonstrating the importance of advanced data analysis in forecasting critical ICU outcomes. This reflects a broader trend in healthcare, where ML and CART are increasingly employed for clinical decision support in ICUs [4]. Liu et al. [29] demonstrated the application of CART in predicting successful extubation in ICU patients, underscoring the algorithm’s ability to support critical clinical decisions based on patient data. Ramos et al. [30] developed an algorithm to aid in ICU admission triage decisions and showcased the practicality and reliability of such methods in managing ICU admissions and patient prioritization effectively. These examples highlight the growing relevance of the CART algorithm in ICU settings, where its ability to process complex data and provide interpretable results is invaluable for patient care and resource management.

**4.1 Limitations**

Our study showed several limitations. First, its retrospective, single-center design may affect the generalizability of our findings. In particular, the choice to exclude patients with a

**Table 3** The table reports the results of multivariate regression analysis for the three models: model 1 (without clustering), model CCI+SAPS II clustering, and model CCI+SOFA clustering

Variable	Model 1 OR (CI <sub>95%</sub> )	Clustering CCI+SAPS II OR (CI <sub>95%</sub> )	Clustering CCI+SOFA OR (CI <sub>95%</sub> )
Cluster, B/ $\beta$	–	4.65 (3.87–5.60)	5.44 (4.50–6.59)
Gender, Male	0.98 (0.82–1.16)	1.06 (0.88–1.26)	1.03 (0.86–1.23)
BMI	1.00 (0.99–1.02)	1.01 (0.99–1.02)	1.01 (0.99–1.02)
Time from in-hospital to ICU admission	1.04 (1.01–1.07)	1.03 (1.00–1.07)	1.03 (1.00–1.06)
Ward			
Emergency	Ref	Ref	Ref
Surgical	1.06 (0.82–1.35)	0.94 (0.72–1.23)	1.03 (0.79–1.34)
Medical	1.40 (1.12–1.83)	1.31 (1.01–1.69)	1.19 (0.91–1.54)
Lung failure	0.66 (0.48–0.90)	0.82 (0.59–1.15)	0.87 (0.62–1.22)
Brain failure	1.64 (1.35–1.97)	1.49 (1.23–1.80)	1.50 (1.24–1.81)
Cardiocirculatory failure	1.55 (1.27–1.90)	1.51 (1.22–1.88)	1.30 (1.05–1.62)
Kidney failure	0.94 (0.76–1.16)	0.74 (0.59–0.92)	0.75 (0.60–0.94)
Metabolic failure	1.28 (1.03–1.60)	1.16 (0.91–1.46)	1.11 (0.87–1.40)
Coagulation failure	1.35 (0.81–2.33)	1.17 (0.68–2.06)	1.30 (0.75–2.32)
Liver failure	1.57 (0.46–7.18)	1.65 (0.44–8.20)	1.45 (0.28–7.28)
Infection			
No	Ref	Ref	Ref
Sepsis	0.49 (0.32–0.74)	0.65 (0.42–1.02)	0.64 (0.41–1.01)
Septic Shock	3.22 (2.10–5.14)	3.18 (2.03–5.16)	2.19 (1.39–3.57)
Trauma	0.70 (0.56–0.88)	0.94 (0.74–1.19)	0.91 (0.72–1.16)
Surgical procedure	0.89 (0.71–1.11)	1.07 (0.84–1.37)	1.00 (0.78–1.27)
Radiological procedure	0.92 (0.69–1.23)	1.30 (0.96–1.78)	1.17 (0.86–1.60)
AIC	3351.22	3070.58	3021.21
BIC	3456.79	3182.02	3132.65

Belonging to cluster B or  $\beta$  is an independent risk factor for mortality compared to cluster A or  $\alpha$  (see table). Both clustering-based models showed statistically significant differences in deviance compared to model 1 (deviance 3315.2). CCI+SAPS II: deviance 3032.6, difference – 282.6, LRT p-value < 0.0001. CCI+SOFA: deviance 2983.2, difference – 332.0, LRT p-value < 0.0001

hospital LOS longer than 15 days before ICU admission may have excluded patients with a higher risk of death. Moreover, the decision-making algorithm needs further validation. To ensure the robustness and applicability of the algorithm in diverse clinical settings, it must undergo rigorous validation using external datasets. This step is crucial to ascertain the algorithm's efficacy and reliability across different ICU environments and patient populations. Hence, while promising, our results should be interpreted cautiously, as they may not accurately reflect the realities of different ICU settings or patient groups.

Second, despite the large sample size, the data selection process and handling of missing data could also have impacted the results, as missing observations were excluded under the missing completely at random assumption.

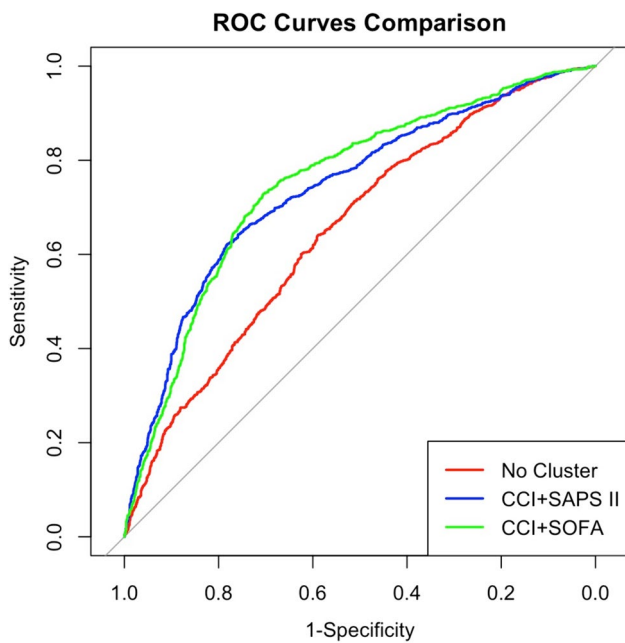
Third, SAPS II and SOFA scores were recorded at admission. This reliance potentially restricts our understanding of the patient's progression in the ICU, as these values do not account for changes in evolving patient conditions.

The dynamic nature of patients' health in the ICU setting suggests that reliance on admission scores alone may not comprehensively represent their clinical trajectory, thus impacting our findings' predictive accuracy and applicability. Furthermore, the use of different scores could lead to different conclusions. Studies are needed to clarify the best combination of scores for clustering.

Fourth, while the use of CART for predicting cluster membership was insightful, the potential overfitting of the model could pose challenges for practical clinical application, especially when referred to other populations.

## 4.2 Strengths

Our study's strengths include the comprehensive utilization of ML based on CA and CART algorithms. Our approach offers a novel perspective on patient categorization and outcome prediction, enhancing the understanding of ICU patient profiles.



**Fig. 4** The figure shows the ROC curves for the considered models. The AUCs (CI<sub>95%</sub>) were: 0.6528 (CI<sub>95%</sub> 0.6317–0.6738, poor accuracy) for model 1, 0.7398 (CI<sub>95%</sub> 0.7201–0.7594, moderate accuracy) for CCI+SAPS II clustered model, and 0.7511 (CI<sub>95%</sub> 0.7320–0.7702, moderate accuracy) for CCI+SOFA clustered model. The adjusted p-values were: <0.001 for Model 1 vs. CCI+SAPS II, <0.001 for Model 1 vs. CCI+SOFA, and 0.2213 for CCI+SAPS II vs. CCI+SOFA

A notable strength of our study is the development of a simple, yet effective decision-making algorithm designed for easy implementation in clinical settings. This simplicity is

a considerable advantage, as it allows for seamless integration into the existing workflows of ICU environments. The algorithm’s user-friendly nature potentially enables a wide range of healthcare professionals to utilize it effectively, enhancing patient care and decision-making efficiency in critical care settings.

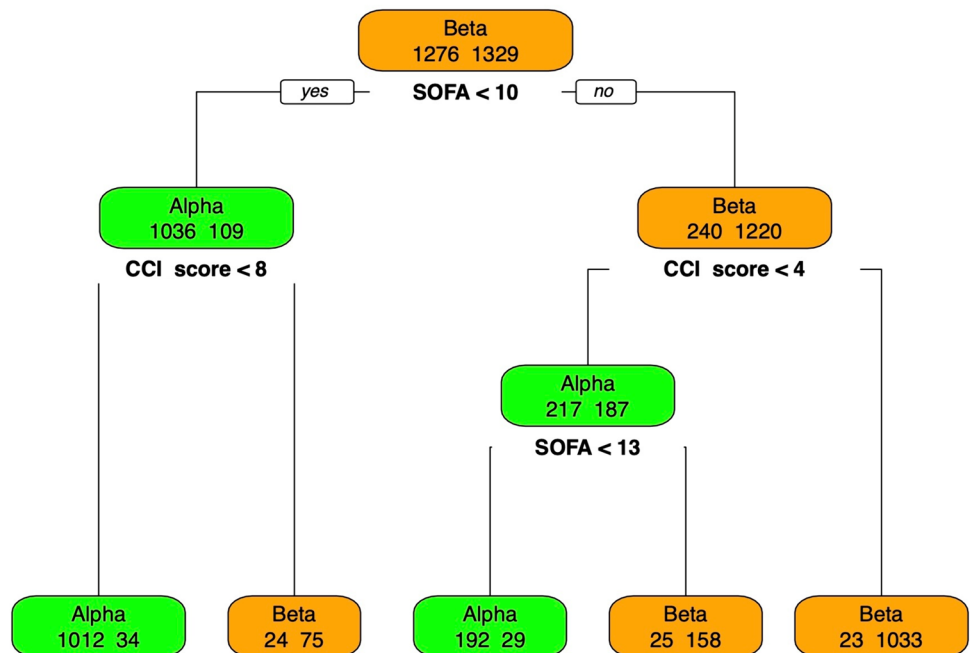
These aspects underline the study’s potential to influence future research and clinical practices in intensive care settings.

### 5 Conclusions

Clustered models significantly improve mortality prediction compared to the no-clustered model and may be more useful in clinical practice for the early identification of high-risk patients. Despite clustered models showing no difference in mortality predictive performance, clustering based on CCI+SOFA showed the best balance between model complexity and data fit and should be preferred to the CCI+SAPS II model. Developing a straightforward, user-friendly decision-making algorithm is a key contribution, offering a practical tool for enhancing ICU performance.

While acknowledging the limitations inherent in our monocentric, retrospective approach, the study opens new avenues for future research and underscores the vital role of ML in improving patient outcomes in critical care. Further multicentric studies are needed to ensure the findings are robust and applicable to a broader patient population.

**Fig. 5** The figure shows the pruned decision tree (3 split nodes) for cluster membership. Each node represents a decision point with branches to the outcomes “Alpha” and “Beta.” The leaves represent the final classification with the absolute numbers of each cluster within that leaf. Under the leaf is reported the splitting criterion. The color of the leaf indicates the most represented cluster: green for “Alpha” and orange for “Beta.”



**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s44231-024-00064-9>.

**Acknowledgements** None.

**Author Contributions** Conceptualisation: AR, SP, RG. Data Curation: SDV, AC, SP, RG. Formal analysis: AR, RG, SP. Methodology: AR, AC, SDV. Project administration: RG. Visualization: AC, SDV, SP. Writing—original draft: AR, RG, SDV. Writing—review & editing: AR, SDV, AC, SP, RG.

**Funding** Not applicable.

**Availability of Data and Materials** Data are available by the corresponding author on a reasonable request.

## Declarations

**Competing Interests** The authors declare that they have no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373–83.
- Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA.* 1993;270(24):2957–63.
- Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707–10.
- Hong N, Liu C, Gao J, Han L, Chang F, Gong M, Su L. State of the art of machine learning-enabled clinical decision support in intensive care units: literature review. *JMIR Med Inform.* 2022;10(3): e28781.
- Berry MW, Mohamed A, Yap BW. *Supervised and unsupervised learning for data science.* Springer; 2019.
- Cidade JP, de Souza Dantas VC, de Figueiredo Thompson A, de Miranda R, Mamfrim R, Caroli H, Escudini G, Oliveira N, Castro T, Povoia P. Identification of distinct clinical phenotypes of critically ill COVID-19 patients: results from a cohort observational study. *J Clin Med.* 2023;12(8):3035.
- Chen X, Li J, Liu G, Chen X, Huang S, Li H, Liu S, Li D, Yang H, Zheng H, et al. Identification of distinct clinical phenotypes of heterogeneous mechanically ventilated ICU patients using cluster analysis. *J Clin Med.* 2023;12(4):1499.
- Breiman L, Friedman J, Olshen R, Stone C. *Classification and Regression Trees.* Wadsworth & Brooks/Cole Advanced Books & Software. Pacific California 1984.
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55–63.
- Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet.* 1974;2(7872):81–4.
- Kuhn M, Johnson K. *Applied predictive modeling,* vol. 26. Springer; 2013.
- Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw.* 2014;61:1–36.
- Kaufman L, Rousseeuw PJ. *Finding groups in data: an introduction to cluster analysis.* John Wiley & Sons; 2009.
- Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions,* vol. 1981218. New York: Wiley; 1981.
- Peto R, Peto J. Asymptotically efficient rank invariant test procedures. *J Royal Stat Soc Ser A (General).* 1972;135(2):185–98.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837–45.
- Breiman L. *Classification and regression trees.* Routledge; 2017.
- Yang M, Mehta HB, Bali V, Gupta P, Wang X, Johnson ML, Aparasu RR. Which risk-adjustment index performs better in predicting 30-day mortality? A systematic review and meta-analysis. *J Eval Clin Pract.* 2015;21(2):292–9.
- Haniffa R, Isaam I, De Silva AP, Dondorp AM, De Keizer NF. Performance of critical care prognostic scoring systems in low and middle-income countries: a systematic review. *Crit Care.* 2018;22(1):18.
- Vranas KC, Jopling JK, Sweeney TE, Ramsey MC, Milstein AS, Slatore CG, Escobar GJ, Liu VX. Identifying distinct subgroups of ICU patients: a machine learning approach. *Crit Care Med.* 2017;45(10):1607–15.
- Liu X, Jiang Y, Jia X, Ma X, Han C, Guo N, Peng Y, Liu H, Ju Y, Luo X. Identification of distinct clinical phenotypes of acute respiratory distress syndrome with differential responses to treatment. *Crit Care.* 2021;25:1–11.
- Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA.* 2019;321(20):2003–17.
- Binois Y, Renaudier M, Dumas F, Youssfi Y, Beganton F, Jost D, Lamhaut L, Marijon E, Jouven X, Cariou A, Bougouin W. Factors associated with circulatory death after out-of-hospital cardiac arrest: a population-based cluster analysis. *Ann Intensive Care.* 2023;13(1):49.
- Castaldi PJ, Benet M, Petersen H, Rafaels N, Finigan J, Paoletti M, Marike Boezen H, Vonk JM, Bowler R, Pistolesi M, et al. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. *Thorax.* 2017;72(11):998–1006.
- Zhang Z, Chen L, Liu X, Yang J, Huang J, Yang Q, Hu Q, Jin K, Celi LA, Hong Y. Exploring disease axes as an alternative to distinct clusters for characterizing sepsis heterogeneity. *Intensive Care Med.* 2023;49(11):1349–59.
- Roberts RJ, Miano TA, Hammond DA, Patel GP, Chen JT, Phillips KM, Lopez N, Kashani K, Qadir N, Cairns CB, et al. Evaluation of vasopressor exposure and mortality in patients with septic shock. *Crit Care Med.* 2020;48(10):1445–53.
- Vallabhajosyula S, Jentzer JC, Kotecha AA, Murphree DH Jr, Barreto EF, Khanna AK, Iyer VN. Development and performance of a novel vasopressor-driven mortality prediction model in septic shock. *Ann Intensive Care.* 2018;8(1):112.

28. Alghatani K, Ammar N, Rezgui A, Shaban-Nejad A. Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation. *JMIR Med Inform.* 2021;9(5): e21347.
29. Liu Y, Mu YU, Li GQ, Yu X, Li PJ, Shen ZQ, Wang HX, Wei LQ. Extubation outcome after a successful spontaneous breathing trial: a multicenter validation of a 3-factor prediction model. *Exp Ther Med.* 2015;10(4):1591–601.
30. Ramos JG, Perondi B, Dias RD, Miranda LC, Cohen C, Carvalho CR, Velasco IT, Forte DN. Development of an algorithm

to aid triage decisions for intensive care unit admission: a clinical vignette and retrospective cohort study. *Crit Care.* 2016;20:81.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.