Journal of Biomedical Informatics 117 (2021) 103768

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Methods and measures to quantify ICU patient heterogeneity

David Cuadrado [a,b], David Riaño [a,*], Josep Gómez [b], Alejandro Rodríguez [b], María Bodí [b]

[a] Universitat Rovira i Virgili, Tarragona, Spain
[b] Intensive Care Unit, University Hospital Joan XXIII, Tarragona, Spain

ARTICLE INFO

Keywords:
Patient heterogeneity
Case-mix
Days to discharge prediction
Intensive care unit
Patient similarity

ABSTRACT

Patients in intensive care units are heterogeneous and the daily prediction of their days to discharge (DTD) a complex task that practitioners and computers are not always able to solve satisfactorily. In order to make more precise DTD predictors, it is necessary to have tools for the analysis of the heterogeneity of the patients. Unfortunately, the number of publications in this field is almost non-existent. In order to alleviate this lack of tools, we propose four methods and their corresponding measures to quantify the heterogeneity of intensive patients in the process of determining the DTD. These new methods and measures have been tested with patients admitted over four years to a tertiary hospital in Spain. The results deepen the understanding of the intensive patient and can serve as a basis for the construction of better DTD predictors.

1. Introduction

Hospital Intensive Care Units (ICU) attend patients with severe health conditions who require continuous care and supervision. These comprise a wide variety of patient types or *case-mix*, and care variability [1]. The identification and quantification of this variety of patients is relevant in order to have a comprehensive and comparative notion of the ICU, its particularities and its management requirements.

Classifying ICU patients in clinical groups (e.g., surgical, medical, trauma), admission groups (scheduled or emergent, or urgent), medical groups (e.g., cardiology, respiratory, or infections), or resource-consumption groups (e.g., diagnostic related groups or DRG) is a common practice, but it does not offer a useful perspective at the time of predicting patient's length of stay (LOS) or patient's days to discharge (DTD). Whereas LOS corresponds to the number of days that a patient stays in the ICU and it is usually predicted within the 24 to 48 h after the patient's admission, DTD measures dynamically the number of days until the patient is discharged and it is predicted in a daily basis.

Both parameters have clinical, management, and administrative impact since they can help clinicians to detect deviation from standards and, therefore, implement corrective actions. DTD prediction brings some additional benefits in the organization, the decision making, and the efficient use of resources in ICUs by better planning of, for example, bed occupancy [2], complex surgeries requiring ICU [3], or transfers to other centers [4], if a saturation is expected.

Predicting LOS with computer technology has received considerable attention [5–7] and their predictive methods have been progressively improved over the years [6,8]. However, the dynamic prediction of inpatients' DTD has awaken much less interest [9,2,10], perhaps due to the added difficulties of collecting ICU data in a daily basis [11,12] or to the sophistication of the technologies for longitudinal data analysis [13].

In ICUs, it is essential to find comprehensive patterns or to develop predictive models to differentiate among patients with different DTDs. But, before we can identify those patterns or make accurate DTD predictors, it is important to know more about ICU patient heterogeneity and gain insight on the clinical parameters that could help identifying patients with different DTDs.

We are particularly interested in the heterogeneity of ICU patients discharged alive. The incorporation of deceased patients in our study could not only unnecessarily distort the studies of heterogeneity and affect the quality of the prediction of DTDs, but we believe also that these cases should be treated with specific methods for mortality prediction.

Focused on patients discharged alive, in this work, we developed four alternative methods and their corresponding measures to quantitatively ascertain the heterogeneity of ICU patients with regard to their DTD values. These methods allow the calculation of heterogeneity values for groups of patients, so that the greater the heterogeneity, the

* Corresponding author.
E-mail address: david.riano@urv.cat (D. Riaño).

https://doi.org/10.1016/j.jbi.2021.103768
Received 9 October 2020; Received in revised form 21 February 2021; Accepted 29 March 2021
Available online 9 April 2021
1532-0464/

more difficult it will be to estimate the DTD of these patients.

The first method (clinical parameter analysis) assumes that, as time passes, the values observed for some (or all) clinical parameters converge to "normality"[1] values, which justify patient's discharge from the ICU. On the other hand, the variability of values of these clinical parameters decreases as the patients approach to their discharge day.

The second method (severity scales analysis) assumes that some clinical scales measuring patient's condition severity or care needs, such as SOFA [14,15], NAS [16], or EMINA [17], tend to normality values as the ICU patient stabilizes and improves along the consecutive days in the ICU.

The third method (confusion analysis) uses a similarity function between patients to calculate how many ICU patients discharged in x days are similar to ICU patients discharged in y days. If these measures are expressed as percentages, we obtain that if $x = y$, 100% of the patients are similar, if $x < y$, the percentage describes the risk of premature discharges, if $x > y$, the risk of overdue discharges, and if $x \neq y$, the risk of wrong patient discharges in x days. Here, patient heterogeneity is seen as the proportion of cases not discharged in x exact days, that are indistinguishable from cases that are discharged exactly in x days.

The fourth method (cluster analysis) borrows internal evaluation methods from the theory of cluster analysis to calculate the quality of DTD groups. Three of the most common indices to calculate grouping quality are Davies-Bouldin [18], Dunn [19], and silhouette [20]. These measures were modified to describe the level of heterogeneity among cases which are clustered according to their DTD values.

The formalization of these four methods and metrics is done in Section 2.1. They were applied to confirm the heterogeneity of patients in the ICU of a tertiary hospital in Spain, previous to the Covid-19 outbreak. The results are exposed and analyzed in Section 3, and a discussion is provided in Section 4. All the studies confirmed the large heterogeneity of ICU patients in this reference hospital.

2. Methods

Four methods to measure ICU patients' heterogeneity with regard to their respective days to discharge were formalized. This formalization was implemented and tested with Python and finally applied to all the ICU patients attended in the ICU of a tertiary hospital in Spain, in the period 2016–2019. In order to illustrate the application of the methods, a small case example is provided as additional material.

2.1. Formal definition of heterogeneity quantification methods and measures

Our analysis of heterogeneity is based on the study of N ICU patients $P_1, \ldots, P_N$ discharged alive, with P_i ($i = 1, \ldots, N$) a sequence of daily descriptions of the patient in terms of m clinical parameters; i.e., $P_i = (d_{i1}, \ldots, d_{i\ell_i})$, with ℓ_i the LOS of the i-th patient, $d_{ij} = (\mathrm{v}_{ij1}, \ldots, \mathrm{v}_{ijm})$ the description of the i-th patient while in her j-th day before ICU discharge (DTD), and v_{ijk} the value of the k-th parameter of patient P_i in her j-th DTD. We assume d_{ij} structures do not contain missing v_{ijk} values. This is a reasonable assumption when dealing with ICU patients, whose clinical parameters are either automatically taken by ICU devices or systematically recorded by ICU staff.

Clinical parameters can be categorical or numerical. Categorical parameters were measured once per day. Numeric parameters which were measured several times along the day were replaced by one or more alternative parameters representing aggregated values such as the minimum, maximum, average, or the standard deviation in the day. In this sense, in this work, heterogeneity is interpreted in terms of the ICU daily variability of the DTD of the current patients. By extension, our measures can also be used to calculate the heterogeneity of any group of ICU patients, based on their respective days to discharge.

All numerical parameters describing ICU patients have a minimum value and a maximum value (see Eqs. (1) and (2)) that were used to normalize the v_{ijk} values in an interval $[0, 1]$ with the min–max normalization (3).

$$min_k = \min_{\substack{i=1,\ldots,N \\ j=1,\ldots,\ell_i}} \{\mathrm{v}_{ijk}\} \tag{1}$$

$$max_k = \max_{\substack{i=1,\ldots,N \\ j=1,\ldots,\ell_i}} \{\mathrm{v}_{ijk}\} \tag{2}$$

$$v_{ijk} = \frac{\mathrm{v}_{ijk} - min_k}{max_k - min_k} \tag{3}$$

Patient descriptions d_{ij} were used to define DTD groups. A DTD group DTD_j is defined as the set of all the clinical descriptions d_{ij} of the patients P_i ($i = 1, \ldots, N$) in their j-th day before discharge; see (4). The number of patient descriptions in a DTD_j group is $n_j = \#\mathrm{DTD}_j$. Note that, if $j < j'$, then $n_j \geqslant n_{j'}$, because some patients could have a LOS between j and j' days.

$$\mathrm{DTD}_j = \bigcup_{i=1,\ldots,N} \{d_{ij}\} \tag{4}$$

The similarity between two normalized values v and v' of a given numerical parameter is $sim_j(v, v') = 1 - (v - v')^2$. If the parameter is categorical, the similarity between two categories v and v' is $sim_j(v, v') = 1$, if $v = v'$, or 0, otherwise.

Using these parameter similarity functions, we defined the similarity between any two patient descriptors d_{ij} and $d_{i'j'}$ as their root mean square resemblance (5).

$$sim(d_{ij}, d_{i'j'}) = \sqrt{\frac{1}{m} \cdot \sum_{k=1}^{m} sim_k(v_{ijk}, v_{i'j'k})} \tag{5}$$

Based on the previous concepts, we can formally describe the four methods and related metrics to calculate DTD-based heterogeneity of patients in an ICU. For the sake of simplicity, we consider DTD groups as $\mathrm{DTD}_i = \{d_{i1}, \ldots, d_{in_i}\}$ ($i = 1, \ldots, \ell$), with ℓ the largest observed LOS, and the patient descriptors in the DTD_i group as $d_{ij} = (v_{ij1}, \ldots, v_{ijm})$, with v_{ijk} the 0–1 normalized value of the k-th clinical parameter for the patient description d_{ij} of a patient that is in her i-th day before discharge.

2.1.1. Method 1: clinical parameters analysis

The clinical parameters should converge to normality values as the patients' day of discharge approaches. Simultaneously, the variability of the observed values should decrease as the discharge day is closer. Heterogeneity here corresponds to how much the average value of the parameters in a DTD_i group deviates from the normality values or normality range and, alternatively, how much the variability of these values for patients within a DTD_i group decreases as i approaches to one (i.e, the day before discharge).

In order to measure these values, we use the mean and standard deviation Eqs. (6) and (7) of the clinical parameter k for the patients in DTD_i, and represent the functions $f_k(x) = mean(k, x)$ and $g_k(x) = stdev(k, x)$, for each one of the $k = 1, \ldots, m$ clinical parameters, and x the days until discharge.

$$mean(k, i) = \frac{1}{n_i} \cdot \sum_{j=1}^{n_k} v_{ijk} \tag{6}$$

[1] Here, "normality" must be understood in the context of an ICU as the range of values that do not justify the patient to stay in the ICU. Therefore, a patient having all the clinical parameters to normality values should be discharged from the ICU.

$$stdev(k,i) = \sqrt{\frac{1}{n_i} \cdot \sum_{j=1}^{n_k} (v_{ijk} - mean(k,i))^2} \tag{7}$$

The graphical representations of $f_k(x)$ and $g_k(x)$ show the evolution of the average value and variability of each clinical parameter as the discharge day is nearer. This allows visual analysis of the patients' heterogeneity with regard to their DTD. In our study, these functions are shown in Figs. 2 and 3.

2.1.2. Method 2: severity scales analysis

Several clinical scales or scores, such as SOFA, NAS, or EMINA, are frequently used in ICU to measure different dimensions of the patient complexity, condition severity, or care requirements. These scales simplify medical interpretation of diagnostic, therapeutic, and prognostic aspects, such as the degree of patient's organ failure (SOFA), the percentage of nursing activity required (NAS), or the risk of developing pressure ulcers as a combined assessment of mental state, mobility, incontinence, nutrition and activity (EMINA).

When treated as clinical parameters, the graphical representation of means and standard deviations of scales in a DTD_i group, as i evolves from n to 1 days before discharge, provides information to analyze patient heterogeneity in terms of the progress of the patient complexity. Intuitively, as a patient approaches to the discharge day, scales should get closer to their respective "normality"[1] values, and the fluctuation of scale values among people in the same DTD group should reduce. Any other behaviour implies a greater heterogeneity between ICU patients.

2.1.3. Method 3: confusion analysis

The patient similarity function in Eq. (5) can be used to identify which patients in DTD_i are similar to which patients in DTD_j ($i,j = 1,\ldots,\ell$).

In Fig. 1, each block represents one of the ℓ DTD groups DTD_ℓ, …, DTD_1, with DTD_i the pivoting group. The dark areas are the patients in each group that are similar to some patients in the pivoting group. The table at the bottom of Fig. 1 shows some information about the DTD groups: the number of patients in the group (n_j) and also the subset ($S_\delta(j,i)$), the number ($s_\delta(j,i)$), and the proportion ($\frac{s_\delta(j,i)}{n_j}$) of patients in the group who are similar to some of the DTD_i cases.

Formally speaking, we use the similarity function defined in (5) in combination with a threshold parameter $\delta \in [0,1]$, such that $S_\delta(i,j)$ in Eq. (8) represents the set of all the patients in DTD_i which have a degree of similarity δ or higher to some patient in DTD_j. As δ approaches to one, similarity is more demanding and less patient descriptions d' in DTD_i are expected to be similar to the pivoting descriptions d in DTD_j. If the number of patient descriptions in $S_\delta(i,j)$ is $n_\delta(i,j)$, the number of patient descriptors similar to patients in DTD_i is calculated as $s(i) = \sum_{j=1}^{\ell} n_\delta(j,i)$.

These values are used to define (1) the proportion of cases in DTD_i which can be confused with patients in DTD_j (i.e., patients that are discharged in i days who are similar to patients discharged in j days); (2) the proportion of cases similar to DTD_i patients which are discharged in less than i days (i.e., premature discharges); (3) the proportion of cases similar to DTD_i patients which are discharged in more than i days (i.e., overdue discharges), and (4) the proportion of cases similar to DTD_i patients which are discharged in a number of days different from i (i.e., feasible discharge errors).

$$S_\delta(i,j) = \bigcup_{d \in DTD_j} \{d' \in DTD_i : sim(d,d') \geqslant \delta\} \tag{8}$$

These confusing discharges describe ICU patient heterogeneity in the sense that they count the number of similar patients with different DTDs. Four confusion ratios are defined in Eqs. (9)–(12) to measure heterogeneity: degree of confusion (c_δ), premature discharge (p_δ), overdue discharge (o_δ), and feasible discharge error (e_δ), respectively.

$$c_\delta(i,j) = \frac{n_\delta(i,j)}{n_i} \tag{9}$$

$$p_\delta(i) = \frac{1}{s(i)} \cdot \sum_{j<i} n_\delta(i,j) \tag{10}$$

$$o_\delta(i) = \frac{1}{s(i)} \cdot \sum_{j>i} n_\delta(i,j) \tag{11}$$

$$e_\delta(i) = \frac{1}{s(i)} \cdot \sum_{j \neq i} n_\delta(i,j) \tag{12}$$

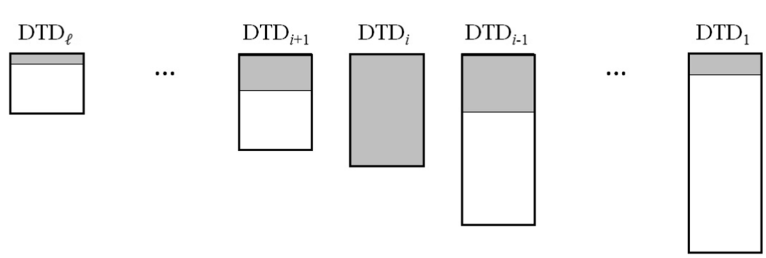
With Eqs. (13) and (14), we can also calculate the average number of patients which are similar to other patients who were discharged n days before (b_δ), or after (a_δ).

$$b_\delta(n) = \frac{1}{\ell - n} \cdot \sum_{i=n+1}^{\ell} \frac{n_\delta(i, i-n)}{n_{i-n}} \tag{13}$$

$$a_\delta(x) = \frac{1}{\ell - n} \cdot \sum_{i=1}^{\ell-n} \frac{n_\delta(i, i+n)}{n_{i+n}} \tag{14}$$

2.1.4. Method 4: cluster analysis

In statistical classification, cluster analysis is based on the premise that elements in the same cluster are similar and elements in different clusters are dissimilar. Combining these intra-cluster similarity and inter-cluster dissimilarity concepts, there are several indices to assess the quality of a given clustering. Among them, Davies-Bouldin [18], Dunn [19], and average silhouette [20] are some of the most used.



#DTD$_i$	$1 \leq n_\ell$ ≤	… ≤	n_{i+1} ≤	n_i ≤	n_{i+1} ≤	… ≤	n_1
Similar to DTD$_i$ (cases or dark areas)	$S_\delta(\ell, i)$	…	$S_\delta(i+1, i)$	$S_\delta(i, i)$	$S_\delta(i\text{-}1, i)$	…	$S_\delta(1, i)$
Similar to DTD$_i$ (number of cases)	$n_\delta(\ell, i)$	…	$n_\delta(i+1, i)$	$n_\delta(i, i)$	$n_\delta(i\text{-}1, i)$	…	$n_\delta(1, i)$
Similar to DTD$_i$ (proportions)	$\frac{n_\delta(\ell, i)}{n_\ell}$	…	$\frac{n_\delta(i+1, i)}{n_{i+1}}$	$\frac{n_\delta(i, i)}{n_{i+1}}$	$\frac{n_\delta(i\text{-}1, i)}{n_{i+1}}$	…	$\frac{n_\delta(1, i)}{n_1}$

Fig. 1. Confusion-based analysis of patient heterogeneity.

Converting these distance-based indices into ICU patient similarity indices is accomplished with Eqs. (15)–(17).

$$DB = \frac{1}{\ell}\sum_{i=1}^{\ell}\max_{1\leqslant j\neq i\leqslant \ell}\frac{2-(m_i(C_i)+m_j(C_j))}{1-sim(C_i,C_j)} \quad (15)$$

$$D = \frac{1-\max\limits_{1\leqslant i<j\leqslant \ell}\{sim(C_i,C_j)\}}{1-\min\limits_{1\leqslant i\leqslant \ell}\{m_i\}} \quad (16)$$

$$silhouette = \frac{1}{\sum_{i=1}^{\ell} n_i}\cdot\sum_{i=1}^{\ell}\sum_{d\in DTD_i}\frac{m_i(d)-m(d)}{1-\min\{m(d),m_i(d)\}} \quad (17)$$

These are based on the identification of one representative patient description C_i for each DTD_i group, with Eq. (18). This representative is the patient in the DTD_i group with the greatest average similarity to the rest of cases in DTD_i.

$$C_i = \mathrm{argmax}_{d\in \mathrm{DTD}_i} m_i(d) \quad (18)$$

The average similarity of any patient description d to a DTD_i group is calculated with Eq. (19), after removing d from DTD_i, if $d \in \mathrm{DTD}_i$. The average similarity within a DTD_i group is defined by Eq. (20), and $m(d) = \max_{j\neq i} m_j(d)$ is the greatest average similarity of any patient description d in DTD_i to any other DTD group.

$$m_i(d) = \frac{1}{n_i}\cdot\sum_{d'\in DTD_i} sim(d,d') \quad (19)$$

$$m_i = \frac{1}{n_i}\cdot\sum_{d\in DTD_i} m_i(d) \quad (20)$$

The Davies-Bouldin, Dunn, and silhouette indices can be used to summarize the quality of the DTD groups according to the similarity of the patient descriptions that they contain, and therefore, provide information on the degree of heterogeneity of the patients in the DTD groups.

2.2. Practical uses of the heterogeneity methods

The methods formalized in 2.1 can be applied in an absolute way to calculate the heterogeneity value of a group of patients in an ICU, but also in a relative way to compare the populations of patients from different ICUs, or to contrast two groups or subgroups of patients (e.g., heterogeneity ratio between scheduled and urgent patients).

In the next section, we describe our application of these methods in the analysis of the heterogeneity of the patients in a concrete ICU.

2.3. UCI patient heterogeneity analysis

All the adult patients admitted in the ICU of the University Hospital Joan XXIII, Spain, in the years 2016–2019 were taken to quantify the patient heterogeneity within that service. Only patients discharged alive were considered. All discharges correspond to patients who were transferred to other non-critical units of the hospital to continue their treatment. None of them were discharged from the hospital. While in the ICU, the daily information of these patients in the 21 days previous to discharge was used in the analysis. Table 1 summarizes the mean, standard deviation, min, and max values of the 19 numerical clinical parameters, and the proportions of the categories of the 18 categorical parameters. A distinction is made between numerical (N), categorical (C), and binary (CB) parameters, and also between demographic (DE), clinical (CP), scales (SC)[2], and administrative (AD) parameters. For example, HR is a numerical clinical parameter (N, CP) containing the daily average heart rate of the patient, NAS is a numerical scale (N, SC) quantifying the daily nursing activities required by the patient, and *AdmType* a binary demographic parameter (CB, DE) informing about the type of ICU admission of the patient (e.g., urgency or scheduled). The information contained in Table 1 is complemented with the box-plots of all the numerical parameters provided as additional material.

A total number of 3,973 patients were involved in the study, with a mean LOS of 8.56 days per patient (6.95 after trimming the data to 21 days). The total number of daily patient descriptions was 27,611.

3. Results

The four methods in Section 2.1 were applied to the data about the ICU patients described in Section 2.3, to measure patient heterogeneity among these patients with regard to their days to discharge. Results are given in their normalized form, for a fair comparison, but some discussions are made in terms of the denormalized values in order to keep the medical sense.

3.1. Patients heterogeneity based on the analysis of clinical parameters

Fig. 2(a) and(b) show the functions $f_k(x)$ and $g_k(x)$ (see Section 2.1.1) that describe, respectively, the evolution of the mean and the standard deviation of all the normalized clinical parameters k in the different DTD_x groups. The horizontal axes represent the number of days to discharge $x = 1, \ldots, 21$. This implies that these functions must be read from right to left if we want to see the evolution of the clinical parameters as the patients approach the day before discharge, when $x = 1$. In Fig. 2(a), most of the parameters remain stable as patients move from DTD_{21} to the day before discharge (DTD_1). Some others, such as SOFA-CNS, SOFA-Cardio, and SOFA-Resp, show a clear decrease from (denormalized) mean values 2, 1, and 1, down to normality value 0, respectively. STRATIFY also moves to normality levels. When denormalized, minimum and maximum mean daily glucose values also decrease from 111–159 to 104–143 mg/dl, and mean heart rate from 118 to 111 bpm. On the contrary, the mean arterial pressure (MAP) increases from 66 to 77 mmHg.

Fig. 2(b) shows the evolution of the variability of the normalized values of the clinical parameters[3]. If you analyze the functions from right to left, as the day of discharge approaches, a decrease of variability is evident for several parameters, particularly in the last 10 days before discharge. For previous days, most of the parameters show continuous oscillations in variational peaks and instability. The largest variability reductions are observed for SOFA-Cardio, SODA-CNS, and SOFA-Resp. These results were confirmed by p-values below 10^{-46}, after statistical Levene's tests. Except for patient temperature, whose variability is always at a reasonable level, and SOFA-Cardio and SOFA-CNS whose variability is above 0.2 until the day before discharge, the rest of clinical parameters show a variability in the last 10 days between 0.1 and 0.25. Among them, SOFA-Resp drastically reduces variability in the last 5 days (p-value ≈ 0), but the rest keep a smooth oscillation that concludes with STRATIFY and Glu-std above 0.2, Glu-max SOFA-Renal, SOFA-Coag, and SOFA-Cardio with 0.17, and HR, Glu-min, SOFA-Resp, MAP, SOFA-Liver between 0.11 and 0.14, the day before discharge.

These are considered high variabilities that explain the great heterogeneity of patients with regard to their clinical parameters, even if we only consider their conditions the day before discharge.

[2] Scales different from SOFA, NAS, and EMINA were considered as clinical parameters in the analysis.

[3] For 0–1 normalized values, the highest standard deviation is 0.5.

Table 1
Statistics on the patient description parameters in the heterogeneity study. The types of parameters are (N: numerical, C: categorical, CB: categorical binary, DE: demographic, CP: clinical parameter, SC: scale, AD: administrative).

(a) Numerical parameters

Name	Type	Mean	Stdev	Min	Max	Explanation
Age	N,DE	59.79	15.44	18	99	Age of the patient at ICU admission
PrevHospDays	N, CP	2.34	5.90	0	35.1	No. of hospital days previous to ICU
APACHEII	N, CP	21.90	8.78	0	51	APACHE-II score at ICU admission
MAP	N, CP	70.43	15.72	20	137	Mean arterial pressure in the day
HR	N, CP	114.61	23.04	45	200	Mean heart rate in the day
Tmp	N, CP	37.01	0.77	32.4	42	Mean body temperature in the day
Glu (min)	N, CP	108.74	26.40	20	200	Minimum glucose value in the day
Glu (max)	N, CP	153.68	47.93	63	330	Maximum glucose value in the day
Glu (stdev)	N, CP	18.91	16.09	0	82.73	Variation (stdev) of glucose in the day
STRATIFY	N, CP	2.78	1.12	0	5	Scale for identifying fall risk factors
SOFA_Cardio	N, CP	0.88	1.27	0	4	SOFA score (cardiovascular system)
SOFA_CNS	N, CP	1.29	1.36	0	4	SOFA score (nrevous system)
SOFA_Coag	N, CP	0.33	0.76	0	4	SOFA score (coagulation)
SOFA_Liver	N, CP	0.16	0.55	0	4	SOFA score (liver)
SOFA_Renal	N, CP	0.28	0.73	0	4	SOFA score (kidneys)
SOFA_Resp	N, CP	0.67	0.93	0	4	SOFA score (respiratory system)
SOFA_Total	N,SC	3.61	2.99	0	20	Addition of all the SOFA scores
NAS	N,SC	53.52	17.57	0	155	Nursing activity score (NAS)
EMINA	N,SC	8.94	2.86	0	15	Risk of pressure ulcers score

(b) Binary and categorical parameters

Name	Type	Proportions	Explanations
Gender	CB, DE	M: 67.3%, F: 32.7%	Gender of the ICU patient
PatType	CB, DE	Medical: 68.2%, Surgical: 31.8%	Type of ICU patient
AdmType	CB, DE	Urgent: 94.4%, Scheduled: 5.6%	ICU admission type
AdmWardGroup	C, AD	ER: 42.6%, Surgery: 19.1%, Scheduled: 18.5%, Other: 19.8%	ICU admission source
APACHE_ADmGroup	C, AD	Postop: 24.7%, HF: 10.1%, Trauma: 11.0%, RF: 10.1%, Other: 44.0%	APACHE admission group
CHE	C, CP	5: 21.7%, 0: 15.3%, 3: 14.1%, 6: 12.4%, Rest: 36.5%	Charlson score
AE	CB, CP	0: 97.2%, : 0.0%, 1-4: 2.8%	Adverse event
CA	CB, CP	N: 62.9%, Y: 37.1%	Arterial catheter Y/N
SU	CB, CP	Y: 80.9%, N: 19.1%	Urinary catheter Y/N
CVC	CB, CP	Y: 51.9%, N: 48.1%	Central venous catheter Y/N
Insuline	CB, CP	N: 53.3%, Y: 46.7%	Insuline Y/N
VA	CB, CP	N: 80.5%, Y: 19.5%	Vasoactive drugs Y/N
SA	CB, CP	N: 58.7%, Y: 41.3%	Sedative analgesics Y/N
ATF	CB, CP	N: 94.1%, Y: 5.9%	Antifungicals Y/N
ATB	CB, CP	N: 48.7%, Y: 51.3%	Taking antibiotics Y/N
VMI	CB, CP	N: 58.5%, Y: 41.5%	Invasive Mech. Vent. Y/N
VMNI	CB, CP	N: 95.7%, Y: 4.3%	Non-invasive Mech. Vent. Y/N
Isol	CB, AD	N: 82.2%, Y: 17.8%	Requires isolation Y/N
LTSV	CB, CP	N: 97.8%, No More: 2.1%, Other: 0.1%	Life support limitation

3.2. Patients heterogeneity based on the analysis of scales

Fig. 3(a) shows the reduction of the mean values of normalized EMINA, NAS, and SOFA-Total scores in the last days before patient discharge, with a trend towards normality values 7, 44, and 2, respectively. These last values have been denormalized to reflect their clinical sense.

Their variability in Fig. 3(b) stabilizes in a decreasing trend in the last five days, with SOFA-Total and NAS reaching the lowest mean variation close to 0.1. For normalized values, such variations cannot be considered low, particularly for patients in their previous day to discharge, and they reflect patient heterogeneity, concerning severity scales.

3.3. Patients heterogeneity based on confusion analysis

Table 2 shows the $n_\delta(i,j)$ matrix and the confusion measures for $\delta = 0.9$. This value of δ states that two patient descriptions are similar only if a similarity of 0.9 or higher is reached on a similarity scale from 0 to 1. The matrix shows that the largest values are in the diagonal $n_\delta(i,i)$, since obviously all the patients in DTD_i are similar to some patients that will be discharged in i days (e.g., themselves). We also observe that the values close to the diagonal are also large and that they decrease as we move away from the diagonal. This means that there is a degree of confusion between patients who were discharged on close days, and that this confusion decreases as the difference between the days until discharge increases. For example, position (4, 5) indicates that 715 patients, out of the 2488 patients in DTD_4, resemble $\delta = 90\%$ to some patients in DTD_5. That is to say, 715 patients who were discharged in

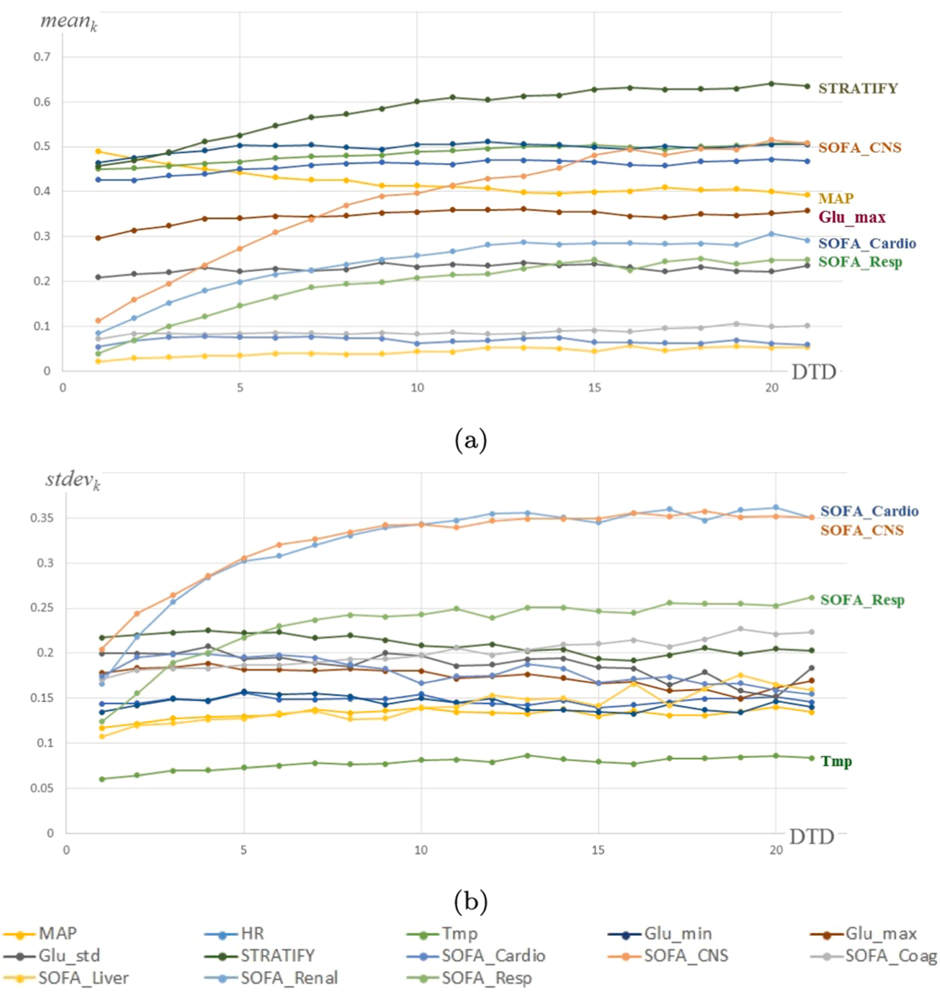


Fig. 2. ICU patient heterogeneity of clinical parameters: (a) and (b) show the respective progression of the mean and the standard deviation of the k numerical clinical parameters, as the days to discharge (DTD) increase from 1 to 21.

four days had a 90% similarity to some others who were discharged in five days. Similarly, there were 416 patients discharged in four days that were similar to some patients discharged in one day (see position (4, 1)).

Columns p_δ and o_δ measure premature and overdue discharges. For example, in average, 19% of the patients who resembled some patients discharged in two days, were in fact discharged in one day, 29% of the cases similar to patients discharged in three days, were discharged in less than three days, and 32% of those similar to patients discharged in four days were discharged in less than four days. These percentages are in rows 2–4 of column p_δ. Similarly for column o_δ, 47% of the patients looking like patients discharged in one day, were in fact, discharged in two or more days, and 35% of patients similar to patients discharged in two days were discharged in more than two days. Adding the values in both columns we calculate the feasible average discharge error, which is represented in column e_δ. That is to say, the number of patients not discharged in i days who resembled some cases discharged in i days. For example, 54% of the cases that were not discharged in two days are in fact very similar to other cases that were.

If we analyze these results in terms of patient heterogeneity, we can confirm that a large proportion of patients (37% on average) look a lot like the patients who were discharged earlier, and that a non inconsiderable number of patients (26% on average) are very similar to patients who were discharged later. In total, 63% of the patients discharged one day resemble patients who where discharged other days, on average. This represents a high heterogeneity in terms of DTD prediction since patients who share practically the same clinical description may have different discharge days.

Finally, columns b_δ and a_δ calculate Eqs. (13) and (14), for $n = 1, \ldots, 10$. They represent the average proportion of patients that resembled other patients that were discharged n days before or after. For example, in average 28% of the patients are similar to patients who are discharged one day before the first ones, and 16% are similar to patients who are discharged two days before. Similarly, 32% of the patients are equivalent to the patients discharged one day later.

3.4. Patients heterogeneity based on the analysis of clusters

All patient descriptions in the dataset define a clustering which is characterized by the days to discharge of each patient description. These clusters correspond to the DTD_i groups ($i = 1, \ldots, 21$). The elements in DTD_i describe patients who are in their i-th day before discharge, and they are expected to be more similar to each other than to patients in other DTD groups.

When calculating Davies-Bouldin, Dunn, and silhouette indices for DTD groups we obtain the values 11.43, 0.037, and -0.054. According to [21–23], these values suggest that the DTD clusters provide ‘no substantial structure’. In other words, patients with a similar DTD value are not necessarily similar, and patients with different DTD values are not necessarily different, in clinical terms. This heterogeneity could be partially attributed to the inclusion of patients with a high DTD in the study. For example, it would be natural for the patients in the DTD_{21} group to be very different from each other and thereby affect the final

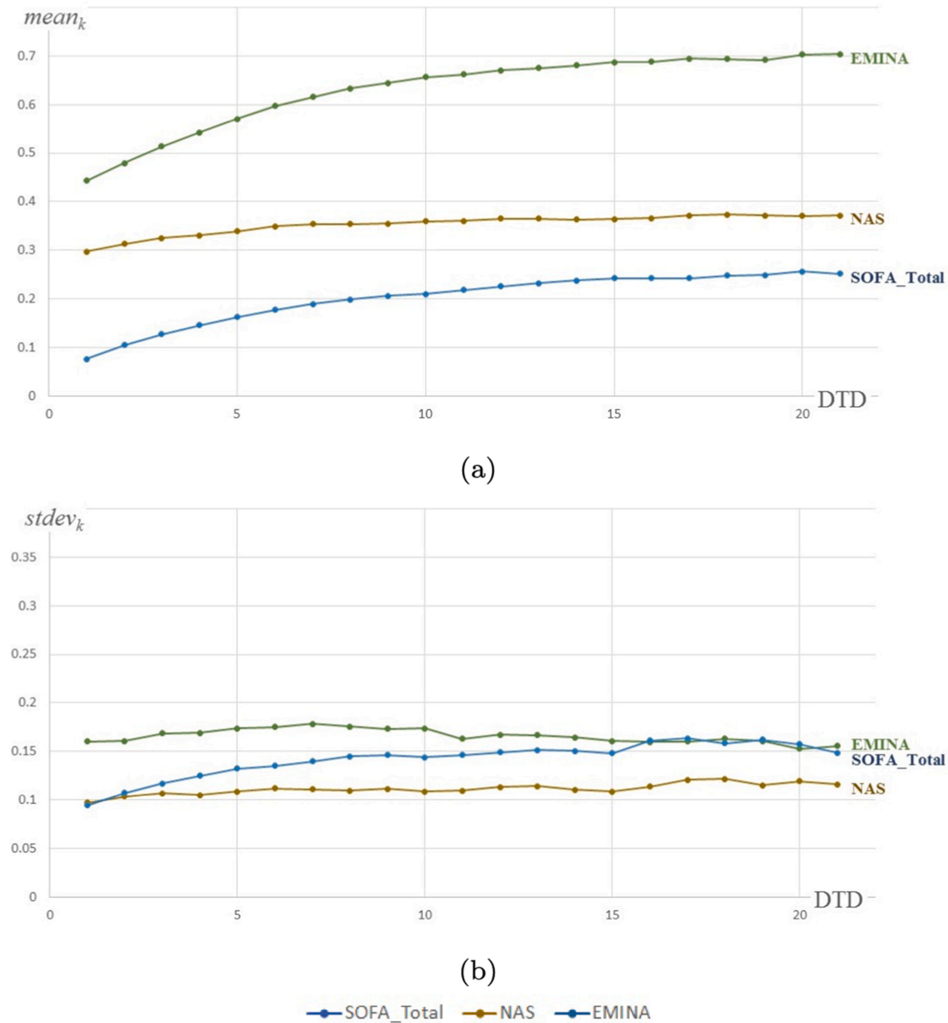


Fig. 3. ICU patient heterogeneity of clinical scales: (a) and (b) show the respective progression of the mean and the standard deviation of the SOFA-Total, NAS and EMINA scales as the discharge day is more distant.

Table 2
Confusion analysis of ICU patient heterogeneity: $n_\delta(i,j)$ matrix, DTD_i cardinalities n_i, and confusion indices $p_\delta(i), o_\delta(i), e_\delta(i), b_\delta(n), a_\delta(n)$ ($i = 1, \ldots, 21; n = 1, \ldots, 10$).

i\j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	n_i	s(i)	p_δ	o_δ	e_δ	b_δ	a_δ
1	3972	1517	835	472	281	177	124	84	69	48	48	32	22	21	9	12	8	9	11	2	3	3972	7463	0.00	0.47	0.47	0.28	0.32
2	1491	3799	1316	632	383	227	143	91	74	49	42	36	22	14	13	8	7	11	9	3	4	3799	8187	0.19	0.35	0.54	0.16	0.19
3	748	1253	3150	902	478	273	175	114	68	47	40	35	24	15	8	10	10	9	7	4	3	3150	7472	0.29	0.29	0.58	0.10	0.14
4	416	583	888	2488	715	363	208	127	92	64	46	41	28	15	13	12	8	11	5	4	5	2488	6208	0.32	0.28	0.60	0.06	0.10
5	254	357	470	709	2009	559	289	164	106	79	56	51	35	19	18	13	9	10	7	3	5	2009	5286	0.35	0.27	0.62	0.05	0.08
6	155	207	268	358	557	1635	417	218	150	99	73	61	34	26	21	17	10	11	6	8	8	1635	4390	0.36	0.26	0.63	0.03	0.06
7	108	125	162	200	287	414	1379	360	195	110	87	68	44	32	28	22	15	12	11	8	8	1379	3744	0.36	0.27	0.63	0.02	0.05
8	73	82	112	122	164	219	360	1169	291	168	113	76	54	43	30	24	19	20	12	6	4	1169	3186	0.36	0.27	0.63	0.02	0.04
9	62	69	67	88	106	151	199	291	1006	295	180	110	68	56	43	35	25	15	11	3	6	1006	2893	0.36	0.29	0.65	0.01	0.03
10	39	44	44	61	78	99	112	168	294	896	278	163	103	67	52	39	27	18	12	9	8	896	2630	0.36	0.29	0.66	0.01	0.03
11	40	36	39	45	58	73	89	114	179	277	809	271	134	93	64	51	32	25	18	9	11	809	2477	0.39	0.28	0.67		
12	28	32	33	38	52	60	68	75	110	163	271	740	223	117	94	59	38	30	15	12	11	740	2282	0.41	0.26	0.68		
13	18	20	23	27	35	33	45	53	66	103	133	223	679	191	122	79	67	43	25	18	17	679	2032	0.39	0.28	0.67		
14	17	13	15	14	19	26	32	43	55	67	93	117	191	619	183	107	65	53	32	22	16	619	1806	0.39	0.26	0.66		
15	8	12	8	13	18	21	28	30	43	52	63	94	122	183	568	174	114	65	41	36	24	568	1720	0.41	0.26	0.67		
16	7	6	10	10	13	17	22	24	35	39	50	58	79	107	174	524	155	95	55	42	33	524	1566	0.42	0.24	0.67		
17	7	7	9	7	9	10	15	19	25	27	32	38	67	65	114	155	493	145	78	51	34	493	1410	0.43	0.22	0.65		
18	7	10	9	9	9	11	12	20	15	18	25	30	43	53	65	95	145	464	154	88	53	464	1341	0.43	0.22	0.65		
19	8	8	7	4	7	6	11	12	11	12	18	15	25	32	41	55	78	154	430	132	73	430	1144	0.44	0.18	0.62		
20	2	3	4	4	3	8	8	6	3	9	9	12	18	22	36	42	51	88	132	406	124	406	990	0.46	0.13	0.59		
21	3	4	3	5	5	8	8	4	6	8	11	11	17	16	24	33	34	53	73	124	376	376	826	0.54	0.00	0.54		
																							mean:	0.37	0.26	0.63		

value of the indices. However, when we consider exclusively the patients during their last week in the ICU, the values obtained are 16.60, 0.03767 and −0.0403. And if only patients in the last three days are considered, the values are 1.806, 4.21 and −0.0133, still very far from values describing a good classification [21–23]. The Davies-Bouldin, Dunn and silhouette indices therefore confirm the heterogeneity of the patients, according to their DTD values.

4. Discussion

Having standard mechanisms to quantify the heterogeneity of patients in an ICU would provide the managers of these units with very useful tools for benchmarking, quality analysis, and improving the care planning of their services. A relevant aspect for this planning is related to the occupation of the unit and the prediction of days to discharge (DTD) of hospitalized patients. In this context, the heterogeneity is given by both the variability of the clinical descriptions of the patients, and the confusion when it comes to adequately predicting the day of discharge when clinically very disparate patients may have equivalent discharge days, and the opposite, patients who are similar but have very distant days to discharge.

Here, we have proposed four methods to calculate the heterogeneity of patients in ICU for the study of the daily prediction of days to discharge. These methods provide different interpretations of the variability of ICU patients and allow different types of analysis to be carried out. Thus, the first method is based on the statistical study of the variation of each clinical parameter in an isolated and continuous manner as the day of discharge approaches. It is based on the assumption that, as the day of discharge approaches, the clinical parameters should converge towards normality values that allow the patient to be discharged. In other words, a greater heterogeneity is expected between patients one or two weeks before discharge than between patients on their day before discharge. This method can be interesting to identify clinical parameters or combinations of clinical parameters to be used as DTD predictive biomarkers. In our study, we observed that (1) STRATIFY, SOFA-CNS, SOFA-Cardio, SOFA-Resp, and maximum glucose reached normality values (in average) as the discharge day approached, and (2) the variability of the clinical parameters remained at similar mean values, regardless of the time to discharge, thus showing a high heterogeneity. Only SOFA-Cardio, SOFA-CNS and SOFA-Resp standard deviations decreased rapidly during the last seven days before discharge, whereas MAP (mean arterial pressure) showed a light increment. However, in the day before discharge, variability of all variables except temperature remained within the range 0.1–0.25. Given that the highest possible variability of normalized data is 0.5, we can conclude that the clinical parameters are highly variable (i.e., their values are highly heterogeneous). Consequently, none of the clinical parameters studied seems to be a good DTD predictor.

The second method is based on the variability of clinical scales and it is useful to assess the predictive capacity of the scales in reference to the DTD. Some scales, such as APACHE II and SOFA, are widely used in the ICU to measure and compare patient complexity. Other scales such as NAS and EMINA evaluate the workload and nursing care. The use of the variability of the scales in the analysis of heterogeneity is based on considering that, as the patient is close to her discharge day, the values of the scales should tend towards mild values and their variability reach low values. Any other behavior represents a high heterogeneity of the scale values and would invalidate them as good DTD indices. In our analysis, none of the scales considered showed a big reduction in the variation which remained stable in the range 0.1–0.2. Again, these values must be interpreted in the context of a maximum variability of 0.5. This high heterogeneity observed in the values of the scales studied limits their quality for determining the DTD of the patients.

Our third method is based on the concept of confusion representing the risk of possible erroneous discharges caused by the lack of alignment between the clinical similarity between patients and the proximity of their DTDs. Based on this, five new metrics were introduced to quantify the risk of premature and overdue discharges, the risk of discharge errors, and the cumulative risks of confusing a patient with other patients who were discharged up to *n* days before or up to *n* days after. This is an interesting method for analyzing the heterogeneity of patients in clinical studies in which a function of distance between patients has been defined (e.g., patient clustering, patient retrieval, or precision medicine [24]). Since all the metrics measure confusion in the interval 0–1, patient heterogeneity can be expressed as a percentage. So, our study showed that the mean confusion error (e_δ) was 0.63, which means that 63% of the patients are similar to other patients with a different DTD. This level of confusion describes a high heterogeneity of the patients from a DTD perspective. As a general rule, we could accept that values greater than 10% define patients with a high heterogeneity.

The forth method adapts the Davies-Bouldin [18], Dunn [19], and silhouette [20] indices to describe patient heterogeneity. If a good clustering seeks to group similar cases in the same cluster, in terms of DTDs, similar patients should be grouped with other patients with the same DTD. Following the interpretation of the values provided by the original indices, our DB and D Eqs. 15,16 measure heterogeneity as a positive number, with 0 the lowest possible heterogeneity, and *silhouette* Eq. (17) in the interval $[-1, +1]$ with lower heterogeneity as the value approaches to $+1$, and values below 0.25 considered high heterogeneity. Our study obtained BD, D, and *silhouette* values that confirmed high heterogeneity, according to the interpretations in [21–23].

We identified some limitations in our current work: (1) the results concern a single ICU. Whereas a multi-center analysis would be preferred, the inclusion of all the clinical cases seen in the ICU of a reference hospital providing ICU services to a population of 750,000 inhabitants and including all the survival cases in four consecutive years, can possibly make the study representative of many other ICUs. (2) Our study involves all the types of ICU patients. ICUs receive a great variety of patient types, so heterogeneity is inherent in intensive patients. Constraining our study to concrete patient types (e.g., surgical, scheduled, or emergency) could affect our results on heterogeneity. In the future, we will do heterogeneity subgroup analysis; however, relevant ICU studies such as [25] use to involve all the ICU population in their works. (3) The confusion analysis is based on the idea that if a patient who is discharged in *i* days is similar to another patient who is discharged in $j \neq i$ days, this is a source of confusion. In future works, we will consider a confusion if the patient is similar to a percentage of the patients discharged in *j* days, and not only if it is similar to one of them. (4) Other factors with a possible influence in the results are the internal organization of the ICU, which may have decided some patient discharges based on organizational reasons rather than pure clinical reasons, or the exclusion of deaths from the study.

Two important issues in the analysis of ICU patient heterogeneity concern the way that time and granularity are considered. In this work, we proposed heterogeneity measures based on the application of temporary windows of one-day granularity. That is to say, we aggregate all the patient information throughout the days in single-day descriptions so that the evolutions of the patients are described longitudinally in units that represent consecutive days. In our context of DTD prediction, this is a convenient approach since discharge decisions in ICUs are made in a daily basis and discharge predictions are also needed in the scale of days. However, other approaches such as time-stamps, time series analysis [26,27] or a finer granularity [26,24] below one day are also possible. These could give rise to other methods and metrics for ICU patient heterogeneity quantification different from ours.

5. Conclusions

When the data on the patients admitted in the ICU of the University Hospital Joan XXIII in Spain are the subject of four different heterogeneity analyses, each one following alternative perspectives of heterogeneity, all confirm a high heterogeneity of conditions of the patients within the same DTD group and also confusion between patients with distant DTDs, which may contain very similar patients. This could explain the complexity of making accurate DTD predictors, and it might be one of the reason for such few published works on this relevant ICU issue.

This paper contributes with a new set of methods and metrics to quantify patient heterogeneity in ICUs with a focus in the prediction of the days to discharge. These methods and metrics come to cover a gap on the current available tools for patient heterogeneity analysis. Although we focused on the study of the heterogeneity of patients in an ICU, the formal methods presented could also be applied to inpatients from any other hospital service.

CRediT authorship contribution statement

DR with the help of DC made the conceptualization of the work, DC and JG took part in the data curation, DC did all the programming with DR's supervision, DR and DC performed the technical analysis of the results, AR and MB contributed with the clinical analysis and validation, DR wrote the initial draft of the paper, which was converted into the final version after a review by all the authors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by the Spanish Ministerio de Ciencia e Innovación, under the project PID2019-105789RB-I00, the Instituto de Salud Carlos III, under the contract FIS PI16/00491, and the project Innòbics 01/2019.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jbi.2021.103768.

References

[1] M.C. Lenert, R.A. Miller, Y. Vorobeychik, C.G. Walsh, A method for analyzing inpatient care variability through physicians' orders, J. Biomed. Inform. 91 (2019) 103111.
[2] J. Ruyssinck, J. van der Herten, R. Houthooft, F. Ongenae, I. Couckuyt, B. Gadeyne, K. Colpaert, J. Decruyenaere, F.D. Turck, T. Dhaene, Random survival forests for predicting the bed occupancy in the intensive care unit, Comput. Math. Methods Med. (2016) 1–7.
[3] Y.U. Bing-Hua, Delayed admission to intensive care unit for critically surgical patients is associated with increased mortality, Am. J. Surgery 208 (2) (2014) 268–274.
[4] J.M. Droogh, M. Smit, A.R. Absalom, J.J. Ligtenberg, J.G. Zijlstra, Transferring the critically ill patient: are we there yet? Crit. Care 19 (1) (2015) 62.
[5] A. Awad, M. Bader-El-Den, J. McNicholas, Patient length of stay and mortality prediction: A survey, Health Services Manage. Res. 30 (2) (2017) 105–120.
[6] I.W.M. Verburg, A. Atashi, S. Eslami, R. Holman, A. Abu-Hanna, E. de Jonge, N. Peek, N.F. de Keizer, Which models can i use to predict adult ICU length of stay? A systematic review, Crit. Care Med. 45 (2) (2017) e222–e231.
[7] I.T. Peres, S. Hamacher, F.L.C. Oliveira, A.M.T. Thomé, F.A. Bozza, What factors predict length of stay in the intensive care unit? J. Crit. Care 60 (2020) 183–194.
[8] G. Gutierrez, Artificial intelligence in the intensive care unit, Crit. Care 24 (101) (2020).
[9] M.W. Temple, C.U. Lehmann, D. Fabbri, Predicting discharge dates from the NICU using progress note data, Pediatrics 136 (2) (2015) e395–e405.
[10] D. Cuadrado, D. Riaño, J. Gómez, M. Bodí, G. Sirgo, F. Esteban, R. Rodríguez, Pursuing optimal prediction of discharge time in ICUs with machine learning methods, in: Artificial Intelligence in Medicine, Vol. 11526, Springer, 2019, pp. 150–154.
[11] K.A. Artis, J. Bordley, V. Mohan, J.A. Gold, Data omission by physician trainees on ICU rounds, Crit. Care Med. 47 (3) (2019) 403–409.
[12] G. Sirgo, F. Esteban, J. Gomez, G. Moreno, A. Rodriguez, L. Blanch, J. Guardiola, R. Gracia, L. De Haro, M. Bodi, Validation of the ICU-DaMa tool for automatically extracting variables for minimum dataset and quality indicators: The importance of data quality assessment, Int. J. Med. Informatics 112 (2018) 166–172.
[13] E.J. Caruana, M. Roman, J. Hernández-Sánchez, P. Solli, Longitudinal studies, J. Thorac. Dis. 7 (11) (2015) E537–E540.
[14] R. Moreno, J. Vincent, R. Matos, A. Mendonça, F. Cantraine, L. Thijs, J. Takala, C. Sprung, M. Antonelli, H. Bruining, S. Willatts, The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. results of a prospective, multicentre study. working group on sepsis related problems of the esicm, Intensive Care Med. 25 (7) (1999) 686–696.
[15] S. Lambden, P.F. Laterre, M.M. Levy, B. Francois, The SOFA score-development, utility and challenges of accurate assessment in clinical trials, Crit Care 23 (1) (2019).
[16] D.R. Miranda, R. Nap, A. de Rijk, W. Schaufeli, L. G, Therapeutic intervention scoring system. Nursing Activities Score, Crit. Care Med. 31 (2003) 374–382.
[17] A. Roca-Biosca, F.P. García-Fernández, S. Chacon, L. Rubio, M. olona cabases, L. Saperas, N. Garcia-Grau, G. Tuset-Garijo, I. De Molina-Fernandez, M. Velasco-Guillen, Validation of EMINA and EVARUCI scales for assessing the risk of developing pressure ulcers in critical patients (in spanish), Enfermería Intensiva 26 (2015) 15–23.
[18] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. 1 (2) (1979) 224–227.
[19] J.C. Dunn, Well-separated clusters and optimal fuzzy partitions, J. Cybernet. 4 (1) (1974) 95–104.
[20] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65.
[21] M. Halkidi, Y. Batistakis, M. Vazirgiannis, On clustering validation techniques, J. Intell. Inf. Syst. 17 (2001) 107–145.
[22] M.K. Pakhira, S. Bandyopadhyay, U. Maulik, Validity index for crisp and fuzzy clusters, Pattern Recogn. 37 (3) (2004) 487–501.
[23] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data. An Introduction to Cluster Analysis, John Wiley and Sons, 1990.
[24] C. Cosgriff, L. Celi, D. Stone, Critical care, critical data, Biomed. Eng. Comput. Biol. 10 (2019), https://doi.org/10.1177/1179597219856564, 1179597219856564.
[25] K. Vranas, J. Jopling, T. Sweeney, M. Ramsey, A. Milstein, C. Slatore, G. Escobar, V. Liu, Identifying distinct subgroups of ICU patients: A machine learning approach, Crit. Care Med. 45 (10) (2017) 1607–1615.
[26] J.A. Silva, E.R. Faria, R.C. Barros, E.R. Hruschka, A.C.P.L.F. d. Carvalho, J. a. Gama, Data stream clustering: A survey, ACM Comput. Surv. 46 (1). doi:10.1145/2522968.2522981.
[27] S. Aghabozorgi, A. Seyed Shirkhorshidi, T. Ying Wah, Time-series clustering – a decade review, Inf. Syst. 53 (2015) 16–38, https://doi.org/10.1016/j.is.2015.04.007, http://www.sciencedirect.com/science/article/pii/S0306437915000733.