The usefulness of machine learning in pediatric intensive care: k-means clustering for shock classification

María Rollán-Martínez-Herrera (meroll16@gmail.com)
Biocruces Bizkaia Health Research Institute
Jon Keretxeta-Sarregi
Vicomtech Foundation
Javier Gil-Antón
Cruces University Hospital
Iván Macia-Oliver
Vicomtech Foundation
Javier Pilar-Orive
Cruces University Hospital

Research Article

Keywords:

Posted Date: April 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1546845/v1

Abstract

Purpose: The use of unsupervised learning techniques based on the multiple variables routinely collected in PICUs could help to find patterns that clinicians may have overlooked, and that may define new classifications of types of shock in the pediatric setting.

Methods: data from 90 patients with shock diagnosis, stored in a third level PICU database, were used. K-means algorithm was applied over the data from the first 24 hours of admission (physiological and analytical variables and the need for devices), without preceding death in less than 48 hours. 3 clusters were created. It was analysed whether these clusters had differences between them, their correlation with diagnosis at discharge, their association with outcomes (mortality and LOS) and whether they were more accurate than classical classification in outcome prediction.

Results: significant differences were found in variables used (e.g. mean diastolic arterial pressure $p < 0.001$, age $p < 0.001$) and not used for training (e.g. $EtCO_2$ min $p < 0.001$, Troponin max $p < 0.01$), diagnosis at discharge ($p < 0.001$) and outcomes ($p < 0.05$). Clustering classification equalled classical classification in its association with LOS ($p = 0.01$) and surpassed it in its association with mortality ($p < 0.04$ vs. $p = 0.16$).

Conclusion: even with such a small sample, k-means clustering method was able to classify shocked pediatric patients with higher outcome correlation than the clinical method. These results support the utility of unsupervised learning algorithms for patient classification in PICU, something that could help to implement guided therapies.

Introduction

Shock is described as an inadequate oxygen supply to the tissues and can be classified in multiple ways. There are four classic pathophysiological variants: cardiogenic, hypovolemic, distributive and obstructive, which explain the hemodynamic situation of the patient and are therefore useful for establishing therapy. Shock can also be classified according to its etiology: septic, anaphylactic, hemorrhagic... The problem with these classifications is they are often confusing, since each pathophysiological type has different etiological varieties and each of the etiological varieties may correspond to different pathophysiological types [1–5].

When faced with a new case of shock, it is necessary to simultaneously evaluate multiple physiological variables to determine the pathophysiological variety, in order to establish a targeted therapy [3]. However, despite the time that has been spent researching about hemodynamic monitoring, there are still no adequate techniques to make an accurate shock classification [4–6], which affect negatively to the clinical management.

Multiple studies have been performed, and classical statistical analysis has been used to determine the type of shock in each patient. However, no single solution to this historical problem has yet been found.

Currently, there are multiple models of biological data analysis using artificial intelligence techniques based on machine learning [7]. The use of supervised learning for patient categorization would imply the assumption that the ground truth of the disease classification is already perfect, but in practical terms this classification may be not optimal or be outdated. Therefore, in clinical practice, similarities between patients are continually encountered, which in the long run sometimes end up defining new categories.

The present study proposes that the use of unsupervised learning techniques based on the multiple variables routinely collected in PICUs (Pediatric Intensive Care Unit) could help to find patterns that clinicians may have overlooked, and that may define new classifications of types of shock in the pediatric setting. The use of clustering algorithms for the classification of shock could not only be applied globally, but could also be used locally. It could be useful to study the characteristics of each hospital and classify patients into clinical groups to study treatment and evolution, similarly to what is done at the level of microbiological flora analysing local behaviour and resistances.

Material And Methods

A retrospective observational study was carried out for the development of a computational model for the classification of types of shock in the PICU of Cruces University Hospital. Clinical and analytical data were collected for all pediatric patients (0 to 14 years) diagnosed with any type of shock since the implementation of "IntelliSpace Critical Care and Anesthesia", in 2012. This study was approved by the local Ethics Committee (Comité de Ética de la Investigación (CEI) OSI Ezkerraldea-Enkarterri-Cruces) with the number CEIC E21/07.

Data were collected in a computerized manner through the program's database: 180 patients had been diagnosed with shock but only 100 had data. Hourly data of physiological, gasometric and analytical variables were obtained. In addition, devices, age, weight, length of stay, diagnostic at discharge and discharge reason, were recorded.

The data were preprocessed, eliminating incorrect values, variables with more than 80% of missing values and variables of no clinical interest (such as number of central venous catheter lumens, urinary catheter size…). After preprocessing, patients with missing values in the heart rate, respiratory rate and pressure columns were eliminated. Afterwards the resulting dataset was of 90 patients. In order to achieve higher clinical value in the classification, only data from the first 24 hours of admission were selected and data preceding death in less than 48 hours were eliminated.

For those variables that depends on age, the values were adjusted with the z-score for that age (weight [8, 9], blood pressure (using the p. 50 for height) [10], and heart and respiratory rate [11]).

After that, the mean, minimum (min) and maximum (max) values of each monitored numerical variable were used. A classification of the patients was performed by clustering.

The following statistical techniques were used. Lillefors test was used to analyse normality. Normal variables were described by mean, non-normal variables by median and qualitative variables by proportion. For quantitative variables with 2 groups, t-test (with Welch correction if homocedastity was not met) or Wilcox test were used depending on normality. In quantitative variables with more than 2 groups, ANOVA or Kruskal-Wallis were used depending on normality and for post-hoc analyses were done with Tukey and the Mann-Whitney test with significance correction respectively. Chi-square was used for qualitative variables. In the case of length of stay, Log-Rank was also used, with post-hoc analysis using Bonferroni correction. A $p < 0.05$ was considered the cut-off point for statistical significance.

K-means was the algorithm selected for clustering [6]. This algorithm classifies individuals into multiple groups, so that individuals within the same group are as similar as possible while individuals from different groups are as different as possible. This similarity is estimated using the Euclidean distance [7], so that if two patients have very different values for a variable, the distance will be high, and vice versa. Each group is represented by its centroid, which corresponds to the mean of all individuals in that group, the initial centroids are randomly selected by the algorithm, and this may influence the results, that is why 25 initial configurations were attempt. As characteristic of k-means algorithm, it is necessary to define both the grouping variables and the number of clusters.

The clustering variables were selected according to their clinical relevance, discarding those with more than 45% of missing values. To study the optimal number of clusters, the elbow method and the average silhouette method were used, and several tests were performed to check the spatial distribution of the groups (Supplemental Figure) and the number of patients per group. Three was taken as the optimal number of clusters.

Once the grouping variables and the number of clusters were selected, the data were prepared: each missing value was estimated by averaging the remaining individuals and then the data were standardized with z-score [8] to make all variables comparable.

Once the clusters were selected, the clinical significance of those clusters was sought as follows:

- The characteristics of each group were studied to determine whether there were differences between them.
- The correlation between the unsupervised classification and the diagnosis at discharge was studied.
- It was assessed whether the classification was related to the outcomes (mortality and length of stay).
- It was tested whether the new classification had a greater association with outcomes than the classic classification.

All analyses were performed with R-Studio; tables and graphs were made with R-studio and Microsoft Excel.

Results

K-means classification was performed using the variables indicated in the Table 1. The following clusters were obtained: 46 patients, 18 patients and 26 patients (Fig. 1).

Table 1
Comparison of variables. Mean (mn) or median (md) and its confidence interval (CI) per group according to the normality of quantitative data, or number (n), proportion (p) and CI of qualitative data, p value of differences and test used for comparison between samples.

	Cluster 1 (n = 46)	Cluster 2 (n = 18)	Cluster 3 (n = 26)	p value	Test
Variables used for clustering	(mn/md/p (CI95%))	(mn/md/p (CI95%))	(mn/md/p (CI95%))		
Female sex	n: 26, p: 0.57 (0.42 : 0.71)	n: 6, p: 0.33 (0.09 : 0.57)	n: 8, p: 0.31 (0.12 : 0.5)	0,06	Chi-square
Age in months	md: 33.63 (14.97 : 55.17)	md: 3.25 (0.2 : 8.77)	md: 137.13 (102.17 : 169.43)	**< 0,001**	Kruskal-Wallis
Weight z-score for age	md: -1.3 (-2.18 : -0.51)	md: -0.93 (-2.07 : 0.16)	md: 0.03 (-0.82 : 2.05)	**0,02**	Kruskal-Wallis
Hearth rate z-score for age (min)	md: -0.32 (-0.62 : 0.35)	md: -2.59 (-5.97 : -0.56)	md: -0.27 (-1.26 : 0.21)	**0,01**	Kruskal-Wallis
Hearth rate z-score for age (mean)	mn: 1.82 (1.4 : 2.24)	mn: 0.77 (-0.11 : 1.64)	mn: 1.99 (1.41 : 2.58)	0,93	ANOVA
Hearth rate z-score for age (max)	md: 3.82 (3.44 : 4.31)	md: 2.7 (1.85 : 4.55)	md: 5.13 (4.39 : 5.38)	**0,002**	Kruskal-Wallis
Respiratory rate z-score for age (min)	md: -2.33 (-2.78 : -1.99)	md: -3.61 (-5.14 : -2.89)	md: -1.06 (-2.33 : -0.63)	**< 0,001**	Kruskal-Wallis
Respiratory rate z-score for age (mean)	mn: 1.07 (0.2 : 1.95)	mn: -1.04 (-2.6 : 0.51)	mn: 2.18 (0.87 : 3.49)	0,32	ANOVA
Respiratory rate z-score for age (max)	mn: 6 (3.49 : 8.51)	mn: 2.63 (-0.38 : 5.63)	mn: 7.49 (5.01 : 9.98)	0,59	ANOVA
Diastolic arterial pressure z-score for age (min)	md: -0.15 (-0.22 : -0.09)	md: -0.09 (0.39)	md: 0.05 (-0.1 : 0.1)	**0,01**	Kruskal-Wallis
Diastolic arterial pressure z-score for age (mean)	mn: 0.24 (0.19 : 0.29)	mn: 0.44 (0.09 : 0.78)	mn: 0.54 (0.48 : 0.61)	**< 0,001**	ANOVA
Diastolic arterial pressure z-score for age (max)	mn: 0.78 (0.7 : 0.87)	mn: 1.23 (0.31 : 2.15)	mn: 1.15 (1.02 : 1.28)	**< 0,001**	ANOVA
Median arterial pressure z-score for age (min)	md: -0.18 (-0.29 : -0.13)	md: -0.72 (-0.3)	md: 0.04 (-0.11 : 0.09)	**< 0,001**	Kruskal-Wallis
Median arterial pressure z-score for age (mean)	mn: 0.15 (0.11 : 0.18)	mn: 0.15 (-0.02 : 0.32)	mn: 0.41 (0.35 : 0.47)	**< 0,001**	ANOVA
Median arterial pressure z-score for age (max)	mn: 0.54 (0.47 : 0.62)	mn: 0.78 (0.06 : 1.5)	mn: 0.85 (0.76 : 0.94)	**< 0,001**	ANOVA

	Cluster 1 (n = 46)	Cluster 2 (n = 18)	Cluster 3 (n = 26)	p value	Test
Systolic arterial pressure z-score for age (min)	mn: -0.21 (-0.28 : -0.14)	mn: -0.35 (-0.58 : -0.12)	mn: -0.01 (-0.07 : 0.06)	**< 0,001**	ANOVA
Systolic arterial pressure z-score for age (mean)	mn: 0.1 (0.05 : 0.14)	mn: -0.07 (-0.16 : 0.02)	mn: 0.32 (0.27 : 0.36)	**< 0,001**	ANOVA
Systolic arterial pressure z-score for age (max)	md: 0.37 (0.28 : 0.43)	md: 0.32 (0.69)	md: 0.61 (0.56 : 0.69)	**< 0,001**	Kruskal-Wallis
Daily diuresis in mL/kg/day	mn: 1.85 (1.16 : 2.54)	mn: 3.4 (0 : 7.27)	mn: 1.51 (0.68 : 2.33)	0,90	ANOVA
Temperature in ºC (min)	mn: 36.74 (36.46 : 37.02)	mn: 34.82 (34.19 : 35.46)	mn: 36.7 (36.22 : 37.17)	0,50	ANOVA
Temperature in ºC (mean)	mn: 37.02 (36.75 : 37.28)	mn: 35.57 (35.15 : 35.98)	mn: 37.15 (36.72 : 37.58)	0,97	ANOVA
Temperature in ºC (max)	mn: 37.31 (36.96 : 37.67)	mn: 36.18 (35.65 : 36.7)	mn: 37.7 (37.18 : 38.21)	0,43	ANOVA
Oxygen saturation in % (min)	md: 94 (93 : 95)	md: 79 (61 : 87)	md: 95.5 (94 : 96)	**0,002**	Kruskal-Wallis
Oxygen saturation in % (mean)	md: 97.88 (96.94 : 98.48)	md: 94 (91.33 : 96.56)	md: 98.27 (97.38 : 98.86)	**0,02**	Kruskal-Wallis
Venous oxygen saturation in % (min)	mn: 56.12 (50.16 : 62.08)	mn: 34.53 (24.92 : 44.14)	mn: 62.96 (57.16 : 68.76)	0,30	ANOVA
Venous oxygen saturation in % (max)	mn: 77.24 (71.62 : 82.85)	mn: 71.73 (59.86 : 83.61)	mn: 78.6 (74.59 : 82.61)	0,81	ANOVA
Carboxyhemoglobin in % (max)	mn: 2.3 (1.89 : 2.71)	mn: 2.49 (2.03 : 2.94)	mn: 2.55 (2.1 : 3)	0,38	ANOVA
Inspirited oxygen fraction in % (max)	mn: 57.78 (46.9 : 68.65)	mn: 88.33 (76.73 : 99.93)	mn: 46.53 (31.11 : 61.96)	0,81	ANOVA
Inspirited oxygen fraction in % (mean)	md: 36.19 (27.5 : 48)	md: 55.59 (40 : 69.76)	md: 26 (22.6 : 40)	**< 0,001**	Kruskal-Wallis
Capillary glucose in mg/dL (min)	mn: 100.11 (90.91 : 109.31)	mn: 75.53 (53.84 : 97.22)	mn: 111.77 (96.74 : 126.81)	0,39	ANOVA
Capillary glucose in mg/dL (max)	md: 156.5 (121 : 193)	md: 156.5 (121 : 193)	md: 156.5 (121 : 193)	**0,01**	Kruskal-Wallis

	Cluster 1 (n = 46)	Cluster 2 (n = 18)	Cluster 3 (n = 26)	p value	Test
Calcium ion in mg/dL (min)	md: 4.6 (4.2 : 4.7)	md: 3.6 (2.9 : 3.8)	md: 4.54 (4.4 : 4.7)	**< 0,001**	Kruskal-Wallis
Creatinine in mg/dL (max)	mn: 3.18 (0 : 7.05)	mn: 3.92 (0 : 10.08)	mn: 3.98 (0 : 7.98)	0,77	ANOVA
Phosphate in mg/dL (max)	md: 4.8 (4.2 : 5.2)	md: 8 (5.8 : 10.7)	md: 4.25 (3.8 : 4.9)	**< 0,001**	Kruskal-Wallis
Reactive C protein in mg/L (max)	md: 55.9 (30.56 : 98.8)	md: 11.4 (7.37 : 35.49)	md: 163.6 (80.73 : 199.21)	**< 0,001**	Kruskal-Wallis
Lymphocytes · 1000/µL (min)	md: 1.4 (0.86 : 2.3)	md: 0.9 (0.5 : 2.29)	md: 0.21 (0.02 : 0.56)	**< 0,001**	Kruskal-Wallis
Lymphocytes · 1000/µL (max)	md: 3.24 (1.9 : 7)	md: 4.42 (1.2 : 21)	md: 0.24 (0.1 : 0.9)	**< 0,001**	Kruskal-Wallis
Neutrophils · 1000/µL (min)	md: 7.25 (5.57 : 11.66)	md: 2.65 (1.9 : 3.77)	md: 1.43 (0.03 : 5.54)	**< 0,001**	Kruskal-Wallis
Neutrophils · 1000/µL (max)	mn: 13.65 (9.61 : 17.69)	mn: 5.53 (3.83 : 7.23)	mn: 6.01 (2.77 : 9.25)	**0,002**	ANOVA
Mechanical ventilation	n: 36, p: 0.78 (0.66 : 0.91)	n: 18, mn: 1	n: 15, p: 0.58 (0.37 : 0.78)	**0,005**	Chi-square
Hemodiafiltration	n: 1, p: 0.02 (0 : 0.07)	n: 5, p: 0.28 (0.05 : 0.51)	n: 1, p: 0.04 (0 : 0.12)	**0,002**	Chi-square
ECMO	n: 1, p: 0.02 (0 : 0.07)	n: 15, p: 0.83 (0.64 : 1.02)	n: 0, p: 0	**< 0,001**	Chi-square
Thermic blanket	n: 5, p: 0.11 (0.02 : 0.2)	n: 1, p: 0.06 (0 : 0.17)	n: 0, p: 0	0,20	Chi-square
Non-clustering variables	(mn/md/p (CI95%))	(mn/md/p (CI95%))	(mn/md/p (CI95%))		
$EtCO_2$ in Torr (min)	md: 30 (27 : 37)	md: 9 (8 : 10)	md: 31 (0 : 38)	**< 0,001**	Kruskal-Wallis
$EtCO_2$ in Torr (mean)	mn: 42.56 (35.27 : 49.86)	mn: 22.33 (16.54 : 28.12)	mn: 35.19 (25.63 : 44.74)	**0,02**	ANOVA
$EtCO_2$ in Torr (max)	mn: 51.35 (43.8 : 58.9)	mn: 33.08 (22.85 : 43.32)	mn: 43.25 (32.67 : 53.83)	**0,04**	ANOVA
End expiratory pressure (mean)	md: 5.5 (5 : 6.57)	md: 6.97 (5 : 7.67)	md: 5.14 (0 : 5.17)	0,13	Kruskal-Wallis

	Cluster 1 (n = 46)	Cluster 2 (n = 18)	Cluster 3 (n = 26)	p value	Test
End expiratory pressure (max)	md: 6 (5 : 7)	md: 9 (5 : 10)	md: 6 (0 : 9)	0,10	Kruskal-Wallis
Cerebral NIRS in % (min)	mn: 49.67 (39.43 : 59.9)	mn: 47.25 (35.27 : 59.23)	mn: 61.25 (45.11 : 77.39)	0,44	ANOVA
Cerebral NIRS in % (mean)	mn: 59.59 (47.6 : 71.59)	mn: 66.11 (59.64 : 72.58)	mn: 70.73 (64.26 : 77.2)	0,12	ANOVA
Troponin in ng/L (max)	md: 249 (0 : 1653)	md: 2732 (329 : 4853)	md: 29 (0 : 35)	**0,01**	Kruskal-Wallis
Cardiogenic shock	n: 9, p: 0.2 (0.08 : 0.31)	n: 13, p: 0.72 (0.49 : 0.95)	n: 2, p: 0.08 (0 : 0.19)	**< 0,001**	Chi-square
Inflammatory shock	n: 24, p: 0.52 (0.37 : 0.67)	n: 3, p: 0.17 (0 : 0.36)	n: 21, p: 0.81 (0.65 : 0.97)	**< 0,001**	Chi-square
Septic shock	n: 23, p: 0.5 (0.35 : 0.65)	n: 3, p: 0.17 (0 : 0.36)	n: 20, p: 0.77 (0.6 : 0.94)	**< 0,001**	Chi-square
Anaphylactic shock	n: 2, p: 0.04 (0 : 0.1)	n: 0, p: 0	n: 0, p: 0	0,38	Chi-square
Hypovolemic shock	n: 7, p: 0.15 (0.04 : 0.26)	n: 5, p: 0.28 (0.05 : 0.51)	n: 3, p: 0.12 (0 : 0.25)	0,34	Chi-square
Hypovolemic secondary to traumatism	n: 3, p: 0.07 (0 : 0.14)	n: 0, mn: 0	n: 1, p: 0.04 (0 : 0.12)	0,52	Chi-square
Hypovolemic secondary to surgery	n: 1, p: 0.02 (0 : 0.07)	n: 1, p: 0.06 (0 : 0.17)	n: 1, p: 0.04 (0 : 0.12)	0,78	Chi-square
Non-specified shock	n: 9, p: 0.2 (0.08 : 0.31)	n: 0, p: 0	n: 1, p: 0.04 (0 : 0.12)	**0,03**	Chi-square
Cardiac surgery	n: 3, p: 0.2 (0.08 : 0.31)	n: 8, p: 0.72 (0.49 : 0.95)	n: 2, p: 0.08 (0 : 0.19)	**< 0,001**	Kruskal-Wallis
Oncologic patients	n: 2, p: 0.04 (0 : 0.1)	n: 1, p: 0.06 (0 : 0.17)	n: 8, p: 0.31 (0.12 : 0.5)	**0,003**	Kruskal-Wallis
Length of stay in days	md: 6 (3 : 9)	md: 11 (4 : 29)	md: 4 (2 : 7)	**0,02**	Kruskal-Wallis
Exitus	n: 3, p: 0.07 (0 : 0.14)	n: 6, p: 0.33 (0.09 : 0.57)	n: 1, p: 0.04 (0 : 0.12)	**0,004**	Chi-square

Analysis of variables used for clustering

Each of the variables was compared between the different clusters (Table 1). Among all the variables, the need for ECMO (ExtraCorporeal Membrane Oxygenation) was the one with the greatest weight, so the clustering was performed again without it. The clusters obtained were 45, 19 and 26 patients, almost the same distribution as before (ECMO: 0%, 79% and 2% respectively, $p < 0.001$).

Cluster 1 was the cluster with the highest proportion of female patients (56%), the one with the lowest weight for age and the one with the lowest mean arterial pressure and diastolic pressure. The patients in this cluster had the lowest carboxyhemoglobin, the highest calcium ion, the lowest creatinine and the highest number of neutrophils. The cluster with the highest proportion of intracranial catheters and thermal blanket and the lowest proportion of patients with hemodiafiltration.

Cluster 2 was characterized as the one with lower mean age, heart and respiratory rate, systolic blood pressure, oxygen saturation, venous oxygen saturation, temperature, and higher inspired oxygen fraction. It was the group of patients with higher daily diuresis, wider capillary glycemia range, lower calcium ion levels, higher phosphate levels, lower C-reactive protein levels, higher lymphocytosis and lower neutrophilia. It was the cluster with the highest proportion of patients with ECMO, mechanical ventilation and hemodiafiltration.

Cluster 3 was the one with older patients and higher weight for age. Highest heart rate and respiratory rate, highest blood pressure, lowest daily diuresis, highest oxygen saturation and lowest inspired oxygen fraction, highest temperature, venous oxygen saturation and carboxyhemoglobin. This was the group with higher capillary glycemia, creatinine, lower phosphate levels, higher C-reactive protein, neutropenia and lymphopenia. In addition, this group had the lowest proportion of patients on mechanical ventilation, ECMO and heat blanket.

Clusters 1 and 3 were the most similar, however, cluster 3 presented higher age ($p < 0.001$), higher weight for age ($p < 0.006$) and higher proportion of males ($p < 0.04$). It presented higher levels in the average, maximum and minimum values of all tensions ($p < 0.001$). Finally, cluster 3 presented lower lymphocytosis and neutrophilia ($p < 0.001$) and higher lymphopenia and neutropenia ($p < 0.001$), in addition to a higher C-reactive protein ($p < 0.001$).

Analysis of variables not used for clustering

Each variable was compared among the three clusters (Table 1). Exhaled CO_2 pressure showed important differences, with maximum levels in cluster 1, followed by cluster 2; cluster 3 presented much lower levels than the other two. Troponin also presented significant differences, with the highest level in cluster 2, followed by cluster 1 and with minimum levels in cluster 3.

Relationship between clustering and the classic classification (Table 1 and Fig. 2)

Cardiogenic shock presented different proportions between groups 1 and 2 ($p < 0.001$) and 2 and 3 ($p < 0.001$), however, groups 1 and 3 did not present differences. Inflammatory shock presented significant

differences between all combinations (1 vs. 2, $p < 0.03$; 2 vs. 3, $p < 0.001$; 3 vs. 4, $p < 0.04$). Hypovolemic shock was not specifically associated with any group. Cluster 2 had the highest number of postoperative cardiac surgery patients (44%, $p < 0.001$), while cluster 3 had the highest number of oncologic patients (31%, $p = 0.003$).

Analysis of outcomes according to clustering

Both length of stay and death were chosen as outcomes. In the case of length of stay (Fig. 3), significant differences were found between the three groups ($p < 0.03$), mainly because of the differences between groups 2 and 3 ($p < 0.02$). Groups 1 and 2 did not present significant differences ($p < 0.11$), and neither did groups 1 and 3 ($p < 0.28$). Using Log-Rank differences were also found between the three groups ($p = 0.01$), mainly due to groups 2 and 3 ($p < 0.02$), between groups 1 and 2 no differences were found ($p = 0.18$), neither between groups 1 and 3 ($p = 0.8$). Significant differences between clusters were objectified in survival ($p < 0.04$), also mainly due to differences with cluster 2. Between cluster 1 and 2 there were significant differences ($p < 0.02$) and the same was true between cluster 2 and 3 ($p < 0.03$). However, between groups 1 and 3 there were no differences ($p = 1$).

Prediction of outcomes by classical classification (Table 2)

Table 2
Association between each type of shock and both length of stay and death.

Types of shock	**No cardiogenic**	**Cardiogenic**	**Wilcox/χ^2 (p value)**	**Log-Rank (p value)**
Median length of stay (days)	5 (3 : 7)	9 (4 : 24)	0.02	0.01
Exitus	0.08 (0.01 : 0.14)	0.21 (0.03 : 0.38)	0.16	-
	No inflammatory	**Inflammatory**	**Wilcox/χ^2 (p value)**	**Log-Rank (p value)**
Median length of stay (days)	7.5 (5 : 13)	3 (2 : 6)	0.01	0.05
Exitus	0.14 (0.03 : 0.25)	0.08 (0 : 0.16)	0.58	-
	No hypovolemic	**Hypovolemic**	**Wilcox/χ^2 (p value)**	**Log-Rank (p value)**
Median length of stay (days)	5 (3 : 7)	7 (5 : 14)	0.17	1
Exitus	0.11 (0.04 : 0.18)	0.13 (0 : 0.33)	1	-

Patients with cardiogenic shock had significantly longer length of stay than patients without cardiogenic shock (Wilcox p-value < 0.02 Log-Rank p-value = 0.01). Patients with inflammatory shock had shorter length of stay than patients without it (Wilcox p-value < 0.02, Log-Rank p-value = 0.05). As for patients with hypovolemic shock, there was no significant association with length of stay.

Cardiogenic and inflammatory shock presented inverse association between them, only one patient presented both types of shock, and none presented neither of them (p < 0.001). Given this circumstance, and the fact that it makes no sense for either type of shock to shorten the length of stay, it is most likely that the true determinant of length of stay is cardiogenic shock.

As for the probability of death, none of the three classic types of shock was associated with the survival rate (Table 2).

Cardiac postoperative patients had significantly longer length of stay than non-postoperative patients (9 vs. 5 days, Wilcox p-value < 0.03, Log-Rank p-value = 0.01), while oncologic patients did not show significant differences in length of stay with respect to non-oncologic patients (3 vs. 6 days, Wilcox p-value = 0.15, Log-Rank p-value = 0.1). Regarding mortality, no significant differences were found between groups in any case (cardiac surgery: 8% vs. non-cardiac surgery: 12%, p = 1; oncologic: 0% vs. non-oncologic: 13%, p = 0.45).

Discussion

It is important to consider that the clustering classification has been done with data of a specific hospital so that it is not generalizable. Post-operative care for congenital heart disease, the absence of cardiac transplantation or the presence of a pediatric oncology unit are some of the particularities that determine the type of patients admitted with shock in this hospital, and therefore the classification made. Therefore, the proposal in this article is not the generalization of the model presented, but the demonstration that clustering algorithms (specifically k-means) can be useful for classifying shocked pediatric patients, and that they can even be more accurate than clinical classification.

Data available were mainly from two groups according to the classical classification, patients with cardiogenic shock and patients with inflammatory shock. However, clustering classification was made in three groups, and it seems optimal in its two-dimensional representation (Figures 1 and 2).

The analysis of the different classificatory variables showed that some were more relevant than others. Specifically, ECMO was the variable with the greatest discriminatory weight, and therefore a new analysis was performed after eliminating this variable. When this variable was eliminated, the groups barely changed, which ruled out the possibility that the good clustering results were due to this variable, and demonstrated the classificatory quality of the algorithm that could discriminate ECMO patients.

Previous studies have shown that clustering algorithms tend to group intensive care patients into physiologically similar groups [15]. Cluster 1 appeared to belong to a group of low weight, hypotensive

infants and young children, mainly with septic shock and some cardiogenic shock. Cluster 2 seemed to correspond to the most severe patients, infants with cardiogenic shock, in large proportion postoperative cardiac patients (44%, $p < 0.001$). Cluster 3 seemed to correspond to the least severe patients, neutropenic children, with apparent distributive septic shock, in large proportion oncologic patients (39%, $p < 0.004$).

The two main classical types of shock were associated with different clusters, cluster 2 had the highest proportion of cardiogenic shock (72%, $p < 0.001$) and inflammatory shock was mainly associated with clusters 1 (52%) and 3 (81%), but with significant differences between them ($p < 0.04$) and with differences between them and cluster 2 ($p < 0.03$). Considering that the clustering classification was performed only using data from the first 24 hours and the classical classification was performed based on the diagnosis at discharge, these results are very promising and they could suggest that clustering is able to discriminate patients earlier than the clinical eye.

As for the assessment of whether this new classification was able to predict outcomes, the length of stay was different among the three clusters (11 vs. 4 days, Log-Rank $p < 0.02$). Graphically (Figure 3), the Kapplan-Meyer curves were different for each cluster of patients, therefore, the absence of differences between groups 1-2 and 2-3 could be due to the small amount of data. The prediction of mortality was also satisfactory, since groups 1 and 2 (7% vs. 33%, $p < 0.02$) and 2 and 3 (33% vs. 4%, $p < 0.03$) presented significant differences, groups 1 and 3, both having such a small proportion of deceased patients, did not present differences. These results are consistent with those of previous analyses, which also suggest that clustering can provide prognostic information [16].

The final objective of this study was to determine whether the use of clustering algorithms could be better than the use of clinical classifications. To do so, it was necessary to compare the outcome prediction ability of the two types of classifications, but is important to bear in mind that the clustering classification was performed with data from the first 24 hours, moreover clustering classification has fewer patients per group than the classical classification, so differences between groups were less likely to be significant.

Cardiogenic shock was associated with longer length of stay (9 vs. 5 days, Log-Rank $p = 0.01$) with the same significance as cluster classification (6 vs. 11 vs. 4 days, Log-Rank $p = 0.01$); therefore, in this scenario clustering did not improve the prediction of length of stay. The association with mortality by classical categorization was not significant in any case (cardiogenic shock 21% vs. non-cardiogenic shock 8%, $p = 0.16$); however, the association with mortality using clustering classification was significant (7% vs. 33% vs. 4%, $p = 0.003$); therefore, classification by clustering improved the prediction of mortality.

It seems that, even with such a small sample, the k-means clustering method is able to classify patients with higher outcome correlation than the clinical method. Other studies have also presented positive results in the application of this type of clustering in other pathologies [16]. However, it is necessary to consider some limitations. When data is collected digitally sometimes presents incorrect recordings, which has to be cautious interpreted [16-18], but this limitation is also present in clinical classification.

Moreover k-means algorithm requires pre-specifying a number of clusters, which means it is not absolutely automatic, it also needs the imputation of missing values [16] and it is very sensitive to outliers. Despite these issues, the present results are still encouraging, and they support the utility of these methods in the classification of patients with shock in pediatric intensive care, something that could help to implement guided therapies [16].

Conclusion

In conclusion, the present study demonstrates the capacity of k-means algorithm to correctly classify pediatric patients with shock and shows a promising future for the use of unsupervised machine learning techniques in pediatric critical care. However, further studies are needed to validate this method on a larger scale.

Declarations

Conflict of Interest Disclosures (includes financial disclosures): The authors disclosed that they do not have any potential conflicts of interest.

Funding/Support: Partially supported by Vicomtech Foundation.

Authors contribution

Dr Rollán-Martínez-Herrera designed the study, collected the data, processed and analysed the data, drafted the initial manuscript, and reviewed and revised the manuscript.
Keretxeta-Sarregi processed and analysed the data and reviewed and revised the manuscript.

Macía-Oliver conceptualized the study, coordinated it and reviewed the manuscript.

Pilar-Orive and Gil-Anton reviewed the design of the study, coordinated the data collection and reviewed the manuscript.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

References

1. Nichols DG, Shaffner DH, Argent AC, et al. Rogers' Textbook of Pediatric Intensive Care. Wolters Kluwer; 2016.
2. Kliegman R, Marcdante KJ. Nelson Essentials of Pediatrics. Elsevier; 2019.
3. Standl T, Annecke T, Cascorbi I, et al. The Nomenclature, Definition and Distinction of Types of Shock. Dtsch Arztebl Int. 2018;115(45):757–68.

4. Kislitsina ON, Rich JD, Wilcox JE, et al. Shock - Classification and Pathophysiological Principles of Therapeutics. Curr Cardiol Rev. 2019;15(2):102–13.
5. Peters MJ, Shipley R. Clinical Classification of Cold and Warm Shock: Is There a Signal in the Noise? Pediatr Crit Care Med. 2020;21(12):1085–7.
6. Tibby SM, Hatherill M, Marsh MJ, et al. Clinicians' abilities to estimate cardiac index in ventilated children and infants. Arch Dis Child. 1997;77(6):516–8.
7. Benke K, Benke G. Artificial Intelligence and Big Data in Public Health. Int J Environ Res Public Health. 2018;15(12):2796.
8. Growth charts -. data table of infant weight-for-age charts. Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/growthcharts/html_charts/wtageinf.htm. Published August 23, 2001. Accessed August 15, 2021.
9. Growth charts -. data table of weight-for-age charts. Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/growthcharts/html_charts/wtage.htm. Published August 23, 2001. Accessed August 15, 2021.
10. Flynn JT, Kaelber DC, Baker-Smith CM, et al. Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents. Pediatrics. 2017;140(3):e20171904.
11. Fleming S, Thompson M, Stevens R, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. Lancet. 2011;377:1011.
12. Likas A, Vlassis N, Verbeek J. J. The global K-means clustering algorithm. Pattern Recogn. 2003;36(2):451–61.
13. Dokmanic I, Parhizkar R, Ranieri J, et al. Euclidean distance matrices: Essential theory, algorithms, and applications. IEEE Signal Process Mag. 2015;32(6):12–30.
14. Mohamad IB, Usman D. Standardization and its effects on K-means clustering algorithm. Res J Appl Sci Eng Technol. 2013;6(17):3299–303.
15. Marlin BM, Kale DC, Khemani RG, et al: Unsupervised pattern discovery in electronic health care data using probabilistic clustering models. Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. New York, NY, ACM, 2012, pp 389–398.
16. Williams JB, Ghosh D, Wetzel RC. Applying Machine Learning to Pediatric Critical Care Data. Pediatr Crit Care Med. 2018 Jul;19(7):599–608.
17. Hug CW, Clifford GD, Reisner AT. Clinician blood pressure documentation of stable intensive care patients: An intelligent archiving agent has a higher association with future hypotension. Crit Care Med. 2011;39:1006–14.
18. Wetzel RC. First get the data – Then do the science. Pediatr Crit Care Med. 2018;19:382–3.

Figures

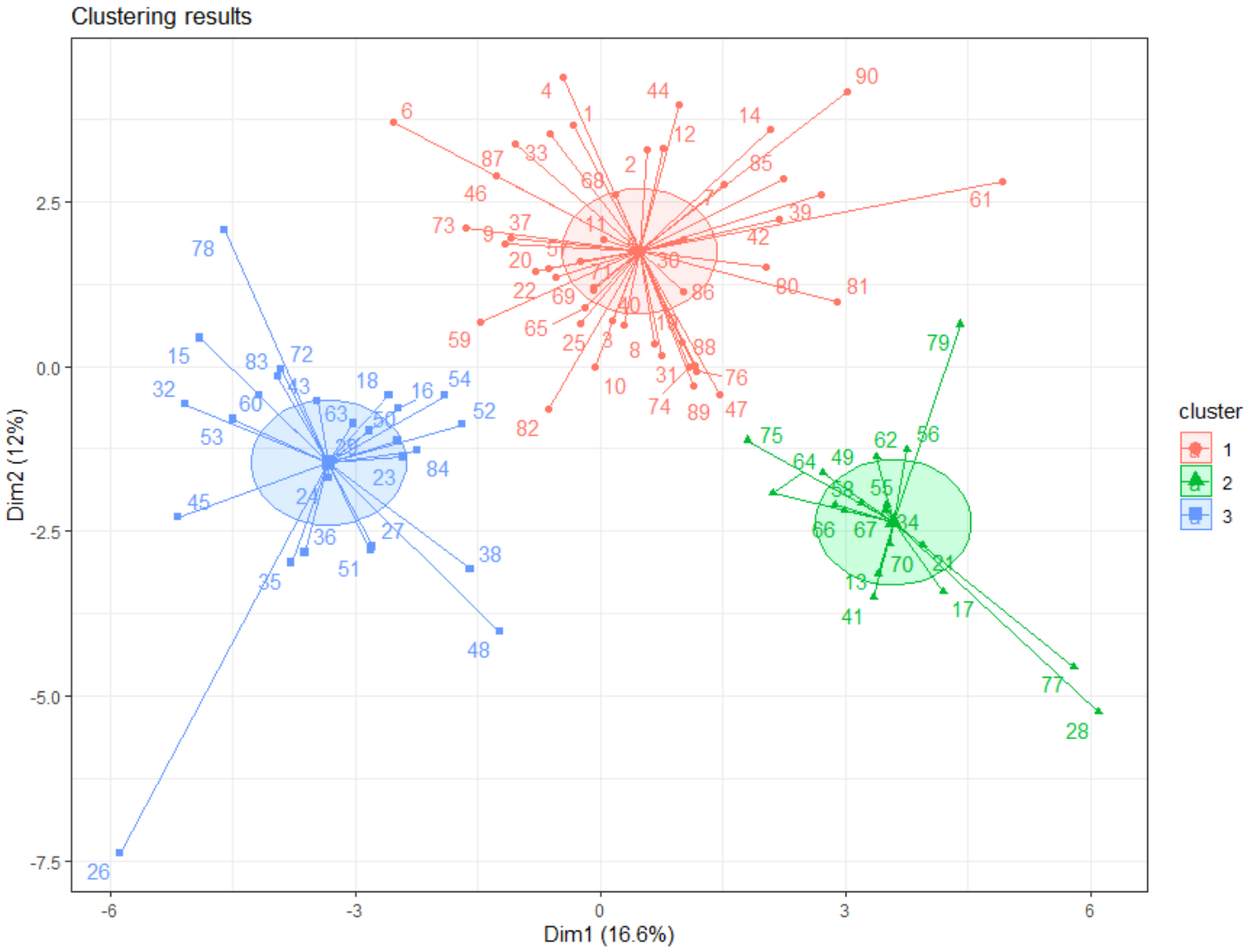

Figure 1

Two-dimensional representation of the clusters based on Principal Component Analyses. Made with R

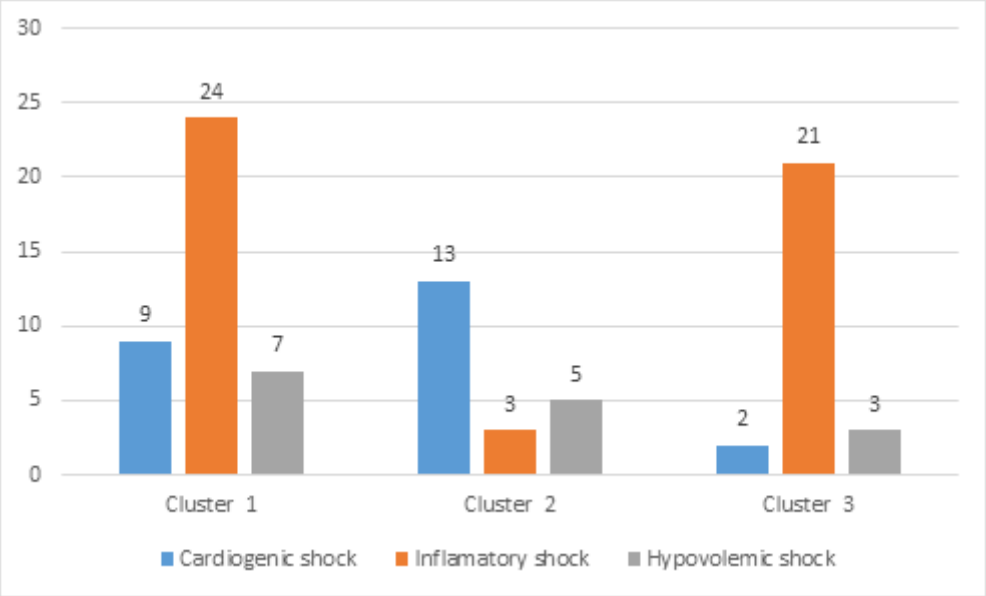

Figure 2

Number of patients with each type of shock in every cluster. Made with Microsoft Excel

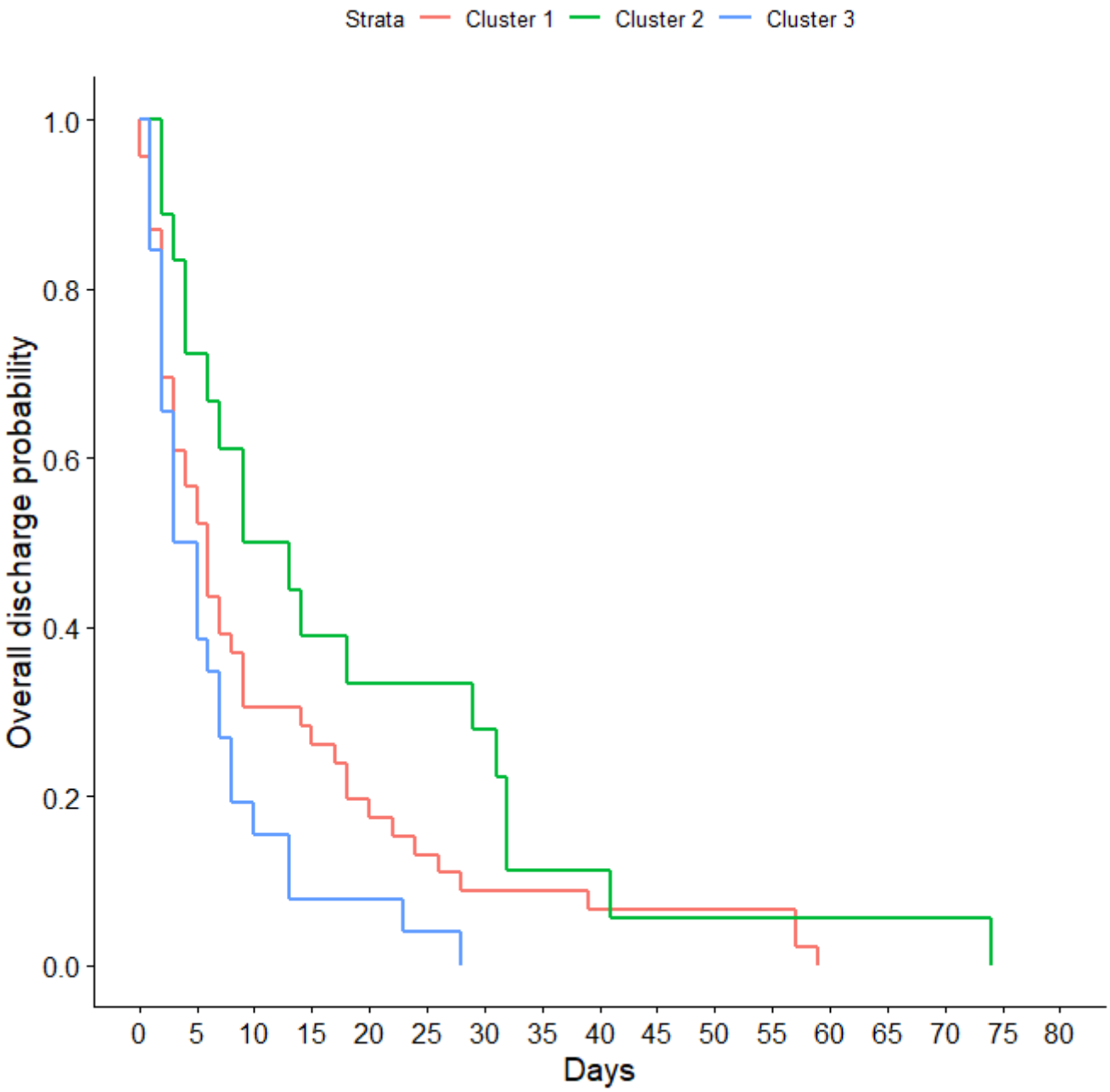

Figure 3

Kaplan-Maier curve for the length of stay in the intensive care unit according to clusters. Made with R

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigure.tif