



Identifying subpopulations of septic patients: A temporal data-driven approach



Anis Sharafoddini^a, Joel A. Dubin^{b,a}, Joon Lee^{c,d,*}

^a School of Public Health and Health Systems, University of Waterloo, 200 University Ave. West, Waterloo, ON, N2L 3G1, Canada

^b Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. West, Waterloo, ON, N2L 3G1, Canada

^c Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, 3280 Hospital Dr NW, Calgary, AB, T2N 4Z6, Canada

^d Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, 3280 Hospital Dr NW, Calgary, AB, T2N 4Z6, Canada

ARTICLE INFO

Keywords:

Hierarchical
DBSCAN
Clustering
Functional data analysis
Longitudinal data
Sepsis
t-SNE

ABSTRACT

Sepsis is one of the deadliest diseases in North America and in spite of the vast amount of research on this topic there is still uncertainty in the outcome of sepsis treatments. This study aimed at investigating the informativeness of temporal electronic health records (EHR) in stratifying septic patients and identifying subpopulations of septic patients with similar trajectories and clinical needs. We performed hierarchical clustering and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) analyses using data from septic patients in the MIMIC III intensive care unit database. The t-Distributed Stochastic Neighbor Embedding (t-SNE) method was utilized to map patients to a two-dimensional space. We utilized silhouette index and cluster-wise stability assessment by resampling to investigate the validity of the clusters. The hierarchical clustering with Euclidean metric identified twelve clinically recognizable subgroups that demonstrated different characteristics in spite of sharing common conditions. Our results demonstrated that data-driven approaches can help in customizing care platforms for septic patients by identifying similar clinically relevant groups.

1. Introduction

In intensive care units (ICUs), it is important to monitor patient health status over time. This necessity results in multiple measurements of a particular clinical variable across a given patient's stay. Due to the time-varying nature of these measurements, they are usually referred to as temporal/longitudinal electronic health records (EHRs). Since temporal EHR data have valuable information about the evolution of patient health status, researchers have been able to improve predictive modeling [1–3] and patient stratification [4] through using these additional data. However, analyzing longitudinal data is accompanied with multiple challenges such as irregular sampling rates and varying lengths of available measurements, as well as the inherent correlation of repeated measurements of a variable over time for the same patient.

The goal of this research is two-fold. First, through an exploratory analysis, we investigated the informativeness of the temporal vital signs data in our septic patient stratification via functional data analysis [5]. Second, we added other clinical information about the patients to the temporal vital signs data and performed cluster analysis to identify

subpopulations of septic patients with similar clinical needs and trajectories in our data. An exploratory analysis of the included variables was used to interpret the identified clusters and derive insights. This information may be used to design customized care platforms for septic patients who share similar needs.

2. Literature review

2.1. Longitudinal data analysis in ICU patients

ICUs have a heterogeneous population with different health status dynamics but similar needs for constant care [6,7]. The heterogeneity in ICUs adds to the importance of finding similar patients and detecting the underlying phenotypic groups. Recently, efforts have been made to employ the temporal information of heterogeneous EHR data to reveal subpopulations.

A large and growing body of literature has investigated vital sign trajectories to discover patient subpopulations in order to identify the underlying pathophysiology of diseases and suggest customized care

* Corresponding author. Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, 3280 Hospital Dr NW, Calgary, AB, T2N 4Z6, Canada.

E-mail addresses: asharafoddini@uwaterloo.ca (A. Sharafoddini), jdubin@uwaterloo.ca (J.A. Dubin), joonwu.lee@ucalgary.ca (J. Lee).

pathways. Pimentel et al. [8] proposed a probabilistic framework for analysis of time series data. They employed an unsupervised Gaussian process model to investigate the evolution of vital signs in post-operative patients. Their study revealed four dominant underlying physiological behaviours in their population and identified the abnormal trajectories.

Lehman et al. [9] investigated the discriminative bivariate dynamics of heart rate and blood pressure in ICU patients, using a switching vector autoregressive process approach. Their study demonstrated that the temporal evolution of the vital signs has additional predictive value for sepsis detection beyond non-temporal approaches. Moreover, their study revealed ten prevalent underlying physiological modes for patient health status, each of which was correlated with different sepsis risk levels. Building on this study, they suggested that the observed patterns are due to a patient's health status, undergoing clinical intervention, and possible measurement artifacts. Therefore, similarity can be investigated based on the underlying dynamics of vital signs and help healthcare providers manage short- and long-term outcomes with appropriate interventions [10]. Lehman et al. [11] extended upon their previous work [9] and compared the changes in vital signs dynamics before and after applying a vasopressor treatment. The study demonstrated distinguishable differences in the dynamics among survivors and non-survivors.

Agarwal et al. [12] used a functional clustering model to find subpopulations in a cohort of patients with chronic kidney disease, based on creatinine measurement trajectories. They found two subpopulations, each with a dominant creatinine trajectory. Exploratory analysis of the clusters revealed various discriminative factors between the clusters, including the presence of comorbidities and adherence to medication.

Considering all this evidence, it seems that longitudinal EHRs are a valuable resource for finding patient subpopulations in order to customize care delivery.

2.2. Sepsis

The Third International Consensus Definitions for Sepsis and Sepsis Shock (Sepsis-3) defined sepsis as life threatening organ dysfunction caused by a dysregulated host response to infection, which costs Canada and the United States more than \$300 million and \$20 billion annually, respectively [13]. A number of interventions are used for treating sepsis. These interventions include: identifying the most appropriate type of antibiotics treatment regimen and administering it promptly to increase survival chance; determining the type, volume and timing of administration of intravenous fluids; and initiation of inotropes or vasopressors. Understanding the various interventions for sepsis, and the administration and outcomes, has been a trending topic [14,15]. However, there is still considerable heterogeneity in the outcomes of sepsis treatments, a phenomenon known as treatment effect heterogeneity [16]. Even after many attempts to explain this heterogeneity, there is no consensus for much of the variability in the outcome of a particular treatment [17]. Many researchers have focused on investigating sepsis-related research questions by using EHR data. Johnson et al. [18], in their comprehensive study, demonstrated that even in one hospital, there are various groups of patients with a diagnosis of sepsis who have highly variable outcomes.

Salgado et al. [19] used fuzzy ensemble models to predict vasopressor dependence. They first found subpopulations in the dataset using an unsupervised clustering method and then trained a fuzzy model on each subpopulation. Researchers have also focused on leveraging temporal data in answering sepsis-related questions. One study utilized the longitudinal measurement of heart rate, mean blood pressure, and respiratory rate for predicting the onset of septic shock with coupled hidden Markov models [20]. They also compared their method to conventional approaches such as support vector machines. According to their results, methods that account for the temporal aspect of data tend to perform better than conventional methods. On the same application, Khoshnevisan et al. [21] demonstrated that using recent temporal patterns with various classification methods consistently outperform atemporal approaches. These results support the idea of leveraging

temporal data when considering septic patients. Fohner et al. [17] employed an unsupervised topic modelling technique (Latent Dirichlet Allocation) on the orders and medications in septic patients EHRs to investigate heterogeneity in sepsis treatment. Their study uncovered 42 topics (treatment patterns) and confirmed the high variability in clinical signatures among these patients. In a recent study, Seymour et al. [22] employed OPTICS (Ordering Points to Identify the Clustering Strategy) and consensus k-means clustering method to identify phenotypes in septic patients cohort. They identified four major phenotypes among Sepsis-3 patients labeled α , β , γ and δ . The α phenotype represented the patients who received the lowest amount of vasopressor. The β phenotype consisted of older patients who had more chronic illness and renal dysfunction. Finally, γ and δ phenotypes had members with 'inflammation and pulmonary dysfunction' and 'liver dysfunction and septic shock', respectively. Whilst their research provided insights into major subpopulations in Sepsis-3 patients, the evolution of patient health status was discarded and their data were limited to the most abnormal value of each included variables (e.g. vital signs, lab results) within the first 6 h of hospital admission.

Building on the previous studies, in this research, functional data analysis and cluster analysis are employed to investigate the presence of similar subpopulations among septic patients, by taking the trajectory of their vital signs into consideration. In this study, we focus on the data from the first 24 h of the onset of clinical concerns for sepsis and account for the temporal aspect of these data. The results may provide a framework for more customized care for each subgroup of septic patients and help to elucidate the underlying reasons for treatment effect heterogeneity.

3. Materials and method

3.1. Study sample

This study utilized a subset of data from patients admitted to the ICUs of the Beth Israel Deaconess Medical Center between 2008 and 2012 (MIMIC III database [23]) provided in [18] and data extraction was done using the code provided by the authors [24]. The data from 2001 to 2007 were excluded to focus on the population of MIMIC which have antibiotic prescription measured. Because the MIMIC database is de-identified and public, patient consent and research ethics approval were waived. From 23,620 ICU admissions that were initially included, three non-adult patients were excluded. 7,536 admissions were excluded to only focus on the first admission of patients with multiple admissions. Patients who were admitted to the cardiothoracic surgical service were also excluded since their postoperative physiologic disorders do not have the same mortality risk as do the other ICU patients (2,298 patients). Moreover, 18 admissions were removed because they had no charted data. Sepsis-3 criteria were used to identify septic patients since it provides an estimate for the start time of clinical concern for sepsis which is pivotal for temporal analysis of data. Out of 5783 septic patients based on Sepsis-3 criteria, 245 patients who had no measurement for their vital signs in the first 24 h after the start time of clinical concern were excluded. Finally, 5539 adult septic patients based on Sepsis-3 criteria were considered in this study (average age, 65.5 [SD, 17.8] years; 3065 [55.3%] male).

3.2. Feature extraction

For each patient, the following sets of predictors from the first 24 h after the start time of clinical concern were extracted:

- Admission and demographic data: ICU service type (coronary care unit, medical, surgical, and trauma/surgical ICUs), admission type (emergency, elective, urgent), gender (female, male), ethnicity and age.

- Minimum and maximum of the following variables: blood urea nitrogen (BUN), hematocrit, creatinine, bicarbonate, lactate, potassium, sodium, glucose, platelets, white blood cells, anion gap, albumin, bilirubin, hemoglobin, chloride, prothrombin time (PT), partial thromboplastin time (PTT), international normalized ratio (INR), bands and Glasgow Coma Scale (GCS).
- Hourly measurement of vital signs: mean blood pressure (MBP), systolic blood pressure (SysBP), heart rate (HR), respiratory rate (RR), oxygen saturation (SPO2) and body temperature (Temp).
- Daily total urine output.
- Interventions: duration and dosage of each of the following vaso-pressor medications: norepinephrine, epinephrine, phenylephrine, vasopressin, dobutamine, and dopamine; the presence of mechanical ventilation and dialysis.
- The International Classification of Diseases 9th Revision (ICD-9) code.

To include the temporal aspect of vital signs, in addition to the simple statistical characteristics (including maximum, minimum, median, mode, mean, standard deviation and number of measurements) that can capture magnitude and variability of variables, functional principal components (FPCs) were used to identify the dominant modes of variation in vital signs. In the last decade, functional principal component analysis (FPCA) has been widely used in the statistics and machine learning community for various applications, including public health and biomedical applications, to reduce the dimensionality of data while preserving the information on variability over time [25]. FPCA was applied for sparsely or densely observed vital signs via the Principal Analysis by Conditional Estimation (PACE) algorithm provided in the R package *fdapace* [26]. In PACE, the expansion for the trajectory $X_i(t)$ for the i th patient when using only the first p eigenfunctions ($\hat{\varphi}_k(t), k = 1, 2, \dots, p$) is

$$\hat{X}_i^p(t) = \hat{\mu}(t) + \sum_{k=1}^p \hat{\xi}_{ik} \hat{\varphi}_k(t) \quad (1)$$

where $\hat{\mu}(t)$ is the estimated mean function, and $\hat{\xi}_{ik}$ represents the FPCs. The first p FPCs were extracted for vital signs such that the components cumulatively explain at least 98% of the total variation in the trajectory of a given vital sign. For laboratory tests, the maximum and minimum measurements during the first 24 h after the start of the clinical concern were used since these values are usually representative of the worst condition and have predictive power [27]. For vaso-pressor administration, the duration and total amount of each variable were extracted for the first 24 h starting at the time of suspected infection. For mechanical ventilation and dialysis, only the presence of the administration (a binary variable) was utilized. Since we treated vital signs as sparse data, no missing data treatment was required for them. However, for laboratory tests, predictive mean matching imputation [28] was employed to address missing data and missingness indicators were used to preserve the missingness pattern. These missing indicators not only have predictive power beyond the observed measurements [29], but also capturing the lab ordering patterns provides insight into physician's opinion about patient health status [30].

3.3. Cluster analysis

3.3.1. Two-dimensional data mapping technique and clustering methods

A common practice when clustering datasets with large numbers of variables is to reduce dimensionality. This study utilized the t-Distributed Stochastic Neighbor Embedding (t-SNE) method [31] to map patients to a two-dimensional space. Since a chosen distance metric has an important role in mapping and clustering, after standardizing the variables, we used two patient similarity metrics (Euclidean distance and cosine similarity) in order to account for the effect of the metric on the derived clusters. After mapping, hierarchical clustering and

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) analyses [32] were used to find subpopulations in the cohort of septic patients. By using two different clustering methods, we would be able to observe different views of data and identify the most robust subpopulations.

3.3.2. Parameters tuning

For tuning the clustering parameters, gap statistic [33] was employed to identify the optimum number of clusters in hierarchical clustering. The parameters of DBSCAN (epsilon [ε] and minimum points [MinPts]) were tuned using the k-distance tuning method. MinPts were set to the dimensionality of data plus one. Then, the distance of each point from its k th ($k = \text{MinPts}$) nearest neighbor was calculated (kdist). After calculating and sorting kdist for all data points, the k-distance graph was plotted and the value for which the graph showed a strong bend the knee point was selected for ε .

3.3.3. Clustering evaluation

The silhouette index [34], which evaluates the suitability of assigning a patient to one group rather than to another was utilized to evaluate the clustering method:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

where $a(i)$ and $b(i)$ represent average dissimilarity of instance i to all other objects in cluster A and B , respectively. The silhouette index is a number between -1 and 1 , where a high value means the patient is strongly matched to its own cluster and weakly matches other clusters. The silhouette index $s(i)$ was calculated for each cluster by averaging the silhouette indices of the patients in that cluster.

In order to evaluate the robustness of the clustering methods, we used the cluster-wise cluster stability assessment by re-sampling [35]. We used the bootstrap method (with replacement) with 100 runs to re-sample data and then Jaccard similarity was employed to assess the similarity of the original clusters to the ones derived from re-sampled data. The mean of these similarity scores was then used as a proxy of the stability of the clusters. A mean similarity score of 0.75 or more represents a very stable solution [36].

In this study, we employed different clustering techniques to assess whether there are clusters in our data regardless of the clustering method being used. However, one can still argue that the clusters found by different methods are not necessarily related. In other words, if there are no inherent clusters in the data, different methods may find different clusters within data based on their algorithms. To address such an argument, we performed an agreement assessment analysis to assess that the clusters found by a given method are very similar to the ones by other methods. Therefore, agreement assessment was included as a step in the clustering pipeline to evaluate consistency in cluster assignments by different methods. While there are techniques available to assess the agreement between specific types of clustering techniques (e.g. dendograms for hierarchical clustering algorithms), comparing the results of a hierarchical method with a non-hierarchical method is challenging. Reilly and colleagues [37] proposed a technique based on Cohen's Kappa statistic to overcome this challenge. This method first creates a contingency table between two clustering methods, then permutes the cluster numbers for one method to maximize Cohen's Kappa statistic and uses the maximum value as a proxy of agreement between two models. However, this method has two shortcomings in our case. First, for a non-square matrix contingency table (i.e. comparing two methods with different numbers of clusters) this technique assumes that the method with a fewer number of clusters actually has as many clusters as the other method but no data points in these clusters (basically adds zero to the matrix). This assumption results in unrealistic values for cases where one clustering method finds subpopulations within a cluster from another technique. Second, this technique is computationally intensive

when we have higher numbers of clusters for each technique. Therefore, we proposed a new technique that overcomes these shortcomings. First, we generated a contingency table for method A and B (in which, element ij indicates how many of the points in cluster i from method A also belongs to cluster j from method B) and then treated the cluster numbers (i.e., i and j) from each method as a nominal variable and calculated the Cramer's V association which gives a value between 0 and 1 (inclusive), where values close to 1 represent high level of agreement.

3.4. Experiments in this study

The experiments in this study were performed in two steps. To determine whether including temporal information of vital signs helps in identifying subpopulations, we first performed cluster analyses solely on vital sign data with and without considering their temporal aspect. Then in the second experiment, along with vital signs, a comprehensive list of variables was added to our analysis to find subpopulations of septic patients. An overview of the experiments is provided in Fig. 1.

3.4.1. Step 1: investigating the informativeness of temporal data over cross-sectional data

In the first step, statistical characteristics (including maximum, minimum, median, mode, mean, standard deviation and number of measurements) and the FPC scores of vital sign measurements were utilized to calculate the pairwise Euclidean distances and cosine similarities. The resulting similarity matrices were used as the input to our mapping algorithm. Then, instead of representing the temporal variation of the vital signs via FPCs, only the average over 24 h for the vital sign measurements was used, in which the variation of the temporal information was lost. The results of these analyses were compared to identify the informativeness of temporal data over cross-sectional data.

3.4.2. Step 2: Finding subpopulations in the septic patient cohort

In the second step, septic patient subpopulations were found using Euclidean and cosine metrics. Then, several exploratory analyses were performed to investigate the clinical interpretation of the clusters and derive insights from them. The patient outcomes (in-hospital, in-ICU, and 30-days mortality rates, and ICU and hospital length of stays [LoSs]), average age, biological sex, the two most-common primary diagnoses categories, and type of admission in each cluster were reported and compared to others to identify distinctive features of each cluster. Moreover, the mean functions for vital signs in each cluster were plotted to examine the variation in vital signs trends in various clusters. The correlation between patient outcomes and various statistical characteristics (minimum, maximum, and average in the first 24 h after clinical concerns) of vital signs and laboratory tests were also investigated to help with clinical interpretation of the clusters and identifying the distinctive features of each cluster. In all correlation analyses, Pearson correlation was used to calculate the correlations. Similar

analyses were performed for the average dosage and duration of different medications. Clusters were also compared in terms of mechanical ventilation and dialysis administration rates. Finally, the average severity of organ dysfunction, level of the inflammatory response of the body, and comorbidity rate in each cluster were quantified using the Sequential Organ Failure Assessment (SOFA) score [38], Quick Sequential Organ Failure Assessment (qSOFA) score [39], Systemic Inflammatory Response Syndrome criteria (SIRS) [40] and Elixhauser score [41], respectively.

4. Results

4.1. FPC score extraction

Applying PACE on vital signs with fraction-of-variance-explained threshold set to 98% resulted in three eigenfunctions for HR and RR, four for MBP, SysBP and SpO₂ and five for Temp. The scores for these eigenfunctions were extracted for each patient to be used as features for the clustering phase.

4.2. Cluster analysis

4.2.1. Step 1: investigating the informativeness of temporal data over cross-sectional data

This section presents the results for hierarchical clustering with Euclidean metric (HE), hierarchical clustering with cosine metric (HC), DBSCAN with Euclidean metric (DE) and DBSCAN with cosine metric (DC). Fig. 2 demonstrates the results for HE and HC in two settings: 1) when only the average of vital signs is used, and 2) after including descriptive statistics and FPC scores. It is worth reiterating here that the objective of this analysis is to explore whether more granular clusters emerge when we transition from atemporal to temporal data where results from atemporal data serve as the baseline. As can be seen, in the first setting (**A** and **C**) regardless of the metric all patients are considered in one cluster and similar to each other. For hierarchical clustering, the tree was cut at the optimum number of clusters achieved by gap statistics. By including temporal information (Fig. 2: **B** and **D**), three and two clusters started emerging with HE and HC, respectively. Although the gap statistic selected a single cluster as optimal in the first setting, we also clustered the average values with the optimum numbers of clusters derived in the second setting as our baseline (i.e., 2 and 3 clusters for HE and HC, respectively). These results are demonstrated in Fig. 3.

For DE and DC (Fig. A.8), two clusters emerged when temporal data were included. It should be noted that DBSCAN does not require specifying the number of clusters; therefore, the optimum number is inferred based on the data after tuning DBSCAN's parameters.

Table 1 shows the evaluation results for baseline and both clustering methods in the second setting. Overall, the average silhouette width for all methods in the second setting were positive and higher than the

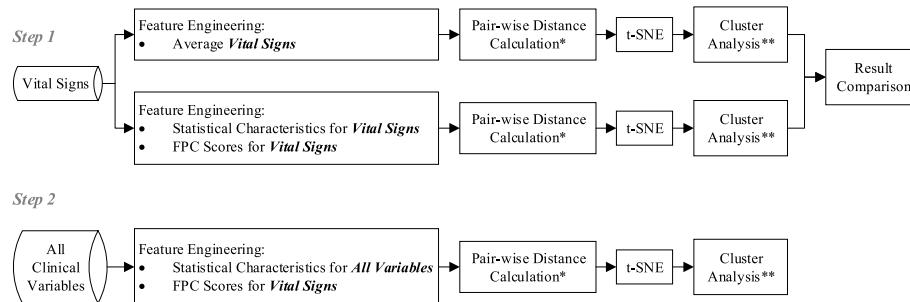


Fig. 1. An overview of experiments.

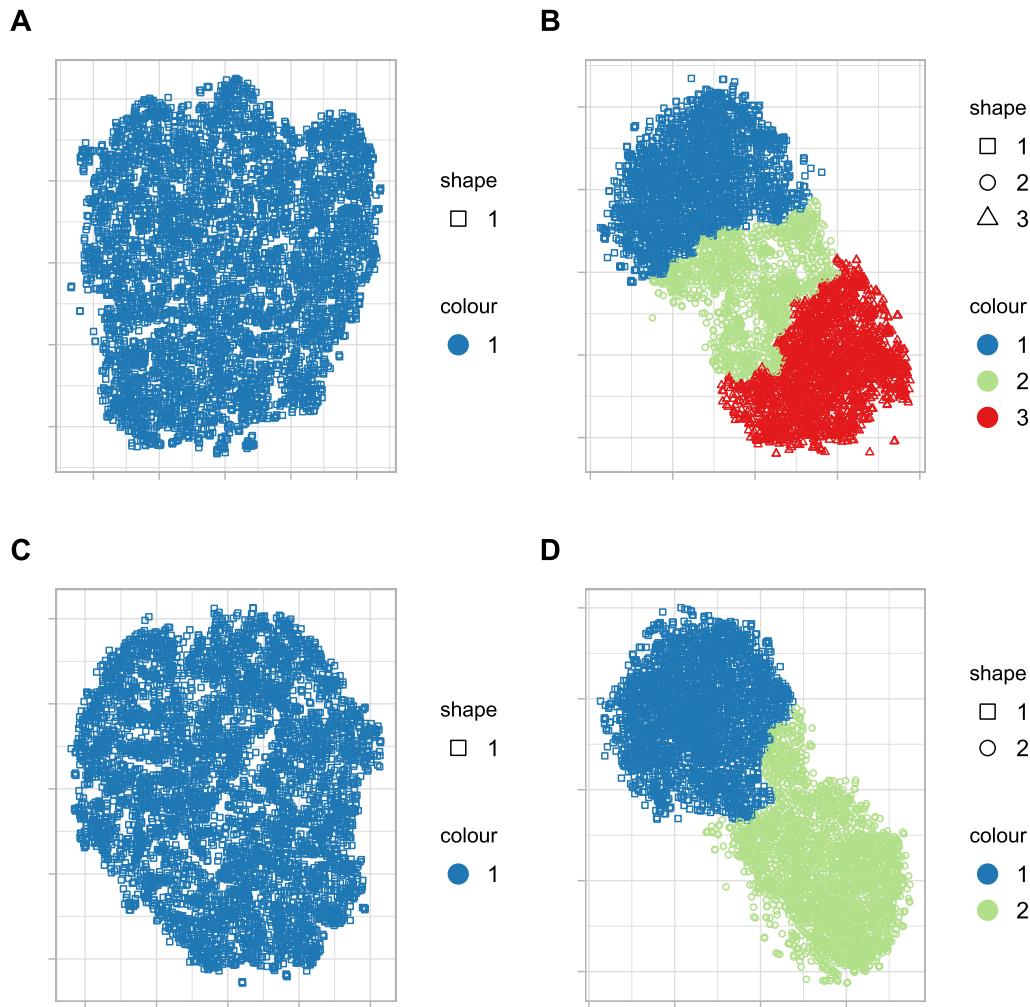


Fig. 2. Hierarchical clustering of vital signs during the first 24 h of the start of the clinical concern, tuned by gap statistics. **A:** Euclidean metric with the average values. **B:** Euclidean metric with statistic characteristics of hourly values along with the FPC scores. **C:** Cosine metric with the average values. **D:** Cosine metric with statistic characteristics of hourly values along with the FPC scores.

baseline which indicates that most of the patients were assigned to the appropriate cluster and including temporal data has improved the clustering results. The mean silhouette value of HC is 0.555 which is higher than the results of DE, DC and HE. Since the baseline was not defined based on the optimum number of clusters derived from gap statistics, it did not receive high score in terms of cluster stability. The stability test indicates moderate (above 0.6) to good (above 0.75) stability for most clusters in the second setting. Overall, HC provides the most stable clusters. These results, while exploratory, suggest that the temporal aspect of vital signs provides additional information that can be helpful in clustering septic patients.

4.2.2. Step 2: Finding subpopulations in the septic patient cohort

HE, HC, DE and DC were implemented on a comprehensive list of information from septic patients in our study sample. Using gap statistics, 12 and 21 clusters were identified as the optimum k for HE (Fig. 4) and HC (Fig. A.9) metrics, respectively.

After tuning the parameters of DBSCAN, regardless of the metric, 15 main clusters were identified in the data (Fig. A.10 and A.11). Cluster 0 represents the outlier samples.

The summary of evaluation results are shown in Table 2. The average silhouette value for all clusters regardless of the method were positive. DBSCAN clustering method achieved a higher average silhouette width

in comparison to hierarchical clustering with almost negligible sensitivity to the metric. In terms of stability, clusters produced by HE were all reproducible and stability values for this method were all above 0.75. Although this observation can imply that there is a high probability that clusters generated by HE represent the true structure in the data, with lack of ground truth for our case, caution must be applied, as high cluster stability may simply be due to the inflexibility of the clustering method.

We also examined the pairwise patient allocation agreement between clustering methods. First, a contingency table was created for the results of each two models and then Cramer's V was used for evaluating the association between the results of two models. As can be seen in Fig. 5, in general, there is a good level of agreement between different methods and the highest level of agreement is between HE and HC.

Having a closer look at the data, we were able to identify clusters in one method that remained almost the same in another method (except for a few patients.) Fig. 6 visualizes the contingency table for HE and HC. As can be seen in Fig. 6, clusters 2, 3, 8, 10, and 11 in HE were the same as clusters 2, 3, 10, 13 and 14 in HC, respectively. Further investigation revealed that some clusters were merged to form one cluster in another method. For instance, in HC, clusters 11 and 18 were merged to form cluster 9 in HE. Similar observations were also made even for HC and DC, for which we had the lowest level of agreement in comparison to other clustering method pairs. These results confirm the presence of

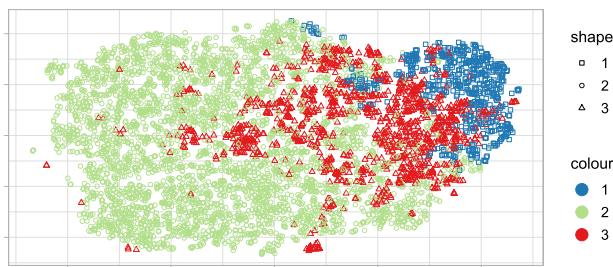
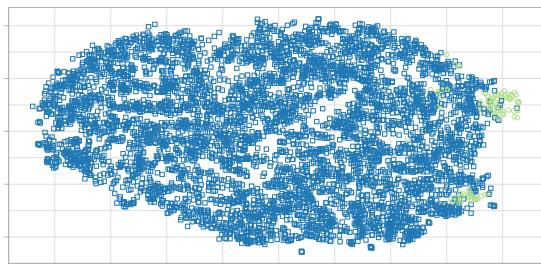
A**B**

Fig. 3. Hierarchical clustering of the average values during the first 24 h of the start of the clinical concern (baseline). **A:** Euclidean metric with the tree cut into 3 clusters. **C:** Cosine metric with the tree cut into 2 clusters.

Table 1
Evaluation results for temporal data informativeness investigation.

Method	Cluster Size	Silhouette Average	Clusterwise Jaccard Bootstrap Mean
Baseline-HE	1: 587	1: 0.07	1: 0.54
	2: 3768	2: 0.04	2: 0.55
	3: 1184	3: 0.18	3: 0.64
Baseline-HC	1: 5423	1: 0.29	1: 0.46
	2: 116	2: 0.21	2: 0.66
HE	1: 2301	1: 0.40	1: 0.79
	2: 1255	2: 0.29	2: 0.72
	3: 1983	3: 0.41	3: 0.36
HC	1: 2818	1: 0.58	1: 0.81
	2: 2721	2: 0.53	2: 0.82
DE	1: 2856	1: 0.35	1: 0.68
	2: 2545	2: 0.46	2: 0.60
DC	1: 2404	1: 0.41	1: 0.40
	2: 2199	2: 0.44	2: 0.41

robust structure within the septic patient cohort.

For HE, HC, DE and DC the minimum size of the clusters was 175, 13, 55 and 15 and the maximum was 869, 496, 1359 and 1307, respectively. Employing cosine metric in both clustering methods resulted in having some very small clusters.

Since HE achieved the highest score in terms of the stability of the clusters, we continue reporting the results for HE and we seek more insights from other methods as needed. In the following subsections, for the sake of brevity we will only highlight the most important results from cluster analysis. We refer the reader to Appendix B for further details (Table B.4 - B.16).

4.2.2.1. Outcomes and demographics. The cluster size, average age, patient outcomes, and top two primary diagnoses are shown in Table 3. In-ICU mortality ([4.28%, 28.03%]), in-hospital mortality ([4.28%, 27.83%]) and 30-day mortality ([4.28%, 31.61%]) rates varied significantly between clusters, implying the heterogeneity in the outcomes of septic patients. Clusters 10, 11 and 8 had the lowest mortality rates and clusters 9 and 2 had the highest rates. Clusters 3, 4 and 5 had very close

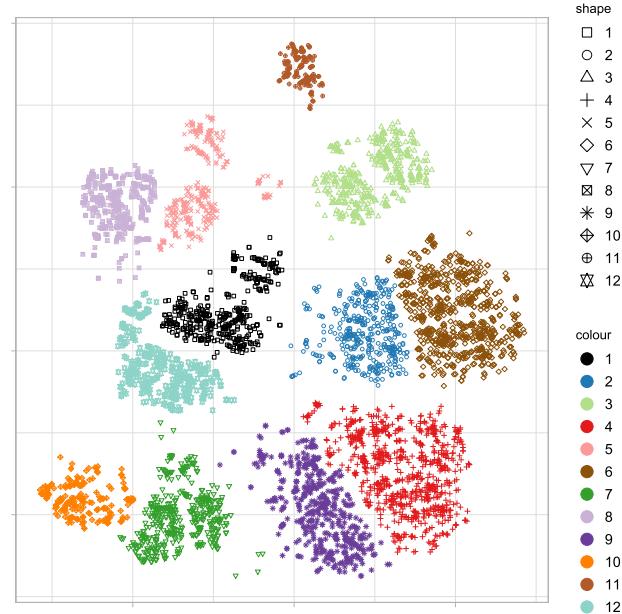


Fig. 4. Hierarchical clustering with Euclidean metric on septic patients data.

Table 2

Summary of evaluation results for septic patients clustering.

Method	Silhouette Average (min, Q1, median, Q3, max) (mean, SD)	Clusterwise Jaccard Bootstrap Mean (min, Q1, median, Q3, max) (mean, SD)
HE	(0.36, 0.40, 0.50, 0.60, 0.83) (0.52, 0.14)	(0.83, 0.88, 0.97, 0.99, 1.00) (0.94, 0.07)
HC	(0.29, 0.39, 0.49, 0.61, 0.81) (0.49, 0.15)	(0.37, 0.79, 0.90, 0.99, 1.00) (0.85, 0.19)
DE	(0.15, 0.34, 0.67, 0.80, 0.88) (0.58, 0.27)	(0.21, 0.72, 0.88, 0.92, 1.00) (0.78, 0.23)
DC	(0.41, 0.45, 0.57, 0.67, 0.83) (0.58, 0.13)	(0.58, 0.94, 0.98, 1.00, 1.00) (0.94, 0.11)

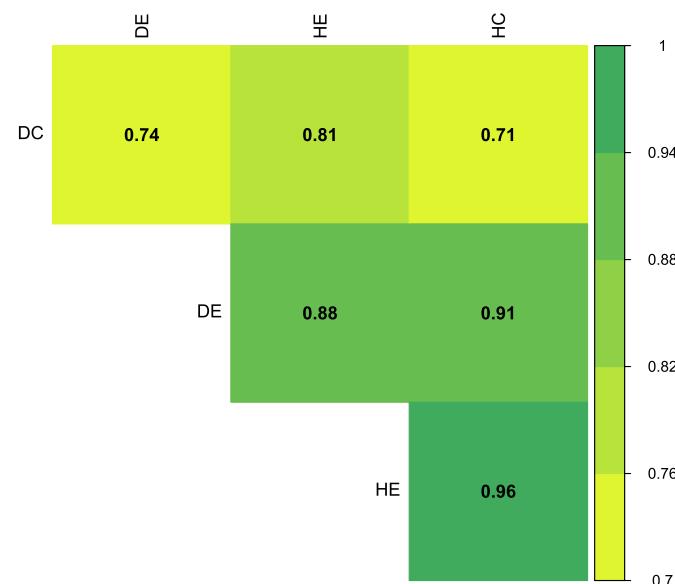


Fig. 5. Pairwise agreement level between different clustering methods.

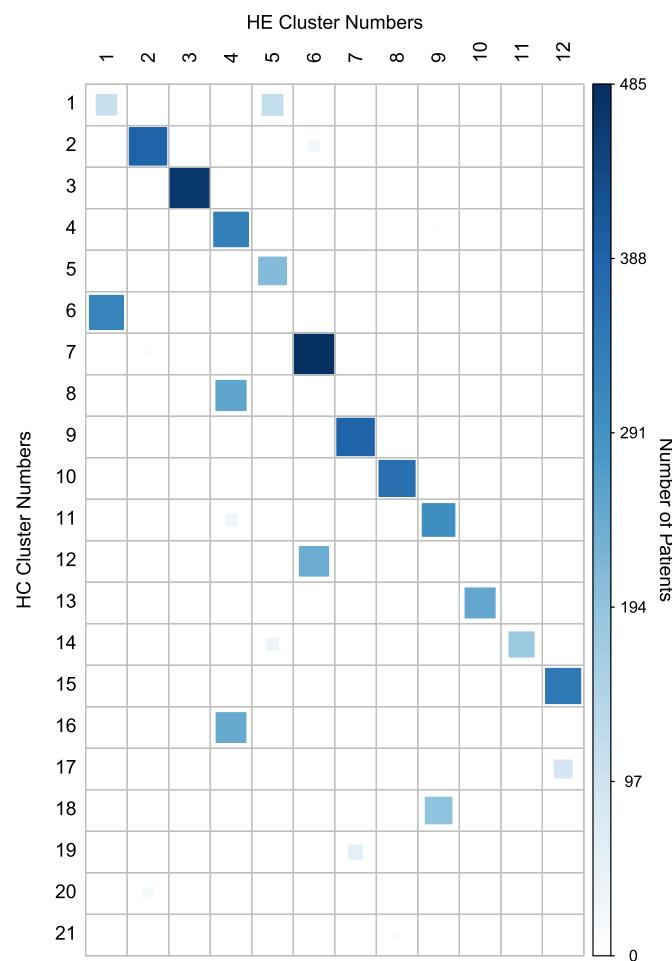


Fig. 6. Visualization of the contingency table between HE and HC.

mortality rates. Cluster 8 had the lowest average age (52.8 years) while cluster 7 had the highest (72.3 years). Cluster 2 and 6 had only male members and patients in clusters 4, 8 and 9 were female. The other clusters had both male and female patients. In terms of admission type, all clusters had members admitted as elective, emergency and urgent while the majority of the patients in all clusters were admitted as emergency. Clusters 2, 3, 4, 6, 9 were dominantly admitted to MICU (>

91%), clusters 1 and 12 to SICU (> 95%), clusters 5 and 8 to TSICU (> 94%), cluster 7 to CCU (> 94%) and cluster 10 to CSRU (> 78%). Cluster 11 had members from all units.

4.2.2.2. Vital signs, laboratory tests and urine output. Fig. 7 illustrates the mean functions for HR, RR and Temp. As can be seen, some clusters have completely different trends in their mean functions for HR, RR and Temp. This observation highlights the importance of including the variation explained by temporal data to capture the evolution of patient health, even short-term dynamics. Clusters 10 and 12 had the highest difference in HR value during the first 24 h. Cluster 10 had the lowest RR value and the largest variation in RR value in 24 h. Clusters 2 and 9 presented a gradual increase in their Temp mean function while clusters 10 and 7 showed considerable fluctuation.

Having a closer look at the statistics of vital signs revealed that clusters 2 and 9 had the highest maximum and average HR, lowest minimum SysBP pressure and highest change in SysBP and RR in the first 24 h after clinical concern. Clusters 10 and 11 had the lowest number of charted measurements for all vital signs, lowest RR and maximum Temp.

The cluster average maximum anion gap cluster was statistically correlated with in-ICU mortality (0.794, p-value = 0.002), in-hospital mortality (0.783, p-value = 0.003) and 30-day mortality (0.863, p-value = 3e-04). This association was not seen for minimum anion gap. The average maximum and minimum albumin was statistically negatively associated with mortality rates among clusters (<-0.73, p-value<0.005.) An interesting observation was the high correlation between glucose variability and mortality rates (>0.81, p-value<0.002). As expected, minimum lactate was associated with mortality rates (>0.65, p-value<0.05). Variation in BUN was also positively associated with mortality rates (>0.62, p-value<0.05). Finally, the minimum WBC was highly correlated with mortality rates (>0.81, p-value<0.001).

4.2.2.3. Medications and interventions. More than 96% of the patients in clusters 2 and 9 and less than 5% of the patients in clusters 4 and 6 received mechanical ventilation. The dialysis rate in all clusters were less than 7%. Clusters 2 and 9 members received the highest average dosage and duration for norepinephrine and vasopressin administration (more than 35% of the members received norepinephrine and vasopressin). Cluster 11 had the lowest average for norepinephrine dosage and duration (less than 2% received it). Dosage and duration of norepinephrine were statistically highly correlated with the mortality rates (0>.87, p-value<0.001). While the administration of epinephrine was very low for many clusters (less than 1.5% received it), it was provided to 10% of the members in cluster 10 patients. Cluster 10

Table 3
Cluster size, age and outcome characteristics.

#	Size	Age (year)	In-icu mortality (%)	In-hospital mortality (%)	30-day mortality (%)	ICU LoS (days)	Hospital LoS (days)	Top two primary diagnoses
1	445	67.68	16.18	15.28	17.75	5.13	10.81	Diseases Of The Circulatory System Injury And Poisoning
2	423	64.53	25.53	25.3	27.9	6.48	10.73	Infectious And Parasitic Diseases Diseases Of The Respiratory System
3	467	66.63	11.13	11.13	15.2	3.25	8.7	Infectious And Parasitic Diseases Diseases Of The Respiratory System
4	869	67.35	11.51	10.93	16.8	3.13	9.3	Infectious And Parasitic Diseases Diseases Of The Digestive System
5	371	63.02	11.32	11.86	13.75	4.71	9.92	Injury And Poisoning Diseases Of The Digestive System
6	756	69.36	12.7	12.7	18.65	3	8.25	Infectious And Parasitic Diseases Diseases Of The Digestive System
7	455	72.26	16.26	16.04	20.44	4.53	8.26	Diseases Of The Circulatory System Diseases Of The Respiratory System
8	380	52.83	10.53	10.53	10.53	6.1	13.62	Injury And Poisoning Diseases Of The Circulatory System
9	503	62.71	28.03	27.83	31.61	7.37	11.81	Infectious And Parasitic Diseases Diseases Of The Respiratory System
10	257	67.7	4.28	4.28	4.28	4.23	10.12	Diseases Of The Circulatory System Injury And Poisoning
11	175	63.61	7.43	7.43	11.43	3	7.58	Neoplasms Diseases Of The Circulatory System
12	438	61.81	13.01	12.79	14.38	6.02	12.82	Diseases Of The Circulatory System Neoplasms

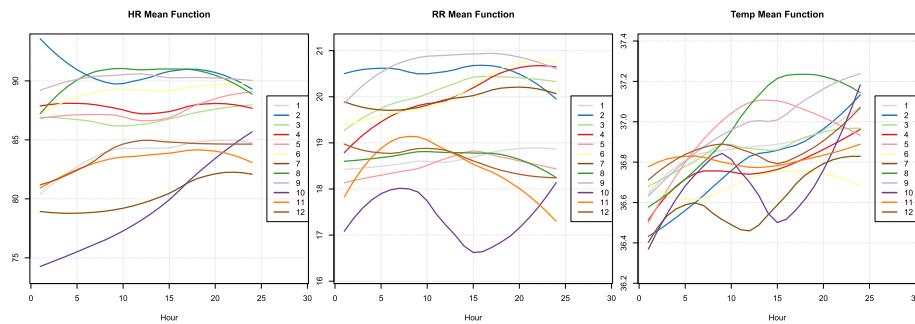


Fig. 7. Mean functions for HR, RR and Temp data in each cluster.

received the highest amount of phenylephrine in comparison to other clusters (more than 66% of the patients received it). Cluster 7 received the highest amount of dopamine (received by more than 18% of members) while in other clusters less than 8% received it. Dobutamine was the least common vasopressor medication in all clusters (no one in cluster 3, 8 and 11 received dobutamine).

4.2.2.4. Severity scores. Clusters 2 and 9 had the highest average SOFA and SIRS at the time of suspected infection while Cluster 11 had the lowest scores. Clusters 4 and 9 had the highest Elixhauser score (5.68 and 5.84) while cluster 10 had the lowest score (0.171). The qSOFA was maximum for clusters 2 and 9 (2.05 and 2.03) and minimum for clusters 11 (1.71).

5. Discussion

This study focused on leveraging temporal vital sign data along with other clinical characteristics in clustering septic patients. Summarizing originally longitudinal vital sign measurements by their average eliminates the temporal dynamics of patient health status changes which can be helpful in phenotyping septic patients. Our results from the exploratory analysis in Step 1 (Section 4.2.1) implied that including information about vital signs trends results in the emergence of subpopulations in our data. Although various techniques have been used to validate the emerged clusters, since there is no ground truth for our problem one can argue that the observed clusters belong to one group. Therefore, these results must be viewed with an exploratory lens.

In Step 2 (section 4.2.2), we performed a clustering analysis using two methods (hierarchical clustering and DBSCAN) with two different metrics (Euclidean and cosine metrics). Our results demonstrated that although the number of subpopulations found by each method was different, the level of agreement between them was high (> 0.7). Changing the similarity metric in DBSCAN clustering did not affect the number of derived clusters, however, the number of clusters in hierarchical clustering was sensitive to the deployed metric. HC found many small clusters, thus, resulting in higher number of clusters in total. Since in other methods, these small clusters were merged together to form larger clusters, a possible explanation for this might be that the cosine metric pays close attention to details when measuring similarities. Although this might be an advantage in personalized prediction applications [27], it results in being sensitive to variation in clustering problems.

In this study, clusters 2 and 9 had the highest mortality rates and the top two primary diagnoses were infectious and parasitic diseases and respiratory system diseases. While cluster 3 had the same primary diagnoses, its mortality rates were half of clusters 2 and 9. The observed increase in mortality rates in clusters 2 and 9 could be attributed to the more severe condition of the patients represented by a higher number of patients receiving mechanical ventilation and vasopressor medications (norepinephrine, phenylephrine, vasopressin). Clusters 11 and 12 consisted of patients with a primary diagnosis of neoplasms or circulatory

system diseases who then became septic. However, cluster 11 had lower mortality rates and shorter LOS in comparison to cluster 12. It seems possible that this difference in outcome is due to the higher amount of norepinephrine, vasopressin and dobutamine administrated in cluster 12 implying the more severe condition for its patients. Clusters 4 and 6 had patients with infectious and parasitic diseases and digestive system diseases. The mortality rates in these two clusters were very close and the level of medical intervention was very similar. Members in clusters 1, 8 and 10 were primarily diagnosed with circulatory system diseases or injury and poisoning, with cluster 1 having the highest mortality rates and cluster 10 having the lowest rates. The data on medical interventions shows cluster 1 received the least amount of vasopressor medications while cluster 10 received the highest amount between these three clusters. The results suggest that septic patients with the same primary diagnosis have different care needs based on the evolution of their health status and therapeutic efficiency. This finding, while preliminary, suggests that temporal aspect of EHR data must be included in the future studies on phenotyping septic patients and treatment effectiveness.

Another important finding was that clusters with a higher maximum anion gap in the first 24 h have higher mortality rates. Although lactate has been widely used for stratifying septic patients in various guidelines, it has not been established as a part of routine workup in some rural or low-volume emergency departments. This limits the ability of clinicians to precisely adhere to the guideline [42,43]. Therefore, in the absence of a lactate test, anion gap can be used as a tool to stratify patients and identify a cohort that needs more aggressive intervention (but not as a tool to diagnose sepsis).

The current study found that the average maximum and average minimum of albumin for each cluster is associated with the cluster mortality rates and can be used for stratifying septic patients. This observation is consistent with the results of other researchers demonstrating the predictive power of albumin trend in mortality prediction for septic patients [44,45].

One interesting finding is that the difference between the average maximum and average minimum value of glucose in the first 24 h of infection onset is positively associated with mortality rates. These results reflect those of Ali et al. [46] who also found that glucose variability was independently associated with septic patients. A possible explanation for these results may be the association of glucose variability with end-organ damages and mortality in critically ill patients [47,48].

In this study, it was observed that the difference between the maximum and minimum value of BUN is correlated with mortality rates. This finding broadly supports the work of Khoury et al. [49] in which BUN variability is a strong predictor of 90-days mortality among patients with heart failure. While the predictive power of BUN at admission for mortality prediction is well established, the exact mechanism behind the association of its variability after admission with mortality is not fully understood.

Although the true clusters in septic patients are unknown due to the

unsupervised nature of these types of studies, in our study, several steps were taken to ensure that the derived clusters are not spurious. First, the Silhouette coefficient and the clusterwise Jaccard bootstrap mean were utilized to investigate the goodness and reproducibility of the clusters, respectively. Then, the consistency between cluster assignments of various clustering methods was quantified. Finally, clinical observations from clusters were cross-checked with the medical literature.

Our study has a few limitations that should be acknowledged. The major limitation of this study is that the data used in this work were from a single, albeit large, medical center in the analysis. Moreover, the scope of this study was limited to temporal data for vital signs and the trends for other variables were not investigated. Another source of weakness in this study, which was mentioned earlier, is the lack of ground truth in our approach. In other words, the nature of the clustering is to explore unknown patterns in the data that may provide insights; however, with respect to the evaluation, in contrast to classification, no true label exists.

Notwithstanding these limitations, this study offers insight into subpopulations within the Sepsis-3 cohort. Although the current study is based on one medical center data, the findings suggest that by identifying groups of septic patients with similar characteristics and need, more customized therapy approaches (e.g. more aggressive care plan for high-risk cohorts) can be considered for septic patients. This research has produced questions in need of further investigation. A further study could assess the informativeness of temporal laboratory tests. Moreover, the identified latent sepsis phenotypes can be used for patient-similarity-based predictive modelling by training a model for each subpopulation

instead of using a one-size-fits-all model.

6. Conclusions

In this paper, we deployed hierarchical clustering and DBSCAN to discover phenotypes in Sepsis-3 patients while including temporal aspect of vital signs. We used t-SNE to construct a two-dimensional mapping of patients based on patient similarities derived by Euclidean and cosine metrics. In the first step, we demonstrated that including temporal characteristics of vital signs independent of clustering method and patient similarity metric helps in stratifying septic patients. In the second step, after including a comprehensive list of variables in our cluster analysis, we identified twelve robust and clinically meaningful clusters with hierarchical clustering and Euclidean distance. The clinical insights from these groups of similar patients pave the way for customizing care platforms for septic patients.

Declaration of competing interest COI

None declared.

Acknowledgement

This study was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants (RGPIN-2014-04743, RGPIN-2020-04382) and an Early Researcher Award from the Ontario Ministry of Research, Innovation and Science (ER15-11-188).

Appendix A

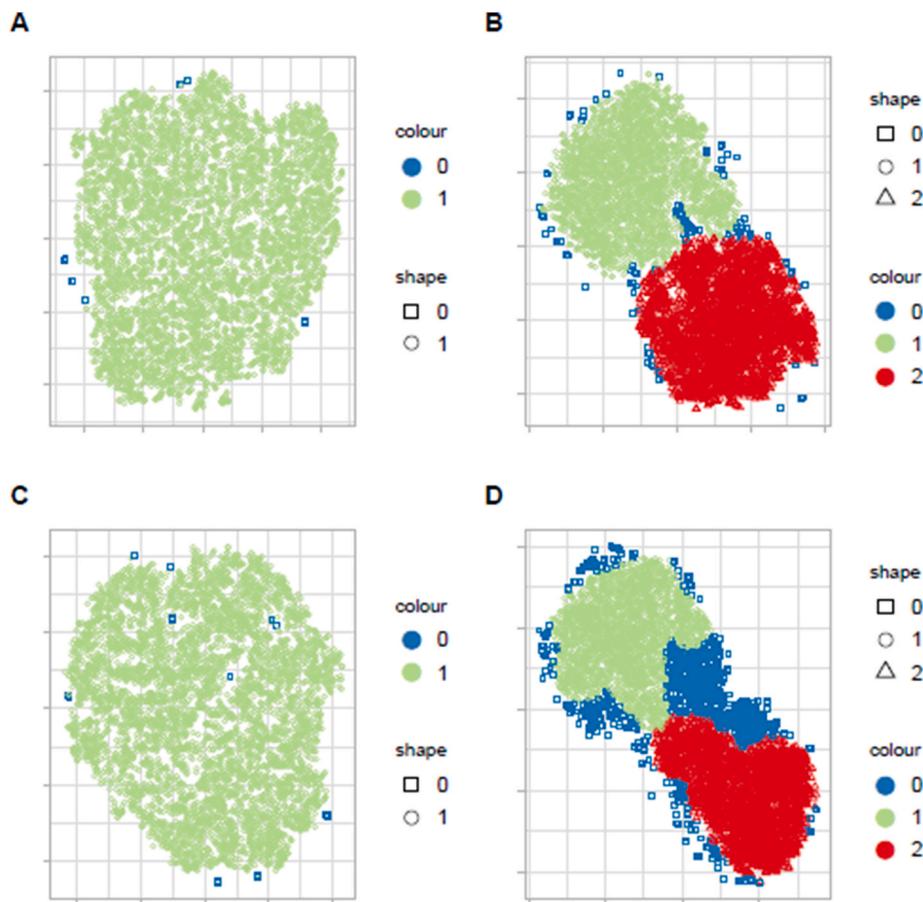


Fig. A.8. DBSCAN clustering of vital signs during the first 24 h of the start of the clinical concern, the k-distance technique (cluster 0 represents outliers.) **A:** Euclidean metric with the average values. **B:** Euclidean metric with statistic characteristics of hourly values along with the FPC scores. **C:** Cosine metric with the average values. **D:** Cosine metric with statistic characteristics of hourly values along with the FPC scores.

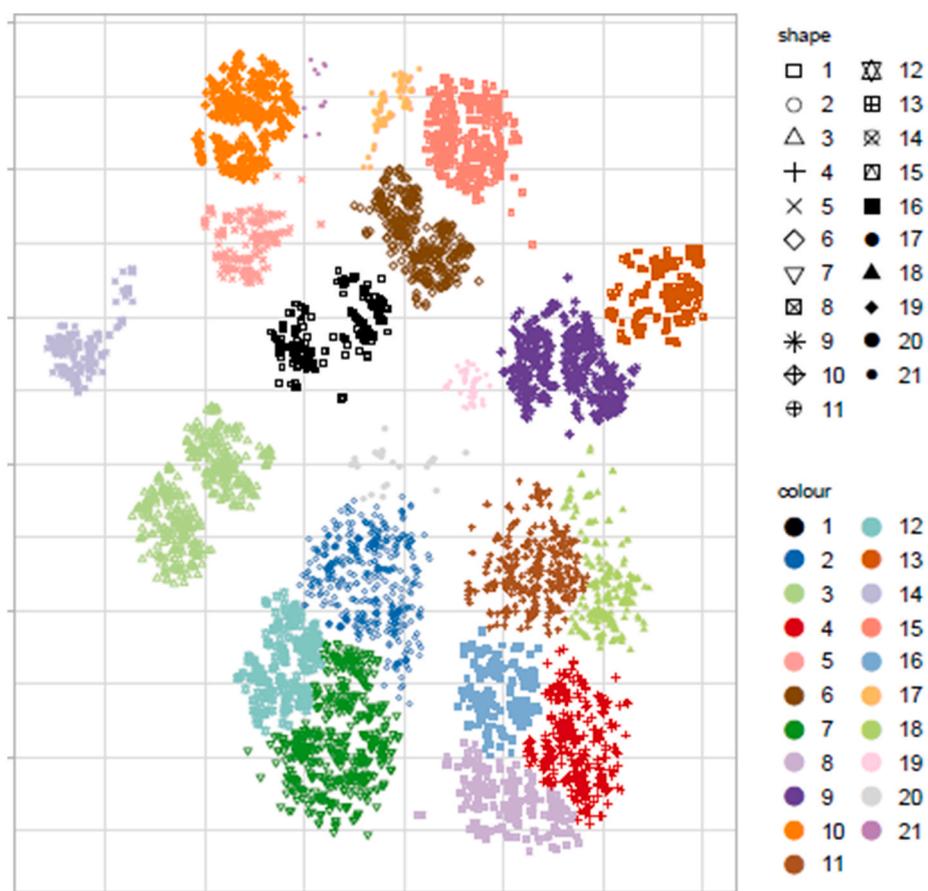


Fig. A.9. Hierarchical clustering with cosine metric on septic patients data.

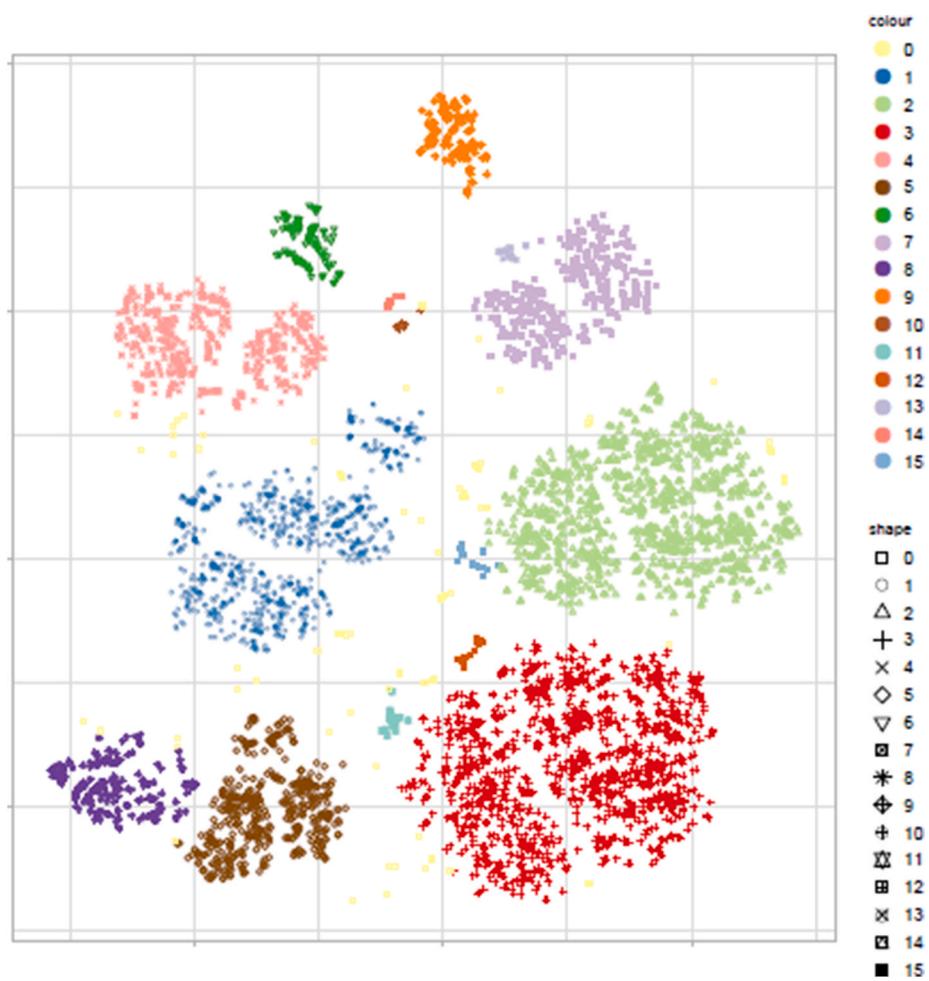


Fig. A.10. DBSCAN clustering with Euclidean metric on septic patients data.

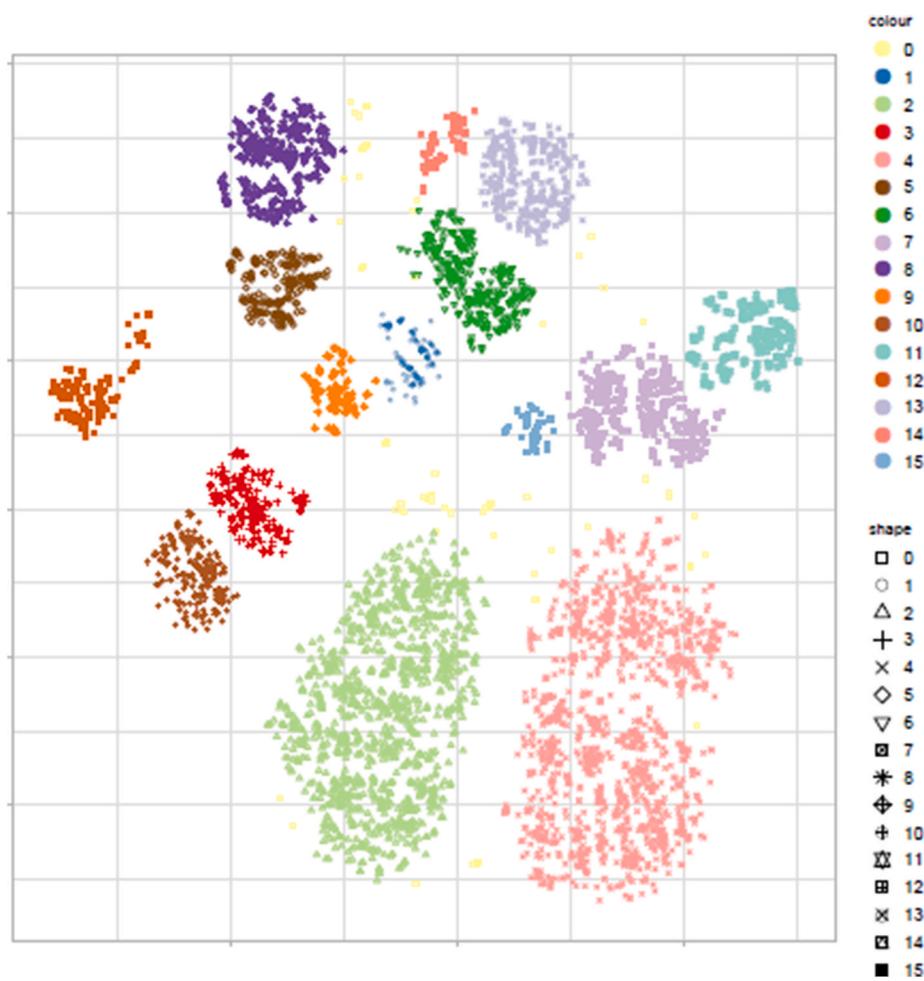


Fig. A.11. DBSCAN clustering with cosine metric on septic patients data.

Appendix B

Table B.4

Summary of maximum values for vital signs in each cluster [mean(SD)].

#">#	Max HR	Max SysBP	Max MBP	Max RR	Max Temp	Max SPO2
1	99.37(18.26)	146.34(21.98)	98.3(18.37)	24.24(5.43)	37.41(0.82)	99.32(1.59)
2	106.49(22.27)	140.99(21.96)	97.98(24.15)	26.51(6.83)	37.38(0.99)	99.34(1.38)
3	100.49(19.86)	138.28(23.49)	94.72(21.27)	26.06(6.54)	37.37(0.91)	98.8(1.6)
4	100.39(19.48)	134.27(21.74)	92.03(19.16)	26.01(6.19)	37.31(0.9)	98.69(1.62)
5	101.13(19.4)	140.47(20.04)	96.15(18.13)	23.54(5.76)	37.44(0.87)	99.32(1.34)
6	102.16(19.88)	134.17(23.98)	90.51(20.18)	26.07(6.18)	37.15(0.84)	98.63(1.92)
7	94.97(21.62)	136.01(23.22)	94.91(25.5)	25.67(5.61)	37.08(1.05)	98.98(1.56)
8	104.3(21.28)	143.45(19.99)	97.32(15.28)	23.41(5.61)	37.62(0.87)	99.53(1.12)
9	105.56(21.44)	139.36(22.95)	96.85(21.87)	26.56(5.97)	37.61(1.12)	99.27(1.77)
10	92.66(13.78)	130.6(17.1)	90.85(17.5)	21.9(4.4)	37.13(0.79)	99.68(0.91)
11	95.33(18.83)	141.09(19.91)	94.99(17.23)	22.39(5.67)	37.15(0.76)	98.92(1.37)
12	97.16(19.64)	145.11(22.71)	99.21(18.66)	23.22(5.93)	37.38(0.93)	99.24(1.84)

Table B.5

Summary of minimum values for vital signs in each cluster [mean(SD)].

#">#	Min HR	Min SysBP	Min MBP	Min RR	Min	Min SPO2
1	73.79(15.97)	102.4(17.26)	62.97(12.71)	14.25(4.18)	36.32(0.78)	94.42(4.01)
2	78.94(18.94)	93.85(18.23)	59.29(12.95)	15.25(4.36)	36.18(1.02)	92.87(7.9)
3	77.4(17)	99.73(18.07)	62.77(13.25)	15.08(4.58)	36.34(0.79)	92.58(5.57)
4	79.32(17.05)	99.92(18.74)	62.97(13.1)	15.66(4.64)	36.3(0.8)	93.2(3.71)

(continued on next page)

Table B.5 (continued)

#">#	Min HR	Min SysBP	Min MBP	Min RR	Min	Min SPO2
5	76.09(15.91)	100.13(17.46)	64.32(11.75)	13.89(3.61)	36.26(0.88)	94.19(6.19)
6	80.71(17.32)	98.95(19.65)	60.41(13.49)	15.98(4.53)	36.26(0.78)	92.88(5.34)
7	69.91(17.59)	95.37(18.78)	59.72(12.56)	14.9(4.03)	36.11(0.96)	92.75(5.18)
8	76.93(17.42)	101.28(16.67)	66.22(11.69)	13.81(3.64)	36.24(0.91)	95.16(4.2)
9	78.46(18.37)	94.1(17.04)	61.51(12.93)	15.55(4.33)	36.31(1.03)	92.75(6.57)
10	72.82(12.55)	96.55(12.76)	65.68(8.8)	13.02(2.86)	36.04(0.67)	95.08(5.08)
11	73.1(15.52)	106.06(17.47)	66.84(13.63)	13.65(3.71)	36.38(0.69)	94.3(3.43)
12	74.2(17.21)	105.39(18.09)	69(12.51)	13.76(3.67)	36.33(0.83)	94.77(5.15)

Table B.6

Summary of maximum values for lab results in each cluster [mean(SD)] - Part 1.

#">#	Anion Gap	Albumin	Bands	Bicarbonate	Bilirubin
1	15.47(3.8)	3.2(0.64)	10.92(10.65)	23.79(4.2)	1.64(3.07)
2	16.37(4.95)	3.08(0.65)	11.05(11.16)	24.15(6.28)	1.79(3.01)
3	15.12(4.37)	3.2(0.64)	9.79(9.65)	24.24(5.01)	1.71(3.28)
4	16.38(4.82)	3.16(0.65)	10.19(10.36)	23.69(4.61)	2.91(5.42)
5	14.75(3.86)	3.2(0.64)	10.46(10.45)	24.09(3.85)	1.17(1.46)
6	16.25(4.6)	3.1(0.66)	11.25(11.09)	23.68(4.85)	2.33(4.73)
7	16.22(4.97)	3.16(0.64)	9.44(10.4)	24.55(4.8)	1.81(4.3)
8	15.14(4.24)	3.25(0.65)	12.34(11.29)	23.79(3.47)	1.72(2.77)
9	16.86(5.92)	3.09(0.69)	10.76(10.6)	23.96(5.43)	2.62(5.01)
10	13.31(3.25)	3.35(0.66)	10.93(9.62)	25.38(3.05)	1.38(1.85)
11	16.45(4.59)	3.29(0.65)	10.55(9.89)	23.93(4.15)	1.65(3.37)
12	15.4(4.4)	3.26(0.69)	10.95(10.7)	24.36(3.77)	1.84(4.12)

Table B.7

Summary of maximum values for lab results in each cluster [mean(SD)] - Part 2.

#">#	Creatinine	Chloride	Glucose	Hematocrit	Hemoglobin
1	1.39(1.55)	106.51(6.41)	159.08(65.88)	34.4(5.47)	11.51(1.93)
2	1.65(1.59)	107.04(7.94)	185.95(124.62)	33.9(6.21)	11.1(2.02)
3	1.65(1.93)	105.31(7.58)	152.98(97.38)	33.22(5.83)	10.98(2)
4	2.11(2.25)	105.05(7.25)	157.98(115.93)	33.45(5.97)	11.12(2.07)
5	1.12(0.78)	106.76(5.99)	162.44(69.4)	34.47(5.28)	11.56(1.87)
6	1.7(1.72)	105.89(7.59)	169.39(115.36)	32.86(5.63)	10.86(1.93)
7	1.8(1.56)	104.15(5.84)	175.25(90.76)	34.32(5.74)	11.35(2.02)
8	1.24(0.72)	107.81(5.12)	161.17(122.52)	36.77(5.9)	12.49(2.07)
9	2.04(1.73)	107.07(7.33)	175.89(99.81)	35.24(6.75)	11.65(2.28)
10	1.19(0.95)	108.8(4.21)	138.33(59.12)	35.58(4.83)	12.11(1.75)
11	1.62(1.16)	105.55(5.36)	175.95(119.25)	34.07(4.85)	11.31(1.75)
12	1.46(1.5)	105.99(5.67)	165.96(73.44)	36.17(5.94)	12.22(2.08)

Table B.8

Summary of maximum values for lab results in each cluster [mean(SD)] - Part 3.

#">#	Lactate	Platelet	Potassium	PTT	INR
1	2.47(1.75)	251.17(120.78)	4.24(0.78)	35.84(22.42)	1.53(1.31)
2	2.99(2.9)	240.63(130.95)	4.49(0.96)	40.15(26.07)	1.63(1.1)
3	2.05(1.36)	244.89(140.55)	4.35(0.87)	38.12(26.88)	1.69(1.76)
4	2.54(1.98)	212.59(138.91)	4.48(0.87)	37.32(21.64)	1.76(1.42)
5	2.59(1.93)	240.97(125.31)	4.29(0.82)	35.19(22.75)	1.43(1.12)
6	2.43(1.72)	239.57(135.78)	4.35(0.95)	36.89(23.09)	1.72(1.61)
7	2.57(2.33)	238.33(120.17)	4.56(0.86)	49.76(36.89)	1.73(1.38)
8	3.33(2.45)	233.75(101.82)	4.49(0.82)	34.35(20.91)	1.42(1.07)
9	3.26(3.06)	218.05(123.64)	4.67(0.96)	41.44(24.95)	1.8(1.59)
10	2.69(1.62)	211.36(74.3)	4.4(0.57)	57.24(33.76)	1.46(0.51)
11	2.3(1.8)	221.26(116.07)	4.46(0.85)	39.92(28.61)	1.66(1.3)
12	2.84(2.77)	240.13(123.33)	4.31(0.69)	34.7(20.01)	1.51(1.42)

Table B.9

Summary of maximum values for lab results in each cluster [mean(SD)] - Part 4.

#">#	PT	Sodium	BUN	WBC	GCS
1	16.5(10.38)	139.84(4.75)	24.88(21.04)	12.95(6.32)	12.17(3.47)
2	17.7(10.66)	140.49(6.08)	32.21(24.37)	14.76(8.5)	10.84(3.68)
3	17.31(11.41)	138.99(6.18)	30.69(26.9)	12.11(15.65)	13.77(2.49)
4	18.89(12.75)	138.61(6.15)	38.41(29.93)	14(16.54)	14.39(1.66)
5	15.6(8.51)	139.86(4.64)	20.73(12.93)	13.34(9.21)	12.33(3.27)

(continued on next page)

Table B.9 (continued)

#">#	PT	Sodium	BUN	WBC	GCS
6	18.07(11.27)	139.21(6.46)	33.72(27)	14.65(15.84)	14.37(1.5)
7	18.45(10.72)	138.72(4.81)	36.8(23.77)	12.53(6.26)	13.08(3.38)
8	15.94(9.34)	140.27(3.99)	20.55(12.78)	14.39(7.02)	11.4(3.51)
9	19(12.02)	140.31(6.03)	38.52(28.31)	15.88(26.01)	9.97(3.85)
10	16.26(4.55)	139.81(3.23)	21.53(13.57)	13.35(4.97)	11.78(4.21)
11	18.2(13.95)	139.06(4.19)	33.51(23.36)	12.74(10.04)	13.68(2.75)
12	16.32(8.89)	139.67(4.56)	24.12(20.21)	13.29(8.3)	11.84(3.76)

Table B.10

Summary of minimum values for lab results in each cluster [mean(SD)] - Part 1.

#">#	Anion Gap	Albumin	Bands	Bicarbonate	Bilirubin
1	13.93(3.23)	3.14(0.65)	9.94(9.55)	22.3(4.59)	1.49(2.77)
2	13.95(3.81)	3.04(0.66)	9.87(10.07)	21.99(6.62)	1.76(3.31)
3	13.82(3.51)	3.18(0.64)	9.31(8.93)	23.14(5.32)	2.02(4.05)
4	14.35(3.7)	3.12(0.64)	9.08(9.11)	22.11(4.87)	2.68(5.13)
5	13.29(3.52)	3.16(0.67)	9.85(9.48)	22.81(4.26)	1.33(1.8)
6	14.18(3.73)	3.05(0.65)	10.13(10.12)	21.97(5.32)	2.26(4.77)
7	14.51(3.89)	3.14(0.63)	8.7(9.34)	22.93(5.29)	1.92(4.06)
8	12.86(3.08)	3.18(0.66)	10.67(10.03)	22.01(4.02)	1.38(2.04)
9	13.96(4.57)	3.04(0.7)	9.58(9.67)	21.57(5.91)	2.32(4.5)
10	11.95(3.09)	3.27(0.68)	9.29(8.78)	23.48(2.9)	1.22(2.87)
11	14.14(3.82)	3.26(0.66)	9.44(8.28)	22.15(4.5)	2.36(4.2)
12	13.71(3.5)	3.21(0.71)	10.08(9.95)	22.88(4.31)	1.83(3.99)

Table B.11

Summary of minimum values for lab results in each cluster [mean(SD)] - Part 2.

#">#	Creatinine	Chloride	Glucose	Hematocrit	Hemoglobin
1	1.26(1.42)	104.24(6.25)	134.49(45.24)	32.32(6.05)	10.88(2.03)
2	1.45(1.4)	103.92(7.86)	133.46(58.77)	31.14(6.41)	10.32(2.11)
3	1.51(1.65)	103.54(7.83)	128.36(59.38)	32.18(6.11)	10.73(2.05)
4	1.86(1.82)	102.48(8)	126.18(51.46)	31.04(6.35)	10.51(2.13)
5	1.02(0.71)	104.77(5.75)	135.31(45.74)	31.8(6.03)	10.91(2)
6	1.52(1.52)	102.82(7.88)	127.93(53.92)	30.63(5.96)	10.22(1.98)
7	1.62(1.35)	102.07(5.99)	138.97(60.28)	32.61(5.97)	10.94(2.03)
8	1.06(0.6)	105.4(5.65)	130.07(36.3)	31.79(6.52)	11.12(2.26)
9	1.73(1.44)	103.78(7.33)	128.72(54.26)	32.08(7.17)	10.8(2.36)
10	1.03(0.67)	105.06(4.46)	129.68(53.52)	28.64(5.38)	9.99(1.82)
11	1.43(0.88)	102.79(5.4)	123.77(47.85)	32.02(5.1)	10.78(1.78)
12	1.3(1.33)	103.82(5.81)	137.94(48.33)	33.88(6.53)	11.55(2.23)

Table B.12

Summary of minimum values for lab results in each cluster [mean(SD)] - Part 3.

#">#	Lactate	Platelet	Potassium	PTT	INR
1	1.74(0.96)	229.21(111.87)	3.89(0.65)	29.95(10.9)	1.34(0.59)
2	1.88(1.52)	214.67(119.82)	3.94(0.72)	33.73(16.92)	1.44(0.83)
3	1.59(0.88)	235.43(135.45)	4.06(0.7)	33.98(19.27)	1.48(0.68)
4	1.85(1.15)	194.39(128.51)	4.08(0.67)	32.89(12.55)	1.6(0.94)
5	1.78(1.14)	217.37(117.25)	3.94(0.56)	29.25(11.2)	1.29(0.42)
6	1.79(1.06)	220.38(127.89)	3.89(0.72)	32.12(13.54)	1.54(0.99)
7	1.88(1.57)	222.14(116.33)	4.1(0.65)	38.02(22.97)	1.57(0.79)
8	1.99(1.21)	193.1(96.19)	4.06(0.61)	28.28(8.76)	1.24(0.3)
9	2.04(1.85)	192.27(114.9)	4.08(0.69)	34.26(13.28)	1.57(1.04)
10	1.47(1.02)	162.19(68.23)	4.09(0.46)	34.99(16.48)	1.24(0.26)
11	1.65(0.9)	204.06(113.13)	4.04(0.72)	33.64(19.12)	1.48(0.75)
12	1.77(1.22)	217.78(116.51)	4(0.56)	29.84(10.57)	1.31(0.44)

Table B.13

Summary of minimum values for lab results in each cluster [mean(SD)] - Part 4.

#">#	PT	Sodium	BUN	WBC	GCS
1	15.05(5.48)	138.22(4.7)	22.75(19.62)	11.34(5.37)	9.44(4.49)
2	15.95(6.92)	138.38(6.08)	28.98(21.46)	12.51(7.43)	7.16(3.52)
3	16.28(6.13)	137.63(6.39)	28.44(24.4)	11.48(15.54)	12.23(3.77)
4	17.38(8.49)	136.76(6.65)	34.85(26.72)	12(10.87)	13.63(2.55)
5	14.52(3.91)	138.29(4.43)	19.15(12.09)	11.46(7.58)	9.66(4.32)
6	16.75(8.37)	137.1(6.92)	30.56(24.13)	12.4(10.88)	13.54(2.29)

(continued on next page)

Table B.13 (continued)

#">#	PT	Sodium	BUN	WBC	GCS
7	17.13(7.14)	136.91(5.08)	33.84(22.15)	11.35(5.7)	11.52(4.45)
8	14.09(2.85)	138.19(5.09)	18.29(12.17)	10.97(5.52)	8.06(4.33)
9	17.01(7.46)	138.04(5.83)	34.16(25.48)	12.99(19.64)	6.34(3.27)
10	14.16(2.44)	138.72(3.34)	19.05(12.25)	9.41(3.75)	6.25(4.61)
11	16.29(6.87)	137.3(4.17)	30.29(21.51)	11.08(8.02)	11.9(3.79)
12	14.79(4.11)	138.09(4.78)	21.85(17.64)	11.26(6.69)	8.66(4.69)

Table B.14

Summary of medication dosage [mean(SD)].

#">#	Norepinephrine	Epinephrine	Phenylephrine	Vasopressin	Dobutamine	Dopamine
1	1.07(4.41)	0.03(0.41)	5.93(26.29)	0.89(6.04)	3.47(34.27)	0.61(11.82)
2	3.62(8.17)	0.09(1.28)	13.77(53.18)	5.27(19.55)	14.89(122.97)	0.61(11.35)
3	0.59(2.96)	0(0)	3.08(22.86)	0.95(8.35)	0.68(10.75)	0(0)
4	0.71(3.11)	0(0)	2.84(29.85)	0.26(4.77)	3.83(39.79)	0.06(1.81)
5	0.72(4.16)	0(0.07)	9.22(39.81)	0.91(7.68)	3.49(30.5)	0.03(0.48)
6	0.86(3.06)	0(0.07)	3.05(25.4)	0.58(5.76)	3.25(27.15)	0.65(17.8)
7	1.55(6.1)	0.03(0.29)	5.69(36.13)	0.46(5.23)	79.01(266.31)	4.32(36.8)
8	1.1(4.85)	0.01(0.18)	15.26(45.44)	1.15(7.57)	1.62(21.05)	0(0)
9	4.47(10.11)	0.01(0.18)	19.2(65.07)	4.32(16.81)	28.79(152.48)	3.49(35.95)
10	0.41(1.95)	0.15(0.76)	13.03(23.16)	0.79(7.13)	25.71(172.95)	0.68(10.48)
11	0.05(0.53)	0(0)	1.29(11.03)	0(0.01)	0.88(6.69)	0(0)
12	1.16(5.49)	0(0.08)	6.78(28.28)	1.22(9.28)	5.62(65.12)	1.13(23.68)

Table B.15

Summary of medication duration [mean(SD)].

#">#	Norepinephrine	Epinephrine	Phenylephrine	Vasopressin	Dobutamine	Dopamine
1	1.45(4.65)	0.05(0.71)	1.07(3.79)	0.42(2.72)	0.1(0.81)	0.03(0.45)
2	4.11(7.17)	0.06(0.68)	1.38(4.41)	2.18(8.02)	0.35(2.19)	0.03(0.51)
3	1.08(3.72)	0(0)	0.35(1.95)	0.4(3.52)	0.02(0.36)	0(0)
4	1.24(4.09)	0(0)	0.3(1.97)	0.11(2.02)	0.1(1.07)	0.01(0.24)
5	0.79(3.53)	0.01(0.18)	1.59(4.69)	0.36(2.98)	0.12(1.18)	0(0.04)
6	1.53(4.54)	0.03(0.65)	0.34(2.14)	0.24(2.4)	0.11(0.89)	0.03(0.84)
7	1.65(5.08)	0.06(0.56)	0.57(2.89)	0.21(2.23)	1.94(5.53)	0.2(1.89)
8	1.06(3.7)	0(0.05)	2.15(5.29)	0.56(3.46)	0.05(0.56)	0(0)
9	4.29(7.16)	0.03(0.48)	1.53(4.26)	1.78(6.89)	0.46(2.05)	0.09(0.98)
10	0.69(3.29)	1.01(4.24)	3.47(4.58)	0.28(2.3)	0.59(3.75)	0.08(1.06)
11	0.21(2.21)	0(0)	0.35(2.52)	0(0.01)	0.06(0.48)	0(0)
12	1.07(4.04)	0.01(0.14)	1.06(3.53)	0.52(3.89)	0.11(1.2)	0.04(0.78)

Table B.16

Summary of severity of illness scores for each cluster [mean(SD)].

#">#	Elixhauser	SOFA	LODS	SIRS	qSOFA
1	2.85(6.11)	4.58(2.81)	4.26(2.6)	2.86(0.92)	1.94(0.75)
2	4.52(7.91)	7.29(3.94)	6.58(3.35)	3.08(0.89)	2.05(0.66)
3	4.01(7.31)	4.48(2.44)	4.2(2.48)	2.77(0.95)	1.86(0.72)
4	5.68(7.29)	5.14(2.68)	4.57(2.38)	2.93(0.93)	1.83(0.69)
5	2.34(6.37)	4.24(2.44)	4.09(2.61)	2.93(0.96)	1.93(0.7)
6	4.38(7.16)	5.03(2.7)	4.63(2.53)	2.96(0.85)	1.95(0.68)
7	2.2(6.4)	5.28(3.2)	5.45(3.14)	2.71(0.97)	1.87(0.68)
8	2.72(5.69)	4.76(2.86)	4.37(2.55)	3.07(0.94)	1.83(0.73)
9	5.84(7.36)	7.99(4.05)	6.91(3.33)	3.13(0.89)	2.03(0.68)
10	0.17(5.38)	5.02(2.47)	4.56(2.28)	2.86(0.97)	1.86(0.6)
11	1.81(6.38)	3.36(1.69)	2.97(2.29)	2.46(0.95)	1.71(0.79)
12	3.59(6.37)	4.9(2.99)	4.43(2.8)	2.86(0.93)	1.81(0.76)

References

- [1] Y. Sha, J. Venugopalan, M.D. Wang, A novel temporal similarity measure for patients based on irregularly measured data in electronic health records, in: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, ACM, 2016, pp. 337–344.
- [2] J. Sun, D. Sow, J. Hu, S. Ebadollahi, Localized supervised metric learning on temporal physiological data, in: Pattern Recognition (ICPR), 2010 20th International Conference on, 2010, pp. 4149–4152, <https://doi.org/10.1109/ICPR.2010.1009>.
- [3] A. Sharafoddini, J.A. Dubin, J. Lee, Patient similarity in prediction models based on health data: a scoping review, JMIR medical informatics 5 (1) (2017) e7.
- [4] A. Singh, G. Nadkarni, O. Gottesman, S.B. Ellis, E.P. Bottinger, J.V. Guttag, Incorporating temporal ehr data in predictive models for risk stratification of renal function deterioration, J. Biomed. Inf. 53 (2015) 220–228, <https://doi.org/10.1016/j.jbi.2014.11.005>.
- [5] J. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer-Verlag, New York, 2005 <https://doi.org/10.1007/b98888>. Springer Series in Statistics.
- [6] Critical care statistics. URL <https://www.sccm.org/Communications/Critical-Care-Statistics>.
- [7] M. Prin, H. Wunsch, International comparisons of intensive care: informing outcomes and improving standards, Curr. Opin. Crit. Care 18 (6) (2012) 700–706, <https://doi.org/10.1097/MCC.0b013e3283591d5>.
- [8] M. Pimentel, D. Clifton, L. Clifton, L. Tarassenko, Modelling patient time-series data from electronic health records using Gaussian processes, in: Advances in

- Neural Information Processing Systems: Workshop on Machine Learning for Clinical Data Analysis, 2013, pp. 1–4.
- [9] L.W.H. Lehman, S. Nemati, G.B. Moody, T. Heldt, R.G. Mark, Uncovering clinical significance of vital sign dynamics in critical care, in: Computing in Cardiology, 2014, pp. 1141–1144, 2014.
- [10] L.W. Lehman, R.P. Adams, L. Mayaud, G.B. Moody, A. Malhotra, R.G. Mark, S. Nemati, A physiological time series dynamics-based approach to patient monitoring and outcome prediction, IEEE J Biomed Health Inform 19 (3) (2015) 1068–1076, <https://doi.org/10.1109/JBHI.2014.2330827>.
- [11] L.W.H. Lehman, S. Nemati, R.G. Mark, Hemodynamic monitoring using switching autoregressive dynamics of multivariate vital sign time series, in: 2015 Computing in Cardiology Conference (CinC), 2015, pp. 1065–1068, <https://doi.org/10.1109/CIC.2015.7411098>.
- [12] V. Agarwal, N.H. Shah, Learning attributes of disease progression from trajectories of sparse lab values, in: PACIFIC SYMPOSIUM ON BIocomputing 2017, World Scientific, 2017, pp. 184–194.
- [13] M. Singer, C.S. Deutszman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.D. Chiche, C.M. Coopersmith, R.S. Hotchkiss, M.M. Levy, J.C. Marshall, G.S. Martin, S.M. Opal, G.D. Rubenfeld, T. van der Poll, J.L. Vincent, D.C. Angus, The third international consensus definitions for sepsis and septic shock (sepsis-3), J. Am. Med. Assoc. 315 (8) (2016) 801–810, <https://doi.org/10.1001/jama.2016.0287>.
- [14] M. Wu, M. Ghassemi, M. Feng, L.A. Celi, P. Szolovits, F. Doshi-Velez, Understanding vasopressor intervention and weaning: risk prediction in a public heterogeneous clinical time series database, J. Am. Med. Inf. Assoc. 24 (3) (2017) 488–495, <https://doi.org/10.1093/jamia/ocw138>.
- [15] A.S. Fialho, L.A. Celi, F. Cismondi, S.M. Vieira, S.R. Reti, J.M. Sousa, S. N. Finkelstein, Disease-based modeling to predict fluid response in intensive care units, Methods Inf. Med. 52 (6) (2013) 494–502, <https://doi.org/10.3414/ME12-01-0093>.
- [16] R.L. Kravitz, N. Duan, J. Braslow, Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages, Milbank Q. 82 (4) (2004) 661–687, <https://doi.org/10.1111/j.0887-378X.2004.00327.x>.
- [17] A.E. Fohner, J.D. Greene, B.L. Lawson, J.H. Chen, P. Kipnis, G.J. Escobar, V.X. Liu, Assessing clinical heterogeneity in sepsis through treatment patterns and machine learning, J. Am. Med. Inf. Assoc. 26 (12) (2019) 1466–1477, <https://doi.org/10.1093/jamia/ocz106>.
- [18] A.E.W. Johnson, J. Aboab, J.D. Raffa, T.J. Pollard, R.O. Deliberato, L.A. Celi, D. J. Stone, A comparative analysis of sepsis identification methods in an electronic database, Crit. Care Med. 46 (4) (2018) 494–499, <https://doi.org/10.1097/CCM.0000000000002965>.
- [19] C.M. Salgado, S.M. Vieira, L.F. Mendona, S. Finkelstein, J.M.C. Sousa, Ensemble fuzzy models in personalized medicine: application to vasopressors administration, Eng. Appl. Artif. Intell. 49 (2016) 141–148, <https://doi.org/10.1016/j.engappai.2015.10.004>.
- [20] S. Ghosh, J. Li, L. Cao, K. Ramamohanarao, Septic shock prediction for icu patients via coupled hmm walking on sequential contrast patterns, J. Biomed. Inf. 66 (2017) 19–31, <https://doi.org/10.1016/j.jbi.2016.12.010>.
- [21] F. Khoshnevisan, J. Ivy, M. Capan, R. Arnold, J. Huddleston, M. Chi, Recent temporal pattern mining for septic shock early prediction, in: 2018 IEEE International Conference on Healthcare Informatics (ICHI), 2018, pp. 229–240, <https://doi.org/10.1109/ICHI.2018.00033>.
- [22] C.W. Seymour, J.N. Kennedy, S. Wang, C.-C.H. Chang, C.F. Elliott, Z. Xu, S. Berry, G. Clermont, G. Cooper, H. Gomez, et al., Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis, Jama 321 (20) (2019) 2003–2017.
- [23] A.E. Johnson, T.J. Pollard, L. Shen, L.W. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L.A. Celi, R.G. Mark, Mimic-iii, a freely accessible critical care database, Sci Data 3 (2016) 160035, <https://doi.org/10.1038/sdata.2016.35>.
- [24] A. Johnson, T. Pollard, alistairewj/sepsis3-mimic: Sepsis-3 Study v1.0.0, May 2018, <https://doi.org/10.5281/zenodo.1256723>. URL.
- [25] S. Ullah, C.F. Finch, Applications of functional data analysis: a systematic review, BMC Med. Res. Methodol. 13 (1) (2013) 43, <https://doi.org/10.1186/1471-2288-13-43>.
- [26] X. Dai, P. Hadjipantelis, H. Ji, H. Mueller, J. Wang, fdapace: functional data analysis and empirical dynamics, R package version 0.3. 0. pp. 36–40.
- [27] J. Lee, D.M. Maslove, J.A. Dubin, Personalized mortality prediction driven by electronic medical data and a patient similarity metric, PloS One 10 (5) (2015), e0127428, <https://doi.org/10.1371/journal.pone.0127428>.
- [28] R.J.A. Little, Missing-data adjustments in large surveys, J. Bus. Econ. Stat. 6 (3) (1988) 287–296, <https://doi.org/10.1080/07350015.1988.10509663>.
- [29] A. Sharafoddini, J.A. Dubin, D.M. Maslove, J. Lee, A new insight into missing data in intensive care unit patient profiles: observational study, JMIR Med Inform 7 (1) (2019), e11605, <https://doi.org/10.2196/11605>.
- [30] A. Sharafoddini, J.A. Dubin, J. Lee, Finding similar patient subpopulations in the icu using laboratory test ordering patterns, in: Proceedings of the 2018 7th International Conference on Bioinformatics and Biomedical Science, ACM, 2018, pp. 72–77.
- [31] L.v. d. Maaten, G. Hinton, Visualizing data using t-sne, J. Mach. Learn. Res. 9 (Nov) (2008) 2579–2605.
- [32] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, Kdd 96 (1996) 226–231.
- [33] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic, J. Roy. Stat. Soc. B 63 (2) (2001) 411–423, arXiv, <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00293>.
- [34] P.J. Rousseeuw, Silhouettes, A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- [35] C. Hennig, Cluster-wise assessment of cluster stability, Comput. Stat. Data Anal. 52 (1) (2007) 258–271. URL, <http://www.sciencedirect.com/science/article/pii/S0167947306004622>.
- [36] C. Hennig, fpc: flexible procedures for clustering, R package version 2 (2) (2010), 0–3.
- [37] C. Reilly, C. Wang, M. Rutherford, A rapid method for the comparison of cluster analyses, Stat. Sin. 15 (1) (2005) 19–33. URL, <http://www.jstor.org/stable/24307237>.
- [38] J.-L. Vincent, A. De Mendonça, F. Cantraine, R. Moreno, J. Takala, P.M. Suter, C. L. Sprung, F. Colardyn, S. Blecher, Use of the sofa score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study, Crit. Care Med. 26 (11) (1998) 1793–1800.
- [39] M. Singer, C.S. Deutszman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.-D. Chiche, C.M. Coopersmith, et al., The third international consensus definitions for sepsis and septic shock (sepsis-3), Jama 315 (8) (2016) 801–810.
- [40] A. C. of Chest Physicians, et al., Society of critical care medicine consensus conference committee: accp/sccm consensus conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis, Crit. Care Med. 20 (1992) 864–874.
- [41] A. Elixhauser, C. Steiner, D.R. Harris, R.M. Coffey, Comorbidity measures for use with administrative data, Med. Care 36 (1) (1998) 8–27, <https://doi.org/10.1097/00005650-19980100-00004>.
- [42] N.M. Mohr, J.P. Vakkalanka, B.A. Faine, B. Skow, K.K. Harland, R. Dick-Perez, B. M. Fuller, A. Ahmed, S.Q. Simson, Serum anion gap predicts lactate poorly, but may be used to identify sepsis patients at risk for death: a cohort study, J. Crit. Care 44 (2018) 223–228.
- [43] Y.J. Park, D.H. Kim, S.C. Kim, T.Y. Kim, C. Kang, S.H. Lee, J.H. Jeong, S.B. Lee, D. Lim, Serum lactate upon emergency department arrival as a predictor of 30-day in-hospital mortality in an unselected population, PloS One 13 (1) (2018) 1–14, <https://doi.org/10.1371/journal.pone.0190519>. URL.
- [44] H. Kendall, E. Abreu, A.-L. Cheng, Serum albumin trend is a predictor of mortality in icu patients with sepsis, Biol. Res. Nurs. 21 (3) (2019) 237–244, <https://doi.org/10.1177/1099800419827600>, PMID: 30722677, arXiv.
- [45] I. Arnau-Barrés, R. Güerri-Fernández, S. Luque, L. Sorli, O. Vázquez, R. Miralles, Serum albumin is a strong predictor of sepsis outcome in elderly patients, Eur. J. Clin. Microbiol. Infect. Dis. 38 (4) (2019) 743–746.
- [46] N.A. Ali, J.M. O'Brien Jr., K. Dungan, G. Phillips, C.B. Marsh, S. Lemeshow, A. F. Connors Jr., J.-C. Preiser, Glucose variability and mortality in patients with sepsis, Crit. Care Med. 36 (8) (2008) 2316.
- [47] L.A. Dossett, H. Cao, N.T. Mowery, M.J. Dortch, J.M. Morris, A.K. May, Blood glucose variability is associated with mortality in the surgical intensive care unit, Am. Surg. 74 (8) (2008) 679–685.
- [48] S.E. Siegelhaar, F. Holleman, J.B.L. Hoekstra, J.H. DeVries, Glucose variability; does it matter? Endocr. Rev. 31 (2) (2010) 171–182, arXiv, <https://academic.oup.com/edrv/article-pdf/31/2/171/10334616/edrv0171.pdf>.
- [49] J. Khoury, F. Bahouth, Y. Stabholz, A. Elias, T. Mashiah, D. Aronson, Z.S. Azzam, Blood urea nitrogen variation upon admission and at discharge in patients with heart failure, ESC heart failure 6 (4) (2019) 809–816.