Unraveling Patient Heterogeneity in ICU With Deep Embedded Clustering Using Co-morbidity, Clinical Examination, and Laboratory Data

José Castela Forte (j.n.alves.castela.cardoso.forte@umcg.nl)
Universitair Medisch Centrum Groningen https://orcid.org/0000-0001-9273-0702
Galiya Yeshmagambetova
University of Groningen: Rijksuniversiteit Groningen
Maureen van der Grinten
University of Groningen: Rijksuniversiteit Groningen
Bart Hiemstra
UMCG: Universitair Medisch Centrum Groningen
Thomas Kaufmann
UMCG: Universitair Medisch Centrum Groningen
Ruben Eck
UMCG: Universitair Medisch Centrum Groningen
Frederik Keus
UMCG: Universitair Medisch Centrum Groningen
Anne Epema
UMCG: Universitair Medisch Centrum Groningen
Marco Wiering
RUG: Rijksuniversiteit Groningen
Iwan van der Horst
Maastricht UMC+: Maastricht Universitair Medisch Centrum+

Research

Keywords: Clustering, Heterogeneity, Machine learning, Patient characterization

Posted Date: December 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-129367/v1

Unraveling patient heterogeneity in ICU with deep embedded clustering using co-morbidity, clinical examination, and laboratory data

First author: José Castela Forte BSc (Hon) [1,2,3] (ORCID ID: 0000-0001-9273-0702)

Other authors: Galiya Yeshmagambetova BSc[3], Maureen L. van der Grinten BSc[3], Bart Hiemstra M.D.,Ph.D.[2], Thomas Kaufmann, M.D., Ph.D. [2], Ruben J. Eck M.D.[4], Frederik Keus M.D.,Ph.D. [5], Anne H. Epema M.D.,Ph.D.[2], Marco A. Wiering Ph.D.[3], Iwan C.C. van der Horst M.D.,Ph.D.[6]

[1] Department of Clinical Pharmacy and Pharmacology University of Groningen, University Medical Center Groningen, the Netherlands

[2] Department of Anesthesiology University of Groningen, University Medical Center Groningen, the Netherlands

[3] Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, the Netherlands

[4] Department of Internal Medicine, University of Groningen, University Medical Center Groningen, the Netherlands

[5] Department of Critical Care, University of Groningen, University Medical Center Groningen, the Netherlands

[6] Department of Intensive Care, Maastricht University Medical Centre+, University Maastricht, Maastricht, the Netherlands

Correspondence to:

José Castela Forte, BSc (Hon)

Department of Pharmacy and Pharmacology

Hanzeplein 1, P.O. Box 30.001

9700 RB Groningen, The Netherlands

Phone: +31628218360

E-mail: j.n.alves.castela.cardoso.forte@umcg.nl

Abstract

Introduction

Despite extensive research, the goal of unravelling patient heterogeneity in critical care remains largely unattained. Combining clustering analysis of routinely collected high-frequency data with the identification of features driving cluster separation may constitute a step towards improving patient characterization.

Methods

In this study, we analysed prospectively collected data from 743 patients including co-morbidities, clinical examination, and laboratory parameters. We compared four clustering methodologies – deep embedded clustering (DEC), hierarchical clustering with and without dynamic time warping, and k-means – and trained a classifier to predict and validate cluster membership. The contribution of different variables to the predicted cluster membership was assessed using SHapley Additive exPlanations values.

Results

DEC yielded better results compared to the traditional clustering algorithms, with the best Jaccard and entropy scores being achieved for 6 clusters. These clusters were characterized as medium to high co-morbidity patients with respiratory pathology and sepsis (cluster 1), patients with primarily acute and chronic cardiac conditions and surgical admission (cluster 2), patients with diverse disease etiology and poor outcomes (cluster 3), low co-morbidity neurological, neurosurgical, and trauma patients (cluster 4), medium co-morbidity patients with cardio-respiratory problems, and neuro-trauma patients with longer length of stay (cluster 5), and patients with sepsis and respiratory infections (cluster 6). All clusters differed in in-ICU, 30-day, and 90-day mortality, as well as incidence of acute kidney injury, and two clusters were categorized as having higher mortality risk, and one cluster as lower mortality risk.

Conclusions

This machine learning methodology, which we made publicly available, is a possible solution to challenges previously encountered by clustering analyses, and may help unravel patient heterogeneity in critical care.

Keywords

Clustering; Heterogeneity; Machine learning; Patient characterization.

Introduction

Critically ill patients constitute a highly heterogeneous population, with high rates of acute and chronic multimorbidity, and different profiles of risk, response to interventions, and outcomes. Despite extensive research, however, the goal of unravelling patient heterogeneity remains largely unattained. Critical care clinicians rely on a combination of laboratory and clinical examination variables, and their own clinical experience (clinical gestalt) and gut feeling to characterize patients [1]. While human beings excel at ascribing meaning to observed patterns once they have seen them, data-driven approaches enabling the combination of diverse data streams can enhance patient characterization within known "sub-phenotypes" or provide insight into new categorizations [2,3].

Different approaches to this challenge of unravelling heterogeneity have shown promise. Clustering analysis has been used for several purposes in critical care research, and is especially appealing because its principles resemble a heuristic which clinicians are familiar with: it finds similarities and differences between patients and divides them into groups [1]. Studies in patients with septic shock or acute respiratory distress syndrome have used clustering analysis to further characterize patients into sub-phenotypes according to different mechanisms of disease [4,5], while other studies identified subgroups with different responses to treatments in patients with asthma and acute kidney injury (AKI) [6,7]. Similarly, prediction models with increasingly higher accuracy and explainability for mortality and organ injury have been suggested, capable of processing high-frequency data [8,9,10]. These two approaches serve different purposes, and as such have different advantages and shortcomings. Prediction models can deal with virtually any data format and provide individual probabilities for an outcome, but are bound by *a priori* hypotheses and can only compute the probability of one specific outcome. On the other hand, traditional

clustering algorithms, are not designed to process the high-frequency, dynamic data collected in ICU [1].

In this study, we sought to develop a novel approach to identify and characterize clusters of patients with similar clinical trajectories and outcomes. To do this, we combined clustering and feature importance analysis to analyze routinely collected data from a minimally selected ICU cohort. Patients were clustered by deep embedded clustering using co-morbidity, clinical examination, and laboratory data. Then, a classifier was trained to predict cluster membership and identify the features driving these predictions.

Materials and methods

Study design and data collection

Data used for this study originated from the prospective, single-centre Simple Intensive Care Studies (SICS) I cohort study. All acutely admitted, critically ill patients included the study underwent clinical examination and critical care ultrasonography (CCUS) within the first 24h of ICU admission. Further details on inclusion criteria, informed consent, and study protocol are available elsewhere [11,12]. The study was approved by the local institutional review board (Medisch Ethische Toetsingscommissie (METc) of the UMCG, M15.168207).

Co-morbidity, clinical examination, and laboratory data

The dataset consisted of patient characteristics including co-morbidities, clinical examination variables including CCUS, vital signs, and urine output, and a time-series of 40 laboratory values measured at least once daily. CCUS measurements were validated by experts, and vital signs were recorded from the bedside monitor [11,12]. Patients with more than 10% missing data (i.e. variables for which no measurements were registered at any moment during ICU stay) were excluded from the analysis. Remaining missing data were imputed using a rolling mean based on ICU-specific values [13]. Feature extraction (mean and variance concatenated over the whole time-series) was employed to represent time-series data. As a measure of co-morbidity, the average number of co-morbid conditions per patient per cluster was calculated based on the information on co-morbidities reported in table 1 [14].

Outcome

To define and assess clinically relevant differences between clusters, mortality at three time points (in-ICU, as well as at 30 and 90 days) was taken as a primary outcome.

Additional outcomes included the development of any stage of AKI, need for vasopressors, ICU length of stay, development of shock, and need for renal replacement therapy.

Development and comparison of different clustering methodologies

Most clustering algorithms, such as k-means clustering and hierarchical clustering, are not designed to process high-frequency, dynamic data [1]. Different strategies have been developed to facilitate this, including combining K-means and HC with a time-series processing methodology such as dynamic time warping (DTW), as well as using clustering algorithms which represent data in a different way, such as deep embedding clustering (DEC) [15,16]. In both these approaches, described in more detail in Additional file 1, features such as mean and variance are extracted from the time-series data and subsequently fed into a clustering algorithm. In this study, we compared a DEC model to two "traditional" clustering algorithms, k-means and hierarchical clustering (HC), as well as a combination of HC and dynamic time warping (HC-DTW).

Deep embedded clustering algorithms utilize autoencoder neural networks to learn a certain representation of the data, and then use this representation to form clusters [15]. The DEC model developed in this study combined a multilayer perceptron (MLP) autoencoder, which is a type of neural network, and a custom clustering layer with the k-means clustering algorithm. The clustering layer reconstructs features created by the MLP autoencoder, and converts it to cluster label probabilities represented by Student's t-distribution. The clustering layer weights represent the cluster centroids and are initialized using k-means algorithm. To improve cluster purity, a centroid-based target distribution is constructed by squaring the encoded vectors and normalizing them by frequency per cluster. Finally, the algorithm is trained to minimize Kullback–Leibler divergence loss for a maximum of 8000 iterations with 0.01 tolerance threshold.

Once clusters were computed for all four algorithms, validity assessments were conducted [17]. Internal validity assesses whether the structure of the clustering is intrinsically appropriate for the data. Patients clustered in the same cluster should have similar data, whereas patients from different clusters should be as distinct as possible from those in other clusters. Here, the Silhouette index was used to internally validate k-means, HC, and HC-DTW. For DEC, cluster-wise stability was computed by resampling the dataset 100 times and computing the Jaccard similarities to the originally defined clusters as well as entropy scores [18,19,20]. External validity assesses whether clustering results match some *a priori* expected data structure. When the true cluster labels are known, this is done by comparing the clustering output to a given "correct" clustering [1,21]. Since no "true" labels are available when attempting to identify new putative patient clusters, the clinical recognizability of these clusters was used as surrogate of external validity.

Cluster membership prediction and feature importance analysis for cluster characterization

A gradient boosting algorithm (XGBoost) was trained to predict cluster membership over 10-folds for each of the 100 clustering configurations resulting from DEC. [22]. Then, SHapley Additive exPlanations (SHAP) values were computed on the run with the highest accuracy to represent the feature importance of each variable in the model. SHAP values are widely used in game theory to determine the contribution of particular features to the difference between the actual and the mean predictions [23]. Positive and negative SHAP values signal that variables contribute positively or negatively to cluster membership, respectively. Finally, clusters were characterized based on the between-cluster differences in input variables and outcomes, and feature importance values. The admission and discharge diagnoses of all patients were analysed and relevant clinical information to aid in the

characterization was extracted. A full schematic overview of the analysis is provided in figure 1.

Statistical analysis

Descriptive characteristics for the study population were reported as means with standard deviations and proportions for continuous and categorical variables, respectively. Differences in input variables and outcomes between clusters were determined using analysis of variance and chi-squared tests. Hazard ratios for mortality per cluster were computed, and the *p*-value for comparison against the full cohort was calculated using the log-rank test. A *p*-value below 0.05 was considered significant. All clustering and further statistical analyses were performed using Python with PyCharm as interface (version 2019.3.5).

Results

Study population

Of the 1075 patients included in SICS-I, 743 had less than 10% missing data and were included in the analysis. Both the numbers of variables and the numbers of measurements per variable varied per patient, with on average 21 measurements of each variable per patient. Patient characteristics, clinical examination variables, and outcomes are reported in table 1. Complete data on laboratory variables and outcomes can be found in Additional file 1 (tables S2-5, figures S9-10).

Performance of different clustering methodologies

Deep embedded clustering yielded better results compared to the traditional clustering algorithms. Both k-means and HC, as well as HC-DTW, grouped most patients in one cluster, and did not identify significant between-cluster differences in clinically relevant variables (Additional file 1: table S1). In contrast, clusters generated by DEC showed balanced cluster membership distribution and identified significant between-cluster differences for the majority of variables (Additional file 1: tables S2-4). Of the seven clustering configurations identified by DEC, stability was highest for six clusters (Additional file 1: figures S1-8). The 10-fold cross-validation XGBoost model predicted cluster membership with 83% accuracy, with sensitivity ranging from 64% to 90% and specificity from 85% to 100% (Additional file 1: table S6, figure S11).

Feature importance analysis and cluster characterization

Sixty-eight patients with high prevalence of respiratory failure or infection (34%), as well as sepsis (21%), were assigned to cluster 1 (figure 3, Additional file 1: table S4). These patients had a long ICU stay, and the highest rate of worsened respiratory condition after 24

hours (figure 2, Additional file 1: tables S2-4 and figure S9). Feature importance analysis identified increased alkaline phosphatase, gamma-GT, bilirubin and lactate as having the greatest impact on cluster membership predictions (Additional file 1: figure S12). Higher values for the former three variables drove predictions towards cluster membership, while a high lactate was associated with non-membership. Patients in this cluster were not at increased mortality risk (ICU, 30-day or 90-day; table 2).

Cluster 2 (n=100) included the highest percentage of surgical patients (48%, figure 2, Additional file 1: table S1), including the largest post-transplant group, and cardiac and vascular procedures (figure 3, Additional file 1: table S6). Almost 40% of patients presented with acute or chronic cardiac condition, with 63% having a low cardiac index (figure 3, Additional file 1: table S6). Accordingly, troponin T, lactate dehydrogenase (LDH), creatine kinase (CK) and inflammatory variables were increased (figure 2, Additional file 1: table S2, figure S9). Patients in this cluster had higher mortality rates (ICU, 30-day, and 90-day; HR 1.55 [95% CI, 1.26-1.91], 1.39 [1.12-1.71], and 1.30 [1.06-1.61], respectively) (figure 3, table 2, Additional file 1: table S6). Higher values for arterial oxygen (pO_2), LDH, lactate, troponin and calcium were associated with cluster membership, while low pO_2, LDH, and ASAT values drove predictions towards non-membership (Additional file 1: figure S13).

Cluster 3 (n=46) consisted of patients with diverse disease etiology, from liver disease and transplant (24%), to cardiac arrest (11%), sepsis (10%) and respiratory failure (10%). These patients were the youngest and had the highest severity scores at admission (figure 2, table S1). Like cluster 2, laboratory variables showed significant elevated troponin T, LDH, CK and inflammatory values (figure 2, Additional file 1: table S2 and figure S10). During clinical examination, this group recorded the lowest urine output values (0.58 ml/kg/h), 43% had delayed capillary refill time, and 11% had severe mottling (figure 2, Additional file 1: table S1). Almost 35% of patients required renal replacement therapy (RRT), and 91%

developed AKI. They also needed the highest vasopressor dose, and had the highest mortality (ICU HR 1.80 [1.33-2.42], 30-day HR 1.66 [1.23-2.23], and 90-day HR 1.58 [1.17-2.12]). High liver enzymes, LDH, and troponin drove predictions towards cluster membership (Additional file 1: figure S14).

One-hundred forty-four patients with low co-morbidity were in cluster 4, having been admitted primarily with neurosurgical or neurological (40%) and trauma-related (17%) diagnoses (figure 3, Additional file 1: table S6). These patients presented with relatively lower APACHE IV and SAPS II scores (62 and 41), and had the shortest ICU stay (2.8 days), lowest AKI incidence (31%), and significantly lower risk of mortality (ICU, 30-day, and 90-day (HR 0.57 [0.48-0.68], 0.75 [0.63-0.89], 0.66 [0.55-0.78], respectively). High albumin, hemoglobin, and pO2 values predicted membership, while high fibrinogen and potassium predicted non-membership (Additional file 1: figure S15).

For cluster 5 (n=290), the main admission diagnoses were respiratory failure (19%), cardiac arrest (12%), neurological causes (11%), or trauma (12%). Patients in this cluster had a medium co-morbidity profile, the longest mean ICU stay (7.7 days), and high rates of delayed CRT at admission (35%). Worsening respiratory condition was frequent (14%), and COPD was a common co-morbidity (17%). High thrombocytes and potassium predicted cluster membership, while high values for pO_2, creatinine, bilirubin, and phosphate drove predictions towards non-membership for some patients (Additional file 1: figure S16). The 95 patients assigned to cluster 6 primarily suffered from sepsis or respiratory infection, having been admitted to the ICU with respiratory (41%) or gastrointestinal diagnoses (20.0%). They also had high rates of distributive shock (38%), with physical examination showing a high cardiac index in 59% of patients, as well as high respiratory rates and heart rates. High values for fibrinogen, creatinine, urea, and CRP drove predictions towards cluster membership

(Additional file 1: figure S17). Patients in these clusters were not at increased or reduced risk of mortality.

Discussion

In this study, we set out to identify patient sub-phenotypes of clinical relevance using time-series laboratory, co-morbidity and clinical examination data. To do this, we compared four different clustering approaches. With a deep clustering algorithm, we identified six sub-phenotypes that capture differences in morbidity and in commonly measured clinical variables. In addition, these sub-phenotypes differed significantly in relevant clinical events rates (such as need for RRT and use of vasopressors) as well as mortality, with one group at low mortality risk, and two at higher mortality risk compared to average.

As the first study evaluating the combination of clustering such a broad range of ICU data including demographic, co-morbidity, clinical examination, and laboratory data with feature importance analysis to characterize patient sub-phenotypes in a minimally-selected ICU population, we draw several conclusions.

We found that traditional clustering algorithms such as k-means and HC were highly susceptible to the variation in data and outliers generated by the inclusion of a large number of laboratory variables. Previous studies using these algorithms with other, mostly low-dimensional datasets had not reported this issue, and do not show the large imbalance in cluster size we observed for k-means and HC in this study [2,4,14]. Interestingly, we identified the same issue with HC-DTW despite the use of a computationally-expensive technique like DTW to make time-series sequences of different length more uniform, and therefore suitable for clustering. Deep embedded clustering, on the other hand, provided a balanced patient distribution across clusters using the extracted features alone. Despite the large amount of data and the moderate cohort size, we managed to achieve stable clusters, and use these labels to train a classifier to identify the features driving cluster membership predictions.

The findings from variable importance analysis provided interesting, adjuvant data for the interpretation of the clusters identified during this analysis. Previous clustering analyses have based interpretation of the phenotypes found on differences in means of the variables measured [2,4], or by listing variables related to each cluster based on relative importance [6]. Here, we complemented the descriptive statistics of each cluster with SHAP values to establish the directionality of the association between high and low values of a variable and cluster membership. For example, membership of clusters with higher mortality risk was associated with increased ASAT and LDH, which are known biomarkers of myocardial ischemia [24,25,26,27] and have also been shown to be positively correlated with ICU mortality [17]. Likewise, patients were more likely to be in the cluster with the lowest mortality risk (cluster 4) when their albumin, hemoglobin, and pO2 values were higher, as well as when their liver function was better. These associations are supported by literature, including the addition of albumin measurements to APACHE scores [28,29,30,31].

The results of our study suggest that this clustering methodology is superior to most widely used approaches for clustering of critically ill patients for several reasons. Firstly, it can process time-series data, unlike k-means, HC, or even HC with dynamic time warping. A recent study using clustering analysis to define cardiovascular phenotypes suggested that incorporating serial measures to study transitions from one phenotype to another during ICU stay would provide additional insight to their analysis, which was limited to static variables [4]. Similarly, clustering studies on treatment response in critically ill patients would benefit substantially from processing time-series data, as opposed to one-time treatment administration [6].

Secondly, departing from a “minimally-selected” patient cohort, the six sub-phenotypes we identified differed significantly in relevant clinical outcomes which were also clinically recognizable and describable. Caution has been advised when interpreting the

results of clustering analyses, especially when identifying “novel” sub-phenotypes, and rightfully so [3]. Clustering algorithms will inevitably partition patients into clusters, and, as with most unsupervised machine learning techniques, it remains hard to establish what variables drove this partition. It was with this in mind that, in this study, rather than trying to identify “new” sub-phenotypes, we set out to cluster patients in a way which maximizes the difference in outcomes and looked at the admission and discharge diagnoses to characterize the clusters. In addition, our combination of clustering, training a classifier on the stable cluster labels, and use of SHAP helped “open the black box” to understand what variables drove the clustering.

Lastly, the inclusion of time-series data into clustering analyses can bring a wide range of benefits to studies aiming to characterize patient sub-phenotypes. It can enable the use of continuous hemodynamic monitoring and laboratory data to detect variations in sensitivity to myocardial ischemia or acute kidney injury, or to identify groups with differential treatment responses over time. To support this we have made the code used in this study open-source at https://github.com/J1C4F8/SICS_DEC and encourage further replication and validation of the algorithm and the findings of this study.

Our study also included some limitations. First, while the goal of the SICS-I study was to collect data from an “unselected-as-possible” clinical population, inclusion criteria did apply which may account for some selection bias [11,12]. For example, patients expected to stay in the ICU for less than 24 hours, due to discharge or extremely dire prognosis, were not included. Second, the six sub-phenotypes identified do not represent an exhaustive classification of critically ill patient subtypes. Information user in previous studies like end-of-life desires, need for life-sustaining therapies, and post-discharge care needs would complement our analysis, which did not include any variables of the disease course after the ICU except for mortality [2]. Lastly, external validation of the six identified clusters in an

independent cohort is necessary. Future studies with larger datasets should look to validate and replicate our findings, and address the possibility of patients belonging to multiple clusters.

Conclusion

Our analysis of a cohort with 743 ICU patients, based on a combination of clustering and feature importance analysis of co-morbidity, clinical examination, and laboratory data identified six patient sub-phenotypes with varying outcomes and clinical trajectories. This machine learning methodology, which we made publicly available, is a possible solution to challenges previously encountered by clustering analyses, and may help unravel patient heterogeneity in critical care.

Data Availability

The datasets generated during and/or analysed during the current study are not publicly available due to containing sensitive patient information but are available from the corresponding author on reasonable request.

Code Availability

Program code available under the GitHub repository https://github.com/J1C4F8/SICS_DEC.

Acknowledgments

None.

Authors' contributions

JCF, GY, MW, and IvdH contributed to study concept and design, analysis and interpretation, and prepared the first draft of the manuscript.

BH, TK, and RE contributed to the original study design and data collection.

JCF, GY, MvdG contributed to processing the data and implementing predictive models.

FK and AE carefully revised the manuscript

Source of funding

This research project received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing Interests

The authors declare that they have no competing interests.

Ethical Approval and Consent to participate

This study was approved by the local institutional review board (Medisch Ethische Toetsingscommissie (METc) of the UMCG, M15.168207).

Consent for publication

Not applicable.

References

1. Castela Forte J, Perner A, van der Horst ICC: The use of clustering algorithms in critical care research to unravel patient heterogeneity. Intensive Care Med. 2019, 45:1025–1028. https://doi.org/10.1007/s00134-019-05631-z

2. Vranas KC, Jopling JK, Sweeney TE, Ramsey MC, Milstein AS, Slatore CG, et al: Identifying Distinct Subgroups of Intensive Care Unit Patients: a Machine Learning Approach. Crit Care Med. 2017, 45(10): 1607–1615. doi: 10.1097/CCM.0000000000002548

3. van Smeden M, Harrell FE, Dahly DL. Novel diabetes subgroups. Lancet Diabetes Endocrinol. 2018, 6(6):439-440. https://doi.org/10.1016/S2213-8587(18)30124-4

4. Geri G, Vignon P, Aubry A, et al: Cardiovascular clusters in septic shock combining clinical and echocardiographic parameters: a post hoc analysis. Intensive Care Med. 2019, 45:657–667. https://doi.org/10.1007/s00134-019-05596-z

5. Sinha P, Delucchi KL, Thompson BT, McAuley DF, Matthay MA, Calfee CS: Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. Intensive Care Med. 2018, 44(11):1859–1869. https://doi.org/10.1007/s00134-018-5378-3

6. Wu W, Bang S, Bleecker ER, Castro M, Denlinger L, Erzurum SC, et al: Multiview Cluster Analysis Identifies Variable Corticosteroid Response Phenotypes in Severe Asthma. Am J Respir Crit Care Med. 2019, 199(11): 1358-1367. https://doi.org/10.1164/rccm.201808-1543OC.

7. Bhatraju PK, Zelnick LR, Herting J, Katz R, Mikacenic C, Kosamo S, et al: Identification of acute kidney injury subphenotypes with differing molecular signatures and responses to vasopressin therapy. Am J Respir Crit Care Med. 2019, 199(7):863–872. https://doi.org/10.1164/rccm.201807-1346OC

8. Nielsen AB, Thorsen-Meyer H-C, Belling K, et al: Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish National Patient Registry and electronic patient records. Lancet Digit Health. 2019, 1:e78-e89. https://doi.org/10.1016/S2589-7500(19)30024-X

9. Thorsen-Meyer H-C, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck, J et al: Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digital Health. 2020, 2(4):e179-e191. https://doi.org/10.1016/S2589-7500(20)30018-2

10. Shickel B, Loftus TJ, Adhikari L et al: DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning. Sci Rep. 2019, 9:1879. https://doi.org/10.1038/s41598-019-38491-0

11. Hiemstra B, Eck RJ, Koster G, et al: Clinical examination, critical care ultrasonography and outcomes in the critically ill: cohort profile of the Simple Intensive Care Studies-I. BMJ Open. 2017, 7:e017170. doi: 10.1136/bmjopen-2017-017170

12. Hiemstra B, Koster G, Wiersema R, et al: The diagnostic accuracy of clinical examination for estimating cardiac index in critically ill patients: the Simple Intensive Care Studies-I. Intensive Care Med. 2019, 45:190–200. doi: 10.1007/s00134-019-05527-y

13. Alkozai EM, Mahmoodi BK, Decruyenaere J, Porte RJ, Lansink-Hartgring AO, Lisman T, et al: Systematic Comparison of Routine Laboratory Measurements With In-Hospital Mortality: ICU-Labome, a Large Cohort Study of Critically Ill Patients. Clin Chem Lab Med. 2018, 56(7):1140-1151. doi: 10.1515/cclm-2016-1028.

14. Azoulay E, Zafrani L, Mirouse A, et al: Clinical phenotypes of critically ill COVID-19 patients. Intensive Care Med. 2020, 46, 1651–1652. https://doi.org/10.1007/s00134-020-06120-4

15. Xie J, Girshick R, Farhadi A: Unsupervised Deep Embedding for Clustering Analysis. International Conference on Machine Learning (ICML). 2016, 48: 478–487. arXiv:1511.06335v2

16. Petitjean P, Ketterlin A, Gancarski P. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition. 2011, 44: 678–693.

17. Halkidi M, Batistakis I, Vazirgiannis M: On clustering validation techniques. J Intell Inf Syst. 2001, 17(2/3):107–14. https://doi.org/10.1023/A:1012801612483

18. Jain AK: Data clustering: 50 years beyond *K*-means. Pattern Recognit Lett. 2010, 31(8):651–666.

19. Handl J, Knowles J, Kell DB: Computational cluster validation in post-genomic data analysis. Bioinformatics. 2005, 21(15):3201–3212. https://doi.org/10.1093/bioinformatics/bti517

20. Hennig C: Cluster-wise assessment of cluster stability. Comput Stat Data Analysis. 2007, 52:258–71. https://doi.org/10.1016/j.csda.2006.11.025

21. Meila M: Comparing clusterings - an information based distance. J Multivar Anal. 2007, 98:873–895. https://doi.org/10.1016/j.jmva.2006.11.013

22. Chen T, Guestrin G: XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 785–794**.** https://doi.org/10.1145/2939672.2939785

23. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018, 2(10):749-760. doi: 10.1038/s41551-018-0304-0

23. Auer J: What does the liver tell us about the failing heart? European Heart Journal. 2013, 34(10): 711–714. doi: https://doi.org/10.1093/eurheartj/ehs440

24. Gao M, Cheng Y, Zheng Y, Zhang W, Wang L, Qin L: Association of serum transaminases with short- and long-term outcomes in patients with ST-elevation myocardial

infarction undergoing primary percutaneous coronary intervention. BMC Cardiovasc Disord. 2017, 28;17(1):43. doi: 10.1186/s12872-017-0485-6.

25. Naschitz JE, Slobodin G, Lewis RJ, Zuckerma E, Yeshurun D: Heart diseases affecting the liver and liver diseases affecting the heart. American Heart Journal. 2000, 140(1): 111–120. doi:10.1067/mhj.2000.107177

26. Schmiechen NJ, Han C, Milzman DP: ED use of rapid lactate to evaluate patients with acute chest pain. Ann Emerg Med. 1997, 30(5):571-7. Doi: 10.1016/s0196-0644(97)70071-4

27. Corti MC, Guralnik JM, Salive ME, Sorkin JD: Serum-albumin level and physical-disability as predictors of mortality in older persons. JAMA. 1994, 272: 1036–1042. doi:10.1001/jama.1994.03520130074036

28. Goldwasser P, Feldman J: Association of serum albumin and mortality risk. J Clin Epidemiol. 1997, 50: 693–703.

29. Vincent J-L, Dubois M-J, Navickis RJ, Wilkes MM: Hypoalbuminemia in Acute Illness: Is There a Rationale for Intervention? A Meta-Analysis of Cohort Studies and Controlled Trials. Ann Surg. 2003, 237(3): 319–334. doi: 10.1097/01.SLA.0000055547.93484.87

30. Vincent JL, Baron J, Reinhart K, Gattinoni L, Thijs L, Webb A, et al: Anemia and Blood Transfusion in Critically Ill Patients. JAMA. 2002, 288(12):1499–1507. doi:10.1001/jama.288.12.1499

Additional files

File name: Additional file 1 (PDF, 1.3 Mb)

File title: Supplementary methods, figures and tables

File description: Supplementary file with additional details on dynamic time warping and deep embedded clustering methodologies, tables and figures representing clustering results, and accuracy and variable importance of the XGBoost classifier.

510 **Table 1.** Clinical characteristics for the 743 patients, including patient characteristics, clinical
511 examination data, co-morbidities and medical history, and outcomes.
512

Patient characteristics	
Age (years)	62 [61, 63]
Gender (% female)	63.3
APACHE IV Score	76.9 [74.8, 79.2]
SAPS II Score	46.9 [45.7, 48.0]
BMI	26.7 [26.4, 27.1]
Surgical patient (%)	35.0
Previous admission to ICU (%)	11.6
Clinical examination	
Cardiac index > 2.2 (%)	42.7
Mottling (% with severe, >4)	3.1
Mechanically ventilated at admission (%)	61.8
Worsened respiratory condition after 24 hours (%)	12.4
Urine output in previous 6 hours (ml/kg/h)	0.9 [0.84,0.95]
CRT prolonged (%)	29.9
EMV score	11.27 [10.9,11.63]
Co-morbidities and medical history	
History of CVD (%)	5.0
History of CKD (%)	6.9
History of cirrhosis (%)	3.4
History of COPD (%)	12.4
Previous dialysis (%)	1.4
History of diabetes (%)	20.3
History of hematological malignancy (%)	4.0
History of metastatic disease (%)	3.6
History of myocardial infarction (%)	8.2
History of respiratory failure (%)	4.9
Admission diagnosis by organ system[#]	
Cardiovascular	32.8
Gastrointestinal	14.2
Genito-urinary	1.2
Haematological	1.4
Metabolic	2.3
Musculoskeletal/skin	0.9
Neurological	15.0
Respiratory	19.5
Transplant	5.1
Trauma	7.7
Outcomes[#]	
In-ICU mortality (%)	19.4

30-day mortality (%)	22.3
90-day mortality (%)	27.6
No AKI (%)	79.7
AKI stage 1 (%)	59.8
AKI stage 2 or 3 (%)	44.3
Required vasoactive medication (%)	48.6
Any type of shock (%)	51.6
Required RRT (%)	10.1
ICU length of stay (days)	6.0 [5.4,6.6]

# Denotes variables not included as input for clustering analysis

Table 2. Mortality rates, and hazard ratios for mortality per cluster

Outcome/cluster		Hazard ratio with 95%CI	p-value
In-ICU Mortality	%		
Cluster 1	19.1	0.99 [0.77-1.26]	0.916
Cluster 2	**30.0**	**1.55 [1.26-1.91]**	**<0.001**
Cluster 3	**34.8**	**1.80 [1.33-2.42]**	**<0.001**
Cluster 4	**11.1**	**0.57 [0.48-0.68]**	**<0.001**
Cluster 5	17.9	0.92 [0.81,1.06]	0.254
Cluster 6	16.8	0.87 [0.70,1.07]	0.191
30-day mortality			
Cluster 1	19.1	0.85 [0.67-1.10]	0.218
Cluster 2	**31.0**	**1.39 [1.12-1.71]**	**0.002**
Cluster 3	**37.0**	**1.66 [1.23-2.23]**	**<0.001**
Cluster 4	**16.7**	**0.75 [0.63-0.89]**	**0.001**
Cluster 5	22.4	1.00 [0.86-1.15]	0.972
Cluster 6	**17.9**	**0.80 [0.65-0.99]**	**0.042**
90-day mortality			
Cluster 1	26.5	0.96 [0.75-1.23]	0.763
Cluster 2	**36.0**	**1.30 [1.06-1.61]**	**0.012**
Cluster 3	**43.5**	**1.58 [1.17-2.12]**	**0.002**
Cluster 4	**18.1**	**0.66 [0.55-0.78]**	**<0.001**
Cluster 5	27.6	1.00 [0.88-1.15]	0.996
Cluster 6	26.3	0.95 [0.77-1.18]	0.673

Hazard ratios were compared using the log-rank test.

Figure 1. Schematic overview of the different steps in the analysis. Patient selection, integration of different data sources, data processing with feature extraction (FE) or dynamic time warping (DTW), comparison of the four clustering algorithms, selection of the best algorithm based on patient distribution and internal validity measures, training of the classifier for attributing true labels to the clusters and calculating feature importance with SHAP, and cluster characterization based on input data from diagnoses, feature importance, and differences in outcomes including mortality, AKI, and other clinical events.

Figure 2. **Heatmap of patient characteristics, clinical examination and co-morbidity data per cluster**. Bars on the right show the colour scale representing the proportion of patients with each characteristic regarding demographics, clinical examination, and co-morbidities. For continuous variables, such as SBP or urine output, it represents a scaled value from highest cluster mean (1.0) to lowest cluster mean (0.0).

Figure 3. Heatmap of outcomes and clinical end-points per cluster. Bars on the right show the colour scale representing the proportion of patients within the cluster with the outcome (upper panel) or the discharge diagnosis (lower panel).

Figures

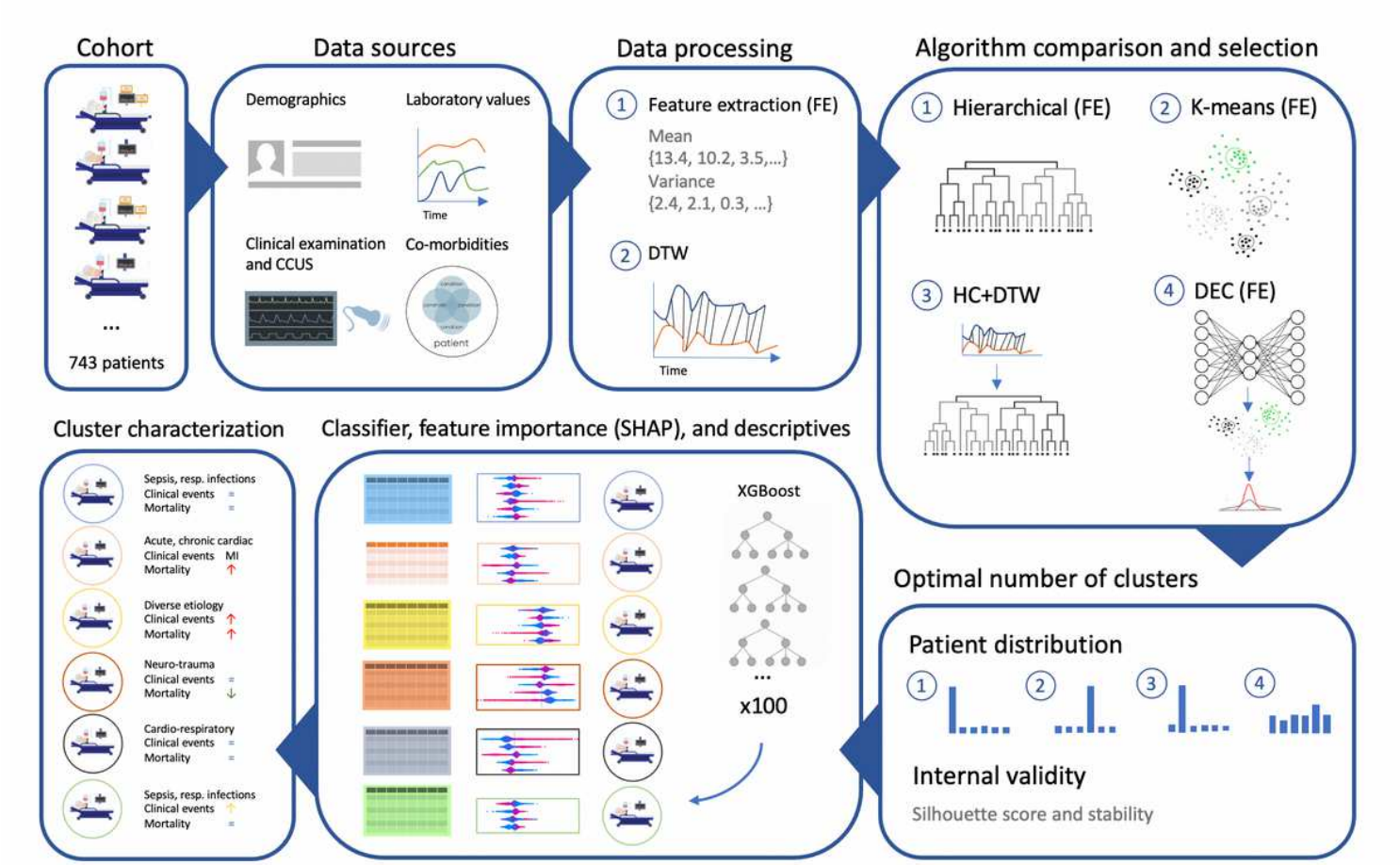

Figure 1

Schematic overview of the different steps in the analysis. Patient selection, integration of different data sources, data processing with feature extraction (FE) or dynamic time warping (DTW), comparison of the four clustering algorithms, selection of the best algorithm based on patient distribution and internal validity measures, training of the classifier for attributing true labels to the clusters and calculating feature importance with SHAP, and cluster characterization based on input data from diagnoses, feature importance, and differences in outcomes including mortality, AKI, and other clinical events.

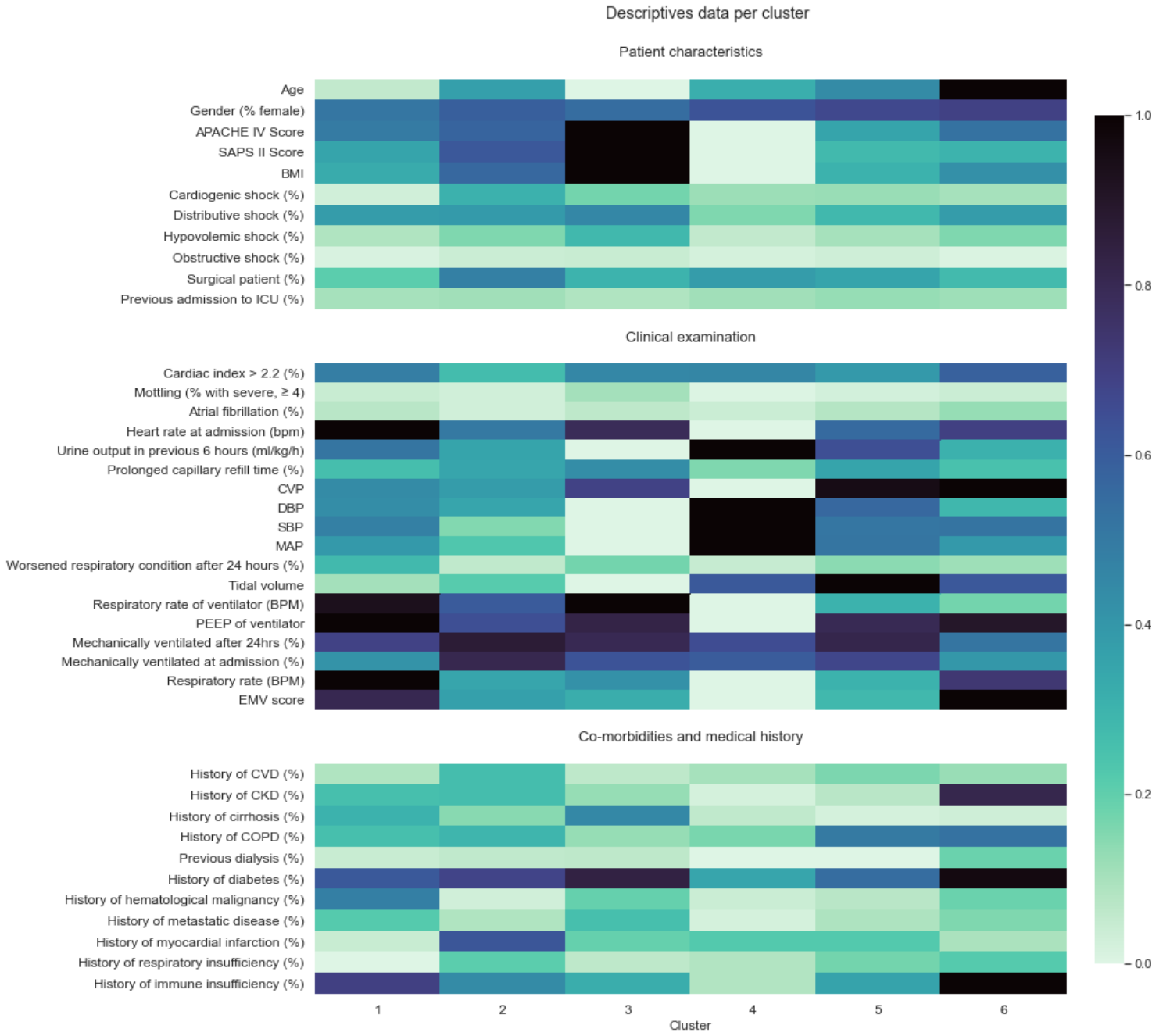

Figure 2

Heatmap of patient characteristics, clinical examination and co-morbidity data per cluster. Bars on the right show the colour scale representing the proportion of patients with each characteristic regarding demographics, clinical examination, and co- morbidities. For continuous variables, such as SBP or urine output, it represents a scaled value from highest cluster mean (1.0) to lowest cluster mean (0.0).

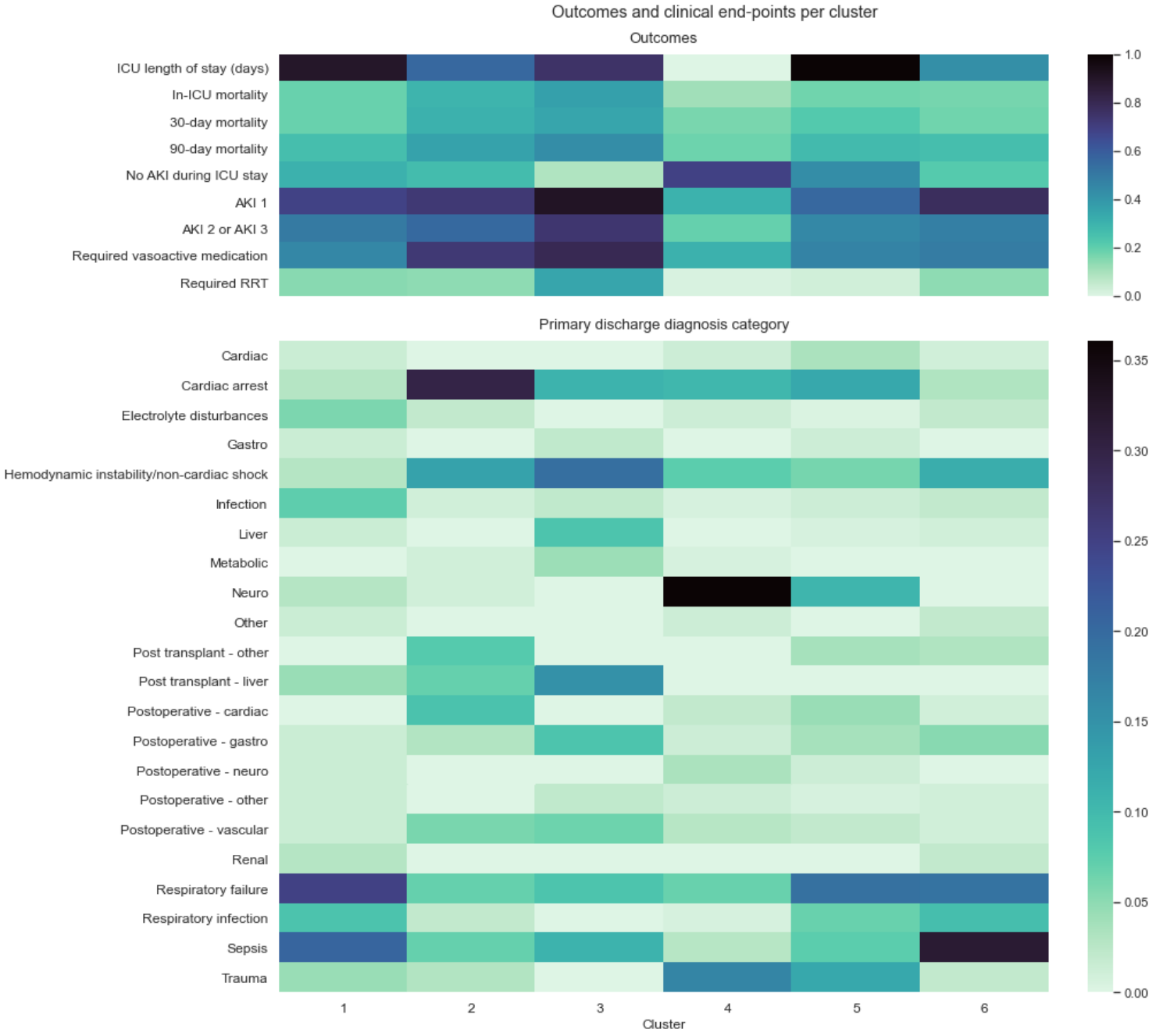

Figure 3

Heatmap of outcomes and clinical end-points per cluster. Bars on the right show the colour scale representing the proportion of patients within the cluster with the outcome (upper panel) or the discharge diagnosis (lower panel).

Supplementary Files

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

- CastelaForteEtAlSupplementarymaterial.pdf