

EDITORIAL



The use of clustering algorithms in critical care research to unravel patient heterogeneity

José Castela Forte^{1,2} , Anders Perner³ and Iwan C. C. van der Horst^{2*}

© 2019 Springer-Verlag GmbH Germany, part of Springer Nature

Critically ill patients constitute the most heterogeneous population in the hospital, with the highest rates of acute and chronic multimorbidity. Daily, two critically ill patients are admitted to the ICU with the same syndrome-based diagnosis, receive similar treatment, and yet have diametrically opposite outcomes. Knowledge regarding disease mechanisms and effectiveness of interventions that could explain this is lacking, but it is increasingly clear that syndrome-based patient categorisation inevitably leads to grouping of patients with different risk profiles, responses to interventions, and outcomes. In septic shock, for instance, cardiovascular sub-phenotypes are insufficiently characterised, which compromises the effectiveness of hemodynamic support [1].

Recently, Geri and colleagues identified five different cardiovascular sub-phenotypes in septic shock patients, using clinical and echocardiographic parameters [1]. Two reflected response to interventions (“well resuscitated”, “still hypovolemic”), while three characterised cardiac function (“left ventricle systolic dysfunction”, “hyperkinetic”, “right ventricle failure”). To personalise and improve treatments, patients must be clustered into novel sub-phenotypes based on clinically objective parameters reflecting disease mechanisms or treatment responses, rather than admission diagnoses. Therefore, we highly commend the authors for taking an innovative, machine learning (ML) based approach in investigating an established problem in critical care and providing a detailed methodological description of

how to incorporate clustering analyses into critical care research in a clinically relevant way.

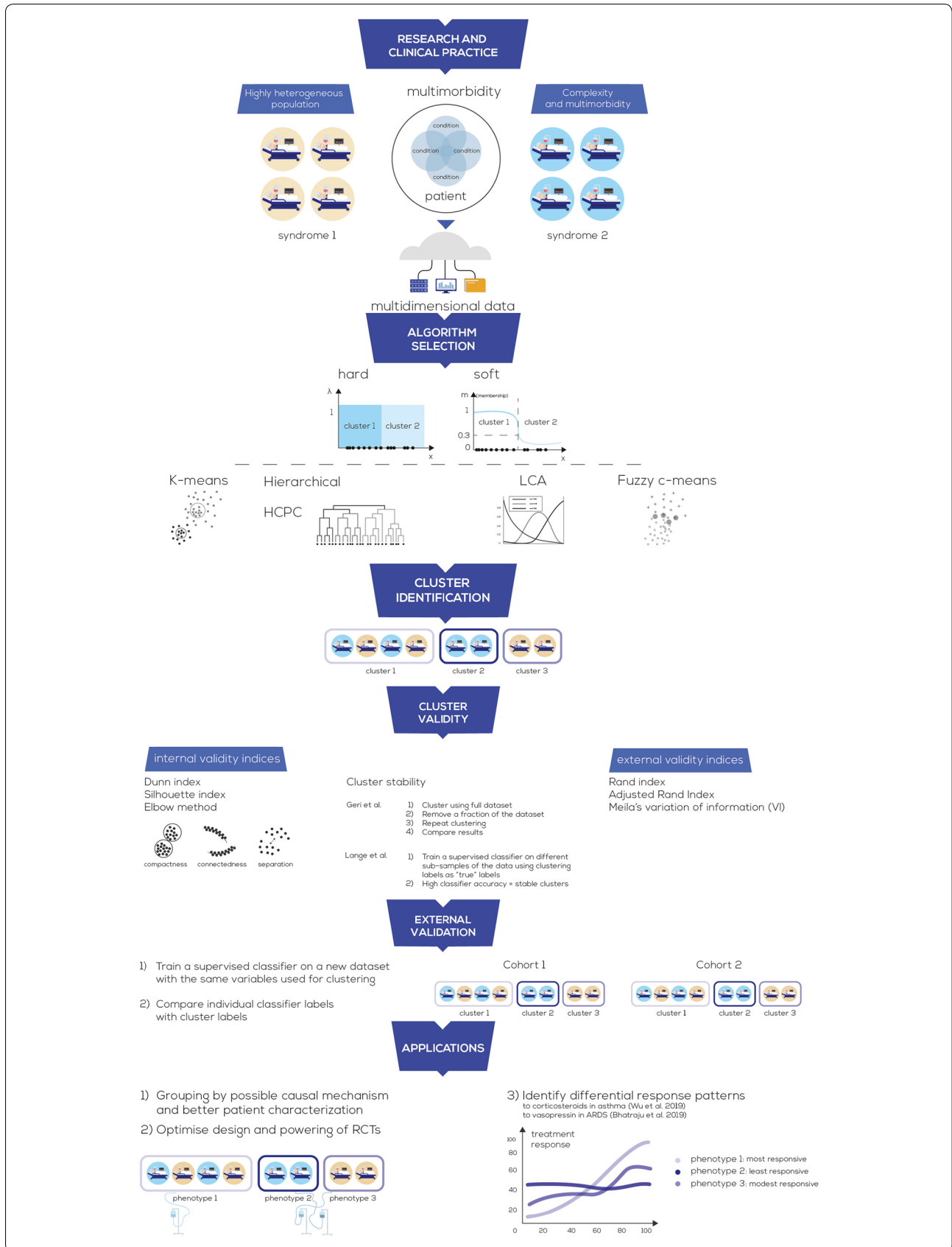
ML applications in critical care are booming and have fuelled ground-breaking research using different methodologies [2–7]. A recent overview by Sinha and Calfee of the advances in identifying homogeneous subgroups and phenotypes in acute respiratory distress syndrome (ARDS), which show divergent clinical characteristics, outcomes, and differential response patterns, highlights how this can add to clinicians’ existing knowledge [7]. With the ever-growing stream of data collected in ICUs generating increasingly high-dimensional and complex datasets, selecting the right analytical tools is more crucial than ever. Figure 1 provides a schematic explanation of how clustering algorithms (a type of unsupervised learning algorithm where no labels are known a priori but get assigned based on inherent similarities between data points) can be applied in exploratory data analyses to identify relevant patient sub-groups.

Geri and colleagues used the hierarchical clustering on principle components (HCPC) algorithm, which is a subtype of hierarchical clustering (HC) [8, 9]. In HC, clusters are visualised in dendrograms that split at different levels based on the similarity between data points [8]. HCPC differs from other HC algorithms in that a principal component analysis (PCA) is conducted before clustering to reduce data dimensionality: after PCA is performed, data are reduced to a few continuous variables (principal components), which contain the most important information in the data. Using these transformed data for clustering can help improve cluster stability in multi-dimensional datasets with multiple continuous variables [9]. *K*-means is another often used algorithm, where data are initially divided into a user-defined number of clusters, which is repeatedly updated until the distance (i.e. difference) between points within a cluster is minimized [8]. Another algorithm particularly successfully applied

*Correspondence: i.c.c.van.der.horst@umcg.nl

² Department of Critical Care, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Full author information is available at the end of the article



(See figure on previous page.)

Fig. 1 Panorama of clustering analyses using heterogeneous ICU data. Clustering analysis of heterogeneous and complex ICU data starts with the selection of a hard or soft algorithm, based on the expected overlap between sub-phenotypes and the size of the dataset. Once clusters have been identified, the validity of the findings must be assessed. Internal validity is assessed by using validity indices such as the Silhouette or Dunn indices. These indices combine multiple measures (compactness, connectedness, and separation) to provide an estimate of whether the structure of the clustering is appropriate for the data. Cluster stability can be assessed in distinct ways, two of which are shown. Lastly, external validity measures such as Rand and Meila's VI indices can confirm whether the clustering results match the a priori expected data structure. To determine whether clustering findings are generalizable, external validation using a similar dataset must follow. Validated findings can be used to better characterize patient sub-phenotypes in terms of clinical characteristics and outcomes, to optimise the design and powering of randomised trials by generating more homogeneous groups, and to identify differential treatment patterns in these randomised trials results by determining the variability in the direction and magnitude of individual treatment effects, both beneficial and adverse. Based on this, patients can be classified as most, moderate, and least responsive according to how well a treatment is expected to work

within critical care research is latent class analysis (LCA) [3, 5, 6]. LCA is an established, model-based statistical technique, that defines the best fitting models for data assumed to contain several unobserved groups. LCA is particularly useful for assessing heterogeneous treatment effects, and ML-based applications may allow its use with increasingly complex and larger datasets [10].

With these and countless other algorithms available, it is important to distinguish between hard and soft clustering algorithms. In hard clustering, each data point can belong to one cluster only, whereas soft clustering algorithms such as fuzzy *c*-means assign data points a membership probability of belonging to one or more clusters [8]. Given the known limitations of hard clustering techniques in dealing with datasets containing static and dynamic variables, soft clustering techniques could come to play a pivotal role in studying overlapping disease mechanisms in heterogeneous populations. For example, replicating Geri et al.'s analysis using soft clustering would provide better patient characterization by allowing mixed clusters including "well resuscitated" and "hyperkinetic" phenotypes, instead of having two and three clusters separately describing treatment response and cardiac function, respectively.

Once clusters have been defined, internal validity (including cluster stability) and external clustering validity have to be tested [11]. Internal validity is assessed by verifying whether the structure of the clustering is intrinsically appropriate for the data: that is, if data points are, simultaneously, similar within the same cluster, and as distinct as possible from those in other clusters. This will define what the ideal number of clusters for the data is, and can be done through indices such as Silhouette and Dunn or by determining cluster stability [8, 12]. This measure represents the cluster variation over different sub-samples of the same input data, and is determined by comparing changes in clusters composition using first the full dataset and then only a fraction thereof, or by training a supervised classifier on different data sub-samples [1, 13]. External validity assesses whether clustering

results match the a priori expected data structure by comparing the clustering output to a given "correct" clustering when "true" class labels are available (Fig. 1) [14]. This is crucial because clustering algorithms will inevitably partition data into clusters irrespective of whether any clusters are indeed present [8]. Lastly, as for all ML-based exploratory studies, external validation should be done to determine the generalizability of the findings. This can be done by using the most relevant variables of a clustering analysis to train a classifier on a new dataset, and then assessing whether individuals are classified into the same groups as during clustering [4].

In conclusion, increasingly flexible and sophisticated clustering techniques are available, which can allow for analyses of higher-dimensional datasets that help better characterize patients, disease mechanisms, and heterogeneous treatment response patterns (Fig. 1). However, before these findings can truly inform the design of multicentre, international prospective studies and trials, efforts to increase the interpretability of the findings are essential. For instance, the Interpretable Clustering via Optimal Trees algorithm developed by Bertsimas et al. provides a clear, tree-based representation of the most important variables and the respective thresholds which led to cluster formation [15].

Author details

¹ Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands. ² Department of Critical Care, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. ³ Department of Intensive Care, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.

Compliance with ethical standards

Conflicts of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 21 March 2019 Accepted: 25 April 2019
 Published online: 6 May 2019

References

1. Geri G, Vignon P, Aubry A, Fedou A-L, Charron C, Silva S et al (2019) Cardiovascular clusters in septic shock combining clinical and echocardiographic parameters: a post-hoc analysis. *Intensive Care Med*. <https://doi.org/10.1007/s00134-019-05596-z>. **(Epub ahead of print)**
2. Vranas KC, Jopling JK, Sweeney TE, Ramsey MC, Milstein AS, Slatore CG et al (2017) Identifying distinct subgroups of ICU patients: a machine learning approach. *Crit Care Med* 45(10):1607–1615. <https://doi.org/10.1097/CCM.0000000000002548>
3. Calfee CS, Delucchi K, Sinha P, Matthay MA, Hackett J, Shankar-Hari M et al (2018) Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med* 6(9):691–698. [https://doi.org/10.1016/S2213-2600\(18\)30177-2](https://doi.org/10.1016/S2213-2600(18)30177-2)
4. Wu W, Bang S, Bleecker ER, Castro M, Denlinger L, Erzurum SC et al (2019) Multiview cluster analysis identifies variable corticosteroid response phenotypes in severe asthma. *Am J Respir Crit Care Med*. <https://doi.org/10.1164/rccm.201808-1543OC>. **(Epub ahead of print)**
5. Bhatraju PK, Zelnick LR, Herting J, Katz R, Mikacenic C, Kosamo S et al (2019) Identification of acute kidney injury subphenotypes with differing molecular signatures and responses to vasopressin therapy. *Am J Respir Crit Care Med* 199(7):863–872. <https://doi.org/10.1164/rccm.201807-1346OC>
6. Sinha P, Delucchi KL, Thompson BT, McAuley DF, Matthay MA, Calfee CS (2018) Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med* 44(11):1859–1869. <https://doi.org/10.1007/s00134-018-5378-3>
7. Sinha P, Calfee CS (2018) Phenotypes in acute respiratory distress syndrome. *Curr Opin Crit Care* 25(1):12–20. <https://doi.org/10.1097/MCC.0000000000000571>
8. Jain AK (2010) Data clustering: 50 years beyond *K*-means. *Pattern Recognit Lett* 31(8):651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
9. Husson F, Josse J, Pagès J (2010) Principal component methods—hierarchical clustering—partitional clustering: why would be need to choose for visualizing data? *Tech Rep* 1–17. Retrieved from: http://factominer.free.fr/more/HCPC_husson_josse.pdf
10. Jo B, Findling RL, Wang C-P, Hastie TJ, Youngstrom EA, Arnold LE et al (2017) Targeted use of growth mixture modeling: a learning perspective. *Stat Med* 36(4):671–686. <https://doi.org/10.1002/sim.7152>
11. Halkidi M, Batistakis I, Vazirgiannis M (2001) On clustering validation techniques. *J Intell Inf Syst* 17(2/3):107–145. <https://doi.org/10.1023/A:1012801612483>
12. Handl J, Knowles J, Kell DB (2005) Computational cluster validation in post-genomic data analysis. *Bioinformatics* 21(15):3201–3212. <https://doi.org/10.1093/bioinformatics/bti517>
13. Lange T, Roth V, Braun M, Buhmann J (2004) Stability-based validation of clustering solutions. *Neural Comput* 16:1299–1323. <https://doi.org/10.1162/089976604773717621>
14. Meila M (2007) Comparing clusterings—an information based distance. *J Multivar Anal* 98:873–895. <https://doi.org/10.1016/j.jmva.2006.11.013>
15. Bertsimas D, Orfanoudaki A, Wiberg A (2018) Interpretable clustering via optimal trees. <https://arxiv.org/pdf/1812.00539.pdf> [stat.ML]