Letter to the Editor

JOURNAL OF HEPATOLOGY

Enhancing clustering validity in MELD subtype analysis: A call for robust statistical methods

To the Editor:

It is crucial to perform error-free calculations to ensure the accuracy and reliability of any analysis. The selection of appropriate statistical tools plays a significant role; using linear models on non-linear data or employing parametric approaches on non-parametric data can lead to distorted outcomes and misleading interpretations. Linear models assume a straightforward relationship, which may overlook the complexities present in non-linear datasets, while parametric methods often rely on specific distributional assumptions that may not be valid for all data. By choosing the right analytical techniques that align with the data's inherent characteristics, researchers can achieve more precise results and derive valid insights. Consequently, a thoughtful approach to tool selection is indispensable for maintaining the integrity of statistical analysis and making well-informed decisions based on the findings.

Takefuji emphasized the importance of recognizing non-linear relationships in biological data, verifying critical model assumptions, and ensuring model adequacy. He also noted that methods for determining feature importance, selection, and reduction often lack definitive ground truth values, advocating instead for non-linear and non-parametric approaches when addressing complex biological data patterns.[1]

Rosenstengle *et al.* explored the variation in intention-to-treat survival based on model for end-stage liver disease (MELD) subtypes, emphasizing that not all models created for end-stage liver disease are equal.[2] In their study, they employed K-means clustering analysis to categorize patients into MELD-Br, MELD-INR, or MELD-Cr groups, depending on the dominant variable contributing to their MELD scores. Their analysis included a total of 39,897 patients who were listed with decompensated cirrhosis and met the specified inclusion criteria.[2]

This letter highlights critical concerns regarding the use of K-means for clustering analysis, particularly in the context of complex non-linear data. K-means clustering relies on linear and parametric assumptions without first verifying that the data exhibit these characteristics. This oversight can lead to misleading outcomes and erroneous conclusions.[3–8] Rosenstengle *et al.* should acknowledge the importance of validating their data thoroughly and consider employing non-linear and non-parametric methods to enhance the robustness and accuracy of their findings.

Linear and parametric methods for clustering validation can distort outcomes when applied to non-linear and non-parametric data. For instance, K-means assumes that clusters are convex and separable with linear boundaries, performing best with linearly separable data. When clusters have non-linear shapes, these methods may misrepresent the actual structure. K-means might inaccurately group points solely based on proximity to centroids, leading to ineffective clustering in complex distributions. Parametric methods often require specific forms of data distribution, such as normality, which can result in misleading outcomes if the actual data does not conform. For example, K-means applied to datasets with elongated or irregular clusters may lead to poor assignments, misclassifying points that are close to a centroid but do not belong to the same cluster. Such oversimplifications can result in a significant underestimation of the data's complexity and heterogeneity.

This paper advocates the adoption of robust, non-linear, and non-parametric statistical methods for evaluating clustering quality. Measures such as the Silhouette value and the Davies-Bouldin index are essential for assessing clustering performance, while the Gap statistic plays a critical role in determining the optimal number of clusters. The Silhouette score, which ranges from -1 to 1, measures how similar an object is to its own cluster compared to other clusters, with higher values indicating better-defined clusters.[9] The Davies-Bouldin index, ranging from 0 to infinity, quantifies cluster similarity, with lower values representing better clustering solutions by evaluating the ratio of within-cluster distances to between-cluster distances.[10] Finally, the Gap statistic provides a measure of the clustering structure's strength relative to random distributions, with values ranging from 0 to infinity; higher scores are indicative of more robust clustering.[11] By utilizing these metrics, researchers can achieve more reliable clustering assessments, yielding insights that are less sensitive to distributional assumptions.

Yoshiyasu Takefuji*

Musashino University, Data Science Department, Japan

*Corresponding author. Address: 3-3-3 Ariake, Koto-Ku, Tokyo 135-8181, Japan.

E-mail address: takefuji@keio.jp

Received 4 February 2025; Received in revised form 13 February 2025; Accepted 13 February 2025; Available online xxx

https://doi.org/10.1016/j.jhep.2025.02.023

Financial support

The author did not receive any financial support to produce this manuscript.

Conflict of interest

The author has no conflict of interest.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Yoshiyasu Takefuji completed this research and wrote this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2025.02.023.

References

[1] Takefuji Y. Revealing bias in feature importance through PLS-DA: a critical examination of machine learning applications in chronic liver disease. J Hepatol 2025. https://doi.org/10.1016/j.jhep.2024.12.021.

[2] Rosenstengle C, Serper M, Asrani SK, et al. Variation in intention-to-treat survival by MELD subtypes: all models created for end-stage liver disease are not equal. J Hepatol 2025;82(2):268–276. https://doi.org/10.1016/j.jhep.2024.08.006.

[3] Raykov YP, Boukouvalas A, Baig F, et al. What to do when K-means clustering fails: a simple yet principled alternative algorithm. PLoS One 2016;11(9):e0162259. https://doi.org/10.1371/journal.pone.0162259. Published 2016 Sep 26.

[4] Wongoutong C. The impact of neglecting feature scaling in k-means clustering. PLoS One 2024;19(12):e0310839. https://doi.org/10.1371/journal.pone.0310839. Published 2024 Dec 6.

[5] Chen YT, Witten DM. Selective inference for k-means clustering. J Mach Learn Res 2023;24:152.

[6] Shen H, Bhamidi S, Liu Y. Statistical significance of clustering with multidimensional scaling. J Comput Graph Stat 2024;33(1):219–230. https://doi.org/10.1080/10618600.2023.2219708.

[7] Burgard JP, Moreira Costa C, Schmidt M. Robustification of the k-means clustering problem and tailored decomposition methods: when more conservative means more accurate. Ann Oper Res 2024;339:1525–1568. https://doi.org/10.1007/s10479-022-04818-w.

[8] Demidenko E. The next-generation K-means algorithm. Stat Anal Data Min 2018;11(4):153–166. https://doi.org/10.1002/sam.11379.

[9] Shahapure KR, Nicholas C. Cluster quality analysis using silhouette score. In: 2020 IEEE 7th international conference on data science and advanced analytics (DSAA). Sydney, NSW, Australia; 2020. p. 747–748. https://doi.org/10.1109/DSAA49011.2020.00096.

[10] Ros F, Riad R, Guillaume S. PDBI: a partitioning Davies-Bouldin index for clustering evaluation. Neurocomputing 2023;528:178–199. https://doi.org/10.1016/j.neucom.2023.01.043.

[11] Arima C, Hakamada K, Okamoto M, et al. Modified fuzzy gap statistic for estimating preferable number of clusters in fuzzy k-means clustering. J Biosci Bioeng 2008;105(3):273–281. https://doi.org/10.1263/jbb.105.273.