Correspondence

Critical evaluation of feature importance assessment in proteomic analysis using skin microdialysis

To the Editor:

Hollstein et al proposed a method for obtaining fluids suitable for proteomic analyses and protein measurements, demonstrating that skin microdialysis can detect distinct immunologic patterns in chronic inflammatory skin diseases.[1] To enhance data integration, feature selection and identification of integration anchors were meticulously applied. Following this, the standard Seurat workflow was used to process the samples, which included principal component analysis (PCA) for identifying 20 relevant dimensions, as determined through an elbow plot. Subsequently, unsupervised clustering was conducted with a resolution of 0.5.[1]

Although Hollstein et al proposed an innovative method for proteomic analysis and protein measurement, their article raises critical concerns regarding the reliance on principal component analysis for assessing feature importance.[1] Principal component analysis is rooted in linear and parametric assumptions, which may not be suitable for all data sets.[2,3] Unless these assumptions are thoroughly investigated and found to hold true, it is imperative for Hollstein et al[1] to consider using nonlinear and nonparametric methods that can mitigate inherent biases in feature importance evaluation.

Although machine learning models generate valuable predictions backed by associated ground truth values for accuracy validation, assessments of feature importance lack such definitive references. This absence can introduce uncertainties in interpreting the relevance of features, potentially compromising the reliability of the findings. To strengthen their analysis, Hollstein et al[1] should consider using statistically robust methods, such as the Spearman correlation with accompanying P values or the Kendall tau with P values. Both approaches are nonlinear and nonparametric, making them well suited for yielding more dependable insights into the relationships among variables within complex data sets.[4,5] By integrating these methodologies, Hollstein et al[1] could enhance the integrity of their results and provide clearer, more meaningful conclusions from their research.

Furthermore, to accurately calculate the associations between the target and the features, 3 essential elements must be addressed: the underlying data distribution, the nature of the relationships between variables, and rigorous statistical validation of findings through P values. Given the absence of ground truth values for feature importance, a comprehensive approach that encompasses these considerations will be crucial for deriving meaningful and unbiased conclusions from their findings. By adopting such an approach, Hollstein et al[1] could significantly enhance the reliability of their results and make valuable contributions to the field of immunologic research in chronic inflammatory skin diseases.

In addition to these considerations, it is imperative to address issues of collinearity and interactions among features before any formal association calculations. By identifying multicollinearity using variance inflation factor analysis, Hollstein et al[1] could remove or appropriately adjust for correlated features. This step is essential to mitigate correlation-induced biases that can distort the interpretation of associations. By ensuring that collinearity and interaction effects are accounted for, Hollstein et al[1] would better position their study to uncover true relationships between the target and features, ultimately advancing the rigor and validity of their research.

DISCLOSURE STATEMENT

Disclosure of potential conflict of interest: The author declares no relevant conflicts of interest.

Yoshiyasu Takefuji, PhD

From the Faculty of Data Science, Musashino University, Tokyo, Japan. E-mail: takefuji@keio.jp.

REFERENCES

1. Hollstein MM, Traidl S, Heetfeld A, Forkel S, Leha A, Alkon N, et al. Skin microdialysis detects distinct immunologic patterns in chronic inflammatory skin diseases. J Allergy Clin Immunol 2024;154:1450-61.
2. Elhaik E. Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. Sci Rep 2022;12:14683.
3. Faber NM, Meinders MJ, Geladi P, Sjöström M, Buydens LMC, Kateman G. Random error bias in principal component analysis. Part II. Application of theoretical predictions to multivariate problems. Anal Chim Acta 1995;304:273-83.
4. Yu H, Hutson AD. A robust Spearman correlation coefficient permutation test. Commun Stat Theory Methods 2024;53:2141-53.
5. Wang JH, Chen YH. Network-adjusted Kendall's tau measure for feature screening with application to high-dimensional survival genomic data. Bioinformatics 2021; 37:2150-6.