Ageing Research Reviews 113 (2026) 102964

Contents lists available at ScienceDirect

Ageing Research Reviews

journal homepage: www.elsevier.com/locate/arr

# Reevaluating principal component analysis in geroscience: A call for nonlinear approaches in AI-based evaluations

Yoshiyasu Takefuji [*,1]

*Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan*

ARTICLE INFO

*Keywords:*
Explainable AI methods
Principal component analysis
Geroscience
Nonlinear models
Statistical methods
Collinearity diagnostics

ABSTRACT

Fuellen et al. (2025) highlighted the essential role of explainable AI methods, particularly principal component analysis (PCA), in evaluating interventions for aging and longevity. However, this paper raises significant concerns regarding PCA's linear and parametric nature, which can misrepresent complex, nonlinear data common in geroscience research. As biological relationships often defy simplistic interpretations, reliance on PCA may obscure vital insights, leading to potential misinterpretations of intervention effects. To enhance accuracy in analyses, this study advocates for the adoption of nonlinear and nonparametric methods, such as Spearman's rank correlation and Kendall's tau. By reconsidering their methodological approaches, researchers can foster more accurate and informed evaluations of aging-related interventions.

Accurate analysis in geroscience is fundamentally reliant on researchers performing error-free calculations, as the robustness of their findings is directly tied to the methodologies employed. This paper aims to highlight specific instances of methodological shortcomings and offers concrete strategies for improvement.

Fuellen et al. (2025) have reviewed the validation requirements for AI-based intervention evaluations in aging and longevity research, emphasizing the importance of explainable AI methods, particularly principal component analysis (PCA). This unsupervised machine learning technique aids in visualizing data within a similarity space, enabling effective representation of both the intervention under investigation and its comparative interventions in two- or three-dimensional formats. The spatial arrangement of these interventions is essential for interpretation, allowing researchers to discern those with comparable characteristics. In a well-structured similarity space, interventions with similar effects and toxicity profiles tend to cluster, providing valuable insights into their interrelations and enhancing the clarity of the evaluation process (Fuellen et al., 2025).

This paper raises significant concerns about the use of principal component analysis (PCA) as an explainable AI method, particularly because of its inherently linear and parametric nature. These characteristics can lead to misleading conclusions when applied to nonlinear and nonparametric data commonly encountered in geroscience (Makrodimitris et al., 2023; Pattaroni et al., 2024; Shen et al., 2024). In a field where intricate biological relationships and aging processes often defy simple linear correlations, relying on PCA may obscure critical nuances and patterns in the data (Chen et al., 2023; Sahu et al., 2020; Schober and Vetter, 2020). Consequently, researchers risk misinterpreting the effects of interventions or overlooking essential variables that could influence outcomes. To ensure robust and accurate analyses in geroscience, alternative approaches that account for the complexity and nonlinearity of biological data should be considered. These alternatives can enhance the interpretability of AI-driven insights, ultimately leading to more informed and effective interventions for aging-related challenges.

PCA faces two critical limitations in biological data analysis: First, as an unsupervised method, it focuses exclusively on feature variance without considering feature-target relationships, potentially missing biologically relevant signals. Second, PCA's underlying assumptions—linear relationships, meaningful correlations between variables, continuous and standardized data, adequate sample size, homoscedasticity, and minimal outliers—are frequently violated by the inherently nonlinear, nonparametric nature of biological data. When applied to such complex datasets, these violations can lead to misleading interpretations and potentially flawed scientific conclusions.

This paper advocates for the adoption of nonlinear and nonparametric robust statistical methods, such as Spearman's rank correlation (Yu and Hutson, 2024) and Kendall's tau (Okoye and Hosseini, 2024),

* Corresponding author.
*E-mail address:* takefuji@keio.jp.
[1] ORCID: 0000-0002-1826-742X.

https://doi.org/10.1016/j.arr.2025.102964
Received 25 January 2025; Received in revised form 16 May 2025; Accepted 23 November 2025
Available online 24 November 2025

both of which provide p-values for assessing significance. This comprehensive approach not only enhances the validity of the findings but also ensures a more accurate interpretation of complex multi-omics data.

These methodological considerations have profound implications for aging research, where the complexity of biological systems demands robust analytical approaches. Fuellen et al. (2025) highlight that AI-based intervention evaluations in aging and longevity research require carefully validated methods that can capture the multidimensional nature of age-related processes. When visualizing intervention similarities in aging research, the limitations of PCA can lead to misleading interpretations of how interventions cluster in relation to their effects on longevity pathways. Alternative methods like nonparametric correlation techniques or hybrid feature selection approaches offer more reliable insights into the complex interactions between genetic, epigenetic, and environmental factors that influence aging. By employing statistically validated feature selection methods, researchers can more accurately identify biomarkers of aging, evaluate intervention efficacy across different biological contexts, and develop more precise personalized interventions that address the heterogeneous nature of aging processes. This methodological refinement is essential for advancing the field toward evidence-based interventions that can meaningfully impact healthspan and lifespan.

## CRediT authorship contribution statement

Yoshiyasu Takefuji completed this research and wrote this article.

## Ethics approval

Not applicable.

## Consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Funding

This research has no fund.

## Conflicts of interest/Competing interests

The author has no conflict of interest.

## Code availability

Not applicable.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.arr.2025.102964.

## Availability of data and material

Not applicable.

## References

Chen, M., Papadikis, K., Jun, C., Macdonald, N., 2023. Linear, nonlinear, parametric, and nonparametric regression models for nonstationary flood frequency analysis. J. Hydrol. 616, 128772. https://doi.org/10.1016/j.jhydrol.2022.128772.

Fuellen, G., Kulaga, A., Lobentanzer, S., Unfried, M., Avelar, R.A., Palmer, D., Kennedy, B.K., 2025. Validation requirements for AI-based intervention-evaluation in aging and longevity research and practice. Ageing Res. Rev. 104, 102617. https://doi.org/10.1016/j.arr.2024.102617.

Makrodimitris, S., Pronk, B., Abdelaal, T., Reinders, M., 2023. An in-depth comparison of linear and non-linear joint embedding methods for bulk and single-cell multi-omics. Brief. Bioinform. 25 (1), bbad416. https://doi.org/10.1093/bib/bbad416.

Okoye, K., Hosseini, S., 2024. Correlation tests in R: Pearson Cor, Kendall's Tau, and Spearman's rho. In: R Programming. Springer, Singapore. https://doi.org/10.1007/978-981-97-3385-9_12.

Pattaroni, C., Begka, C., Cardwell, B., Jaffar, J., Macowan, M., Harris, N.L., Westall, G.P., Marsland, B.J., 2024. Multi-omics integration reveals a nonlinear signature that precedes progression of lung fibrosis. Clin. Transl. Immunol. 13 (1), e1485. https://doi.org/10.1002/cti2.1485.

Sahu, P., Kang, J., Erdemci-Tandogan, G., Manning, M.L., 2020. Linear and nonlinear mechanical responses can be quite different in models for biological tissues. Soft Matter 16 (7), 1850–1856. https://doi.org/10.1039/c9sm01068h.

Schober, P., Vetter, T.R., 2020. Nonparametric statistical methods in medical research. Anesth. Analg. 131 (6), 1862–1863. https://doi.org/10.1213/ANE.0000000000005101.

Shen, X., Wang, C., Zhou, X., Zhou, W., Hornburg, D., Wu, S., Snyder, M.P., 2024. Nonlinear dynamics of multi-omics profiles during human aging. Nat. Aging 4 (11), 1619–1634. https://doi.org/10.1038/s43587-024-00692-2.

Yu, H., Hutson, A.D., 2024. A robust Spearman correlation coefficient permutation test. Commun. Stat. - Theory Methods 53 (6), 2141–2153. https://doi.org/10.1080/03610926.2022.2121144.