Trends in Food Science & Technology 157 (2025) 104919

Contents lists available at ScienceDirect

Trends in Food Science & Technology

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

Beyond principal component analysis: Enhancing feature reduction in electronic noses through robust statistical methods

ARTICLE INFO

Handling Editor: Dr I Oey

Keywords:
Electronic noses
Principal component analysis
Feature reduction
Nonparametric methods
Nonlinear data
Statistical significance

ABSTRACT

Background: Zheng et al. (2025) provided a comprehensive review of advancements in electronic noses used for detecting alcoholic beverages. Their work highlights the critical role of Principal Component Analysis (PCA) in feature reduction, which enhances the accuracy of various analytical methods such as linear discriminant analysis (LDA), random forest (RF), convolutional neural networks (CNN), and back propagation neural networks (BPNN). While PCA is a widely used technique, its application in electronic nose technologies necessitates a closer examination of its limitations.

Scope and approach: This paper critically evaluates the limitations of PCA when applied to nonlinear and nonparametric data, emphasizing the potential for distorted conclusions that can arise from its use. Through an extensive literature review, the paper discusses the implications of PCA within electronic nose applications. Key areas of focus include the importance of assessing data distribution, understanding statistical relationships, and validating significance using p-values. Additionally, the paper advocates for the adoption of nonparametric statistical methods, such as Spearman's correlation and Kendall's tau, to enhance the reliability of the analyses conducted.

Key findings and conclusion: The review reveals that the linear assumptions underlying PCA may misrepresent variance in nonlinear datasets, leading to misleading projections that obscure structural information. PCA's focus on global patterns can also overlook significant local variations, potentially causing overlaps among distinct classes within high-dimensional data. These limitations necessitate caution when utilizing PCA in electronic nose technologies. Therefore, to ensure valid and reliable results in this rapidly advancing field, it is essential to adopt robust statistical methods and conduct thorough preliminary analyses that account for the specific characteristics of the data. Mitigating the risks of distorted conclusions will improve the accuracy and credibility of findings in this area of research.

Consent to participate

Not applicable.

Ethics approval

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Not applicable.

Code availability

Not applicable.

Authors' contributions

Yoshiyasu Takefuji completed this research and wrote this article.

Zheng et al. (2025) conducted a comprehensive review of the principles and advancements in electronic noses for detecting alcoholic beverages. They highlighted the importance of principal component analysis (PCA) as a key technique for feature reduction, which streamlines data before applying methods such as linear discriminant analysis (LDA), random forest (RF), convolutional neural networks (CNN), and back propagation neural networks (BPNN). By effectively employing PCA, the functionality and accuracy of these analytical approaches are significantly enhanced, contributing to improved performance in electronic nose applications.

However, while Zheng et al. presented innovative insights into recent progress in electronic nose technologies, this paper raises critical concerns regarding the potential limitations of using PCA for feature reduction. Specifically, it addresses the issues that arise when relying on linear and parametric approaches to analyze nonlinear and nonparametric data, which can ultimately lead to distorted conclusions (Chen, 2023; Tian, 2024). Recognizing these challenges is essential for ensuring the reliability and validity of results in this evolving field.

Principal Component Analysis (PCA) is a widely used technique for

https://doi.org/10.1016/j.tifs.2025.104919
Received 7 January 2025; Received in revised form 8 February 2025; Accepted 11 February 2025
Available online 17 February 2025

feature reduction; however, its application to nonlinear or nonparametric data often results in distortions due to several inherent limitations (Chen, 2023; Tian, 2024). Primarily, PCA operates as a linear method, based on the assumption that relationships within the data can be effectively captured through linear equations (Pongpiachan et al., 2024). This assumption presents challenges when the underlying structure of the data is nonlinear, as PCA may struggle to accurately represent variance, ultimately leading to misleading projections. For instance, in datasets characterized by circular distributions, PCA attempts to fit a linear line, which can significantly obscure critical information regarding the data's structure.

Furthermore, PCA focuses on capturing global patterns by maximizing variance across the entire dataset while neglecting local variations that are often crucial for understanding intrinsic patterns. As a result, when applied to datasets with significant local structures, PCA can distort these relationships by overlooking them entirely. Additionally, the projection process in PCA can cause overlaps between points from different classes or clusters in high-dimensional space, obscuring their distinct identities.

While traditional machine learning methods may rely on ground truth values for accuracy validation, feature reduction techniques like PCA lack such benchmarks for assessing the accuracy of feature importance. This absence of ground truth introduces inherent distortions in the feature reduction process. To accurately capture true associations between variables, three key components should be carefully considered: data distribution, the statistical relationship between variables, and the validation of statistical significance via p-values. This paper advocates for the adoption of robust, bias-free statistical methods (Okoye & Hosseini, 2024), including nonparametric techniques such as Spearman's correlation (Yu & Hutson, 2024) and Kendall's tau (Chen, 2022), both supplemented with p-values. These methods offer advantages in handling non-normal data distributions and can provide more reliable insights into variable relationships. However, before applying these statistical methods, it is crucial to conduct a Variance Inflation Factor (VIF) analysis to eliminate collinearity and interaction effects among features, thus preventing inflation in feature importance assessments (Salmerón-Gómez, 2024).

Feature reduction offers numerous benefits for researchers, including the reduction of computational complexity, minimization of overfitting, and enhanced interpretability of data. Additionally, it can lead to improved model performance by eliminating irrelevant or redundant features, accelerating training times, and facilitating easier visualization of high-dimensional data. By focusing on the most significant features, researchers can also gain deeper insights into the underlying patterns within their datasets, ultimately leading to more robust and generalizable models.

In the context of high-dimensional data, such as multi-omics datasets, PCA is often employed as a preprocessing step to mitigate the "curse of dimensionality." HoweverZheng et al. should acknowledge that the linearity assumption inherent in PCA can lead to the unintended exclusion of significant nonlinear relationships among features. This may result in a loss of valuable information when translating into insights regarding the underlying biological processes.

The paper will incorporate a critical discussion addressing this issue, emphasizing that while PCA can provide benefits in dimensionality reduction, it may not always be the most suitable choice for preserving meaningful relationships in nonlinear contexts. Furthermore, Zheng et al. will explore alternative feature selection methods that may better retain important non-linear characteristics, ensuring a more comprehensive analysis of the data.

It is frequently noted that the End-To-End approach is particularly effective in deep learning, as it allows for automatic feature extraction (Talaei Khoei, 2023; Mohd Noor, 2024). In such cases, utilizing the raw input data is often preferable to pre-processing the input values. This further supports the argument against the need to linearize data prior to its use in nonlinear models. Relevant references are provided below.

In conclusion, while Zheng et al. (2025) commendably highlight Principal Component Analysis (PCA) as a crucial tool for feature reduction in electronic noses, significant limitations must be acknowledged. PCA's linearity assumes that data relationships can be adequately captured through linear equations, which can distort analyses when applied to nonlinear or nonparametric datasets. This misrepresentation of variance can obscure essential local structures and lead to misleading conclusions. To enhance the reliability and validity of results in electronic nose applications, it is imperative to employ robust, bias-free statistical methods, such as nonparametric techniques including Spearman's correlation and Kendall's tau. Additionally, a thorough variance inflation factor analysis should precede these methods to mitigate collinearity and interaction effects, ensuring accurate assessment of feature importance and preserving the integrity of analytical findings in this advancing field.

Funding

This research has no fund.

Conflicts of interest/competing interests

The author has no conflict of interest.

Data availability

No data was used for the research described in the article.

References

Chen, S., Ghadami, A., & Epureanu, B. I. (2022). Practical guide to using Kendall's τ in the context of forecasting critical transitions. *Royal Society Open Science, 9*(7), Article 211346. https://doi.org/10.1098/rsos.211346

Chen, X., Yi, L., & Liu, R. (2023). FEDA: A nonlinear subspace projection approach for electronic nose data classification. *IEEE Transactions on Instrumentation and Measurement, 72*, 1–11. https://doi.org/10.1109/TIM.2022.3224521

Mohd Noor, M. H., & Ige, A. O. (2024). A survey on state-of-the-art deep learning applications and challenges. *arXiv*. https://doi.org/10.48550/arXiv.2403.17561

Okoye, K., & Hosseini, S. (2024). Correlation Tests in R: Pearson cor, Kendall's tau, and Spearman's Rho. In *R programming*. Singapore: Springer. https://doi.org/10.1007/978-981-97-3385-9_12.

Pongpiachan, S., Wang, Q., Apiratikul, R., Tipmanee, D., Li, L., Xing, L., Mao, X., Li, G., Han, Y., Cao, J., Surapipith, V., Aekakkararungroj, A., & Poshyachinda, S. (2024). Combined use of principal component analysis/multiple linear regression analysis and artificial neural network to assess the impact of meteorological parameters on fluctuation of selected PM2.5-bound elements. *PLoS One, 19*(3), Article e0287187. https://doi.org/10.1371/journal.pone.0287187

Salmerón-Gómez, R., García-García, C. B., & García-Pérez, J. (2024). A redefined variance inflation factor: Overcoming the limitations of the variance inflation factor. *Computational Economics*. https://doi.org/10.1007/s10614-024-10575-8

Talaei Khoei, T., Ould Slimane, H., & Kaabouch, N. (2023). Deep learning: Systematic review, models, challenges, and research directions. *Neural Computing & Applications, 35*(23), 23103–23124. https://doi.org/10.1007/s00521-023-08957-4

Tian, T., Liu, T., Li, H., Yang, H., & Yan, J. (2024). A nonlinear manifold embedding extreme learning machine to improve the gas recognition of electronic noses. *IEEE Sensors Journal, 24*(9), 15046–15054, 10.1109/.

Yu, H., & Hutson, A. D. (2024). A robust Spearman correlation coefficient permutation test. *Communications in Statistics - Theory and Methods, 53*(6), 2141–2153. https://doi.org/10.1080/03610926.2022.2121144

Zheng, Z., Liu, K., Zhou, Y., Debliquy, M., Bittencourt, C., & Zhang, C. (2025). A comprehensive overview of the principles and advances in electronic noses for the detection of alcoholic beverages. *Trends in Food Science & Technology, 156*, Article 104862. https://doi.org/10.1016/j.tifs.2024.104862

Yoshiyasu Takefuji*

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

* Corresponding author.
E-mail address: takefuji@keio.jp.