Trends in Food Science & Technology 157 (2025) 104918

Contents lists available at ScienceDirect

Trends in Food Science & Technology

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

Response to letter to the editor from Y. Takefuji on "Beyond principal component analysis: Enhancing feature reduction in electronic noses through robust statistical methods"

Zichen Zheng [a,b], Kewei Liu [a,c], Yiwen Zhou [a], Marc Debliquy [c], Carla Bittencourt [b], Chao Zhang [a,*]

[a] College of Mechanical Engineering, Yangzhou University, Yangzhou, 225127, PR China
[b] Chimie des Interactions Plasma-Surface, University of Mons, 20 Place du Parc, 7000, Mons, Belgium
[c] Service de Science des Matériaux, Faculté Polytechnique, University of Mons, 7000, Mons, Belgium

ARTICLE INFO

Handling Editor: Dr I Oey

Keywords
Principal component analysis
Electronic nose
Data processing method
Variance inflation factor
Nonlinear

ABSTRACT

Background: Principal Component Analysis (PCA) is extensively utilized in Electronic Nose (E-nose) research for dimensionality reduction, allowing simplification of high-dimensional data and enhancing computational efficiency. However, its dependency on linear assumptions and sensitivity to outliers pose significant challenges, particularly when faced with nonlinear or overlapping datasets.

Scope and approach: This paper explores advanced methods such as Kernel PCA (KPCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP), which are more adept at managing nonlinear data, as well as nonparametric methods like Spearman's correlation and Kendall's tau to illuminate sensor data relationships. Furthermore, we examine the necessity of Variance Inflation Factor (VIF) analysis in addressing multicollinearity, highlighting hybrid approaches like Random Forest-VIF and PCA-VIF that enhance model stability and interpretability.

Key findings and conclusions: The original article by Zheng et al. (2025) demonstrates the broad applicability of PCA in detecting alcoholic beverages using E-noses, yet emphasizes the requirement for further research into its limitations. While PCA is foundational, its shortcomings call for the integration of advanced methodologies that cater to the complexities of E-nose data. Future research should focus on refining preprocessing protocols, utilizing nonlinear techniques, and managing data variability to improve accuracy and robustness, ultimately expanding E-nose applications across various domains and ensuring reliable performance.

We appreciate the opportunity this letter gives us to further elaborate on the discussion initiated in our recent publication (Zheng et al., 2025). This article was written as a commentary, "a forum for personal opinions, observations or hypotheses, to present new perspectives and advance understanding of controversial issues by provoking debate and comment" according to the Guide for Authors of Trends in Food Science and Technology. The prompt engagement following its publication underscores the academic relevance of the topic and reflects the presence of diverse perspectives within the field.

Principal Component Analysis (PCA) remains a fundamental dimensional reduction technique and a widely applied chemometric tool in Electronic Nose (E-nose) research. It is valued for its role in data simplification and improving analytical precision. However, its reliance on linear assumptions and limitations with nonlinear data can undermine reliability in practical applications. Addressing these challenges is crucial to improving the robustness and real-world performance of e-nose systems. In this paper, Prof. Takefuji identifies three main concerns. Firstly, he highlights the shortcomings of PCA in analyzing nonlinear and nonparametric datasets, emphasizing that its linear assumptions can result in distorted projections, overlooked local structures, and overlapping classes. Secondly, they highlight the lack of ground truth benchmarks in PCA and advocate for nonparametric methods like Spearman's correlation and Kendall's tau, supplemented with p-values, to improve the accuracy of feature importance assessments. Lastly, he emphasizes the importance of performing a Variance Inflation Factor (VIF) analysis before applying statistical methods to detect and mitigate

* Corresponding author. College of Mechanical Engineering, Yangzhou University, Huayang West Road 196, Yangzhou, 225127, Jiangsu Province, PR China.
E-mail address: zhangc@yzu.edu.cn (C. Zhang).

https://doi.org/10.1016/j.tifs.2025.104918
Received 22 January 2025; Received in revised form 9 February 2025; Accepted 11 February 2025
Available online 12 February 2025
0924-2244/

collinearity and interaction effects, ensuring a more reliable evaluation of feature importance in complex datasets. Each point is addressed in detail in the following subsections, and a final section has been included to provide concluding remarks on this matter.

1. The limitations of PCA in E-nose research

The manuscript by (Zheng et al., 2025) aimed to provide an up-to-date review of the present applications and technological progress in the E-nose technique for alcoholic beverages detection. Data processing algorithms were also explored as a supplementary topic to ensure a more rigorous and comprehensive review. Section 3 of our article (Zheng et al., 2025) highlighted PCA as the most frequently employed data analysis method in alcohol detection by presenting a table featuring a selection of recent examples. However, we merely summarize the current applications of PCA without delving deeply into the inherent limitations of this data processing method or their implications for related research. PCA indeed faces some notable challenges, especially in the application of E-nose and their real-world detection scenarios, as pointed out by Prof. Takefuji in his letter.

In the application of E-nose sensors, the high-dimensional data generated typically comprises a substantial number of correlated features. PCA is employed to reduce dimensionality through linear transformations, effectively removing redundant features, suppressing noise, and significantly optimizing computational efficiency and storage requirements (Bax et al., 2022; Piłat-Rożek et al., 2023; Tan et al., 2025). Therefore, PCA as a conventional algorithm is recommended for straightforward scenarios owing to its efficiency in rapid prediction and suitability for hardware implementation (Hotel et al., 2018; Peng et al., 2009; Tang et al., 2010). However, there are certain limitations to the PCA. First, the primary limitation of PCA lies in its assumption of linear separability, which can lead to suboptimal performance when applied to inherently nonlinear datasets, such as those characterized by circular or manifold-like distributions. In such cases, PCA's linear projections are unable to accurately capture the true variance and underlying relationships within the data (Labrín & Urdinez, 2020). Second, PCA is highly sensitive to outliers, as well as the scale and distribution of the input data, which can impact its effectiveness. Third, interpreting principal components becomes increasingly challenging when dealing with many features, limiting its utility in complex datasets (Mahanti et al., 2024). Fourth, PCA has limitations when handling overlapping data, as it focuses solely on maximizing variance without considering class separability. This can result in poor differentiation between overlapping classes, as PCA does not account for inter-class variance. Additionally, important class-specific features may be overlooked if they contribute minimally to overall variance. For better performance with overlapping datasets, supervised techniques like Linear Discriminant Analysis (LDA) or advanced methods such as KPCA and t-SNE, which consider class relationships, are more suitable. To address this, researchers have developed alternative or hybrid approaches such as Kernel PCA (KPCA), which extends PCA by utilizing kernel functions to map data into higher-dimensional spaces for better analysis of nonlinear structures (Li, Wang, et al., 2024). Advanced techniques like t-Distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) are specifically designed to manage nonlinear data, offering more precise representations of complex patterns (Dang et al., 2024; Jiang et al., 2025). Integrating these methods into E-nose systems or similar applications can mitigate PCA's limitations while preserving the advantages of dimensionality reduction.

It is undeniable that, despite its inherent limitations, PCA remains the primary data processing method in E-nose detection consistent with our article, as supported by some other review studies (Yanchen Li et al., 2024; Mahanti et al., 2024; Tan et al., 2025). We align with Prof. Takefuji's letter and anticipate that future advancements in the field will address PCA's drawbacks by integrating more advanced techniques, such as machine learning and deep learning, which are better suited for handling complex, nonlinear, and high-dimensional data. Additionally, the exploration of novel dimensionality reduction methods tailored to the specific characteristics of E-nose data could significantly enhance detection accuracy and robustness. Interdisciplinary efforts combining materials science, sensor technology, and data analytics will likely drive further innovation and broaden the applications of E-nose systems.

2. Nonlinear and nonparametric methods

The data from an E-nose is typically nonlinear and often considered nonparametric, because it lacks predefined structures or distributions, making flexible, data-driven analytical methods essential for effective processing and interpretation (Ma et al., 2024). Prof. Takefuji suggested that Spearman's correlation and Kendall's tau were supplemented with p-values. These nonparametric methods are typically used to evaluate relationships between variables, particularly in cases involving non-normally distributed data or outliers (Hettmansperger & McKean, 2010; Perreault, 2024). In the context of E-nose systems, we agree these methods could be valuable for analyzing correlations between sensor responses and specific odor characteristics or environmental variables. Evidence in the literature explicitly demonstrates the application of Spearman's correlation and Kendall's tau in E-nose research (Wang et al., 2022; Xiong et al., 2023). However, there are limited reports on E-nose detection for alcoholic beverages, which are mainly focused on the last few years, proving that these data processing methods are gradually becoming a research hotspot (Hou et al., 2024; Huang et al., 2025; Zhang et al., 2024).

Spearman's rank-order correlation coefficient correlation and Kendall's tau, supplemented with p-values, measure the strength and direction of monotonic relationships between two variables while providing a statistical significance test to determine the likelihood that the observed correlation occurred by chance (Rezapour & Balakrishnan, 2013; Wu & Wang, 2022). Unlike Pearson's correlation, which relies on assumptions of linearity and normal distribution, these approaches are more robust against outliers, non-linear patterns, and heteroscedasticity, making them well-suited for real-world data without introducing bias (Profillidis & Botzoris, 2019). Spearman's correlation is highly effective in identifying monotonic relationships, making it ideal for large-scale data analysis, feature selection in machine learning, and environmental monitoring (Awasthi et al., 2025; Jiang et al., 2024). In contrast, Kendall's tau offers superior performance in handling tied ranks and demonstrates strong robustness with small sample sizes, making it preferable in fields such as psychometrics, healthcare analytics, and real-time sensor data processing (Chen, Ghadami, & Epureanu, 2022; Wang et al., 2020). However, several factors could be attributed to the limited combination of Spearman's correlation and Kendall's tau with E-nose research. E-nose research often emphasizes predictive accuracy in odor classification and detection, favoring dimensionality reduction techniques like PCA and classification algorithms such as SVM or Random Forest to handle high-dimensional data. Spearman's and Kendall's methods, while effective for exploring relationships, are not specifically designed for these purposes. Additionally, the nonlinear and multivariate characteristics of E-nose sensor data make advanced machine learning approaches more suitable than simpler correlation metrics. These methods are also sometimes regarded as basic statistical tools, leading to their potential for exploratory analysis, feature ranking, or validation being underutilized. Furthermore, the lack of practical examples in the literature demonstrating their application to E-nose systems reduces awareness and adoption.

Numerous nonlinear and nonparametric methods have been employed in recent literature to overcome the limitations of linear techniques and analyze complex datasets, including those in E-nose research. Common machine learning approaches, such as Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN), are widely utilized for classification and regression tasks (Anwar &

Anwar, 2025). Deep learning models, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), excel at capturing nonlinear patterns (Ni et al., 2024). Dimensionality reduction techniques like t-SNE, UMAP, and KPCA are frequently applied to manage high-dimensional data effectively. Statistical methods, such as Gaussian Process Regression (GPR) and quantile regression, along with clustering techniques like DBSCAN and K-Means with nonlinear kernels, are also extensively used (Yoon & Lee, 2024). Furthermore, methods for nonlinear time-series analysis, including Dynamic Time Warping (DTW) and autoregressive models, are particularly effective for addressing temporal variations in sensor data (Zhu et al., 2024). The integration of these advanced techniques with traditional methods has significantly improved the analysis and interpretation of complex data, especially in E-nose research. However, the drawbacks of commonly-used machine learning models cannot be overlooked either. (i) While models can be assessed based on known target outcomes (i.e., ground truth for predictions), there is no equivalent "ground truth" for determining feature importance. Consequently, the importance scores generated by different models are inherently model-specific and may vary depending on the methodology used. (ii) Feature importance scores can be distorted by overfitting, where the model places excessive emphasis on noise or specific patterns in the training data. The inclusion of irrelevant or redundant features can also skew these scores. Furthermore, variations in model sensitivity to hyperparameter configurations can destabilize importance estimates, amplifying potential biases. For instance, random forests may not be reliable when predictor variables differ in measurement scale or the number of categories, which often overestimate the importance of variables with many categories due to their reliance on impurity-based splitting measures (Strobl et al., 2007). In contrast, linear models might undervalue features with nonlinear effects due to their linear assumptions. Similarly, deep learning models, often seen as "black boxes", complicate the accurate assessment of the true impact of individual features.

Given the lack of ground truth values for establishing true associations between targets and features, three key factors must be carefully considered. First, understanding the data distribution pattern is crucial, as it influences the choice of statistical tests and models. Many statistical methods assume specific distributions (e.g., normality in parametric tests), and deviations from these assumptions may require alternative approaches. Second, analyzing statistical relationships between variables, such as through correlation, regression, or causal inference, helps uncover patterns and dependencies, which are essential for predictive modeling and mechanistic understanding. Third, p-values assess the statistical significance of observed results, indicating whether they are likely due to chance. A commonly used threshold is $p < 0.05$, but statistical significance should always be interpreted alongside effect sizes, confidence intervals, and domain-specific considerations to ensure robust conclusions (Andrade, 2019).

Therefore, since our article focuses only on the application and research of E-nose in the context of alcoholic beverages, it does not provide an in-depth or comprehensive discussion of data processing methods across the entire field of E-nose research. As a result, certain data processing techniques were not mentioned or elaborated upon. We think Prof. Takefuji offered a research-worthy thought that given the complexity of E-nose data, applying Spearman's correlation and Kendall's tau could offer valuable insights for addressing data variability challenges and improving accuracy in odor detection and classification. The complementary strengths of these two methods are particularly valuable in the E-nose domain, where sensor arrays often produce complex, non-linear, and noisy data. Spearman's correlation excels at detecting monotonic trends between gas concentration levels and sensor responses, facilitating feature extraction and enhancing gas classification accuracy. Meanwhile, Kendall's tau improves data reliability by effectively managing frequent tied sensor outputs, thereby enhancing the system's sensitivity to subtle variations in gas mixtures. This synergistic application of Spearman's correlation and Kendall's tau contributes to the development of more sensitive, selective, and stable E-nose systems, supporting their widespread use in environmental monitoring, food safety, and medical diagnostics.

3. Variance inflation factor analysis

Variance Inflation Factor (VIF) analysis is essential in data preprocessing, particularly for regression models or statistical methods requiring independent features. It measures the ratio of variance caused by multicollinearity among explanatory variables compared to the variance without multicollinearity (Vu et al., 2015). VIF analysis is crucial for detecting and addressing multicollinearity, which occurs when predictor variables are highly correlated. This issue can lead to unstable coefficient estimates, higher standard errors, and reduced model interpretability. By identifying and resolving redundant variables, VIF analysis improves the accuracy of statistical conclusions, ensures reliable evaluation of feature importance, and enhances model stability and generalizability in complex datasets (Cheng, Sun, Yao, Xu, & Cao, 2022). VIF plays a supporting role in the E-nose field (Chen, Yang, et al., 2022). Mutual Information-VIF (MI-VIF) was used to improve model interpretability and performance in the E-nose system (Cheng, Sun, et al., 2022). Many hybrid methods incorporating VIF into broader analytical frameworks hold promise for future E-nose data analysis, such as RF-VIF (Random Forest-VIF) (Yu et al., 2025), PCA-VIF (Toosi et al., 2022), GA-VIF (Genetic Algorithm-VIF) (Monday et al., 2025), Lasso-VIF (Zhao et al., 2025), ReliefF (Al-Haddad et al., 2024), etc.

Our article mentions VIF only in the context of examples but does not explain it (Zhang et al., 2020). We fully agree with Prof. Takefuji on the significance of integrating VIF into data preprocessing to enhance model performance and interpretability. Nonetheless, in the specific area of E-nose applications for alcohol beverage detection, which is the main focus of our article, this method is rarely discussed in existing literature. We encourage future research to systematically evaluate the effectiveness of VIF-based preprocessing methods in addressing multicollinearity within E-nose datasets for alcohol beverage detection.

4. Concluding remarks

In summary, while PCA's advantages and widespread application are well-recognized, it is insufficient to overlook its limitations, as an increasing body of research has identified its drawbacks and proposed alternative approaches to address them. This highlights an undeniable need for further refinement and innovation in data analysis methodologies. The article by (Zheng et al., 2025) provides numerous examples of PCA applied to E-nose detection of alcoholic beverages. However, these instances do not fully encompass the broader scope of research on PCA and other data processing techniques within the E-nose field. We concur that it is essential to discuss the advantages and disadvantages of various data processing methods, including PCA, to provide valuable insights for the future development of E-nose technology. By critically evaluating these aspects, researchers can make informed decisions when selecting appropriate methods or developing new approaches tailored to the specific challenges in E-nose applications. We hope this response successfully highlights the limitations of PCA in E-nose applications, particularly its inability to adequately represent nonlinear and nonparametric data structures, which can misrepresent variance and obscure key local patterns in sensor data. We also encourage exploring and integrating existing technologies, such as nonparametric techniques, advanced dimensionality reduction methods tailored for nonlinear datasets, and robust multicollinearity mitigation approaches, to create a comprehensive framework for data analysis. Emphasizing the importance of thorough preprocessing, including feature selection and VIF analysis, is critical to ensure the reliability and accuracy of results. Lastly, we advocate for the continued development and application of advanced technologies in E-noses, incorporating these improved methods to enhance their performance further and expand their practical applications across diverse fields.

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.

Acknowledgments

This work was supported by the Outstanding Youth Foundation of Jiangsu Province of China (Grant No. BK20211548), the Yangzhou Science and Technology Plan Project (Grant No. YZ2023246), the China Scholarship Council (Grant No. 202308320445) and Postgraduate Research and Practice Innovation Program of Jiangsu Province of China (No. KYCX23_3551).

Data availability

Data will be made available on request.

References

Al-Haddad, L. A., Alawee, W. H., & Basem, A. (2024). Advancing task recognition towards artificial limbs control with ReliefF-based deep neural network extreme learning. *Computers in Biology and Medicine, 169*, Article 107894. https://doi.org/10.1016/j.compbiomed.2023.107894
Andrade, C. (2019). The P value and statistical significance: Misunderstandings, explanations, challenges, and alternatives. *Indian Journal of Psychological Medicine, 41*, 210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19, 0253-7176 (Print).
Anwar, H., & Anwar, T. (2025). Quality assessment of chicken using machine learning and electronic nose. *Sensing and Bio-Sensing Research, 47*, Article 100739. https://doi.org/10.1016/j.sbsr.2025.100739
Awasthi, M. P., Pant, R. R., Bishwakarma, K., Raut, G. K., Paudel, G., Kafle, S., et al. (2025). Integrating multivariate statistical techniques and geochemical indices for water quality assessment in pond ecosystems of Kirtipur Municipality, Kathmandu, Nepal. *Sustainable Water Resources Management, 11*(2), 13. https://doi.org/10.1007/s40899-024-01182-4
Bax, C., Robbiani, S., Zannin, E., Capelli, L., Ratti, C., Bonetti, S., et al. (2022). An experimental apparatus for e-nose breath analysis in respiratory failure patients. *Diagnostics, 12*(4), 776.
Chen, S. A.-O., Ghadami, A. A.-O., & Epureanu, B. I. (2022). Practical guide to using Kendall's τ in the context of forecasting critical transitions. *Royal Society Open Science, 9*, 2054–5703. https://doi.org/10.1098/rsos.211346 (Print), 211346.
Chen, J., Yang, Y., Deng, Y., Liu, Z., Xie, J., Shen, S., et al. (2022). Aroma quality evaluation of Dianhong black tea infusions by the combination of rapid gas phase electronic nose and multivariate statistical analysis. *LWT, 153*, Article 112496. https://doi.org/10.1016/j.lwt.2021.112496
Cheng, J., Sun, J., Yao, K., Xu, M., & Cao, Y. (2022). A variable selection method based on mutual information and variance inflation factor. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 268*, Article 120652. https://doi.org/10.1016/j.saa.2021.120652
Cheng, J., Sun, J., Yao, K., Xu, M., Tian, Y., & Dai, C. (2022). A decision fusion method based on hyperspectral imaging and electronic nose techniques for moisture content prediction in frozen-thawed pork. *LWT, 165*, Article 113778. https://doi.org/10.1016/j.lwt.2022.113778
Dang, Y., Reddy, Y. V. M., & Cheffena, M. (2024). Facile E-nose based on single antenna and graphene oxide for sensing volatile organic compound gases with ultrahigh selectivity and accuracy. *Sensors and Actuators B: Chemical, 419*, Article 136409. https://doi.org/10.1016/j.snb.2024.136409
Hettmansperger, T. P., & McKean, J. W. (2010). *Robust nonparametric statistical methods*. CRC press. https://doi.org/10.1201/b10451
Hotel, O., Poli, J.-P., Mer-Calfati, C., Scorsone, E., & Saada, S. (2018). A review of algorithms for SAW sensors e-nose based volatile compound identification. *Sensors and Actuators B: Chemical, 255*, 2472–2482. https://doi.org/10.1016/j.snb.2017.09.040
Hou, Q., Wang, Y., Qu, D., Zhao, H., Tian, L., Zhou, J., et al. (2024). Microbial communities, functional, and flavor differences among three different-colored high-temperature Daqu: A comprehensive metagenomic, physicochemical, and electronic sensory analysis. *Food Research International, 184*, Article 114257. https://doi.org/10.1016/j.foodres.2024.114257
Huang, Q., Qiu, X., Guan, T., Yu, J., Mao, Y., Li, Y., et al. (2025). The influence of different geographical origins and grades on the flavor of Qingxiangxing Baijiu: An integrated analysis using descriptive sensory evaluation, GC×GC–MS, E-nose, and ICP-MS. *Food Chemistry X, 25*, Article 102141. https://doi.org/10.1016/j.fochx.2024.102141
Jiang, Y., Wang, Z., Hu, Y., Lin, W., Yao, L., Xiao, W., et al. (2025). Intelligent sniffer: A chip-based portable e-nose for accurate and fast essence evaluation. *Sensors and Actuators B: Chemical, 426*, Article 136989. https://doi.org/10.1016/j.snb.2024.136989
Jiang, J., Zhang, X., & Yuan, Z. (2024). Feature selection for classification with Spearman's rank correlation coefficient-based self-information in divergence-based fuzzy rough sets. *Expert Systems with Applications, 249*, Article 123633. https://doi.org/10.1016/j.eswa.2024.123633
Labrín, C., & Urdinez, F. (2020). Principal component analysis. In *R for political data science* (pp. 375–393). Chapman and Hall/CRC.
Li, Y., Wang, Z., Zhao, T., Li, H., Jiang, J., & Ye, J. (2024). Electronic nose for the detection and discrimination of volatile organic compounds: Application, challenges, and perspectives. *TrAC, Trends in Analytical Chemistry, 180*, Article 117958. https://doi.org/10.1016/j.trac.2024.117958
Li, W., Xu, J., Yang, W., Liu, F., Zhou, H., & Yan, Z. (2024). Approach and application of extracting matching features from E-nose signals for AI tasks. *Biomedical Signal Processing and Control, 90*, Article 105869. https://doi.org/10.1016/j.bspc.2023.105869
Ma, X., Wu, F., Yan, J., Duan, S., & Peng, X. (2024). TF-TCN: A time-frequency combined gas concentration prediction model for E-nose data. *Sensors and Actuators A: Physical, 376*, Article 115654. https://doi.org/10.1016/j.sna.2024.115654
Mahanti, N. K., Shivashankar, S., Chhetri, K. B., Kumar, A., Rao, B. B., Aravind, J., et al. (2024). Enhancing food authentication through E-nose and E-tongue technologies: Current trends and future directions. *Trends in Food Science & Technology, 150*, Article 104574. https://doi.org/10.1016/j.tifs.2024.104574
Monday, C., Zaghloul, M. S., Krishnamurthy, D., & Achari, G. (2025). Incremental machine learning and genetic algorithm for optimization and dynamic aeration control in wastewater treatment plants. *Journal of Water Process Engineering, 69*, Article 106600. https://doi.org/10.1016/j.jwpe.2024.106600
Ni, W., Wang, T., Wu, Y., Liu, X., Li, Z., Yang, R., et al. (2024). Multi-task deep learning model for quantitative volatile organic compounds analysis by feature fusion of electronic nose sensing. *Sensors and Actuators B: Chemical, 417*, Article 136206. https://doi.org/10.1016/j.snb.2024.136206
Peng, G., Tisch, U., Adams, O., Hakim, M., Shehada, N., Broza, Y. Y., et al. (2009). Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nature Nanotechnology, 4*(10), 669–673. https://doi.org/10.1038/nnano.2009.235
Perreault, S. (2024). Simultaneous computation of Kendall's tau and its jackknife variance. *Statistics & Probability Letters, 213*, Article 110181. https://doi.org/10.1016/j.spl.2024.110181
Piłat-Rożek, M., Łazuka, E., Majerek, D., Szeląg, B., Duda-Saternus, S., & Łagód, G. (2023). Application of machine learning methods for an analysis of e-nose multidimensional signals in wastewater treatment. *Sensors, 23*(1), 487. https://doi.org/10.3390/s23010487
Profillidis, V. A., & Botzoris, G. N. (2019). Chapter 5 - statistical methods for transport demand modeling. In V. A. Profillidis, & G. N. Botzoris (Eds.), *Modeling of transport demand* (pp. 163–224). Elsevier. https://doi.org/10.1016/B978-0-12-811513-8.00005-4.
Rezapour, M., & Balakrishnan, N. (2013). Estimators based on trimmed Kendall's tau in multivariate copula models. *Statistical Methodology, 15*, 55–72. https://doi.org/10.1016/j.stamet.2013.05.001
Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), 25. https://doi.org/10.1186/1471-2105-8-25
Tan, Y., Chen, Y., Zhao, Y., Liu, M., Wang, Z., Du, L., et al. (2025). Recent advances in signal processing algorithms for electronic noses. *Talanta, 283*, Article 127140. https://doi.org/10.1016/j.talanta.2024.127140
Tang, K.-T., Chiu, S.-W., Pan, C.-H., Hsieh, H.-Y., Liang, Y.-S., & Liu, S.-C. (2010). Development of a portable electronic nose system for the detection and classification of fruity odors. *Sensors, 10*, 9179–9193.
Toosi, N. B., Soffianian, A. R., Fakheran, S., & Waser, L. T. (2022). Mapping disturbance in mangrove ecosystems: Incorporating landscape metrics and PCA-based spatial analysis. *Ecological Indicators, 136*, Article 108718. https://doi.org/10.1016/j.ecolind.2022.108718
Vu, D. H., Muttaqi, K. M., & Agalgaonkar, A. P. (2015). A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy, 140*, 385–394. https://doi.org/10.1016/j.apenergy.2014.12.011
Wang, S., Hu, X. Z., Liu, Y. Y., Tao, N. P., Lu, Y., Wang, X. C., et al. (2022). Direct authentication and composition quantitation of red wines based on Tri-step infrared spectroscopy and multivariate data fusion. *Food Chemistry, 372*, Article 131259. https://doi.org/10.1016/j.foodchem.2021.131259
Wang, B. J., Ma, S. Y., Pei, S. T., Xu, X. L., Cao, P. F., Zhang, J. L., et al. (2020). High specific surface area SnO_2 prepared by calcining Sn-MOFs and their formaldehyde-sensing characteristics. *Sensors and Actuators B: Chemical, 321*, Article 128560. https://doi.org/10.1016/j.snb.2020.128560
Wu, Z., & Wang, C. (2022). Limiting spectral distribution of large dimensional Spearman's rank correlation matrices. *Journal of Multivariate Analysis, 191*, Article 105011. https://doi.org/10.1016/j.jmva.2022.105011
Xiong, L., He, M., Hu, C., Hou, Y., Han, S., & Tang, X. (2023). Image presentation and effective classification of odor intensity levels using multi-channel electronic nose technology combined with GASF and CNN. *Sensors and Actuators B: Chemical, 395*, Article 134492. https://doi.org/10.1016/j.snb.2023.134492
Yoon, D., & Lee, C.-H. (2024). A disturbance rejection control for urban air mobility using artificial sensor-based Gaussian process regression. *Aerospace Science and Technology, 147*, Article 109055. https://doi.org/10.1016/j.ast.2024.109055

Yu, Z., Kanwal, Q., Wang, M., Nurdiawati, A., & Al-Ghamdi, S. G. (2025). Spatiotemporal dynamics and key drivers of carbon emissions in regional construction sectors: Insights from a random forest model. *Cleaner Environmental Systems*. , Article 100257. https://doi.org/10.1016/j.cesys.2025.100257

Zhang, B., Lin, L., Zheng, C., Liu, X., Cui, W., Li, X., et al. (2024). Using in situ untargeted flavoromics analysis to unravel the empty cup aroma of Jiangxiang-type Baijiu: A novel strategy for geographical origin traceability. *Food Chemistry, 438*, Article 137932. https://doi.org/10.1016/j.foodchem.2023.137932

Zhang, H., Shao, W., Qiu, S., Wang, J., & Wei, Z. (2020). Collaborative analysis on the marked ages of rice wines by electronic tongue and nose based on different feature data sets. *Sensors, 20*.

Zhao, Y., Zhao, Y., Liao, H., Pan, S., & Zheng, Y. (2025). Interpreting LASSO regression model by feature space matching analysis for spatio-temporal correlation based wind power forecasting. *Applied Energy, 380*, Article 124954. https://doi.org/10.1016/j.apenergy.2024.124954

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*. https://doi.org/10.1016/j.tifs.2024.104862

Zhu, Z., Jiang, Q., Wang, M., Xu, M., Zhang, Y., Shuang, F., et al. (2024). A CO concentration prediction method for electronic nose based on TrellisNet with gated recurrent unit and dilated convolution. *Microchemical Journal, 199*, Article 110014. https://doi.org/10.1016/j.microc.2024.110014