ARTICLE IN PRESS

Trends in Food Science & Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Trends in Food Science & Technology

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

Reevaluating feature importance in machine learning for food authentication: Addressing bias and enhancing methodological rigor

ARTICLE INFO

Handling Editor: Dr I Oey

Keywords:
Food authentication
Machine learning
Artificial intelligence
Feature importance
Bias assessment
Statistical methods
Robust analysis

ABSTRACT

Background: Bhat et al. (2025) highlight the significant role of artificial intelligence (AI) and machine learning (ML) in food authentication through advanced algorithms that analyze large datasets for patterns associated with food fraud.

Objective: This paper aims to critically assess the approach of Bhat et al., with a specific focus on model-based feature importance and the biases related to traditional machine learning methods.

Methods: The paper distinguishes between machine learning target predictions and feature importances, advocating for the rigorous application of robust statistical techniques, including Spearman's correlation and p-values, to accurately reveal genuine associations among variables.

Results: The analysis emphasizes the necessity for researchers to comprehend the foundational principles of AI and ML to avoid misapplication of these technologies.

Conclusion: The paper recommends integrating both nonparametric and nonlinear methods to effectively reduce bias and improve the reliability of feature importance assessments in food authentication.

Bhat et al. (2024) reported on advanced methods in food authentication, emphasizing the significant roles of artificial intelligence (AI) and machine learning (ML). These technologies utilize sophisticated algorithms to analyze extensive datasets, effectively identifying patterns that indicate potential food fraud. In the context of spectral data analysis, variable selection is crucial, as it not only reduces measurement costs but also enhances model performance and facilitates clearer interpretations. Popular chemometric techniques for analyzing spectral data include genetic algorithms (GA), interval partial least squares regression (iPLS), model evaluation and factor analysis, as well as **model-based feature importance** statistics (Bhat, 2024).

While this paper acknowledges the advancements in smart food authentication aimed at enhancing safety and quality, as reviewed by Bhat et al., it raises critical concerns regarding model-based feature importance. Numerous studies, including over 100 peer-reviewed articles, have underscored the prevalence of biased feature importances generated by machine learning models, leading to potentially inaccurate conclusions across various applications in general, including food authentication (Asilian Bdgoli, 2022; Curchoe, 2020; Demircioğlu, 2021; Feng, 2024; Grandhi & Singh, 2024; Krawczuk & Łukaszuk, 2016; Shiue, Guh, & Tseng, 2009). While machine learning target predictions are based on ground truth values, feature importances from machine learning models do not possess equivalent reference points for validation. Consequently, there is a need for bias-free and robust statistical methods such as Spearman's correlation with p-values (Eden, Li, & Shepherd, 2022; Liu, Li, Wanga, & Shepherd, 2018; Yu & Hutson, 2024).

Notably, Bhat et al. did not clearly distinguish between machine learning target predictions and feature importances, which is crucial for a comprehensive understanding of the methodologies involved (Asilian Bidgoli, 2022; Demircioğlu, 2021; Grandhi & Singh, 2024; Krawczuk & Łukaszuk, 2016). This paper advocates for the use of machine learning target predictions while strongly discouraging the reliance on feature importances derived from these models. By promoting a clearer differentiation between these two aspects, this paper aims to improve the rigor and reliability of machine learning applications in food authentication.

The misuse of AI and ML methods is widespread across various disciplines (Takefuji, 2024a, 2024b; 2024c; 2024d; 2024e), largely due to the fact that researchers often have deep expertise in their specific domains but may lack proficiency in numerical algorithms and bias assessment techniques. This gap in knowledge can lead to the inadvertent application of these sophisticated tools without fully understanding their limitations and potential pitfalls. Addressing this issue is crucial for improving the reliability of results in food authentication and ensuring that AI and ML are utilized appropriately and effectively in this critical area.

One critical issue is that feature importances derived from machine learning models should not be used for food authentication due to their model-specific nature. This variability indicates that different machine learning algorithms employ distinct methodologies for calculating feature importance, resulting in disparate degrees of bias and potential misinterpretation of the data. This paper advocates for the use of true associations between the target and features employing robust bias-free statistical methods such as Spearman's correlation with p-values (Eden et al., 2022; Liu et al., 2018; Yu & Hutson, 2024).

Moreover, it is essential for practitioners utilizing AI and ML to comprehend the fundamental principles underlying these techniques (Chu et al., 2012; Demircioğlu, 2021; Epstein, Nallapareddy, & Ray, 2023; Haury, Gestraud, & Vert, 2011; Montesinos-López et al., 2023; Shim, Lee, & Hwang, 2021; Smialowski, Frishman, & Kramer, 2010;

https://doi.org/10.1016/j.tifs.2024.104853
Received 30 November 2024; Received in revised form 6 December 2024; Accepted 20 December 2024
Available online 12 February 2025

Please cite this article as: Yoshiyasu Takefuji, *Trends in Food Science & Technology*, https://doi.org/10.1016/j.tifs.2024.104853

Yousef, SaçarDemirci, Khalifa, & Allmer, 2016). Unfortunately, many researchers fail to grasp these critical concepts. When analyzing the relationship between target variables and features, several key elements come into play. These include understanding the data distribution, examining the statistical relationships among variables, and validating statistical significance through p-values. Cross-validation is only effective for target prediction accuracy, not for feature importance accuracy.

To mitigate the impact of bias and enhance the reliability of feature importance assessments, it is essential to employ appropriate linear or non-linear, as well as parametric or nonparametric methods, alongside comprehensive p-value calculations (Antonelli, 2016; Cole, Edwards, Breskin, & Hudgens, 2021; ConzueloRodriguez et al., 2022; Hade & Lu, 2014; Nazer et al., 2023; Pérez-Rodríguez, 2012; Vrbin, 2022). This multi-faceted approach not only reduces potential biases but also improves the robustness of the analyses, leading to more accurate interpretations and conclusions in the realm of feature importance.

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.

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

Antonelli, J., Trippa, L., & Haneuse, S. (2016). Mitigating bias in generalized linear mixed models: The case for bayesian nonparametrics. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics, 31*(1), 80–95. https://doi.org/10.1214/15-STS533

Asilian Bidgoli, A., Rahnamayan, S., Dehkharghanian, T., Grami, A., & Tizhoosh, H. R. (2022). Bias reduction in representation of histopathology images using deep feature selection. *Scientific Reports, 12*(1), Article 19994. https://doi.org/10.1038/s41598-022-24317-z

Bhat, M. A., Rather, M. Y., Singh, P., Hassan, S., & Hussain, N. (2024). Advances in smart food authentication for enhanced safety and quality. *Trends in Food Science & Technology, 155*. https://doi.org/10.1016/j.tifs.2024.104800. Article 104800.

Chu, C., Hsu, A. L., Chou, K. H., Bandettini, P., Lin, C., & Alzheimer's Disease Neuroimaging Initiative. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *NeuroImage, 60*(1), 59–70. https://doi.org/10.1016/j.neuroimage.2011.11.066

Cole, S. R., Edwards, J. K., Breskin, A., & Hudgens, M. G. (2021). Comparing parametric, nonparametric, and semiparametric estimators: The weibull trials. *American Journal of Epidemiology, 190*(8), 1643–1651. https://doi.org/10.1093/aje/kwab024

Conzuelo Rodriguez, G., Bodnar, L. M., Brooks, M. M., Wahed, A., Kennedy, E. H., Schisterman, E., & Naimi, A. I. (2022). Performance evaluation of parametric and nonparametric methods when assessing effect measure modification. *American Journal of Epidemiology, 191*(1), 198–207. https://doi.org/10.1093/aje/kwab220

Curchoe, C. L. (2020). All models are wrong, but some are useful. *Journal of Assisted Reproduction and Genetics, 37*(10), 2389–2391. https://doi.org/10.1007/s10815-020-01895-3

Demircioğlu, A. (2021). Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. *Insights into imaging, 12*(1), 172. https://doi.org/10.1186/s13244-021-01115-1

Eden, S. K., Li, C., & Shepherd, B. E. (2022). Nonparametric estimation of Spearman's rank correlation with bivariate survival data. *Biometrics, 78*(2), 421–434. https://doi.org/10.1111/biom.13453

Epstein, E., Nallapareddy, N., & Ray, S. (2023). On the relationship between feature selection metrics and accuracy. *Entropy, 25*(12), 1646. https://doi.org/10.3390/e25121646

Feng, G. (2024). Feature selection algorithm based on optimized genetic algorithm and its application in high-dimensional data processing. *PLoS One, 19*(5), Article e0303088. https://doi.org/10.1371/journal.pone.0303088

Grandhi, A., & Singh, S. K. (2024). Interrelated dynamic biased feature selection and classification model using enhanced gorilla troops optimizer for intrusion detection. *Alexandria Engineering Journal, 114*, 312–330. https://doi.org/10.1016/j.aej.2024.10.100

Hade, E. M., & Lu, B. (2014). Bias associated with using the estimated propensity score as a regression covariate. *Statistics in Medicine, 33*(1), 74–87. https://doi.org/10.1002/sim.5884

Haury, A. C., Gestraud, P., & Vert, J. P. (2011). The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *PLoS One, 6*(12), Article e28210. https://doi.org/10.1371/journal.pone.0028210

Krawczuk, J., & Łukaszuk, T. (2016). The feature selection bias problem in relation to high-dimensional gene data. *Artificial Intelligence in Medicine, 66*, 63–71. https://doi.org/10.1016/j.artmed.2015.11.001

Liu, Q., Li, C., Wanga, V., & Shepherd, B. E. (2018). Covariate-adjusted Spearman's rank correlation with probability-scale residuals. *Biometrics, 74*(2), 595–605. https://doi.org/10.1111/biom.12812

Montesinos-López, O. A., Crespo-Herrera, L., Saint Pierre, C., Bentley, A. R., de la Rosa-Santamaria, R., Ascencio-Laguna, J. A., Agbona, A., Gerard, G. S., Montesinos-López, A., & Crossa, J. (2023). Do feature selection methods for selecting environmental covariables enhance genomic prediction accuracy? *Frontiers in Genetics, 14*, Article 1209275. https://doi.org/10.3389/fgene.2023.1209275

Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS digital health, 2*(6), Article e0000278. https://doi.org/10.1371/journal.pdig.0000278

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., & Dreisigacker, S. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 (Bethesda, Md.), 2*(12), 1595–1605. https://doi.org/10.1534/g3.112.003665

Shim, M., Lee, S. H., & Hwang, H. J. (2021). Inflated prediction accuracy of neuropsychiatric biomarkers caused by data leakage in feature selection. *Scientific Reports, 11*(1), 7980. https://doi.org/10.1038/s41598-021-87157-3

Shiue, Y. R., Guh, R. S., & Tseng, T. Y. (2009). GA-based learning bias selection mechanism for real-time scheduling systems. *Expert Systems with Applications, 36*(9), 11451–11460. https://doi.org/10.1016/j.eswa.2009.03.019

Smialowski, P., Frishman, D., & Kramer, S. (2010). Pitfalls of supervised feature selection. *Bioinformatics, 26*(3), 440–443. https://doi.org/10.1093/bioinformatics/btp621

Takefuji, Y. (2024a). Chi-squared and p-values vs. machine learning feature selection. *Annals of Oncology*. https://doi.org/10.1016/j.annonc.2024.10.013

Takefuji, Y. (2024b). Unveiling feature importance biases in linear regression: Implications for protein-centric cardiovascular research. *Atherosclerosis, 119049*. https://doi.org/10.1016/j.atherosclerosis.2024.119049

Takefuji, Y. (2024c). Reassessing feature importance biases in machine learning models for infection analysis. *Journal of Infection, 89*(6), Article 106357. https://doi.org/10.1016/j.jinf.2024.106357

Takefuji, Y. (2024d). Evaluating feature importance biases in logistic regression: Recommendations for robust statistical methods. *European Journal of Internal Medicine*. https://doi.org/10.1016/j.ejim.2024.11.022

Takefuji, Y. (2024e). Unveiling hidden biases in machine learning feature importance. *Journal of Energy Chemistry, 102*, 49–51. https://doi.org/10.1016/j.jechem.2024.10.032

Vrbin, C. M. (2022). Parametric or nonparametric statistical tests: Considerations when choosing the most appropriate option for your data. *Cytopathology: Official Journal of the British Society for Clinical Cytology, 33*(6), 663–667. https://doi.org/10.1111/cyt.13174

Yousef, M., Saçar Demirci, M. D., Khalifa, W., & Allmer, J. (2016). Feature selection has a large impact on one-class classification accuracy for MicroRNAs in plants. *Advances in bioinformatics, 2016*, Article 5670851. https://doi.org/10.1155/2016/5670851

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

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

E-mail address: takefuji@keio.jp.