ARTICLE IN PRESS

Journal of Cardiothoracic and Vascular Anesthesia 000 (2025) 1−2

Contents lists available at ScienceDirect

Journal of Cardiothoracic and Vascular Anesthesia

journal homepage: www.jcvaonline.com

Letter to the Editor

Clinical Machine Learning Pitfalls: Reliability of Feature Importance in Prediction of Continuous Renal Replacement Therapy in Acute Type A Aortic Dissection Assessment

To the Editor:

Li et al. proposed a machine learning framework using Extreme Gradient Boosting (XGBoost) and SHapley Additive exPlanation (SHAP) to predict the need for continuous renal replacement therapy after acute type A aortic dissection repair.[1] They developed a predictive model for continuous renal replacement therapy using multiple clinical and laboratory variables derived from perioperative data from 588 patients from a single center. After comparative evaluation, XGBoost was selected over other machine learning algorithms due to its superior performance. Subsequently, SHAP values were used to evaluate feature importance, identifying key predictors including intraoperative peak lactate level, red blood cell transfusion volume, renal artery involvement status, and myoglobin, cystatin C, and creatine kinase-myocardial band concentrations. Despite the widespread adoption of such frameworks, it is essential to recognize that high predictive performance does not guarantee reliable feature rankings. We believe that inherent biases raise concerns about the reliability of feature importance. Several studies have highlighted the discordance between predictive performance and meaningful attribution, which reinforces the need for rigorous, model-independent frameworks to support clinically relevant interpretation. Feature importance rankings often reflect artifacts of the prediction process rather than genuine causal relationships. Indeed, overreliance on predictive accuracy to justify feature relevance is a well-documented issue, the details of which are discussed in the Supplementary Material in detail.[2-5]

Tree-based machine learning models such as XGBoost are known to introduce biases in feature importance estimation, often favoring variables involved in early decision splits.[6-10] These importance scores are further shaped by the model's internal logic, feature interactions, and multicollinearity. SHAP, a widely used explainable artificial intelligence technique, inherits these model-specific biases and may even amplify them due to its dependence on the model's output structure.[11,12] Consequently, the combined use of XGBoost and SHAP can result in interpretability challenges. This issue may be particularly pronounced in high-dimensional datasets with complex and correlated features, where overfitting becomes more likely and model reliability diminishes. We believe that claims of successful feature identification under such conditions, especially when accounting for nonlinearity and interactions, should be carefully examined in light of these limitations. Moreover, validating feature importance remains inherently difficult due to the lack of ground truth, which leads to inconsistent rankings across models. In Li et al.'s study, the presence of intricate feature relationships and data sensitivity further complicates interpretation, highlighting the need for cautious analysis and robust methodological design.

To ensure accurate interpretations in health risk assessment, a robust analytical framework is essential. This approach should incorporate methodologies better suited for capturing complex associations within health data, such as unsupervised learning techniques including feature agglomeration and highly variable gene selection.[13,14] The use of nonparametric statistical methods like Spearman's rho or Kendall's tau would most likely also be highly beneficial.[15,16] These methods can detect various types of relationships, offering enhanced interpretability crucial for translating findings into actionable clinical insights. Ultimately, this multifaceted approach is indispensable for generating accurate, reproducible, and clinically relevant information that can truly advance health risk assessment.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Statement

No new data were generated or analyzed in support of this research.

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.

CRediT authorship contribution statement

Souichi Oka: Writing – original draft, Conceptualization. **Yoshiyasu Takefuji:** Writing – review & editing, Supervision, Project administration.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1053/j.jvca.2025.08.035.

References

1 Li K, Li Y, Gao Q, et al. Machine learning in risk prediction of continuous renal replacement therapy after surgical repair of acute type A aortic dissection. J Cardiothorac Vasc Anesth 2025;0:6–28. https://doi.org/10.1053/j.jvca.2025.06.028.
2 Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. J Mach Learn Res 2019;20:177. https://doi.org/10.48550/arXiv.1801.01489.
3 Lipton ZC. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. ACM Queue 2018;16:31–57. https://doi.org/10.1145/3236386.3241340.
4 Lones MA. Avoiding common machine learning pitfalls. Patterns 2024;5:101046. https://doi.org/10.1016/j.patter.2024.101046.
5 Musolf AM, Holzinger ER, Malley JD, et al. What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Hum Genet 2022;141:1515–28. https://doi.org/10.1007/s00439-021-02402-z.
6 Alaimo Di Loro P, Scacciatelli D, Tagliaferri G. 2-Step gradient boosting approach to selectivity bias correction in tax audit: An application to the VAT gap in Italy. Stat Methods Appl 2023;32:237–70. https://doi.org/10.1007/s10260-022-00643-4.
7 Adler AI, Painsky A. Feature importance in gradient boosting trees with cross-validation feature selection. Entropy 2022;24:687. https://doi.org/10.3390/e24050687.
8 Huti M, Lee T, Sawyer E, et al. An investigation into race bias in random forest models based on breast DCE-MRI derived radiomics features. Clin Image-Based Proced. Fairness AI Med Imaging 2023;225:225–34. https://doi.org/10.1007/978-3-031-45249-9_22.
9 Salles T, Rocha L, Gonçalves M. A bias-variance analysis of state-of-the-art random forest text classifiers. Adv Data Anal Classif 2021;15:379–405. https://doi.org/10.1007/s11634-020-00409-4.
10 Touw WG, Bayjanov JR, Overmars L, et al. Data mining in the life sciences with random forest: A walk in the park or lost in the jungle. Brief Bioinform 2013;14:315–26. https://doi.org/10.1093/bib/bbs034.
11 Huang X, Marques-Silva J. On the failings of Shapley values for explainability. Int J Approx Reason 2024;171:109112. https://doi.org/10.1016/j.ijar.2023.109112.
12 Kumar I, Scheidegger C, Venkatasubramanian S, et al. Shapley residuals: Quantifying the limits of the Shapley value for explanations. Adv Neural Inf Process Syst 2021;34:26598–608. https://doi.org/10.48550/arXiv.2106.10860.
13 Zhang J, Wu X, Hoi SCH, et al. Feature agglomeration networks for single stage face detection. Neurocomputing 2020;380:180–9. https://doi.org/10.1016/j.neucom.2019.10.087.
14 Xie Y, Jing Z, Pan H, et al. Redefining the high variable genes by optimized LOESS regression with positive ratio. BMC Bioinformatics 2025;26:104. https://doi.org/10.1186/s12859-025-06112-5.
15 Yu H, Hutson AD. A robust Spearman correlation coefficient permutation test. Commun Stat Theory Methods 2024;53:2141–53. https://doi.org/10.1080/03610926.2022.2121144.
16 Okoye K, Hosseini S. Correlation tests in R: Pearson Cor, Kendall's Tau, and Spearman's Rho. R Programming: Statistical data analysis in research. Springer Nature 2024:247–77. https://doi.org/10.1007/978-981-97-3385-9_12.

Souichi Oka*,[1]
Yoshiyasu Takefuji†
*Science Park Corporation, Kanagawa, Japan
†Musashino University, Tokyo, Japan

https://doi.org/10.1053/j.jvca.2025.08.035