European Journal of Surgical Oncology 51 (2025) 110025

Contents lists available at ScienceDirect

European Journal of Surgical Oncology

journal homepage: www.ejso.com

Comment on "Optimized machine learning model for predicting unplanned reoperation after rectal cancer anterior resection"

ARTICLE INFO

Keywords:
Feature selection
Machine learning
Biases
True associations

Letter to the Editor

We examined the recent study by Yang et al., entitled "Optimized machine learning model for predicting unplanned reoperation after rectal cancer anterior resection," which presents critical points that warrant further discussion [1]. Their objective was to predict the risk of unplanned reoperation (URO) following anterior resection in rectal cancer patients. They conducted feature selection for 14 variables using both least absolute shrinkage and selection operator (LASSO) regression and the Boruta algorithm. To interpret the feature selection, they utilized the SHapley Additive exPlanations (SHAP) method, which revealed that tumor location, previous abdominal surgery, and operative time were the most significant factors influencing the risk of URO. Despite achieving impressive evaluation metrics (accuracy of 0.842 and AUC of 0.889), their model highlights critical issues with relying on machine learning for feature selection and the inherent biases in methods such as LASSO, Boruta, and SHAP.

Firstly, LASSO eliminates significant nonlinear features due to its linear and parametric method, introducing critical biases [2]. It tends to select only one variable from highly correlated groups, potentially missing other important predictors. Additionally, its tendency to shrink coefficients to zero can oversimplify the model by excluding relevant variables. LASSO is also sensitive to the regularization parameter, and improper tuning can significantly impact performance, introducing biases based on specific configurations.

Secondly, while Boruta is effective in identifying important features, it can also exhibit biases similar to LASSO. Its reliance on random forest algorithms can lead to dataset-specific biases, especially with imbalanced data. Additionally, Boruta's iterative process may overly emphasize certain features, skewing model predictions. It is also sensitive to the parameters of the underlying random forest, and improper tuning can introduce biases based on specific configurations. Over 100 peer-reviewed articles have documented non-negligible biases in feature importances from machine learning models including LASSO and Boruta.

Lastly, SHAP inherits and can amplify biases from the machine learning models, distorting interpretations and conclusions drawn from the analysis [3,4]. The function form 'explain = SHAP(model)' serves as evidence of SHAP's dependency on the model. As SHAP relies on the outputs of these models to explain feature importance, it is subject to the same significant biases inherent in the models. Consequently, this reliance can lead to erroneous conclusions and reduce the reliability of the findings. Machine learning models often tend to maximize prediction accuracy, which can result in overfitting. Therefore, high target prediction accuracy does not guarantee reliable feature importances.

Numerous peer-reviewed studies have highlighted significant biases in these models. Due to the absence of ground truth values, different models utilize distinct methodologies for calculating feature importances. This variability necessitates careful interpretation, as feature importances may not accurately reflect true relationships within the dataset. To uncover genuine relationships between the target variable and its features, consider three critical aspects: data distribution, statistical relationships between variables, and statistical validation through p-values. Understanding data distribution is vital for appropriate modeling techniques. Overlooking nonlinear relationships may obscure significant patterns. Investigating statistical interactions between target and predictor variables is essential. Nonparametric approaches can capture these complexities more accurately. Additionally, incorporating statistical validation techniques, such as hypothesis testing and p-value analysis, is crucial to substantiate claims about feature importance and ensure observed relationships are not due to random variation.

Instead of relying on LASSO, Boruta, and SHAP for feature selection, this paper advocates for using unbiased, robust statistical methods such as Spearman's rho and Kendall's tau with p-values. Prior to applying these methods, performing a Variance Inflation Factor (VIF) analysis is crucial to identify and remove features with collinearity and interaction effects. These nonlinear and nonparametric approaches provide valuable insights into variable relationships while addressing data complexity. The paper recommends that Yang et al. reassess their analysis using these robust techniques to ensure more reliable and valid outcomes. By focusing on genuine associations, they can enhance the integrity of their findings and gain a deeper understanding of the underlying relationships in their data.

In summary, while machine learning methods like LASSO, Boruta,

https://doi.org/10.1016/j.ejso.2025.110025
Received 19 February 2025; Accepted 7 April 2025

and SHAP are widely used for feature selection, they have biases which can result in deviations from the true association. Incorporating robust statistical methods and thorough validation techniques can lead to more accurate insights and trustworthy conclusions.

Author contributions

Souichi Oka: Conceptualization and Writing - original draft. Yoshiyasu Takefuji: Supervision and Writing - review & editing.

Funding

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

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. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejso.2025.110025.

References

[1] Yang S, Li Y, Yang W, Luo X, Chen L. Optimized machine learning model for predicting unplanned reoperation after rectal cancer anterior resection. Eur J Surg Oncol 2024;50(12):108703. https://doi.org/10.1016/j.ejso.2024.108703.
[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] Bilodeau B, Jaques N, Koh PW, Kim B. Impossibility theorems for feature attribution. Proc Natl Acad Sci USA 2024;121(2):e2304406120. https://doi.org/10.1073/pnas.2304406120.
[4] 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.

Souichi Oka*
Science Park Corporation, 3-24-9 Iriya-Nishi Zama-shi, Kanagawa, 252-0029, Japan

Yoshiyasu Takefuji
Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo, 135-8181, Japan
E-mail address: takefuji@keio.jp.

* Corresponding author.
E-mail address: souichi.oka@sciencepark.co.jp (S. Oka).