ARTICLE IN PRESS

Radiotherapy and Oncology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Letter to the Editor

Towards reliable feature interpretation in machine learning-based acute diarrhoea toxicity assessment

ARTICLE INFO

Keywords:
Rectal cancer
Radiotherapy
Predictive model
Machine learning
Reproducibility of results

Letter to the Editor

Zhang et al. developed machine learning (ML) models to predict acute diarrhoea in rectal cancer patients undergoing chemoradiotherapy, warranting further discussion regarding interpretability and methodological bias [1]. Zhang et al. used logistic regression to predict a $\Delta G \geq 2$ increase in acute diarrhoea toxicity and applied SHapley Additive exPlanations (SHAP) to identify key predictors, including small bowel V_{10Gy}, treatment arm, age, and performance status. While the study offers valuable insights, its methodology raises important concerns regarding interpretability and bias. In the 'Methods and Materials' section, Zhang et al. explicitly state, "Logistic regression was selected for $\Delta G \geq 2$ prediction due to its simplicity and easy interpretation." However, this choice raises two critical methodological concerns.

First, achieving high predictive accuracy for $\Delta G \geq 2$ (AUROC = 0.71) does not guarantee reliable or consistent feature importance rankings. As emphasized in numerous studies, predictive performance alone does not guarantee that the features identified are meaningful or reliable [2–5]. Overreliance on accuracy metrics can obscure the interpretive validity of the model's outputs. Supplementary material includes more discussion and references. Second, logistic regression is a logit-linear and parametric model, assuming fixed functional forms and logit-linear relationships between predictors and the log-odds of the outcome [6–8]. While its simplicity and interpretability are appealing, these characteristics make it ill-suited for modeling complex biological phenomena, which often involve non-linear interactions and non-parametric distributions. As a result, genuine non-linear patterns may be entirely overlooked, leading to an incomplete or even misleading understanding of acute diarrhoea risk.

Additionally, although SHAP values aim to clarify feature importance, they can inherit and sometimes amplify biases from the underlying model [9–11]. This dependency is evident in the formulation 'explain = SHAP(model)', which ties the explanation directly to the model's logic. In Zhang et al.'s pipeline, both logistic regression and SHAP are susceptible to bias, and their combination illustrates a common pitfall in ML-based biomarker discovery. When linearity assumptions fail to capture biological complexity, feature importance estimates become unstable. The cascading amplification of bias raises concerns about the interpretive validity of identified features. Generally, validating feature importance is difficult without ground truth; high dimensionality and multicollinearity amplify model-specific biases, destabilizing rankings and obscuring biological relevance.

To improve acute diarrhoea risk assessment reliability, a comprehensive analytical framework is needed. This framework should handle complex clinical and biomarker data, incorporating methods for non-linear patterns and multicollinearity reduction. Unsupervised techniques such as Feature Agglomeration (FA) and Highly Variable Gene Selection (HVGS) offer valuable alternatives for dimensionality reduction and feature prioritization [12,13]. Non-parametric statistical methods like Spearman's rho and Kendall's tau can detect monotonic associations without linearity, enhancing both precision and interpretability [14,15]. These approaches suit translational biomarker research, providing clear and trustworthy insights for clinical decision-making and stakeholder communication.

In conclusion, while Zhang et al.'s study contributes to predictive oncology, its reliance on linear modeling and SHAP interpretation limits biological insight. Future research should prioritize robust, non-parametric approaches to ensure interpretive validity, especially in biomarker discovery where clinical decisions depend on trustworthy feature relevance.

Data statement

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

Funding source

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

DOI of original article: https://doi.org/10.1016/j.radonc.2025.111032.

https://doi.org/10.1016/j.radonc.2025.111140
Received 10 August 2025; Accepted 15 August 2025
Available online 21 September 2025

Please cite this article as: Souichi Oka et al., *Radiotherapy and Oncology*, https://doi.org/10.1016/j.radonc.2025.111140

CRediT authorship contribution statement

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

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.

Acknowledgements

We extend our sincere gratitude to Nobuko Inoue and Kiyo Yoshida of Science Park, Inc. for their invaluable assistance with the extensive literature review.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2025.111140.

References

[1] Zhang Y, Brand D, Shen Z, Simard M, Hindocha S, Chohan O, et al. Predicting acute diarrhoea in rectal cancer chemoradiotherapy: Secondary analysis of the phase III ARISTOTLE trial. Radiother Oncol 2025;210:111032. https://doi.org/10.1016/j.radonc.2025.111032.

[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] Lenhof K, Eckhart L, Rolli LM, Lenhof HP. Trust me if you can: a survey on reliability and interpretability of machine learning approaches for drug sensitivity prediction in cancer. Brief Bioinform 2024;25:bbae379. https://doi.org/10.1093/bib/bbae379.

[5] Musolf AM, Holzinger ER, Malley JD, Bailey-Wilson JE. 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] Dey D, Haque MS, Islam MM, Aishi UI, Shammy SS, Mayen MSA, et al. The proper application of logistic regression model in complex survey data: a systematic review. BMC Med Res Method 2025;25:15. https://doi.org/10.1186/s12874-024-02454-5.

[7] Pinheiro-Guedes L, Martinho C, Martins MR. Logistic regression: limitations in the estimation of measures of association with binary health outcomes. Acta Med Port 2024;37:697–705. https://doi.org/10.20344/amp.21435.

[8] Wang T, Tang W, Lin Y, Su W. Semi-supervised inference for nonparametric logistic regression. Stat Med 2023;42:2573–89. https://doi.org/10.1002/sim.9737.

[9] Bilodeau B, Jaques N, Koh PW, Kim B. Impossibility theorems for feature attribution. Proc Natl Acad Sci 2024;121:e2304406120. https://doi.org/10.1073/pnas.2304406120.

[10] 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.

[11] Hooshyar D, Yang Y. Problems with SHAP and LIME in interpretable AI for education: a comparative study of post-hoc explanations and neural-symbolic rule extraction. IEEE Access 2024;12:137472–90. https://doi.org/10.1109/ACCESS.2024.3463948.

[12] Zhang J, Wu X, Hoi SCH, Zhu J. Feature agglomeration networks for single stage face detection. Neurocomputing 2020;380:180–9. https://doi.org/10.1016/j.neucom.2019.10.087.

[13] Xie Y, Jing Z, Pan H, Xu X, Fang Q. Redefining the high variable genes by optimized LOESS regression with positive ratio. BMC Bioinf 2025;26:104. https://doi.org/10.1186/s12859-025-06112-5.

[14] Okoye K, Hosseini S. Correlation tests in R: Pearson Cor, Kendall's Tau, and Spearman's Rho. In: Okoye K, Hosseini S, editors. R programming: statistical data analysis in research. Nature: Springer; 2024. p. 247–77. https://doi.org/10.1007/978-981-97-3385-9_12.

[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.

Souichi Oka[a,*], Yoshiki Takahashi[a], Yoshiyasu Takefuji[b]
[a] *Science Park Corporation, 3-24-9 Iriya-Nishi Zama-shi, Kanagawa 252-0029, Japan*
[b] *Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan*

[*] Corresponding author.
E-mail addresses: souichi.oka@sciencepark.co.jp (S. Oka), ytakahashi@sciencepark.co.jp (Y. Takahashi), takefuji@keio.jp (Y. Takefuji).