Mitigating feature importance bias in regression models for clinical decision-making

TO THE EDITORS: Brennand et al proposed a novel approach allowing patients to choose between uterine-preserving surgery and hysterectomy for pelvic organ prolapse.[1] They utilized inverse probability of treatment weighting in linear regression and modified Poisson regression to estimate adjusted mean differences and relative risks, respectively. A logistic regression algorithm helped identify and prioritize covariates based on their associations with treatment exposure and outcomes.[1]

However, concerns arise regarding logistic regression's use for generating feature importance metrics for covariate selection due to significant biases.[2,3] This paper emphasizes that feature importance derived from such models can lead to misleading conclusions due to inherent biases. This paper advocates for robust statistical methods,[4] such as chi-squared tests and Spearman's correlation, to accurately identify true associations between targets and features. By emphasizing reliable statistical approaches, their study seeks to strengthen the validity of its findings and facilitate informed treatment decisions. In essence, researchers should prioritize true associations over biased feature importances in their studies.

Linear and logistic regression are common statistical tools in machine learning and data analysis for understanding variable relationships and making predictions. Both methods can introduce biases in determining feature importance due to their handling of variables.[2,3] Linear regression predicts a continuous output via a linear combination of input features. Feature importance is inherently assessed through model coefficients, which can misrepresent relationships—especially with nonlinear interactions or when multicollinearity obscures individual contributions among correlated features.

In contrast, logistic regression, used for binary classification, calculates the probability of a binary outcome based on features. Like linear regression, it typically derives feature importance from coefficients but suffers from class imbalance and independence assumptions. Overrepresentation of one class can skew importance toward predictive features of the majority class, overshadowing those relevant to minorities. High feature correlation can further dilute accurate attribution of importance among predictors.

Despite offering valuable insights, users must interpret feature importance derived from these models cautiously. Assumptions about relationships, multicollinearity, and class imbalance can distort the perceived significance of features. Researchers should clearly understand the distinctions among machine learning prediction and feature importance calculations. Although machine learning primarily aims for accurate predictions, feature importances derived from these models consistently reflect model-specific biases rather than true associations. This paper advocates focusing on genuine relationships to enhance the validity of findings, promoting a nuanced understanding of influencing factors. By doing so, researchers can improve the reliability of predictions and inform better decision-making across various applications.

Data availability

Not applicable. ■

Yoshiyasu Takefuji, PhD
Faculty of Data Science
Musashino University
Tokyo, Japan
takefuji@keio.jp

The author reports no conflict of interest.

This research has no fund.

REFERENCES

1. Brennand EA, Scime NV, Huang B, et al. Hysterectomy versus uterine preservation for pelvic organ prolapse surgery (HUPPS): a prospective cohort study. Am J Obstet Gynecol 2024. https://doi.org/10.1016/j.ajog.2024.10.021.

2. Chen J, Ooi LQR, Tan TWK, et al. Relationship between prediction accuracy and feature importance reliability: an empirical and theoretical study. Neuroimage 2023;274:120115. https://doi.org/10.1016/j.neuroimage.2023.120115.

3. Thakur D, Biswas S. Permutation importance based modified guided regularized random forest in human activity recognition with smartphone. Eng Appl Artif Intell 2024;129:107681. https://doi.org/10.1016/j.engappai.2023.107681.

4. Alam R, Srivastava A, Patel HD. Chapter 24 - categorical variable analyses: chi-square, Fisher exact, and mantel–haenszel. In: Eltorai AEM, Arab A, Atala A, Siddiqui MM, eds. Handbook for designing and conducting clinical and translational research. Translational urology. Cambridge, Massachusetts: Academic Press; 2025:115–20. https://doi.org/10.1016/B978-0-323-90186-4.00035-3.

 https://doi.org/10.1016/j.ajog.2024.12.010