7. Murad MH, et al. J Clin Epidemiol 2009;62:865–867.
8. Guyatt GH, et al. J Clin Epidemiol 2011;64:1283–1293.
9. Higgins JPT, et al, eds. www.training.cochrane.org/handbook. Updated August 22, 2024.

Conflicts of interest
This author discloses the following: Siddharth Singh has received institutional grants from Pfizer. The remaining authors disclose no conflicts.

Most current article

https://doi.org/10.1053/j.gastro.2025.01.012

Enhancing Colorectal Cancer Subtyping: Addressing Limitations of L1-Penalized Estimation in Alternative Splicing Analysis

Dear Editors:

Ambeskovic et al[1] introduced an innovative colorectal cancer subtyping framework based on alternative splicing, which is particularly sensitive to variations in exon usage and has potential for clinical application. This framework uses a feature selection process that combines bootstrapping with L1-penalized estimation to improve the robustness and accuracy of subtyping. Ambeskovic et al[1] filtered percent spliced-in values using F statistics and calculated stability scores for exon-skipping events by assessing the maximum selection frequencies of each event through bootstrapping and L1-penalized estimation. This approach has demonstrated a significant reduction in false positives compared to traditional cross-validation methods.[1]

However, although Ambeskovic et al[1] used L1-penalized estimation within their framework, there is an inherent risk of overlooking important features due to the limitations associated with linear and parametric modeling.[2–5] This concern is particularly relevant when the analysis involves correlated or nonlinear relationships among features. Specifically, L1-penalized estimation can lead to the omission of significant predictors as it tends to favor sparsity and may arbitrarily select one feature over others. As a result, this could compromise the overall effectiveness of the model by failing to capture essential interactions and dependencies, ultimately impacting the robustness of the subtyping framework.

To address these limitations, it is recommended that the investigators from Ambeskovic et al[1] consider incorporating nonlinear and nonparametric methods, such as Spearman's correlation with *P* values[6,7] or Kendall's tau with *P* values[8] in addition to their existing reliance on L1-penalized estimation. These alternative approaches can provide a more comprehensive evaluation of feature importance as they are specifically designed to capture the complexities inherent in relationships within high-dimensional data. By implementing these strategies, the investigators could greatly enhance feature selection and optimize the overall performance of their colorectal cancer subtyping framework, ultimately yielding more accurate and clinically relevant insights. Furthermore, to accurately assess associations between variables, it is crucial to evaluate data distribution, investigate the relationships among variables, and conduct rigorous statistical validation through *P* values. This holistic approach will strengthen the reliability and interpretability of the subtyping results.

YOSHIYASU TAKEFUJI
Faculty of Data Science
Musashino University
Tokyo, Japan

References

1. Ambeskovic A, et al. Gastroenterology 2024;167:1358–1370.e12.
2. Banerjee I, et al. J Am Coll Radiol 2023;20:842–851.
3. Wang H, et al. Bioinformatics 2012;28:2834–2842.
4. Fakhraei S, et al. Expert Syst Appl 2014;14:6945–6958.
5. Mabrouk OM, et al. Sci Rep 2024;14:28430.
6. Yu H, Hutson AD. Commun Stat Theory Methods 2024;53:2141–2153.
7. Eden SK, et al. Biometrics 2022;78:421–434.
8. Wang JH, Chen YH. Bioinformatics 2021;37:2150–2156.

Conflicts of interest
The author discloses no conflicts.

Most current article

https://doi.org/10.1053/j.gastro.2025.01.013

Reply. We appreciate the letter from Dr Takefuji[1] commenting on our recently published study in *Gastroenterology*[2] describing an exon-skipping–based methodology for subtyping colorectal cancers. The comments emphasize the potential omission of important correlated features within the feature selection process of our methodology that combines bootstrapping with L1-penalized estimation. Additionally, the incorporation of nonlinear and nonparametric methods such as Spearman's correlation for enhancing feature selection reliability and interpretability of the subtyping results are proposed in the comments.

In our study, we aimed to develop an interpretable low-complexity colorectal subtype identifier that could be adapted for clinical application. Although we agree that L1-penalized estimation may select only one feature among a set of highly correlated variables, due to the enforced sparsity, it also ensures selection of features that carry unique and nonredundant signals. When several features are highly correlated, selecting just one from the group typically maintains most of the predictive performance.[3]

A limitation of feature selection methods that use L1 regularization is their tendency to select different subsets of features in response to small variations in the training data.[4]