American Journal of Transplantation xxx (xxxx) xxx

Contents lists available at ScienceDirect

American Journal of Transplantation

journal homepage: www.amjtransplant.org

Letter to the Editor

Beyond least absolute shrinkage and selection operator: A comprehensive framework for analyzing complex biological data using nonlinear and nonparametric methods

Yoshiyasu Takefuji *

Faculty of Data Science, Musashino University, Tokyo, Japan

Ma et al[1] introduced an innovative approach by incorporating donor-lung computed tomographic images into machine learning models for predicting severe primary graft dysfunction after lung transplantation. Their methodology employed least absolute shrinkage and selection operator (LASSO) regression to enhance variable selection and model interpretability, using 100 iterative runs to identify consistently significant clinical features as reliable predictors.

This implementation, however, raises important concerns about the application of LASSO to biological data characterized by nonlinear behaviors. The fundamental reliance of LASSO on linear parametric assumptions may exclude essential nonlinear features, leading to significant oversights in capturing biological interactions. Studies demonstrate that when analyzing data with nonlinear characteristics, LASSO can distort outcomes,[2-4] highlighting the need for alternative methodologies better suited to the inherent complexities of biological systems.

Although Akaike's Information Criterion effectively balances model fit and complexity when overfitting risks exist, assessing LASSO-derived feature importances presents unique challenges due to the absence of direct benchmarks. Inconsistent methodologies across different models can result in biased significance assessments.[2-4] Importantly, high prediction accuracy does not guarantee valid feature importances, as these measures can diverge significantly in practice. This disconnect underscores the necessity for methodologies providing more reliable assessments of feature relevance in biological data analysis.

Recent research has further highlighted the limitations of LASSO with nonlinear or nonparametric data.[2-4] This fundamental misalignment often leads to violations of the core assumptions of LASSO, resulting in skewed outcomes and erroneous interpretations. Despite its popularity, the effectiveness of LASSO diminishes considerably when ground truth values are not available to validate feature importance rankings.

Radiomics data sets typically exhibit complex, nonlinear relationships between features and clinical outcomes. Applying linear methods such as LASSO to data violating underlying assumptions risks producing misleading models by discarding important nonlinear patterns. Researchers should first confirm their data satisfies linearity requirements before applying such techniques. LASSO may overlook critical biological relationships including higher-order interactions among genes within signaling pathways and complex dependencies in transcription regulation networks.

To address these limitations, I propose a comprehensive analytical framework better suited to biological data complexities. This includes employing Random Forests for target predictions, as

DOI of original article: https://doi.org/10.1016/j.ajt.2025.01.039.

* Corresponding author. 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan.
E-mail address: takefuji@keio.jp.

https://doi.org/10.1016/j.ajt.2025.06.021
Received 3 April 2025; Received in revised form 2 June 2025; Accepted 17 June 2025
Available online xxxx

Please cite this article as: Takefuji Y, Beyond least absolute shrinkage and selection operator: A comprehensive framework for analyzing complex biological data using nonlinear and nonparametric methods, American Journal of Transplantation, https://doi.org/10.1016/j.ajt.2025.06.021

they more effectively manage nonlinear relationships between variables. For detecting true associations, nonlinear nonparametric methods such as Spearman's correlation and Kendall's tau offer greater robustness to outliers without assuming linear relationships. For complex nonmonotonic interactions, information-theoretic approaches such as Mutual Information analysis and Effective Transfer Entropy can capture intricate dependencies without presuming specific relationship structures.[5] This framework aims to provide more accurate analysis of biological data while minimizing the risk of overlooking crucial relationships or drawing incorrect conclusions about feature importance.

Finally, both LASSO and Random Forests provide feature importance measures that reflect the contributions of variables to predictive accuracy within each algorithm's framework—not absolute measures of causal effect. These rankings require cautious interpretation and corroboration through domain knowledge and targeted experiments.

To balance interpretability with the capacity to capture complex nonlinear relationships, I propose a 2-stage workflow. First, use mutual information and transfer entropy to screen dependencies among features and outcomes, identifying informative variables and potential nonlinear interactions. Second, build a sparse model (LASSO or generalized additive model) using only top-ranked features and identified interaction terms. This creates an interpretable predictor complemented by a network visualizing underlying interdependencies, combining the clarity of linear models with the flexibility of information-theoretic methods.

When relationships are approximately linear and parametric assumptions hold, simpler models often remain optimal. Linear methods offer transparent effect estimates, established inference procedures, and protection against overfitting. Before employing nonlinear tools, researchers should conduct exploratory analyses to assess linearity and distributional assumptions. If diagnostics confirm a parsimonious linear model fits well, simplicity should prevail. Nonlinear methods add value only when genuine complexity exists, following the principle that model selection should balance the characteristics of the data with predictive accuracy and interpretability requirements.

Author contributions

Y.T. completed this research and wrote this article.

Funding

The author received no funding for this study.

Declaration of competing interest

The author of this manuscript has no conflicts of interest to disclose as described by *American Journal of Transplantation*.

ORCiD

Yoshiyasu Takefuji https://orcid.org/0000-0002-1826-742X

References

1. Ma W, Oh I, Luo Y, et al. Developing approaches to incorporate donor-lung computed tomography images into machine learning models to predict severe primary graft dysfunction after lung transplantation. *Am J Transplant*. 2025;25(6):1339–1349. https://doi.org/10.1016/j.ajt.2025.01.039.
2. Tibshirani RJ. The lasso problem and uniqueness. *Electron J Stat*. 2013;7:1456–1490. https://doi.org/10.1214/13-EJS815.
3. Climente-González H, Azencott CA, Kaski S, Yamada M. Block HSIC Lasso: model-free biomarker detection for ultra-high dimensional data. *Bioinformatics*. 2019;35(14):i427–i435. https://doi.org/10.1093/bioinformatics/btz333.
4. He D, Sun S, Xie L. Multi-target HSIC-Lasso via exploiting target correlations. *Expert Syst Appl*. 2024;242:122845. https://doi.org/10.1016/j.eswa.2023.122845.
5. Wang J, Xu Y, Liu L, et al. Comparison of LASSO and random forest models for predicting the risk of premature coronary artery disease. *BMC Med Inform Decis Mak*. 2023;23(1):297. https://doi.org/10.1186/s12911-023-02407-w.