European Neuropsychopharmacology 101 (2025) 18–19

Contents lists available at ScienceDirect

European Neuropsychopharmacology

journal homepage: www.sciencedirect.com/journal/european-neuropsychopharmacology

A call for more robust and interpretable models in predicting treatment-resistant depression

ARTICLE INFO

Keywords:
Major depression
Machine learning
Feature importance
Model bias
Statistical validation

Letter to the editor

Serretti et al. (2025) proposed a machine learning framework using eXtreme Gradient Boosting (XGBoost) and SHapley Additive exPlanation (SHAP) to predict treatment-resistant depression (TRD) in a large, multicenter cohort of 2953 patients. They utilized eXtreme Gradient Boosting (XGBoost), achieving a ROC AUC of 0.80, and subsequently applied SHapley Additive exPlanation (SHAP) to assess feature importance. This analysis highlighted key predictors such as Duration of Current Episode, Duration of Disease, and other indicators of illness chronicity. However, while the model demonstrated strong predictive performance, it is important to note that high accuracy in target prediction does not necessarily validate the reliability of feature importance rankings. This distinction raises concerns about potential biases in the analytical pipeline, which may undermine the clinical translatability of the identified predictors. Overreliance on predictive accuracy to justify feature relevance is a well-known issue (Fisher et al., 2019). A significant body of literature has highlighted that strong prediction does not ensure meaningful attribution. Feature importance rankings often reflect model artifacts rather than true causal relationships.

Tree-based machine learning models, including XGBoost, are known to exhibit biases in feature importance estimation—particularly a tendency to favor features that enable early splits in the decision tree. This bias is especially pronounced in high-dimensional clinical datasets, where complex correlations and multicollinearity are common, and the risk of overfitting is elevated (Ugirumurera et al., 2024). In Serretti et al.'s study, several top-ranked predictors fall into this category, raising concerns about potential multicollinearity and interpretability. Given the absence of a definitive ground truth for feature relevance, and the challenges in disentangling correlated effects, feature importance rankings in such models should be interpreted with caution, as they may not reliably reflect underlying causal relationships.

Furthermore, SHAP's reliance on the model's internal logic means its attributions are shaped by the structural assumptions and biases of the algorithm it explains. In the case of XGBoost, the presence of correlated features can lead SHAP to distribute importance arbitrarily, resulting in unstable and potentially misleading rankings. SHAP does not correct for these distortions; its outputs reflect the model's internal heuristics rather than the independent clinical relevance of the predictors. Therefore, interpreting SHAP-derived feature importance requires model-agnostic validation and careful scrutiny.

To address these limitations, a more robust analytical framework is needed—one that goes beyond model-dependent attribution and accounts for the structural complexity of clinical data. Techniques such as Feature Agglomeration (Zhang et al., 2020) and Highly Variable Gene Selection (Xie et al., 2025) offer unsupervised approaches to dimensionality reduction that can mitigate issues arising from multicollinearity and feature redundancy. Complementing these with non-parametric statistical methods like Spearman's rho or Kendall's tau enables the detection of non-linear and rank-based associations, enhancing interpretability. Together, these strategies provide a more stable foundation for identifying clinically meaningful predictors and improving the translatability of machine learning findings in psychiatric research.

CRediT authorship contribution statement

Souichi Oka: Writing – original draft, Conceptualization. Kota Takemura: Investigation. Yoshiyasu Takefuji: Writing – review & editing, Supervision, Project administration.

Funding sources

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.

Data availability

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

https://doi.org/10.1016/j.euroneuro.2025.09.012
Received 19 September 2025; Received in revised form 23 September 2025; Accepted 28 September 2025

References

Fisher, A., Rudin, C., Dominici, F., 2019. 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. 20, 177.
Serretti, A., Kasper, S., Bartova, L., Zohar, J., Souery, D., Montgomery, S., Ferentinos, P., Rujescu, D., Kautzky, A., Attanasio, F., Zanardi, R., Fabbri, C., Baune, B.T., Ferri, R., Mendlewicz, J., 2025. Clinical predictors of treatment resistant depression. Eur. Neuropsychopharmacol. 98, 26–34.
Ugirumurera, J., Bensen, E.A., Severino, J., Sanyal, J., 2024. Addressing bias in bagging and boosting regression models. Sci. Rep. 14, 18452.
Xie, Y., Jing, Z., Pan, H., Liu, S., Li, J., 2025. Redefining the high variable genes by optimized LOESS regression with positive ratio. BMC Bioinformat. 26, 104.
Zhang, J., Wu, X., Hoi, S.C.H., Wang, L., Zhu, J., 2020. Feature agglomeration networks for single stage face detection. Neurocomputing 380, 180–189.

Souichi Oka[a,*], Kota Takemura[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 at: SciencePark Corporation, 3-24-9 Iriya-Nishi Zama-shi, Kanagawa 252-0029, Japan.
E-mail addresses: souichi.oka@sciencepark.co.jp (S. Oka), ktakemura@sciencepark.co.jp (K. Takemura), takefuji@keio.jp (Y. Takefuji).