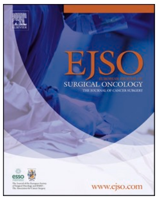
European Journal of Surgical Oncology 51 (2025) 109687

Contents lists available at ScienceDirect

European Journal of Surgical Oncology

journal homepage: www.ejso.com



Enhancing feature importance analysis with Spearman’s correlation with p-values: Recommendations for improving PHLF prediction

ARTICLE INFO

Keywords:
Post-hepatectomy liver failure (PHLF)
Machine learning models
Feature importance
XGBoost
Random forest
SHAP (shapley additive explanations)

Famularo et al. employed several machine learning models to predict the risk of post-hepatectomy liver failure (PHLF) in hepatocellular carcinoma (HCC) patients [1]. They used Extreme Gradient Boosting (XGB), Random Forest (RF), and Support Vector Machine (SVM) models, trained with clinical and radiomic features. Hyperparameters were optimized through grid search, and model performance was evaluated using AUC, accuracy, sensitivity, and specificity. The feature importance of the models was assessed, and SHAP (Shapley Additive Explanations) was applied to the XGBoost and Random Forest models to explain the contribution of individual features to the model’s predictions.

Machine learning techniques like XGBoost are widely used, but different machine learning models employ distinct methodologies for calculating feature importance, and this can lead to varying degrees of bias. Feature importance from machine learning models is always inherently biased, which is a significant issue [2,3]. Therefore, although Famularo et al. possess considerable expertise, it is equally necessary to fully consider the complexities of algorithmic calculations and their associated biases. To avoid non-negligible biased feature importances, it is recommended to rely on robust statistical methods that assess genuine associations between variables, such as Spearman’s correlation with P-values, non-linear and nonparametric approaches [4].

XGBoost (Extreme Gradient Boosting) is a decision tree-based ensemble algorithm known for its effectiveness but may introduce biases in feature importance calculations. It uses metrics like gain (accuracy improvement contribution), coverage (relative observation association), and frequency (feature usage). These metrics are model-specific, meaning feature importance rankings vary with model configurations, making them non-absolute. Correlated features can distort importance evaluations, leading to underestimation or overestimation of contributions. Feature interactions further complicate assessments, and overfitting can inflate the importance of features that perform poorly on new data. The sequential tree-building approach may disproportionately emphasize features that correct previous errors [2].

Random Forest, another powerful machine learning algorithm, excels in feature selection and prediction. It constructs numerous decision trees and combines their predictions to improve stability and accuracy. However, Random Forest can introduce bias with categorical features having many unique values, which might give high importance to features with little predictive impact. Additionally, it may fail to capture non-linear relationships, underestimating feature importance in such cases. Class imbalance can also skew feature importance towards features prevalent in the majority class, neglecting crucial features in the minority class [2].

SHAP is a reliable explanation framework, but it is not without biases [3]. Its dependence on the model means SHAP values can inherit and amplify biases present in the model. Moreover, SHAP assumes that features are independent when assessing their contributions, which may not always hold true. When features are correlated, their combined impact may distort the overall interpretation. SHAP values can also be particularly sensitive to outliers or noise in the dataset, which can result in misleading evaluations of feature importance. Features that are typically of low importance can be disproportionately influenced by outliers, causing skewed importance metrics. Furthermore, SHAP may misrepresent contributions when features have non-linear relationships with the target, particularly in complex models like XGBoost.

In conclusion, while both XGBoost and Random Forest are widely used methods, caution is needed when interpreting feature importance due to the potential biases inherent in these models. Additionally, while SHAP values provide a structured approach to explaining model predictions, they are inherently influenced by the model’s biases and underlying assumptions, particularly regarding feature independence. Given these considerations, the authors may want to consider incorporating robust statistical methods, such as Spearman’s correlation with p-values [4], to supplement machine learning-derived feature importance. This would provide more reliable insights into the true associations between features and outcomes, thereby enhancing the validity of their findings. By doing so, they can uncover genuine relationships between features and outcomes, ultimately advancing knowledge in predicting PHLF occurrence preoperatively, while mitigating the risks associated with biased interpretations.

https://doi.org/10.1016/j.ejso.2025.109687
Received 16 January 2025; Accepted 10 February 2025

Consent to participate

Not applicable.

Ethics approval

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Not applicable.

Code availability

Not applicable.

Authors' contributions

Haoqian Pan investigated and wrote this article, Yoshiyasu Takefuji supervised and wrote this article.

Funding

This research has no fund.

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.

References

[1] Famularo S, Maino C, Milana F, Ardito F, Rompianesi G, Ciulli C, et al. Preoperative prediction of post hepatectomy liver failure after surgery for hepatocellular carcinoma on CT-scan by machine learning and radiomics analyses. Eur J Surg Oncol 2024:109462. https://doi.org/10.1016/j.ejso.2024.109462.

[2] Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources, and a solution. BMC Bioinf 2007;8(1):25. https://doi.org/10.1186/1471-2105-8-25.

[3] Bilodeau B, Jaques N, Koh PW, Kim B. Impossibility theorems for feature attribution. Proc Natl Acad Sci USA 2024;121(2):e2304406120. https://doi.org/10.1073/pnas.2304406120.

[4] Andréas H, Alfonso V. The Kendall and spearman rank correlations of the bivariate skew normal distribution. Scand J Stat 2022;49:1669–98. https://doi.org/10.1111/sjos.12587.

Haoqian Pan, Yoshiyasu Takefuji*
Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku,
Tokyo, 135-8181, Japan

* Corresponding author.
E-mail addresses: g2450009@stu.musashino-u.ac.jp (H. Pan),
takefuji@keio.jp (Y. Takefuji).