Beyond predictive accuracy: interpreting feature importances in pulmonary endarterectomy risk models with total correlation analysis

Received: 25 April 2025
Accepted: 6 May 2025

To the Editor:

Liley *et al.* [1] developed and internally validated an open-source risk-assessment tool for pulmonary endarterectomy by comparing three modelling approaches: ordinary linear regression, penalised linear regression and random forest. For each surgical outcome, models were calibrated, fitted and evaluated *via* k-fold cross-validation to ensure robustness. Across all end-points, the random forest models delivered the highest predictive accuracy and highlighted the key clinical and haemodynamic predictors of perioperative risk in chronic thromboembolic pulmonary hypertension [1].

However, despite demonstrating excellent predictive performance, relying on conventional machine learning-derived feature importances carries substantial interpretative risks. Unlike prediction accuracy, which can be directly measured against known outcomes, importance scores from different algorithms have no external "ground truth" for validation. Ordinary linear regression ranks features by coefficient magnitude, penalised regression shrinks some coefficients toward zero, and random forests use metrics such as split frequency or impurity reduction. Because each method embodies its own bias and heuristics, they often produce conflicting importance rankings that may reflect idiosyncrasies of the model rather than true underlying relationships [2–5].

Both ordinary and penalised linear regressions impose a fixed, additive functional form: each outcome is modelled as a linear combination of predictors, and the only flexibility comes from shrinking coefficients toward zero. In contrast, biological and haemodynamic data frequently violate these assumptions, exhibiting threshold effects, saturations and intricate synergies among variables. When you fit a strictly linear, parametric model to such nonparametric, nonlinear relationships, the result is systematic under-fitting: true signal is smoothed over, variance is misallocated, and coefficient estimates become distorted. In practice, this mis-specification not only diminishes predictive fidelity but also produces misleading feature rankings, overemphasising variables that happen to align with the imposed linear structure, while hiding those that drive risk through more complex, nonlinear pathways.

Extensive evidence – spanning more than 100 peer-reviewed studies – shows that algorithm-specific feature-importance measures often mislead: they can exaggerate the contribution of highly correlated covariates, ignore subtler but synergistic predictors, mistake spurious associations for causal drivers, and fail to detect variables whose effects emerge only through complex, higher-order interactions. While Liley *et al.* [1] acknowledge the power of machine learning to deliver high predictive accuracy, they caution that model-derived importance scores lack an external "ground truth" and thus should not be taken at face value. Instead, such metrics must be interpreted with great care – or better yet, validated against independent, nonparametric approaches – before being used to inform clinical decisions or infer underlying pathophysiology.

To overcome these limitations, this paper advocates augmenting their modelling framework with a nonlinear, nonparametric information-theoretic approach – total correlation (TC) analysis – for uncovering complex interactions and nonmonotonic patterns. "Complex interactions" occur when the combined effect

Shareable abstract (@ERSpublications)
Predictive accuracy in machine learning does not ensure reliable feature importance; it demands validation *via* ground-truth-free non-parametric methods. Total correlation reveals higher-order interactions and non-monotonic patterns in clinical risk. https://bit.ly/3YDGTQC

Cite this article as: Takefuji Y. Beyond predictive accuracy: interpreting feature importances in pulmonary endarterectomy risk models with total correlation analysis. *Eur Respir J* 2025; 66: 2500893 [DOI: 10.1183/13993003.00893-2025].

of two or more predictors differs from the sum of their individual effects, for example, when a haemodynamic measurement only increases surgical risk above a critical threshold of exercise capacity. "Nonmonotonic patterns" arise when a predictor's association with the outcome changes direction across its range, such as a U-shaped relationship between mean pulmonary arterial pressure and postoperative complications. Traditional regression models assume linear (or simple penalised) relationships and thus struggle to accommodate threshold or bidirectional effects, while random forests, though flexible, partition the predictor space *via* axis-aligned splits and can miss subtle, higher-order dependencies. In contrast, TC quantifies the total shared information among any set of variables without presuming linearity or additivity, enabling simultaneous detection of synergy, redundancy and higher-order interactions that drive perioperative risk in chronic thromboembolic pulmonary hypertension.

To assist researchers in dissecting these intricate dependencies, clinical data analysis tools such as multivariate mutual information (MI) and its advanced extension, TC, have proven invaluable [6–10]. MI measures the shared information between multiple variables with nonmonotonic patterns and the outcome, capturing nonlinear and nonadditive dependencies. However, MI alone cannot distinguish whether this information arises from redundant signals, where different predictors convey overlapping content, or from synergistic effects that only emerge when variables act jointly. TC, also known as multi-information, extends MI by quantifying the total correlation within a predictor set and decomposing it into redundancy and synergy. By first using MI to flag variable groups exhibiting nonmonotonic associations and then applying TC to unravel their internal information structure, investigators can detect complex, higher-order interactions that linear regressions and axis-aligned tree splits typically miss.

Yoshiyasu Takefuji

Faculty of Data Science, Musashino University, Tokyo, Japan.

Corresponding author: Yoshiyasu Takefuji (takefuji@keio.jp)

Author contributions: Y. Takefuji completed this research and wrote this article.

Conflicts of interest: The author has no potential conflicts of interest to disclose.

References

1 Liley J, Bunclark K, Newnham M, *et al.* Development of an open-source tool for risk assessment in pulmonary endarterectomy. *Eur Respir J* 2025; 65: 2401001.
2 Zuur AF, Ieno EN, Walker NJ, *et al.* Limitations of linear regression applied on ecological data. *In:* Zuur AF, Ieno EN, Walker NJ, *et al.*, eds. Mixed Effects Models and Extensions in Ecology with R. Statistics for Biology and Health. New York, Springer, 2009; pp. 11–33.
3 Anandhi P, Nathiya E. Application of linear regression with their advantages, disadvantages, assumption and limitations. *Int J Stat Appl Math* 2023; 8: 133–137.
4 Wallace ML, Mentch L, Wheeler BJ, *et al.* Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction. *BMC Med Res Methodol* 2023; 23: 144.
5 Salles T, Rocha L, Gonçalves M. A bias-variance analysis of state-of-the-art random forest text classifiers. *Adv Data Anal Classif* 2021; 15: 379–405.
6 Rayan RA, Suruliandi A, Raja SP. Modified mutual information feature selection algorithm to predict COVID-19 using clinical data. *Comput Methods Biomech Biomed Engin* 2024; in press [https://doi.org/10.1080/10255842.2024.2429012].
7 Zou J, Li Z, Carleton N, *et al.* Mutual information for detecting multi-class biomarkers when integrating multiple bulk or single-cell transcriptomic studies. *Bioinformatics* 2024; 40: btae696.
8 Aboonq MS, Alqahtani SA. Leveraging multivariate analysis and adjusted mutual information to improve stroke prediction and interpretability. *Neurosciences (Riyadh)* 2024; 29: 190–196.
9 Xiong Y, Wang J, Shang X, *et al.* Efficient and scalable construction of clinical variable networks for complex diseases with RAMEN. *Cell Rep Methods* 2025; 5: 101022.
10 Adamis D, Singh J, Coada I, *et al.* Measuring clinical outcomes in adult ADHD clinics: psychometrics of a new scale, the adult ADHD Clinical Outcome Scale. *BJPsych Open* 2024; 10: e180.