BJA

British Journal of Anaesthesia, xxx (xxx): xxx (xxxx)

CORRESPONDENCE

Preoperative risk prediction of major cardiovascular events in noncardiac surgery using the 12-lead electrocardiogram: convolutional neural network prediction. Comment on *Br J Anaesth* 2025; 135: 1161–71

Yoshiyasu Takefuji*

Faculty of Data Science, Musashino University, Tokyo, Japan

E-mail: takefuji@keio.jp

Keywords: convolutional neural network; counterfactual ECGs; feature agglomeration; highly variable feature selection; Spearman correlation; stability of feature rankings

Editor—Harris and colleagues[1] examined whether routine preoperative electrocardiograms (ECGs) can improve prediction of major cardiovascular complications after noncardiac surgery, and whether combining ECGs with standard clinical information adds value. They found that convolutional neural network (CNN) models using ECG waveforms modestly outperformed a widely used clinical risk score, indicating that ECGs carry clinically meaningful prognostic signals. Crucially, they provided clear explanations of which ECG features were associated with higher risk, helping clinicians understand why certain patients were flagged. For a clinical readership, the practical takeaways are that preoperative ECGs could offer incremental prognostic benefit, integration with routine variables can further enhance prediction, and interpretable outputs could support trust and real-world use, all directly relevant to refining perioperative risk assessment and patient counselling.[1]

However, a prediction-driven interpretability approach based on counterfactual ECGs is not recommended. High target prediction accuracy in supervised models such as CNNs does not guarantee reliable explanations because interpretation is model specific and can be misleading.[2–10] Supervised models involve two distinct notions of accuracy: target prediction accuracy, which can be validated against ground-truth labels, and feature importance, which generally lacks ground truth for accuracy validation. Thus, strong cross-validated area under the receiver operating characteristic curve (AUROC) and convincing counterfactuals do not ensure trustworthy feature attributions owing to the absence of ground truth in interpretation.[11–19] In other words, high target prediction accuracy does not guarantee reliable feature importances owing to the absence of ground truth and label-driven errors.

Given the lack of ground truth for feature importance, more robust, multifaceted strategies are warranted. We recommend using CNN models for target prediction and performing feature interpretation with unsupervised methods, specifically feature agglomeration and highly variable feature selection, followed by non-target, nonlinear, nonparametric analyses such as Spearman correlation with *P*-values. Compared with label-driven supervised models, which often yield unstable and model-specific feature rankings, these structure-discovering techniques and monotonic association tests provide more stable and reproducible feature prioritisation.

In short, Harris and colleagues[1] advocate CNN for prediction and a complementary unsupervised, nonparametric framework for interpretation, rather than relying on prediction-driven counterfactual ECGs. Stability in feature rankings can be evaluated simply: select the top n features from the full set (set 1), remove the highest-ranked feature to create a reduced dataset, then reselect the top n−1 features (set 2) and compare the ranking orders between set 1 and set 2.

Declaration of interest

The author declares that they have no conflict of interest.

References

1. Harris C, Pimpalkar A, Aggarwal A, et al. Preoperative risk prediction of major cardiovascular events in noncardiac surgery using the 12-lead electrocardiogram: an

DOI of original article: 10.1016/j.bja.2025.07.085.

explainable deep learning approach. *Br J Anaesth* 2025; **135**: 1161–71
2. Antamis T, Drosou A, Vafeiadis T, Nizamis A, Ioannidis D, Tzovaras D. Interpretability of deep neural networks: a review of methods, classification and hardware. *Neurocomputing* 2024; **601**, 128204
3. Fan FL, Xiong J, Li M, Wang G. on interpretability of artificial neural networks: a survey. *IEEE Trans Radiat Plasma Med Sci* 2021; **5**: 741–60
4. Liu Z, Xu F. Interpretable neural networks: principles and applications. *Front Artif Intell* 2023; **6**, 974295
5. van Hilten A, Katz S, Saccenti E, Niessen WJ, Roshchupkin GV. Designing interpretable deep learning applications for functional genomics: a quantitative analysis. *Brief Bioinform* 2024; **25**. bbae449
6. Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci U S A* 2019; **116**: 22071–80
7. Makkeh A, Graetz M, Schneider AC, Ehrlich DA, Priesemann V, Wibral M. A general framework for interpretable neural learning based on local information-theoretic goal functions. *Proc Natl Acad Sci U S A* 2025; **122**, e2408125122
8. Coupland H, Scheidwasser N, Katsiferis A, et al. Exploring the potential and limitations of deep learning and explainable AI for longitudinal life course analysis. *BMC Public Health* 2025; **25**: 1520
9. Seo B, Li J. Explainable machine learning by SEE-Net: closing the gap between interpretable models and DNNs. *Sci Rep* 2024; **14**, 26302
10. Budhkar A, Song Q, Su J, Zhang X. Demystifying the black box: a survey on explainable artificial intelligence (XAI) in bioinformatics. *Comput Struct Biotechnol J* 2025; **27**: 346–59
11. Parr T, Hamrick J, Wilson JD. Nonparametric feature impact and importance. *Inf Sci* 2024; **653**, 119563
12. Watson DS, Wright MN. Testing conditional independence in supervised learning algorithms. *Mach Learn* 2021; **110**: 2107–29
13. Lipton ZC. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. *Queue* 2018; **16**: 31–57
14. Fisher A, Rudin C, Dominici F. 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* 2019; **20**: 177
15. Lenhof K, Eckhart L, Rolli LM, Lenhof HP. Trust me if you can: a survey on reliability and interpretability of machine learning approaches for drug sensitivity prediction in cancer. *Brief Bioinform* 2024; **25**. bbae379
16. Mandler H, Weigand B. A review and benchmark of feature importance methods for neural networks. *ACM Comput Surv* 2024; **56**: 318
17. Potharlanka JL, Bhat MN. Feature importance feedback with Deep Q process in ensemble-based metaheuristic feature selection algorithms. *Sci Rep* 2024; **14**: 2923
18. Wood D, Papamarkou T, Benatan M, et al. Model-agnostic variable importance for predictive uncertainty: an entropy-based approach. *Data Min Knowledge Discov* 2024; **38**: 4184–216
19. Molnar C, König G, Herbinger J, et al. General pitfalls of model-agnostic interpretation methods for machine learning models. In: Holzinger A, Goebel R, Fong R, Moon T, Müller K-R, Samek W, editors. *xxAI – Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*. Cham, Switzerland: Springer International Publishing; 2022. p. 39–68

doi: 10.1016/j.bja.2025.10.054