

Comment to the article: Machine-learning model for postoperative atrial fibrillation

Makaleye yorum: Postoperatif atriyal fibrilasyon için makine-öğrenimi modeli

Fatih Yiğit¹, Taylan Adademir¹, Kaan Kırallı¹

Department of Cardiovascular Surgery, Koşuyolu High-Specialization Training and Research Hospital, İstanbul, Türkiye

We read with great interest the study by Akbulut et al. on machine-learning (ML) prediction of postoperative atrial fibrillation (POAF) after isolated CABG.^[1] Although the integration of artificial intelligence into peri-operative care is welcome, several methodological issues may overstate the model's reliability and clinical value.

Insufficient sample size and events-per-variable (EPV)

Although the initial model considered 91 candidate predictors, only ≈ 15 features appear to have been retained after Boruta-based selection. With 50 POAF events in 100 patients, the resulting EPV is approximately 3.3-well below the level required for model stability.

A model with 15 predictors and an anticipated outcome incidence of $\sim 22\%$ would require at least 550-700 participants to ensure reliable parameter estimation, prevent overfitting, and achieve a shrinkage factor ≥ 0.9 .^[2,3]

The current sample therefore falls far short of the minimum sample size required for internally valid model development under modern reporting standards.

Artificially balanced outcome prevalence

The cohort was sub-sampled to a 50/50 POAF-non-POAF split, creating an artificial prevalence of 0.50. All reported performance metrics in Table 3-such as sensitivity, specificity,

precision, F1 score, accuracy, and Cohen's κ -reflect this engineered balance of 50% rather than the true clinical incidence of POAF, which is $\approx 22\%$. Recent simulation work shows that such balancing inflates apparent accuracy and produces severe mis-calibration when the model is applied to real-world data.^[4]

Limited test set and absence of external validation

Only 20 patients comprised the hold-out test set; misclassifying a single case alters accuracy by five percentage points. No geographic or temporal validation was reported, contrary to TRIPOD-AI reporting guidance.^[2]

Choice of algorithms and interpretability

The core model utilizes a Probabilistic Data Association (PDA) classifier-originally designed for radar and sonar tracking, not for clinical binary classification. Moreover, no model explainability method (e.g., SHAP) was provided, despite transparent interpretation being essential to clinical applicability and trust.

Moreover, the confusion matrix in Figure 4 is inconsistent with the sensitivity, specificity and precision values reported in the text (e.g., TP=10, FP=1 yield sensitivity=1.00 and specificity ≈ 0.90 , not vice versa), indicating a reporting error and raising doubt about reliability of the performance estimates.

Corresponding author: Fatih Yiğit.

E-mail: faithygt90@hotmail.com

Doi: 10.5606/tgkdc.dergisi.2025.28291

Received: July 02, 2025

Accepted: September 24, 2025

Published online: October 20, 2025

Cite this article as: Yiğit F, Adademir T, Kırallı K. Comment to the article: Machine-learning model for postoperative atrial fibrillation. Turk Gogus Kalp Dama 2025;33(4):590-591. doi: 10.5606/tgkdc.dergisi.2025.28291.

©2025 All right reserved by the Turkish Society of Cardiovascular Surgery.



This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes (<http://creativecommons.org/licenses/by-nc/4.0/>).

In summary, while Akbulut et al.'s work represents an important step toward incorporating artificial intelligence into cardiac surgical care, several methodological limitations-particularly regarding sample size, outcome balancing, validation, and reporting standards-warrant cautious interpretation. Recognizing and addressing these limitations in future research will be essential to building robust, generalizable, and clinically trustworthy ML tools for peri-operative risk stratification.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: All authors contributed equally to this article.

Conflict of Interest: The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding: The authors received no financial support for the research and/or authorship of this article.

REFERENCES

1. Akbulut B, Çakır M, Sarıkaya MG, Oral O, Yılmaz M, Aykal G. Artificial intelligence to predict biomarkers for new-onset atrial fibrillation after coronary artery bypass grafting. *Turk Gogus Kalp Dama* 2025;33:144-53. doi: 10.5606/tgkdc.dergisi.2025.27304.
2. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. doi: 10.1136/bmj-2023-078378.
3. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;38:1276-1296. doi: 10.1002/sim.7992.
4. Carriero A, Luijken K, de Hond A, Moons KGM, van Calster B, van Smeden M. The harms of class imbalance corrections for machine learning based prediction models: A simulation study. *Stat Med* 2025;44:e10320. doi: 10.1002/sim.10320.