Article

Not peer-reviewed version

An Interpretable Center-Specific Machine Learning Model for Risk Stratification Following Mitral Valve Surgery: A Pilot Study Center-Specific AI in Mitral Valve Surgery

Aleksandra Stańska * , Miriam Kilarska , Mateusz Janeczek , Wojciech Karolak , Andrzej Klapkowski

Posted Date: 1 June 2026

doi: 10.20944/preprints202606.0051.v1

Keywords: machine learning; artificial intelligence; risk stratification; mitral valve surgery; cardiac surgery; explainable artificial intelligence; clinical decision support; predictive modeling

Article

An Interpretable Center-Specific Machine Learning Model for Risk Stratification Following Mitral Valve Surgery: A Pilot Study Center-Specific AI in Mitral Valve Surgery

Aleksandra Stańska [1,*], Miriam Kilarska [2], Mateusz Janeczek [3], Wojciech Karolak [2] and Andrzej Klapkowski [2]

[1] Division of Quality of Life Research, Department of Psychology, Faculty of Health Sciences, Medical University of Gdańsk, Marii Skłodowskiej-Curie Street 3A, 80-210 Gdańsk, Poland
[2] Department of Cardiac & Vascular Surgery, Faculty of Medicine, Medical University of Gdańsk, Gdańsk, Poland
[3] Faculty of Medicine, Medical University of Gdańsk, Gdańsk, Poland
* Correspondence: astanska@gumed.edu.pl; Tel.: +48-601-180-636

Abstract

Background/Objectives: Mitral valve surgery is associated with substantial perioperative heterogeneity and risk of postoperative complications. Although established risk scores such as EuroSCORE II provide population-level prognostic estimates, their performance may be limited in specific surgical populations and institutional settings. This pilot study aimed to develop and internally validate an interpretable center-specific machine learning model for perioperative risk stratification following mitral valve surgery and to explore its translational implementation through a prototype clinical application. **Methods:** A retrospective single-center study was conducted including 211 consecutive patients undergoing mitral valve surgery with ring implantation. Routinely available demographic, laboratory, and perioperative variables were evaluated as candidate predictors. The primary endpoint was a composite of major postoperative complications, including in-hospital mortality, stroke, conversion to sternotomy, and rethoracotomy. Predictive approaches included logistic regression, LASSO regression, and random forest classification. Internal validation was performed using 5-fold cross-validation and bootstrap resampling. Model explainability was assessed using regression coefficients and SHAP (SHapley Additive exPlanations) analysis. **Results:** The composite endpoint occurred in 34 patients (16.1%). Among evaluated approaches, the simplified logistic regression model demonstrated the most favorable balance between interpretability and predictive performance, achieving a test-set AUC of 0.67 and a bootstrap-validated AUC of 0.69. Five-fold cross-validation yielded a mean AUC of 0.75. LASSO regression achieved the highest cross-validated discrimination (AUC 0.78) while retaining only a limited subset of predictors, with cardiopulmonary bypass time emerging as the dominant variable. Across models, higher age, serum creatinine concentration, cardiopulmonary bypass duration, and cross-clamp time were associated with increased complication risk, whereas higher hemoglobin levels were associated with lower risk. **Conclusions:** This pilot study demonstrates the feasibility of developing interpretable center-specific machine learning models for perioperative risk stratification following mitral valve surgery. Simplified regression-based approaches provided clinically transparent predictions with moderate discriminatory performance, while penalized models showed potential for improved generalizability. Further multicenter validation is required before clinical implementation.

Keywords: machine learning; artificial intelligence; risk stratification; mitral valve surgery; cardiac surgery; explainable artificial intelligence; clinical decision support; predictive modeling

Background

Mitral valve surgery remains one of the most commonly performed procedures in contemporary cardiac surgery and is associated with substantial perioperative heterogeneity regarding patient characteristics, operative complexity, and postoperative outcomes. Despite ongoing advances in surgical techniques, perioperative care, minimally invasive approaches, and postoperative monitoring, major postoperative complications continue to represent an important clinical challenge.

Current perioperative risk assessment in cardiac surgery relies predominantly on generalized population-based scoring systems such as EuroSCORE II and the Society of Thoracic Surgeons (STS) risk models [1,2]. Although these tools provide valuable prognostic information at the population level, their predictive performance may be limited in specific procedural subgroups, including mitral valve surgery, particularly within evolving minimally invasive and institution-specific surgical pathways [3,4]. Furthermore, generalized risk calculators may insufficiently capture local perioperative practices, patient selection patterns, and center-specific procedural variability.

In recent years, machine learning approaches have gained increasing attention in perioperative medicine due to their potential ability to identify complex multidimensional relationships between clinical variables and postoperative outcomes [5,6]. Several studies have explored artificial intelligence-based prediction models in cardiac surgery; however, many available approaches remain limited by insufficient interpretability, lack of external validation, substantial model complexity, or reduced clinical applicability in routine perioperative workflows [5,7].

An additional challenge in perioperative artificial intelligence implementation involves balancing predictive performance with clinical interpretability. While highly complex machine learning algorithms may achieve strong predictive discrimination in selected datasets, limited transparency and "black-box" behavior remain substantial barriers to clinician trust and practical bedside implementation [5,8].

Institution-specific perioperative prediction models may represent a complementary approach to generalized population-based risk scores. Local surgical techniques, perioperative management strategies, postoperative care pathways, and institutional patient characteristics may substantially influence complication patterns and predictive relationships [3]. Consequently, continuously updated center-specific models may better reflect real-world perioperative practice within individual cardiac surgery centers.

The present pilot study aimed to develop and internally validate an interpretable institution-specific perioperative risk stratification model for patients undergoing mitral valve surgery using routinely available perioperative and laboratory variables. Additionally, the study explored practical translational implementation through development of a lightweight browser-based prototype application designed to demonstrate the feasibility of rapid perioperative bedside risk estimation.

Methods

Study Design and Population

This retrospective single-center pilot study was conducted at the Department of Cardiac Surgery, University Clinical Centre in Gdańsk, Poland. The study included consecutive adult patients undergoing mitral valve surgery with ring implantation. Clinical, perioperative, and postoperative data were retrospectively extracted from institutional databases and operative records.

The primary aim of the study was to develop and internally validate an exploratory machine learning-based perioperative risk stratification model for major postoperative complications following mitral valve surgery. Due to the pilot and exploratory nature of the study, no formal sample size calculation was performed.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2026 doi:10.20944/preprints202606.0051.v1

Variables and Preprocessing

Demographic, perioperative, and laboratory variables routinely available in clinical practice were considered as candidate predictors. Initial exploratory analyses included:

- age,
- hemoglobin concentration,
- serum creatinine,
- cardiopulmonary bypass time,
- aortic cross-clamp time,
- left ventricular ejection fraction (EF),
- ring size,
- atrial fibrillation status,
- NYHA class,
- sex,
- procedural characteristics.

Continuous variables were inspected for implausible and non-physiological values. Cross-clamp time values equal to 0 minutes were considered invalid entries and treated as missing values. Similarly, creatinine values equal to 0 were excluded from analyses as implausible measurements.

Hemoglobin values were converted from string-based decimal formatting into numeric representation prior to analysis. Complete-case analysis was applied for model development.

After exploratory model comparisons, a simplified clinically interpretable model was selected, including:

- age,
- hemoglobin,
- serum creatinine,
- cardiopulmonary bypass time,
- cross-clamp time.

Outcome Definition

The primary endpoint was the occurrence of major postoperative complications. A composite binary endpoint was created based on the presence of at least one severe postoperative adverse event recorded during hospitalization, including:

- in-hospital mortality,
- stroke,
- conversion to sternotomy,
- Rethoracotomy.

Composite cardiovascular endpoints are commonly used to improve statistical efficiency in studies with limited event counts, although heterogeneity between endpoint components should be considered when interpreting results [18].

Statistical Analysis

Descriptive statistics were calculated for all study variables. Continuous variables were summarized as mean ± standard deviation or median with interquartile range, depending on data distribution. Categorical variables were presented as counts and percentages.

All analyses were performed in Python 3.12 (Python Software Foundation, Wilmington, DE, USA) using Jupyter Notebook. Data preprocessing was performed using pandas and NumPy, while predictive modeling utilized scikit-learn. Model explainability analyses were conducted using SHAP.

Machine Learning Model Development

Several predictive approaches were evaluated, including:

 doi:10.20944/preprints202606.0051.v1

- logistic regression,
- penalized logistic regression (LASSO),
- random forest classification.

The dataset was divided into training and testing subsets using stratified random sampling (80/20 split). Class imbalance was addressed using balanced class weighting.

Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC-ROC).

To estimate model robustness and reduce the risk of overfitting, internal validation procedures included:

- 5-fold cross-validation,
- bootstrap resampling.

Appropriate internal validation is considered essential for prediction model development, particularly in exploratory datasets with limited sample sizes [19].

Model Interpretability

Feature importance and explainability were evaluated using:

- regression coefficients,
- random forest feature importance,
- SHAP (SHapley Additive exPlanations) analysis [17].

SHAP summary plots were generated to assess the relative contribution and directionality of variables influencing model predictions. Emphasis on explainability and model transparency is increasingly recognized as an important prerequisite for clinical implementation of artificial intelligence systems in medicine [20].

Prototype Implementation

To demonstrate practical translational applicability, a lightweight browser-based prototype application (“Mitral Valve Risk AI”) was developed using HTML and JavaScript. The application enabled rapid estimation of postoperative complication risk using routinely available perioperative variables.

Ethical Considerations

The study was conducted in accordance with institutional standards for retrospective observational research using anonymized routinely collected clinical data.

Results

Study Population

The final study cohort consisted of 211 patients undergoing mitral valve surgery with ring implantation. Baseline demographic, perioperative, and postoperative characteristics are summarized in Table 1.

The mean age of the cohort was 60.4 ± 13.7 years, and 57.8% of patients were male. Mean left ventricular ejection fraction was 53.8 ± 8.9%. Mean cardiopulmonary bypass time was 149.8 ± 47.5 minutes, while median cross-clamp time was 92 minutes (IQR 72–109).

Mitral valve repair was performed in 72.5% of procedures, whereas mitral valve replacement accounted for 27.5% of cases. Thoracotomy access was used in 17.1% of operations.

Postoperative atrial fibrillation occurred in 55.9% of patients. The observed in-hospital mortality rate was 4.7%, while stroke occurred in 1.9% of patients. Additional postoperative outcomes included rethoracotomy in 8.1% of cases, conversion to sternotomy in 2.8%, and permanent pacemaker implantation in 3.3% of patients.

 doi:10.20944/preprints202606.0051.v1

The composite endpoint of major postoperative complications, defined as the occurrence of in-hospital mortality, stroke, conversion to sternotomy, or rethoracotomy, occurred in 34 patients (16.1%).

Table 1. Baseline characteristics of the study cohort.

Variable	Value
Age, years	60.4 ± 13.7
Hemoglobin, g/dL	13.7 ± 1.6
Creatinine, mg/dL	1.0 ± 0.4
Cardiopulmonary bypass time, min	149.8 ± 47.5
Cross-clamp time, min	93.7 ± 28.8
Ejection fraction, %	53.8 ± 8.9
Ring size	33.5 ± 4.2
Male sex	122 (57.8%)
Thoracotomy access	36 (17.1%)
Mitral valve repair	153 (72.5%)
Mitral valve replacement	58 (27.5%)
Preoperative atrial fibrillation	98 (46.4%)
Postoperative atrial fibrillation	118 (55.9%)
Rethoracotomy	17 (8.1%)
Stroke	4 (1.9%)
Conversion to sternotomy	6 (2.8%)
Pacemaker implantation	7 (3.3%)
In-hospital mortality	10 (4.7%)
Major postoperative complication composite endpoint	34 (16.1%)

Predictive Model Performance

Comparative performance of evaluated predictive models is presented in Table 2.

Among evaluated approaches, the simplified logistic regression model demonstrated the best balance between interpretability, coefficient transparency, and stable predictive performance. The model achieved a test-set AUC of 0.67, while bootstrap validation yielded a mean AUC of 0.69. Five-fold cross-validation demonstrated a mean AUC of 0.75, although moderate variability across folds

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2026 doi:10.20944/preprints202606.0051.v1

was observed, likely reflecting the relatively limited number of adverse events and the exploratory nature of the dataset.

Expanded multivariable models containing larger numbers of perioperative and categorical predictors demonstrated lower generalizability and reduced predictive performance, suggesting overfitting in the context of a relatively small single-center cohort.

Random forest classification models demonstrated lower overall discrimination and greater instability across validation folds compared with regularized regression approaches.

Interestingly, LASSO regression achieved a mean cross-validated AUC of 0.78 (SD 0.18) despite lower single test-set discrimination. The penalized model retained only a limited subset of predictors, with cardiopulmonary bypass time emerging as the dominant retained variable, suggesting that sparse regularized models may improve generalizability in smaller perioperative datasets.

Table 2. Comparative performance of predictive models.

Model	Test AUC	Mean CV AUC	Comment
Logistic regression (simplified)	0.67	0.75	Interpretable model with stable performance
Logistic regression (expanded)	0.45	0.53	Lower generalizability and possible overfitting
Random forest	0.56	0.56	Moderate and less stable discrimination
LASSO regression	0.51	0.78	Strong cross-validation performance with sparse feature selection

Feature Importance and Explainability

Across evaluated models, higher:

- age,
- serum creatinine,
- cardiopulmonary bypass time,
- cross-clamp time

were consistently associated with increased predicted complication risk.

Higher hemoglobin concentration was associated with lower predicted risk.

SHAP analyses confirmed the relative importance of perioperative burden and baseline physiological reserve in driving model predictions. The most influential variables identified across explainability analyses included:

- cardiopulmonary bypass time,
- hemoglobin,
- age,
- cross-clamp time,
- creatinine concentration.

The final simplified logistic regression model additionally demonstrated clinically plausible coefficient directionality across included predictors.

Calibration and Internal Validation

Calibration analysis demonstrated moderate agreement between predicted and observed event probabilities, although instability was observed at higher predicted probability ranges due to the limited number of major adverse events. Calibration has been recognized as a critical component of clinical prediction model evaluation and may substantially influence real-world model reliability [21].

Bootstrap validation demonstrated moderate overall robustness but relatively broad confidence intervals, reflecting the exploratory nature of the study and limited event count.

Lowering the classification threshold improved sensitivity for detecting complications at the cost of increased false-positive classifications consistent with known trade-offs in ROC-based threshold optimization [22].

Prototype Application

A lightweight browser-based prototype application ("Mitral Valve Risk AI") was successfully developed to demonstrate the potential translational implementation of the predictive model. The prototype enabled rapid individualized perioperative risk estimation using routinely available clinical variables.

The application was developed exclusively as a proof-of-concept research tool and should not be interpreted as a clinically validated decision-support system.

Discussion

The present pilot study demonstrates the feasibility of developing institution-specific perioperative risk stratification models for patients undergoing mitral valve surgery using routinely available clinical and laboratory variables. Despite the relatively limited cohort size, several machine learning and statistical approaches demonstrated moderate predictive capability for major postoperative complications.

Interestingly, model performance varied substantially depending on the degree of model complexity and regularization. While the simplified logistic regression model demonstrated the best balance between interpretability, clinical transparency, and overall consistency, LASSO regularization achieved the highest cross-validated performance. This finding suggests that feature reduction and penalization may improve model parsimony, although prediction models developed in smaller datasets may remain vulnerable to instability and overfitting [9,10].

In contrast, expanded multivariable models and random forest approaches demonstrated lower stability and weaker overall discrimination. These findings are consistent with evidence suggesting that higher-complexity machine learning algorithms do not universally outperform simpler statistical approaches, particularly in relatively small clinical datasets vulnerable to overfitting [9,10].

Several predictors demonstrated clinically plausible associations with postoperative risk. Older age, impaired renal function, prolonged cardiopulmonary bypass exposure, and longer cross-clamp time are established markers of increased perioperative burden and physiological vulnerability in cardiac surgery populations [11,12]. In contrast, higher hemoglobin concentration appeared protective, potentially reflecting greater baseline physiological reserve and reduced perioperative frailty; this is consistent with evidence linking preoperative anemia with worse outcomes after cardiac surgery [13].

An important aspect of the present study is the emphasis on explainability and clinical interpretability. Rather than relying exclusively on black-box algorithms, the study incorporated transparent coefficient-based modeling alongside SHAP explainability analyses [17]. This approach may be particularly relevant in perioperative medicine, where clinician trust, transparency, reporting quality, and interpretability remain major barriers to broader implementation of artificial intelligence-based tools [5,8,14,15].

The study additionally explored practical translational implementation through development of a lightweight browser-based prototype application. Although preliminary, this proof-of-concept demonstrates how institution-specific perioperative prediction tools could potentially be integrated into routine clinical workflows. Similar recent work in cardiac surgery has highlighted the

 doi:10.20944/preprints202606.0051.v1

translational potential of machine-learning-based prediction models and clinically oriented digital applications [15,16].

The findings may support the broader concept of center-specific perioperative risk stratification systems in cardiac surgery. Institutional differences in surgical techniques, perioperative management, patient selection, and postoperative care may substantially influence complication patterns and predictive relationships. Consequently, locally developed and continuously updated predictive models may provide clinically relevant complementary information beyond generalized population-based risk scores [3,4,15].

Limitations

Several limitations should be acknowledged. First, this was a retrospective single-center study conducted at a tertiary academic cardiac surgery center, which may limit the generalizability of the findings and introduce institution-specific bias related to patient selection, surgical strategy, and perioperative management.

Second, the cohort size and number of adverse events were relatively limited, increasing the risk of model instability and overfitting. Smaller clinical datasets may reduce the reliability and transportability of predictive models, particularly when multiple candidate predictors are evaluated [23]. Although internal validation procedures including cross-validation and bootstrap resampling were performed, the models should still be interpreted as exploratory and hypothesis-generating.

Third, some variables required manual preprocessing and cleaning, including removal of implausible entries. Additionally, the composite endpoint was based on retrospectively collected institutional data and may not fully capture the heterogeneity of clinically relevant postoperative complications.

Finally, external multicenter validation will be necessary before any clinical implementation can be considered.

Conclusions

This pilot study demonstrates the feasibility of developing interpretable institution-specific perioperative risk prediction models for mitral valve surgery using routinely available clinical variables. Simplified logistic regression models provided strong interpretability and stable performance, while LASSO regularization achieved the highest cross-validated discrimination, suggesting that sparse penalized models may improve generalizability in smaller perioperative datasets. Further multicenter validation and prospective refinement are required before clinical deployment.

Author Contributions: Conceptualization, A.S.; methodology, A.S.; software, A.S.; validation, A.S., M.K. and M.J.; formal analysis, A.S.; investigation, M.K., M.J., W.K. and A.K.; resources, W.K. and A.K.; data curation, A.S., M.K. and M.J.; writing - original draft preparation, A.S.; writing - review and editing, A.S., M.K., M.J., W.K. and A.K.; visualization, A.S.; supervision, W.K. and A.K.; project administration, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This retrospective study utilized anonymized routinely collected clinical data and was conducted in accordance with institutional regulations of the University Clinical Centre in Gdańsk. According to local institutional policies governing retrospective analyses of fully anonymized clinical datasets, formal ethics committee approval was not required.

Informed Consent Statement: The requirement for informed consent was waived due to the retrospective design of the study and the use of fully anonymized routinely collected clinical data.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and institutional restrictions related to patient-level clinical data.

Acknowledgments: The authors would like to thank the clinical and administrative staff of the Department of Cardiac Surgery, University Clinical Centre in Gdańsk, for their support in data collection and clinical documentation.

Conflicts of Interest: The authors declare no conflicts of interest.; Use of Artificial Intelligence Tools; During the preparation of this manuscript, artificial intelligence tools were used for language editing, text structuring, and assistance in drafting selected manuscript sections. The authors reviewed and edited all generated content and take full responsibility for the final version of the manuscript.

References

1. Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. *Eur J Cardiothorac Surg*. 2012;41(4):734-44. doi:10.1093/ejcts/ezs043.
2. O'Brien SM, Feng L, He X, Xian Y, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. *Ann Thorac Surg*. 2018;105(5):1419-28. doi:10.1016/j.athoracsur.2018.03.003.
3. Mastroiacovo G, Paladini A, Guarracino F, et al. Is EuroSCORE II still a reliable predictor for cardiac surgery risk? *Eur J Cardiothorac Surg*. 2023;64(3):ezad294. doi:10.1093/ejcts/ezad294.
4. Di Bacco L, Speziale G, Nasso G, et al. Reliability of EuroSCORE II on prediction of thirty-day and long-term mortality in minimally invasive cardiac surgery. *J Clin Med*. 2024;13(13):3986. doi:10.3390/jcm13133986.
5. Miles TJ, Arsalan M, Bavaria JE. Machine learning in cardiac surgery: a narrative review. *J Thorac Dis*. 2024;16(4):2553-67. doi:10.21037/jtd-23-1659.
6. Benedetto U, Dimagli A, Sinha S, et al. Machine learning improves mortality risk prediction after cardiac surgery: systematic review and meta-analysis. *J Thorac Cardiovasc Surg*. 2022;163(6):2075-87.e9. doi:10.1016/j.jtcvs.2020.07.105.
7. Sinha S, Dong T, Dimagli A, Vohra HA, Holmes C, Benedetto U, et al. Comparison of machine learning techniques in prediction of mortality following cardiac surgery: analysis of over 220 000 patients from a large national database. *Eur J Cardiothorac Surg*. 2023;63(6):ezad183. doi:10.1093/ejcts/ezad183.
8. Rellum SR, Schuurmans J, van der Ven WH, Eberl S, Driessen AHG, Vlaar APJ, Veelo DP. Machine learning methods for perioperative anesthetic management in cardiac surgery patients: a scoping review. *J Thorac Dis*. 2021;13(12):6976-6993. doi:10.21037/jtd-21-765.
9. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004.
10. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *J Clin Epidemiol*. 2021;132:88-96. doi:10.1016/j.jclinepi.2020.12.005.
11. Salis S, Mazzanti VV, Merli G, Salvi L, Tedesco CC, Veglia F, et al. Cardiopulmonary bypass duration is an independent predictor of morbidity and mortality after cardiac surgery. *J Cardiothorac Vasc Anesth*. 2008;22(6):814-22. doi:10.1053/j.jvca.2008.08.004.
12. Udesh R, Mehta A, Gleason TG, Wechsler LR, Thirumala PD. Perioperative strokes and early outcomes in mitral valve surgery: a nationwide analysis. *J Cardiothorac Vasc Anesth*. 2017;31(2):529-36. doi:10.1053/j.jvca.2016.12.006.
13. Padmanabhan H, Siau K, Curtis J, Ng A, Menon S, Luckraz H, et al. Preoperative anemia and outcomes in cardiovascular surgery: systematic review and meta-analysis. *Ann Thorac Surg*. 2019;108(6):1840-8. doi:10.1016/j.athoracsur.2019.04.108.
14. 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.

 doi:10.20944/preprints202606.0051.v1

15. Abbasi A, Li C, Dekle M, Bermudez CA, Brodie D, Sellke FW, et al. Interpretable machine learning-based predictive modeling of patient outcomes following cardiac surgery. *J Thorac Cardiovasc Surg*. 2025;169(1):114-123.e28. doi:10.1016/j.jtcvs.2023.11.034.
16. Miguel-Medina T, Góngora Alonso S, de la Torre Díez I, Blanco Sáez M, Elissalt HLA, Ruigómez Noriega A, et al. Development of a machine learning-based predictive model and clinically oriented web application for 30-day mortality following cardiac surgery. *Sensors (Basel)*. 2026;26(5):1656. doi:10.3390/s26051656.
17. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-4774.
18. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007;334(7597):786. doi:10.1136/bmj.39136.682083.AE.
19. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005.
20. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9(4). doi:10.1002/widm.1312.
21. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7.
22. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol. 2006;163(7):670-675. doi:10.1093/aje/kwj063.
23. van Smeden M, Moons KGM, de Groot JAH, Collins GS, Altman DG, Eijkemans MJC, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. Stat Methods Med Res. 2019;28(8):2455-2474. doi: 10.1177/0962280218784726.