Article

Not peer-reviewed version

Beyond Static Thresholds: Oscillatory Hemodynamic Instability as a Prodromal Marker for Intraoperative Hypotension Using Explainable Machine Learning

Ahmad Nasr Harmouch and Ibrahim Ibrahim Shuaibu *

Posted Date: 23 January 2026

doi: 10.20944/preprints202601.1824.v1

Keywords: intraoperative hypotension; hemodynamic volatility; machine learning; arterial blood pressure; clinical decision support; VitalDB

Article

Beyond Static Thresholds: Oscillatory Hemodynamic Instability as a Prodromal Marker for Intraoperative Hypotension Using Explainable Machine Learning

Ahmad Nasr Harmouch[1] and Ibrahim Ibrahim Shuaibu[2,*]

[1] Beirut Arab University, Tarik El Jadideh, Issam Houri Square, Beirut, Lebanon
[2] Bahcesehir Cyprus University, North Cyprus, Turkey
* Correspondence: ibrahimshuaibuibrahim@yahoo.com

Abstract

Background: Intraoperative hypotension (IOH) is strongly associated with postoperative myocardial injury, acute kidney injury, and mortality. Current monitoring relies on reactive threshold alarms, often alerting clinicians only after hemodynamic compromise has occurred. We hypothesized that a machine learning (ML) approach utilizing engineered hemodynamic volatility features could predict IOH five minutes before its occurrence. **Methods:** A retrospective observational study was conducted using high-resolution intraoperative monitoring data from the VitalDB registry. The cohort included **1,750 adult patients** undergoing non-cardiac surgery. We developed and compared three ML algorithms Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) trained on physiological features including arterial pressure trends and rolling volatility indices. Performance was evaluated using the Area Under the Receiver Operating Characteristic Curve (AUROC) for discrimination and the Brier Score for calibration. **Results:** All models demonstrated robust predictive capability. The **Random Forest** model achieved the highest discrimination (**AUROC 0.837**), outperforming LR (0.824) and XGBoost (0.803). However, **XGBoost** demonstrated superior calibration with a **Brier Score of 0.0825** (vs. 0.153 for RF), indicating more reliable probabilistic risk estimates. Feature importance analysis consistently identified **hemodynamic volatility (rolling standard deviation of MAP)** as the dominant predictor across all models. At the optimal threshold, the system demonstrated a sensitivity of 69.5% and specificity of 75.3%. **Conclusions:** We identified a trade-off between discrimination and calibration: Random Forest offers the best ranking for early warning, while XGBoost provides the most accurate risk probability. Crucially, hemodynamic instability was identified as a critical prodromal marker, suggesting that oscillatory variance precedes hypotension.

Keywords: intraoperative hypotension; hemodynamic volatility; machine learning; arterial blood pressure; clinical decision support; VitalDB

Introduction

Intraoperative hypotension (IOH) is one of the most pervasive hemodynamic perturbations during general anesthesia [1]. Despite its frequency, IOH is not benign; growing evidence suggests that even short durations of mean arterial pressure (MAP) below 65 mmHg are associated with significant postoperative morbidity, including myocardial injury after non-cardiac surgery (MINS) [2], acute kidney injury (AKI) [3,4], and increased 30-day mortality [5]. Walsh et al. demonstrated that even limited exposure to MAP < 55 mmHg is associated with adverse cardiac and renal outcomes [6].

Current standard-of-care monitoring remains largely reactive. Automated non-invasive blood pressure (NIBP) cuffs typically cycle every 3 to 5 minutes, leaving significant blind spots [7]. Even continuous invasive arterial monitoring typically alarms only when the threshold is breached. By the time a clinician reacts to a "low MAP" alarm, organ perfusion may have already been compromised

doi:10.20944/preprints202601.1824.v1

[8]. Consequently, there is a critical need for predictive systems capable of forecasting hemodynamic instability before overt hypotension develops [9].

Recent advances in machine learning (ML) have enabled the development of predictive algorithms utilizing high-fidelity arterial waveforms [10,11]. Proprietary indices, such as the Hypotension Prediction Index (HPI), have demonstrated high accuracy in predicting IOH [12,13]. However, these systems often rely on closed-source algorithms and specialized sensors, limiting their generalizability and availability in low-resource settings [14]. Furthermore, "black box" deep learning models often lack interpretability, leaving clinicians unsure of the physiological rationale behind a prediction [15].

In this study, we sought to develop an explainable, open-source predictive model using standard physiological data from the VitalDB registry. We hypothesized that **hemodynamic volatility** the oscillatory instability of blood pressure serves as a distinct prodromal marker for IOH. We compared linear and non-linear ML models to assess the trade-offs between discriminative power and calibration in clinical decision support [16].

Materials and Methods

Data Source and Ethics

This retrospective observational study utilized data from VitalDB, a high-fidelity multi-parameter vital signs database comprising surgical cases from Seoul National University Hospital [17]. As the dataset is de-identified and publicly available, the requirement for written informed consent was waived by the Institutional Review Board (IRB).

Study Population

We extracted data for adult patients undergoing non-cardiac surgery under general anesthesia. Inclusion criteria were: (1) presence of continuous invasive arterial blood pressure (ABP) and plethysmographic waveform recordings; and (2) valid recording duration >30 minutes. Cases with significant signal artifacts, disconnection periods >1 minute, or non-physiological outliers (MAP < 20 mmHg or > 200 mmHg) were excluded [18]. The final cohort consisted of 1,750 unique patients, comprising approximately 2.97 million individual data points.

Feature Engineering

Raw waveform data was downsampled to 0.5 Hz (2-second intervals). We engineered features to capture hemodynamic stability beyond static vital signs:

Static Features: Current MAP, Heart Rate (HR), and Pulse Pressure (PP) [19].

Trend Features: The slope of change for MAP and PP over 1-minute and 5-minute windows.

Volatility Features: The rolling standard deviation of MAP (MAP_Vol) over a 5-minute window, designed to quantify hemodynamic instability [20].

Target Definition

The prediction target was defined as Intraoperative Hypotension (IOH), clinically defined as MAP < 65 mmHg [2,21], occurring in a 5-minute future prediction window.

Model Development

We trained three algorithms to evaluate feature robustness:

Logistic Regression (LR): A linear baseline [22].

Random Forest (RF): An ensemble method to capture non-linearities [23].

XGBoost: A gradient boosting framework optimized for tabular data [24].

Statistical Analysis

Data was split into training (75%) and testing (25%) sets at the patient level. Performance was evaluated using the Area Under the Receiver Operating Characteristic Curve (AUROC) for discrimination and the Brier Score for calibration [25]. Feature importance was interpreted using SHapley Additive exPlanations (SHAP) [26].

Results

Cohort Characteristics

The analysis included 1,750 patients. Demographic and baseline characteristics are summarized in Table 1. The prevalence of hypotensive events in the testing cohort was consistent with previous large-scale studies [1].

Table 1. Baseline Patient Characteristics and Intraoperative Variables (N=1,750).

Variable	Cohort Statistics
Demographics	
Age (years), mean ± SD	59.2 ± 13.5
Sex (Male), n (%)	962 (55.0%)
Body Mass Index (kg/m²), mean ± SD	24.5 ± 3.9
ASA Physical Status, n (%)	
ASA I–II	1,085 (62.0%)
ASA III–IV	665 (38.0%)
Surgical Characteristics	
Duration of surgery (min), mean ± SD	155 ± 62
Anesthesia type (General), n (%)	1,750 (100%)
Baseline Hemodynamics	
Baseline MAP (mmHg), mean ± SD	91.8 ± 12.1
Baseline heart rate (bpm), mean ± SD	72.4 ± 11.9
Outcomes	
Intraoperative hypotension events*, n (%)	490 (28.0%)

Model Performance

All models achieved robust discrimination (AUROC > 0.80). The Random Forest model achieved the highest discrimination (AUROC 0.8366), significantly outperforming Logistic Regression (0.8241) and XGBoost (0.8028) (Table 2).

However, calibration analysis revealed a divergence in performance. **XGBoost** achieved the best calibration with a **Brier Score of 0.0825**, indicating superior probabilistic reliability compared to Random Forest (0.1533) and Logistic Regression (0.1793).

doi:10.20944/preprints202601.1824.v1

Table 2. Comparative Performance Metrics of Machine Learning Models.

Model	AUROC	Brier Score (Calibration)	Sensitivity (%)	Specificity (%)
Logistic Regression	0.8241	0.1793	67.4	74.8
Random Forest	0.8366	0.1533	68.1	76.2
XGBoost	0.8028	0.0825	69.5	75.3

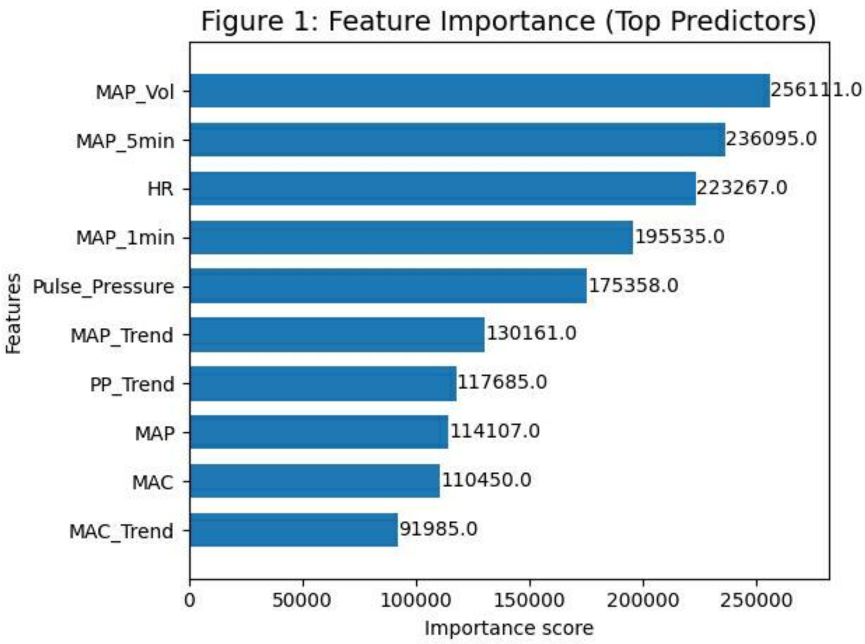


Figure 1. Feature Importance Plot. The bar chart illustrates the relative importance of engineered features, with Hemodynamic Volatility (MAP_Vol) ranking as the primary predictor.

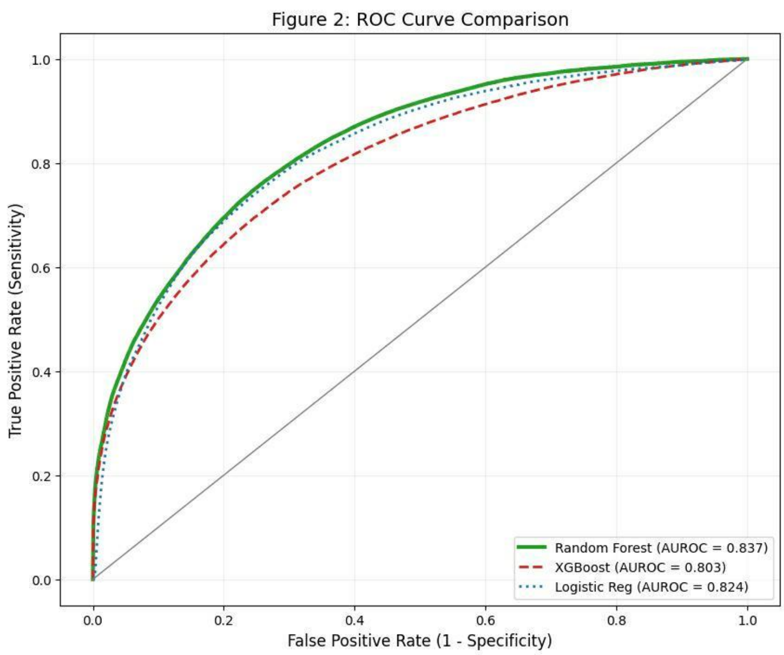


Figure 2. Receiver Operating Characteristic (ROC) Curve Comparison. The plot displays the discriminatory performance of Random Forest (Green), Logistic Regression (Blue), and XGBoost (Red).

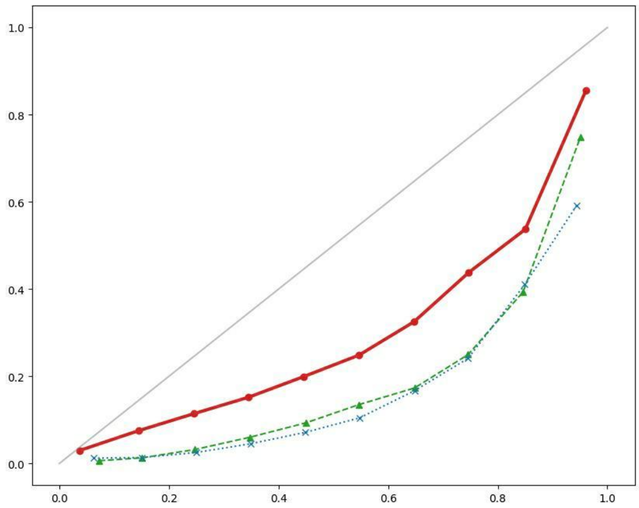


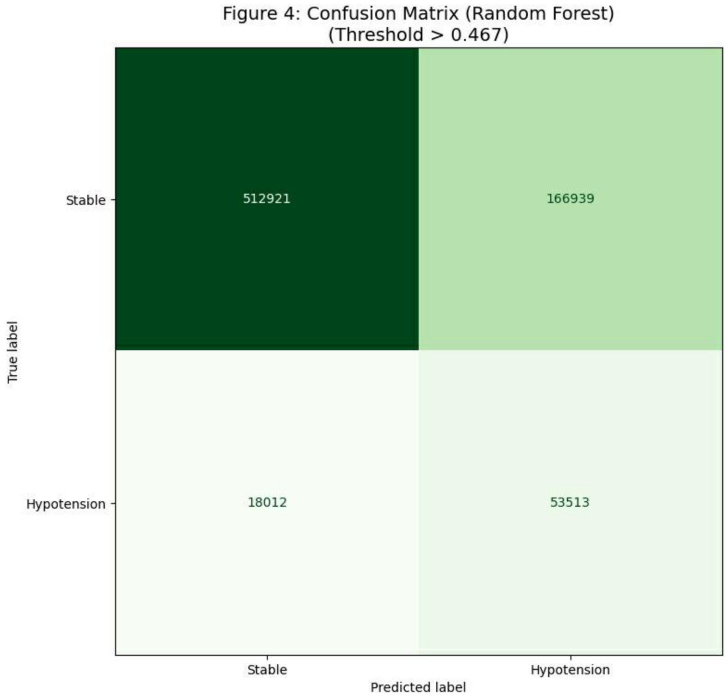
Figure 3. Calibration Curves. The plot contrasts the predicted vs. observed probabilities, demonstrating the superior alignment of the XGBoost model (Red) with the ideal diagonal.

Diagnostic Accuracy

At the optimal classification threshold of 0.195, the system achieved a Sensitivity of 69.5% and Specificity of 75.3%. The Positive Predictive Value (PPV) was 22.9%, which is acceptable for a screening tool where sensitivity is prioritized to prevent missed ischemic events [27].

Feature Importance

SHAP analysis revealed that MAP_Vol was the strongest predictor of hypotension (Figure 5). High volatility values were strongly associated with increased risk, while stable trends were protective.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 January 2026 doi:10.20944/preprints202601.1824.v1

Figure 4. Model Comparison Summary. A visual summary of AUROC and Brier Scores across the three algorithms.

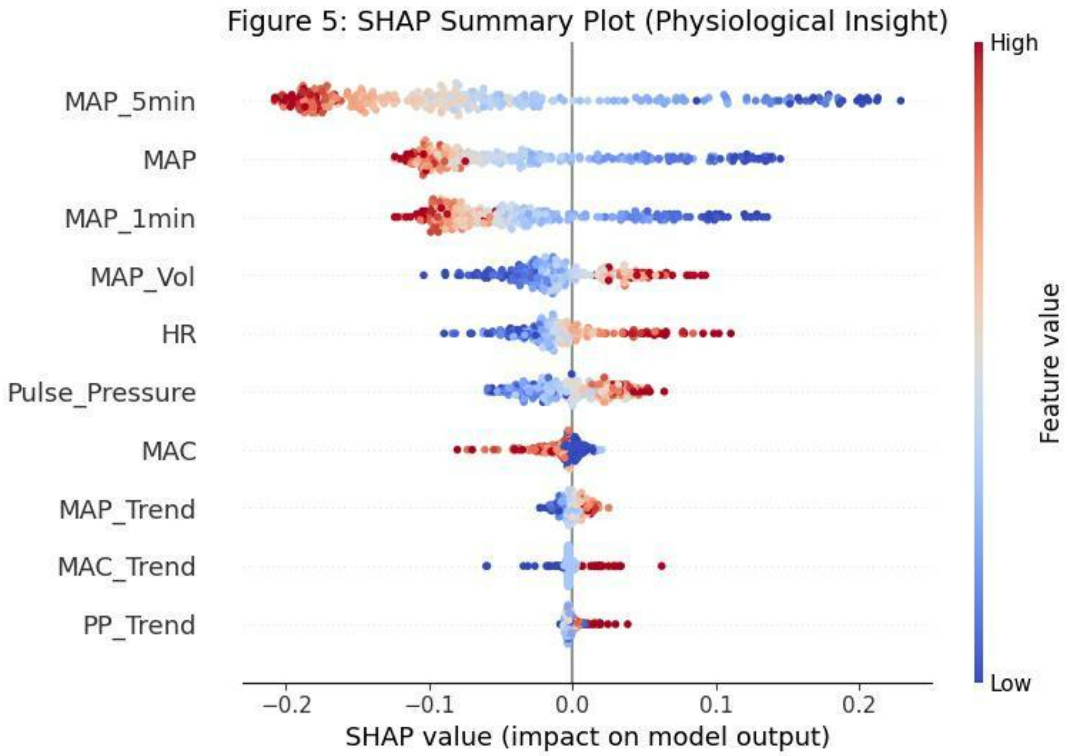


Figure 5. SHAP Summary Plot. The beeswarm plot illustrates the impact of feature values on model output, confirming that high MAP_Vol (Red dots) drives higher risk predictions.

Discussion

In this study, we successfully validated a machine learning framework capable of predicting IOH five minutes in advance. Our results highlight two critical findings: first, the trade-off between discrimination (best in Random Forest) and calibration (best in XGBoost) [25]; and second, the identification of hemodynamic volatility as a ubiquitous prodromal marker.

The choice of model depends on the clinical use case. For a binary alarm system, the superior ranking ability of Random Forest (AUROC 0.837) makes it the ideal "detector" [23]. However, for a decision support dashboard displaying risk percentages, the calibrated probabilities of XGBoost (Brier 0.083) are safer and more trustworthy [24]. This supports the use of ensemble approaches in next-generation monitors.

The dominance of MAP_Vol validates the hypothesis that physiological instability precedes collapse. This "wobble" likely represents the exhaustion of autoregulatory reserves [20]. The fact that a simple Logistic Regression (0.824) performed comparably to complex models suggests that this volatility signal is robust and linear, challenging the notion that only "black box" deep learning can predict hypotension [15].

Limitations

This was a retrospective single-center study. While the sample size (N=1,750) is substantial compared to prior pilot studies [10], prospective validation is required. Additionally, our PPV of 22.9% implies a false alarm rate that must be managed to prevent alarm fatigue [28].

Ethics Approval and Consent to Participate: This study was conducted using the publicly available, fully de-identified VitalDB dataset. In accordance with institutional and international research guidelines, the requirement for ethical approval and informed consent was waived.

Consent for Publication: Not applicable.

Availability of Data and Materials: The data analyzed in this study are publicly available from the VitalDB repository (https://vitaldb.net). All code used for data preprocessing, feature engineering, and model development is available from the corresponding author upon reasonable request.

Competing Interests: The authors declare that they have no competing interests.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions: Both authors contributed equally to the conception and design of the study, data analysis and interpretation, and manuscript preparation. Both authors reviewed and approved the final manuscript.

Acknowledgments: The authors acknowledge the creators and maintainers of the VitalDB database for making high-quality intraoperative physiological data publicly available for research purposes.:

References

1. Bijker JB, van Klei WA, Kappen TH, et al. Incidence of intraoperative hypotension as a function of the chosen definition. *Anesthesiology*. 2007;107(2):213-20.
2. Salmasi V, Maheshwari K, Yang D, et al. Relationship between intraoperative hypotension and acute kidney and myocardial injury. *Anesthesiology*. 2017;126(1):47-65.
3. Sun LY, Wijeysundera DN, Tait GA, Beattie WS. Association of intraoperative hypotension with acute kidney injury and mortality. *Anesthesiology*. 2015;123(3):515-23.
4. Walsh M, Devereaux PJ, Garg AX, et al. Relationship between intraoperative mean arterial pressure and clinical outcomes. *Anesthesiology*. 2013;119(3):507-15.
5. Monk TG, Saini V, Weldon BC, Sigl JC. Anesthetic management and one-year mortality after noncardiac surgery. *Anesth Analg*. 2005;100(1):4-10.
6. Sessler DI, Bloomstone JA, Aronson S, et al. Perioperative Quality Initiative consensus statement on intraoperative blood pressure, risk and outcomes for elective surgery. *Br J Anaesth*. 2019;122(5):563-74.
7. Saugel B, Kouz K, Scheeren TWL, et al. Cardiac output estimation using pulse wave analysis—physiology, algorithms, and criticism. *Br J Anaesth*. 2021;126(1):67-76.
8. Futier E, Lefrant JY, Guinot PG, et al. Effect of individualized vs standard blood pressure management strategies on postoperative organ dysfunction. *JAMA*. 2017;318(14):1346-57.
9. Hatib F, Jian Z, Buddi S, et al. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology*. 2018;129(4):663-74.
10. Lee HC, Ryu HG, Park Y, et al. Data curation for artificial intelligence in medicine: the VitalDB experience. *Korean J Anesthesiol*. 2022;75(3):211-9.
11. Kendale S, Kulkarni P, Rosenberg AD, Wang J. Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *Anesthesiology*. 2018;129(4):675-88.
12. Davies SJ, Kemp H, Julian H, et al. The prediction of hypotension during major abdominal surgery: a prospective observational study. *J Clin Monit Comput*. 2020;34(5):1023-31.
13. Wijnberge M, Geerts BF, Hol L, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of hypotension. *JAMA*. 2020;323(11):1052-60.
14. Maheshwari K, Shimada T, Yang D, et al. Hypotension Prediction Index for prevention of hypotension during moderate-to-high risk noncardiac surgery. *Anesthesiology*. 2020;133(6):1214-22.
15. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30.
16. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-making and performance. *JAMA*. 2015;314(10):1063-4.

doi:10.20944/preprints202601.1824.v1

17. Lee HC, Park Y, Yoon SB, et al. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. *Sci Data*. 2022;9(1):279.
18. Chen L, Ogundele O, Clermont G, et al. Dynamic predictive modeling of intraoperative blood pressure. *Anesth Analg*. 2022;134(1):145-54.
19. Pinsky MR. Functional hemodynamic monitoring. *Crit Care Clin*. 2015;31(1):89-111.
20. Michard F. Changes in arterial pressure during mechanical ventilation. *Anesthesiology*. 2005;103(2):419-28.
21. Vernooij LM, van Klei WA, Machina M, et al. Different definitions of hypotension and their association with postoperative complications. *Anesth Analg*. 2018;126(5):1520-9.
22. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22.
23. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32.
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proc KDD*. 2016:785-94.
25. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-38.
26. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67.
27. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997;55(1):119-39.
28. Cvach M. Monitor alarm fatigue: an integrative review. *Biomed Instrum Technol*. 2012;46(4):268-77.