

Robust preoperative and holistic prediction of hypoxemia in gastrointestinal endoscopy under sedation using machine learning: a retrospective study

JianJun Zou

`zoujianjun100@126.com`

Nanjing First Hospital, Nanjing Medical University

JiaFang Duan

Taikang Xianlin Drum Tower Hospital, Affiliated to Nanjing University School of Medicine

YanNa Si

Affiliated Nanjing Brain Hospital, Nanjing Medical University

PiHong Hou

Nanjing First Hospital, Nanjing Medical University

ShiTing Chen

China Pharmaceutical University

YeWen Zhang

China Pharmaceutical University

Zhou Zhou

Nanjing First Hospital, Nanjing Medical University

KaiZong Huang

China Pharmaceutical University

Rui Ding

Nanjing First Hospital, Nanjing Medical University

SiRuo Tang

Nanjing First Hospital, Nanjing Medical University

Yuan Zhang

Nanjing First Hospital, Nanjing Medical University

Research Article

Keywords: Gastrointestinal endoscopy, machine learning, Hypoxemia, Prediction, Outpatient

Posted Date: June 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9685417/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

In gastrointestinal endoscopy (GIE), the growing use of sedation and anesthesia often leads to intraoperative hypoxemia, escalating the risk of respiratory, renal, and cardiovascular complications. Providing anesthesiologists with a machine learning (ML) model tailored for GIE patients is essential for mitigating these risks.

Methods

A cohort of patients who underwent painless GIE were retrospective collected in the Nanjing First Hospital. We developed five ML models separately and selected ones with best performance as the final preoperative and holistic model. To verify the generalization ability of the final models, an independent cohort was collected from the Fourth Affiliated Hospital of Nanjing Medical University to conduct external validation. In addition, the Shapley additive explanations method was applied to interpret the optimal models.

Results

According to the inclusion and exclusion standard, we included a total of 10170 patients in our study. In both preoperative and holistic models, the XGBoost models were selected as the final models because of excellent performance (AUROC was 0.875 in preoperative model and 0.941 in holistic model). In external validation including 995 patients, the model's the model showed consistent performance, supporting its robustness and generalization.

Conclusions

Our project proposed preoperative and holistic ML models for predicting hypoxemia during outpatient GIE operations, ensuring patient safety during surgery, and providing them with safe and effective care. Factors such as hemoglobin levels, diabetes presence, nasopharyngeal ventilation, and high-flow oxygen use significantly influence hypoxemia in these patients.

Background

Sedation is increasingly administered during gastrointestinal endoscopy (GIE) to alleviate patient anxiety and discomfort, improve examination outcomes, and minimize procedural recall^[1]. However, these practices carry an inherent risk of intraoperative hypoxemia, a frequent adverse event during sedated GIE, with a reported incidence ranging from 1.4 to 51%^[2-4]. Severe hypoxemia necessitates temporary withdrawal of the endoscope, followed by mask ventilation or, if hypoxemia persists, urgent tracheal intubation. Hypoxemia during GIE, regardless of duration, carries clinical consequences. Prolonged episodes may result in myocardial ischemia^[5, 6], while even a brief desaturation ($\text{SpO}_2 < 90\%$) is associated with an elevated risk of

respiratory, renal, and cardiovascular complications^[7]. Nonetheless, the high patient volume in outpatient GIE, coupled with limited manpower and resources, often leads to suboptimal perioperative management and adverse outcomes. In this context, efficient identification of high-risk patients becomes critical, as it enables targeted escalation of sedation, monitoring, and preventive interventions for those most likely to benefit, while maintaining standard perioperative care for all, thereby reducing complications and improving patient outcomes.

Machine learning (ML) offers a powerful approach to address the challenges outlined above. By enabling complex multi-parametric decision algorithms and capturing nonlinear relationships, ML has proven effective in identifying high-risk patients in various clinical settings^[8, 9]. In the field of anesthesiology, Palla and colleagues demonstrate that integrating ML models during perioperative care improves clinicians' predictive accuracy^[10]. Extending this paradigm to gastrointestinal endoscopy under sedation, a tailored ML model could similarly assist anesthesiologists in anticipating and preventing hypoxemia.

However, existing prediction models fall short in addressing this need for outpatient GIE under sedation. First, Lundberg et al. developed three ML models for hypoxemia prevention during surgery under general anaesthesia^[11], but these are derived from inpatient populations and are not directly applicable to outpatient GIE settings. Second, Geng et al. proposed an artificial neural network (ANN) model for sedated GIE endoscopy using only three variables^[12]. Although it achieves an AUC of 0.80, its small sample size (n=220) limits generalizability and raises concerns about overfitting. Third, Geng et al. also developed a logistic regression (LR) model for GIE hypoxemia^[13], as LR is inherently limited to linear relationships, potentially overlooking complex, nonlinear interactions among clinical variables^[14]. In summary, a robust, interpretable ML model specifically tailored to outpatient GIE patients, with adequate sample size and validated across multiple centers, remains lacking.

Given that risk factors for hypoxemia span both preprocedural patient characteristics and intraoperative variables, a comprehensive prediction strategy should address both stages. Therefore, this study aimed to identify clinical predictors of hypoxemia in outpatient GIE, and develop and validate two ML models, a preoperative model based solely on preprocedural data and a holistic (intraoperative) model incorporating both preoperative and intraoperative variables, to enable timely risk assessment and intervention, ultimately enhancing patient safety and quality of care.

Methods

Study population

Patients who underwent painless GIE in our study were collected in the Nanjing First Hospital from October 2021 to February 2023. The Ethics Committee of Nanjing First Hospital granted approval for our study protocol, which was conducted in accordance with the principles of the Declaration of Helsinki [document number: KY20220509-01-KS-01]. At the same time, informed consent requirement was waived because that all data was anonymized for investigators.

The inclusion criteria were adults aged 18 years or older, with the American Society of Anesthesiologists (ASA) class I–II, receiving anesthesia-assisted gastrointestinal endoscopy (upper, lower, or both) in outpatient. While the exclusion criteria were as followings: (1) patients under 18 years old; (2) with severe cardiopulmonary diseases; (3) pregnancy, breastfeeding, or positive pregnancy test; (4) hypersensitivity to anesthetic agents; (5) therapeutic endoscopy; (6) having a serious lack of clinical data.

Anesthesia regimen and sedation monitoring

During the evaluation in anesthesia outpatient, patients with a high risk of hypoxemia were estimated by anesthesiologists empirically, such as obese patients and elderly patients. To lessen the probability of hypoxemia, preventive interventions like high flow oxygen (gas flow 60L/min) before anesthesia induction or nasopharyngeal airway at the beginning of GIE were applied for high-risk patients. In the procedure room, oxygen supplementation of 3 L/min was delivered via a nasopharynx tube for other common patients 3 minutes before anesthesia induction, which enhanced the basal pulse oximeter oxygen saturation (SpO₂). For anesthesia induction, patients were used propofol in combination with remifentanyl or along with low-dose etomidate via intravenous bolus, etomidate was particularly used for elderly as well as coronary heart disease patients^[15]. The MOAA/S score, which measures sedation on a scale of 0 to 5, was used to determine the degree of sedation reached at the time of GIE. Throughout the procedure, the deep sedation situation was maintained by giving a certain amount of propofol combined with remifentanyl. Standard monitoring, including electrocardiogram, blood pressure, and SpO₂ was measured and recorded in all patients. Once respiratory depression occurred during operation, the protective procedures were taken as follows: (1) 90% ≤ SpO₂ < 95%: increasing the oxygen flow from 3 to 6 L/min, (2) 75% ≤ SpO₂ < 90% for < 60 seconds: increasing the oxygen flow to 6 L/min and opening the airway by the jaw-thrust maneuver at same time, (3) 75% ≤ SpO₂ < 90% for ≥ 60 seconds or SpO₂ < 75%: performing mask ventilation, and implementing tracheal intubation in case the hypoxemia still could not be corrected. After the procedure, patients continued to be monitored in the post-anesthesia care unit until MOAA/S score was ≥ 4.

Study outcomes and definition

The primary endpoint in this study was the incidence of hypoxemia, which was defined as a pulse oximetry reading < 90% for any duration during the procedure and without probe misalignment, regardless of episode duration.

Data collection

Several common clinical variables were considered to collect, including: demographic characteristics, such as age, sex, and body mass index (BMI); medical history, such as hypertension, diabetes, hypothyroidism, obstructive sleep apnea-hypopnea syndrome (OSAHS) and chronic obstructive pulmonary disease (COPD); operation history, such as history of lung surgery and history of gastrointestinal surgery; airway assessment indicators, such as neck circumference, thyromental distance (TMD), thyromental height (TMH), sternomental distance (SMD), Inter-incisor distance, modified Mallampati class and short neck; physiological state, such as American Society of Anesthesiologists (ASA) class, hemoglobin and Baseline SpO₂; data recorded during operation, such as procedure time, propofol dose, remifentanyl dose and usage of etomidate; preventive interventions, such as high flow oxygen and nasopharyngeal airway.

Statistical analysis

For baseline characteristics, we used the Shapiro-Wilk test to judge whether continuous variables are normally distributed. Continuous variables were described as means with standard deviation or a median value with an interquartile range in accordance with the results of the Shapiro-Wilk test and the differences between groups were explored with Student's T-test or Mann-Whitney U-test. Categorical variables were presented as the total number of incidents and their percentage and Chi-squared or Fisher's exact test was used to assess the differences between groups. The above statistics and were conducted with SPSS version 25.0 (IBM Corporation, Armonk, NY, USA), and p-values <0.05 were considered statistically significant.

Data preparation for machine learning

Firstly, the entire database was randomly divided into a training set and a testing set at an 8:2 ratio. After splitting the database, the multi-classification variables were subjected to the One-Hot encoding, which resulting in dummy variables. Furthermore, the z-score, which is produced by deducting the sample mean from each individual value and then dividing the result by the sample standard deviation, was used for continuous variables to normalize. It is worth noting that we use training set to fit the normalized model, and then apply it to training data and test data, this step is crucial to preventing data leakage.

Model development

In order to meet different needs, facilitating clinicians to identify and prepare for the risk of hypoxemia in patients before procedure, we have divided the statistically significant variables of univariate analysis into two modules: preoperative variables and intraoperative variables. The least absolute selection and shrinkage operator (LASSO) algorithm was performed to select features in preoperative variables and all variables (preoperative variables and intraoperative variables) respectively, which was supposed to avoid model overfitting as well as increase interpretability and prediction accuracy^[16].

All selected features were included as predictors to construct five ML algorithms for both preoperative model and holistic model: (1) logistic regression, (2) random forests classifier (RFC), (3) support vector machines (SVM), (4) eXtreme Gradient Boosting (XGBoost), and (5) multi-layer perceptron (MLP). 10-fold cross-validation was carried out in the training dataset to prevent overfitting, and for each ML technique, the validation set was utilized to determine the ideal set of hyperparameters. In addition, considering the potential impact of data imbalance, the class weight hyperparameters was set up in every ML algorithm. All implementations of the models were based on software python (version 3.9.12; <https://www.python.org/>).

Model validation

In our study, the testing set served as data for internal validation to preliminarily prove the generalization of the ML model that developed according to the training set. Moreover, an independent cohort was collected from the Fourth Affiliated Hospital of Nanjing Medical University in March 2023, patients who received GIE under sedation were included as the external validation set to further evaluate the generalization ability and usability of the model.

Model evaluation

We used the area under receiver operating characteristic curve (AUROC) to evaluate the performance of models, which is the most commonly used indicator for evaluating prediction models. However, as the more precise gauge for the prediction power of models with unbalanced data, the area under precision-recall curve (AUPRC) was still recorded in our study. Brier score (The closer the score is to 0, the better), a means to quantify the discrepancy between an algorithm's projected probability and the actual outcome, was used to assess how well a model is calibrated. For each model, the optimal decision threshold was determined in the training set by maximizing the Youden index and was then applied unchanged to the testing set for calculation of threshold-dependent metrics, including sensitivity, specificity, accuracy, F1 score, positive predictive value, and negative predictive value. For the final XGBoost models, the corresponding thresholds were also applied unchanged to the external validation cohort. Model comparison was primarily based on AUROC, AUPRC, and the Brier score, whereas threshold-dependent metrics were reported to describe performance at the selected operating point. The final preoperative and holistic models were selected by comprehensively considering discrimination, calibration, and threshold-dependent clinical performance.

Model interpretation

Due to the black-box nature of ML models, its application in clinical environment is limited. However, transparency, rationality, and understanding of decision-making processes are important in the medical field, consequently we aim to overcome this flaw by applying Shapley additive explanations (SHAP) based interpretation methods. SHAP is a practical tool that helps increase the model's trust and acceptability in clinical practice, which could visualize the contribution and direction to the risk of hypoxemia of every clinical feature (A unified approach to interpreting model predictions^[11]). The impact of features on the prediction outcome increases with the absolute value of SHAP. A positive value indicates the positive impact of the feature on the outcome, whereas a negative value represents a negative impact^[17]. Besides, for further facilitate clinical use, especially in busy outpatient, we developed two user-friendly online web calculators stem from the XGBoost model. Clinicians or patients only need to input corresponding data based on the actual situation on the interface to obtain the risk of hypoxemia in painless GIE, which is simple, quick, and accessible risk estimation.

Results

Baseline characteristics

A total of 10170 eligible patients underwent painless GIE were selected in our study (47.2% males, median age of 52 years). Table 1 shows the detailed information on study population. In general, hypoxemia was observed in 1069 patients (10.5%, median age: 58 years, number of male patients: 592), on the contrary, 9101 patients smoothly went through the procedure (89.5%, median age: 50 years, number of male patients: 4209). The training set and the testing set were similar in baseline characteristics. The flow chart of our study is shown in Fig. 1.

Table 1
Baseline characteristics of the control group patients

Variables	Non-hypoxemia (N = 9101)	Hypoxemia (N = 1069)	P-values
Demographics			
Age, years, median (IQR)	50.0 [39.0, 60.0]	58.0 [53.0, 64.0]	< 0.001*
BMI, kg/m ² , median (IQR)	23.2 [21.3, 25.4]	26.3 [24.1, 28.0]	< 0.001*
Sex, female, n (%)	4892 (53.8)	477 (44.6)	< 0.001*
Medical History, n (%)			
Hypertension	2769 (30.4)	540 (50.5)	< 0.001*
Diabetes	1406 (15.4)	726 (67.9)	< 0.001*
Cerebral infarction	507 (5.6)	230 (21.5)	< 0.001*
Coronary heart disease	388 (4.3)	101 (9.4)	< 0.001*
Asthma	164 (1.8)	111 (10.4)	< 0.001*
Hypothyroidism	313 (3.4)	138 (12.9)	< 0.001*
COPD	87 (1.0)	90 (8.4)	< 0.001*
Allergic	293 (3.2)	151 (14.1)	< 0.001*
Snore	1491 (16.4)	432 (40.4)	< 0.001*
History of lung surgery	188 (2.1)	242 (22.6)	< 0.001*
History of gastrointestinal surgery			< 0.001*
No	8745 (96.1)	994 (93.0)	
Esophagogastric surgery	334 (3.7)	73 (6.8)	
Colorectal surgery	22 (0.2)	2 (0.2)	
Smoking			< 0.001*
No	7911 (86.9)	817 (76.4)	
Current	1118 (12.3)	242 (22.6)	
Ex	72 (0.8)	10 (0.9)	
Drinking	1578 (17.3)	290 (27.1)	< 0.001*
ASA score, n (%)			< 0.001*
Abbreviations: ASA, American Society of Anesthesiologists; BMI, body mass index; COPD, chronic obstructive pulmonary disease; RHSMD, height-to-sternomental distance; RHTMD, height-to-thyromental distance; TMD, thyromental distance; *, P < 0.05.			

Variables	Non-hypoxemia (N = 9101)	Hypoxemia (N = 1069)	P-values
☒	1395 (15.3)	36 (3.4)	
☒	7471 (82.1)	884 (82.7)	
☒	230 (2.5)	146 (13.7)	
☒	5 (0.1)	3 (0.3)	
Neck circumference, cm, median (IQR)	35.0 [34.0, 37.0]	37.0 [35.0, 40.0]	< 0.001*
RHTMD, median (IQR)	22.1 [21.2, 22.9]	22.5 [21.7, 23.3]	< 0.001*
TMH, cm, median (IQR)	4.7 [4.4, 5.0]	4.7 [4.5, 4.9]	0.012*
RHSMD, median (IQR)	10.7 [10.3, 11.2]	10.8 [10.4, 11.1]	0.027*
Inter-incisor distance, n (%)			< 0.001*
One finger	5 (0.1)	0 (0.0)	
Two fingers	226 (2.5)	53 (5.0)	
Three fingers	8866 (97.5)	1016 (95.0)	
Modified Mallampati class, n (%)			< 0.001*
☒	1659 (18.2)	47 (4.4)	
☒	6823 (75.0)	866 (81.0)	
☒	619 (6.8)	156 (14.6)	
Anodontia/Agomphosis, n (%)	431 (4.7)	75 (7.0)	0.002*
Micromandible, n(%)	87 (1.0)	27 (2.5)	< 0.001*
Short neck, n (%)	592 (6.5)	220 (20.6)	< 0.001*
Endoscopy, n (%)			< 0.001*
Gastroscopy	2807 (30.8)	306 (28.6)	
Colonoscopy	567 (6.2)	16 (1.5)	
both	5727 (62.9)	747 (69.9)	
Hemoglobin, g/L, median (IQR)	135.0 [127.0, 143.0]	127.0 [120.0, 135.0]	< 0.001*
Baseline SpO ₂ , %, median (IQR)	98.0 [97.0, 99.0]	97.0 [96.0, 97.0]	< 0.001*
Failed endoscope intubating times			< 0.001*
Abbreviations: ASA, American Society of Anesthesiologists; BMI, body mass index; COPD, chronic obstructive pulmonary disease; RHSMD, height-to-sternomental distance; RHTMD, height-to-thyromental distance; TMD, thyromental distance; *, P < 0.05.			

Variables	Non-hypoxemia (N = 9101)	Hypoxemia (N = 1069)	P-values
0	551 (6.1)	11 (1.0)	
1	8343 (91.7)	1022 (95.6)	
2	207 (2.3)	34 (3.2)	
3	0 (0.0)	2 (0.2)	
Procedure time, min, median (IQR)	19.0 [9.0, 23.0]	21.0 [10.0, 23.0]	< 0.001*
Induction propofol dose, mg, median (IQR)	125.0 [115.0, 135.0]	134.0 [125.0, 135.0]	< 0.001*
Total remifentanyl dose, µg, median (IQR)	45.0 [20.0, 55.0]	45.0 [20.0, 55.0]	0.704
Total propofol dose, mg, median (IQR)	245.0 [155.0, 285.0]	285.0 [185.0, 315.0]	< 0.001*
The use of etomidate, n (%)	990 (10.9)	56 (5.2)	< 0.001*
Nasopharyngeal airway, n(%)	899 (9.9)	50 (4.7)	< 0.001*
High flow oxygen, n (%)	636 (7.0)	13 (1.2)	< 0.001*
Abbreviations: ASA, American Society of Anesthesiologists; BMI, body mass index; COPD, chronic obstructive pulmonary disease; RHSMD, height-to-sternomental distance; RHTMD, height-to-thyromental distance; TMD, thyromental distance; *, P < 0.05.			

Feature selection

After the univariate analysis, 36 variables were included in the LASSO regression analysis. Finally, we identified 6 clinical features to build the preoperative models and 11 clinical features for the holistic models. The clinical features of preoperative models included: BMI, ASA class, neck circumference, hemoglobin, baseline SpO2 and diabetes. The clinical features incorporated in holistic models included: BMI, ASA class, Ratio of Height to Thyromental Distance (RHTMD), neck circumference, hemoglobin, baseline SpO2, induction propofol dose, diabetes, history of lung surgery, nasopharyngeal airway and high flow oxygen. The detailed coefficients of all non-zero coefficient variables after LASSO were shown in Supplementary Table 1.

Model performance

On the whole, both preoperative models and holistic models achieved good performance. AUROCs based on testing set of five models in preoperative models were LR (0.881, 95%CI: 0.872–0.891), RFC (0.913, 95%CI: 0.905–0.921), SVM (0.904, 95%CI: 0.895–0.913), XGBoost (0.918, 95%CI: 0.910–0.927), MLP (0.880, 95%CI: 0.870–0.890). As for the holistic model, the AUROCs in testing set were as following: LR (0.936, 95%CI: 0.918–0.954), RFC (0.939, 95%CI: 0.924–0.955), SVM (0.941, 95%CI: 0.923–0.959), XGBoost (0.941, 95%CI: 0.924–0.957), MLP (0.935, 95%CI: 0.916–0.955). Because of the minor differences of AUROCs between models (p-values in Delong test when XGBoost compared to other models were shown in Supplementary Tables 2 and 3), we comprehensively considered other evaluating indicators (Table 2) and finally XGBoost was deemed the optimal model for the preoperative model as well as the holistic model. Supplementary

Figs. 1 and 2 present the receiver operating characteristic (ROC) curve and the precision-recall curve (PRC) of each model. The calibration curve (Fig. 2) illustrates that results calculated from the developed ML models have good consistency with actual results. In addition, the decision curve analysis (DCA) is also shown in Fig. 2.

Table 2
A The performance of the five preoperative models based on the control group

	Threshold	Dataset	Accuracy	Sensitivity	Specificity	F1	Brier score	NPV	PPV
LR	0.489	training	0.800	0.823	0.797	0.459	0.145	0.975	0.318
		testing	0.800	0.823	0.797	0.483	0.143	0.972	0.342
RFC	0.555	training	0.834	0.837	0.834	0.509	0.130	0.978	0.366
		testing	0.833	0.818	0.835	0.527	0.130	0.973	0.389
SVM	0.224	training	0.856	0.802	0.862	0.534	0.067	0.974	0.400
		testing	0.855	0.779	0.865	0.550	0.072	0.968	0.425
XGB	0.141	training	0.822	0.872	0.817	0.503	0.061	0.982	0.353
		testing	0.823	0.853	0.819	0.522	0.068	0.977	0.376
MLP	0.105	training	0.803	0.805	0.803	0.458	0.072	0.973	0.320
		testing	0.801	0.810	0.800	0.481	0.077	0.970	0.342

Table 2
B The performance of the five holistic models based on the control group

	Threshold	Dataset	Accuracy	Sensitivity	Specificity	F1	Brier score	NPV	PPV
LR	0.519	training	0.862	0.866	0.862	0.573	0.107	0.982	0.428
		testing	0.857	0.861	0.857	0.545	0.105	0.982	0.399
RFC	0.517	training	0.868	0.884	0.866	0.587	0.103	0.984	0.440
		testing	0.855	0.856	0.855	0.541	0.108	0.982	0.395
SVM	0.117	training	0.869	0.869	0.869	0.586	0.050	0.982	0.442
		testing	0.870	0.886	0.868	0.575	0.046	0.986	0.425
XGB	0.155	training	0.884	0.880	0.885	0.618	0.044	0.984	0.477
		testing	0.879	0.851	0.882	0.582	0.046	0.982	0.442
MLP	0.133	training	0.874	0.849	0.877	0.589	0.052	0.980	0.451
		testing	0.874	0.847	0.877	0.572	0.047	0.981	0.432
Abbreviations: LR, logistic regression; RFC, random forest classifier; SVM, support vector machine; XGB, eXtreme Gradient Boosting; MLP, multi-layer perceptron; NPV, negative predictive value; PPV, positive predictive value.									

External validation

Through the inclusion and exclusion criteria mentioned above, a total of 995 patients were included in the externally validated cohort. According to the predefined XGBoost model, the AUROC was 0.875 (95%CI: 0.844–0.905) in preoperative model and was 0.935 (95%CI: 0.913–0.958) in holistic model. Using the decision threshold determined by the Youden index, the preoperative XGBoost model yielded a sensitivity of 79.1%, specificity of 78.0%, PPV of 31.9%, NPV of 96.6%; in holistic XGBoost model, we reached a sensitivity of 86.1%, specificity of 84.2%, PPV of 41.6%, NPV of 97.9%. More detailed evaluation metrics were listed in Table 3.

Table 3
The performance of the XGBoost algorithm in the external validation cohort

AUROC (95%CI)	preoperative model	holistic model
	0.875 (0.844–0.905)	0.935 (0.913–0.958)
AUPRC (95%CI)	0.497 (0.407–0.588)	0.726 (0.637–0.799)
Accuracy	0.781	0.844
Sensitivity	0.791	0.861
Specificity	0.780	0.842
F1	0.455	0.561
Brier score	0.077	0.060
NPV	0.966	0.979
PPV	0.319	0.416
Abbreviations: AUROC, the area under the receiver operating characteristic curve; AUPRC, the area under the precision recall curve; CI, Confidence Interval; NPV, negative predictive value; PPV, positive predictive value.		

Model interpretation

We rank the importance of features included in the XGBoost model based on the mean absolute SHAP value (Fig. 3A, 3C). As the figures show, diabetes and baseline SpO2 are the most important features in preoperative and holistic models respectively. In the SHAP summary plot (Fig. 3B, 3D), each spot represents a SHAP value for a feature and an instance. At the same time, the spots were colored according to the value of the feature; red color indicates direct link with hypoxemia while blue shows an inverse association. Therefore, diabetes, BMI, neck circumference and ASA class are risk factors in preoperative XGBoost model while hemoglobin as well as baseline SpO2 are protective factors. Similarly, in the holistic XGBoost model, diabetes, BMI, induction propofol dose, history of lung surgery, neck circumference, ASA class, and RHTMD were associated with higher predicted risk of hypoxemia, whereas baseline SpO2, hemoglobin, nasopharyngeal airway use, and high-flow oxygen use were associated with lower model-predicted risk. Notably, nasopharyngeal airway and high-flow oxygen were intervention-related variables; therefore, their SHAP patterns should be interpreted as model-based associations rather than evidence of independent causal protection.

Discussion

In this study, we developed and validated five machine learning models based on a large cohort of 10,170 patients undergoing GIE under sedation. These models were used to construct a preoperative model and a holistic model for predicting the risk of hypoxemia. Among the evaluated algorithms, XGBoost showed the best performance, with AUROC values of 0.918 for the preoperative model and 0.941 for the holistic model. Importantly, the main contribution of this study was not limited to improved predictive performance. We established a dual-model sequential prediction framework that covered both preoperative screening and

intraoperative reassessment. This framework provided a new approach for continuous perioperative risk assessment and clinical application of hypoxemia prediction.

Hypoxemia risk assessment was a stage-specific process rather than a single-time-point decision. The preoperative model was primarily used for initial risk assessment upon patient admission. By relying solely on variables available before surgery, the model enabled early risk stratification. It could help the anesthesia team optimize monitoring intensity, oxygenation support, and sedation strategies prior to the start of the procedure. In contrast, the holistic model incorporated intraoperative information. This allowed risk estimates to be dynamically updated during the procedure as new data became available. Such dynamic updating supported real-time adjustments to sedation depth, airway management, and oxygenation intervention strategies. The dual-model system developed in this study enabled sequential prediction from static preoperative assessment to dynamic intraoperative updating. As a result, hypoxemia risk assessment could extend throughout the perioperative period and better reflected real-world clinical management.

DCA provided a quantitative assessment of the model's potential clinical utility. At a prediction threshold of 0.1, the XGBoost model identified an additional 80 hypoxemia events per 1,000 patients. At the same threshold, its net benefit was 6.4-fold higher than that of the "treat-all" strategy, and this advantage became more pronounced at higher thresholds. These findings indicated that the model not only demonstrated good predictive performance, but also had the potential to improve the efficiency of high-risk patient identification in clinical practice. It could also provide an important basis for subsequent intervention planning.

To facilitate clinical implementation, we developed a web-based calculator based on the best-performing XGBoost model (<https://hypoxemia-gfkc8d2cwxn4tlv3j6ysk.streamlit.app/>). The tool was accessible via mobile phones or computers, which lowered the barrier to model use and enabled rapid risk assessment within routine clinical workflows. Specifically, the preoperative model was designed for early risk stratification before the procedure and to assist clinicians in developing individualized management strategies. During the procedure, the holistic model incorporated newly available information to dynamically update risk estimates and support continuous monitoring and timely intervention. Embedding the model into an operational tool allowed this study to translate predictive outputs into a practical pathway for clinical implementation. This approach was particularly relevant to high-throughput settings, such as outpatient and day-surgery endoscopy units.

Building on previous evidence, our study identified several hypoxemia-related factors, with practical implications for risk identification and targeted clinical intervention^[18–23]. From the patient's perspective, BMI and neck circumference were important risk factors identified in this study^[12, 13, 24–27]. These factors mainly reflected obesity-related changes in upper airway structure and reduced airway reserve^[28]. Previous studies have shown that obesity increases the load on upper airway soft tissues and reduces airway stability. As a result, airway collapse and inadequate ventilation were more likely to occur under sedation. For patients with a high BMI or increased neck circumference, greater attention should be paid to airway risk before the procedure. Additional airway assessment and anticipatory planning may be needed when appropriate. When feasible, long-term weight management should be encouraged to reduce the risk of perioperative hypoxemia. Diabetes may also increase the risk of hypoxemia^[29]. Therefore, good glycemic control was an important goal of chronic disease management and could also play a role in perioperative safety^[30]. For patients with

diabetes or poor glycemic control, blood glucose levels should be optimized as much as possible preoperatively to reduce the risk of hypoxemia^[31–38].

From the perspective of clinical interventions, this study further confirmed the important role of various intraoperative management strategies in reducing the risk of hypoxemia. The use of a nasopharyngeal airway and high-flow oxygen therapy^[39], both included in the holistic model, was significantly associated with a lower risk of hypoxemia. This finding was consistent with previous studies^[40–41]. These interventions may have reduced airway collapse during sedation by maintaining upper airway patency and improving oxygen reserve. They may also have counteracted the effects of reduced muscle tone and posterior displacement of the tongue^[42]. For patients identified as high risk before the procedure, anesthesiologists could prioritize enhanced oxygenation support and airway management during the procedure. Early use of a nasopharyngeal airway or high-flow oxygen therapy may help achieve targeted risk control^[43–45]. This strategy extended from preoperative risk optimization to dynamic intraoperative intervention. In this way, model outputs could be translated into specific clinical actions and may help improve perioperative patient safety.

More importantly, GIE under sedation was routinely performed on a large scale and served a broad patient population, particularly in primary care settings. Given its widespread use, even a modest incidence of hypoxemia could impose a substantial safety burden at the population level. Accordingly, an efficient, accessible, and dynamically updated risk assessment tool was of practical value in this setting. By developing a dual-model sequential prediction system and embedding it into a practical web-based tool, this study enabled hypoxemia risk to be identified before the procedure, updated during the procedure, and managed through targeted interventions. This approach may help reduce perioperative adverse events and improve patient safety.

This study had several limitations. First, its retrospective design may have introduced bias in data collection, data entry, and confounding control. Second, the external validation cohort was derived from the same geographic region and did not include populations from other regions or countries, which limited the assessment of model generalizability. Third, although the final XGBoost models showed favorable AUROC values and high negative predictive values in the external validation cohort, their positive predictive values were modest, with PPVs of 31.9% for the preoperative model and 41.6% for the holistic model. This finding may be partly related to the relatively low incidence of hypoxemia and the use of a Youden-index-based threshold, which was selected to balance sensitivity and specificity rather than to maximize PPV. Therefore, the models should be interpreted as risk-screening and decision-support tools rather than stand-alone triggers for intensive preventive interventions. Finally, the included variables were selected according to clinical experience and previous literature. Some relevant factors may not have been captured, leading to potential unmeasured confounding. Future multicenter, prospective studies are needed to validate the applicability of these models across different populations and to optimize decision thresholds according to local clinical workflows and resource availability.

Conclusion

In conclusion, we developed preoperative and holistic machine learning models for predicting hypoxemia in patients undergoing sedated gastrointestinal endoscopy. XGBoost is the best-performing algorithm among

the five models evaluated, External validation confirms the robustness of our approach, showcasing its potential clinical utility. These models may assist anesthesiologists in detecting and preventing hypoxemia, thereby enhancing patient safety during surgery and care during GIE procedures.

Abbreviations

ANN

artificial neural network

ASA

American Society of Anesthesiologists class

AUPRC

area under precision-recall curve

AUROC

area under receiver operating characteristic curve

BMI

body mass index

COPD

chronic obstructive pulmonary disease

ERV

expiratory reserve volume

FRC

functional residual capacity

GEPs

gastrointestinal endoscopic procedures

GIE

gastrointestinal endoscopy

HFNC

high-flow nasal cannula

HFNO

high flow nasal oxygenation

LASSO

least absolute selection and shrinkage operator

LR

logistic regression

ML

machine learning

MLP

multi-layer perceptron

NC

nasopharyngeal catheter

NP

nasal prongs

NPV
negative predictive value
OSA
obstructive sleep apnea
OSAHS
obstructive sleep apnea-hypopnea syndrome
PPV
positive predictive value
PRC
precision-recall curve
RFC
random forests classifier
ROC
receiver operating characteristic curve
SHAP
Shapley additive explanations
SMD
sternomental distance
SNC
standard nasal cannula
SpO₂
basal pulse oximeter oxygen saturation
SVM
support vector machines
TMD
thyromental distance
TMH
thyromental height
XGBoost
eXtreme Gradient Boosting
sleep apnea syndrome (SAS)
RHTMD
Ratio of Height to Thyromental Distance

Declarations

Ethics approval and consent to participate: The Ethics Committee of Nanjing First Hospital approved our study protocols. Moreover, informed permission was exempted due to the retrospective study design and anonymous data collection.

Clinical trial number: Not applicable.

Availability of data and materials: The datasets used and analysed during the current study cannot be shared publicly due to ethical reasons but are available from the corresponding author on reasonable request.

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

Funding: This work was supported by National Natural Science Foundation of China (81873954, 82173899), the Six Talent Peaks Project of Jiangsu (WSW-106), Nanjing Medical Science and Technical Development Foundation (ZKX22030) and Jiangsu Pharmaceutical Association (H202108, A2021024, Q202202, JY202207).

Authors' contributions: JJZ, YNS and JFD conceived, designed and supervised the study; STC and YWZ analyzed data and prepared results; PHH, STC, YWZ and ZZ wrote the paper with input from all authors; KZH revised the paper; RD, SRT and YZ were involved in data acquisition. All authors finally approved the paper. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Supplementary Information: Supplementary file

References

1. Early DS, Lightdale JR, Vargo JJ, 2nd, Acosta RD, Chandrasekhara V, et al. Guidelines for sedation and anesthesia in GI endoscopy. *Gastrointest Endosc.* 2018;87(2):327-37.
2. Gouda B, Gouda G, Borle A, Singh A, Sinha A, et al. Safety of non-anesthesia provider administered propofol sedation in non-advanced gastrointestinal endoscopic procedures: A meta-analysis. *Saudi J Gastroenterol.* 2017;23(3):133-43.
3. Qadeer MA, Rocio Lopez A, Dumot JA, Vargo JJ. Risk factors for hypoxemia during ambulatory gastrointestinal endoscopy in ASA I-II patients. *Dig Dis Sci.* 2009;54(5):1035-40.
4. Dumonceau JM, Riphaus A, Schreiber F, Vilmann P, Beilenhoff U, et al. Non-anesthesiologist administration of propofol for gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy, European Society of Gastroenterology and Endoscopy Nurses and Associates Guideline--Updated June 2015. *Endoscopy.* 2015;47(12):1175-89.
5. Fisher L, Fisher A, Thomson A. Cardiopulmonary complications of ERCP in older patients. *Gastrointest Endosc.* 2006;63(7):948-55.
6. Holm C, Christensen M, Rasmussen V, Schulze S, Rosenberg J. Hypoxaemia and myocardial ischaemia during colonoscopy. *Scand J Gastroenterol.* 1998;33(7):769-72.
7. Rostin P, Teja BJ, Friedrich S, Shaefi S, Murugappan KR, et al. The association of early postoperative desaturation in the operating theatre with hospital discharge to a skilled nursing or long-term care facility. *Anaesthesia.* 2019;74(4):457-67.
8. Lee SM, Hwangbo S, Norwitz ER, Koo JN, Oh IH, et al. Nonalcoholic fatty liver disease and early prediction of gestational diabetes mellitus using machine learning methods. *Clin Mol Hepatol.* 2022;28(1):105-16.

9. Venerito V, Lopalco G, Abbruzzese A, Colella S, Morrone M, et al. A Machine Learning Approach to Predict Remission in Patients With Psoriatic Arthritis on Treatment With Secukinumab. *Front Immunol.* 2022;13:917939.
10. Palla K, Hyland SL, Posner K, Ghosh P, Nair B, et al. Intraoperative prediction of postanaesthesia care unit hypotension. *Br J Anaesth.* 2022;128(4):623-35.
11. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2018;2(10):749-60.
12. Geng W, Tang H, Sharma A, Zhao Y, Yan Y, et al. An artificial neural network model for prediction of hypoxemia during sedation for gastrointestinal endoscopy. *J Int Med Res.* 2019;47(5):2097-103.
13. Geng W, Jia D, Wang Y, Jin S, Ren Y, et al. A prediction model for hypoxemia during routine sedation for gastrointestinal endoscopy. *Clinics (Sao Paulo).* 2018;73:e513.
14. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med.* 2016;375(13):1216-9.
15. Meng QT, Cao C, Liu HM, Xia ZY, Li W, et al. Safety and efficacy of etomidate and propofol anesthesia in elderly patients undergoing gastroscopy: A double-blind randomized clinical study. *Exp Ther Med.* 2016;12(3):1515-24.
16. Sun Y, Hu P, Wang J, Shen L, Xia F, et al. Radiomic features of pretreatment MRI could identify T stage in patients with rectal cancer: Preliminary findings. *J Magn Reson Imaging.* 2018.
17. Zhou SN, Jv DW, Meng XF, Zhang JJ, Liu C, et al. Feasibility of machine learning-based modeling and prediction using multiple centers data to assess intrahepatic cholangiocarcinoma outcomes. *Ann Med.* 2023;55(1):215-23.
18. Oknińska M, Mackiewicz U, Zajda K, Kieda C, Mączewski M. New potential treatment for cardiovascular disease through modulation of hemoglobin oxygen binding curve: Myo-inositol trispyrophosphate (ITPP), from cancer to cardiovascular disease. *Biomed Pharmacother.* 2022;154:113544.
19. Xiong W, Xu M, Pudasaini B, Guo X, Liu J. The influence of anemia on one-year exacerbation rate of patients with COPD-PH. *BMC Pulm Med.* 2018;18(1):143.
20. Ramanan M, Fisher N. The Association between Arterial Oxygen Tension, Hemoglobin Concentration, and Mortality in Mechanically Ventilated Critically Ill Patients. *Indian J Crit Care Med.* 2018;22(7):477-84.
21. Dhariwal A, Plevris JN, Lo NT, Finlayson ND, Heading RC, et al. Age, anemia, and obesity-associated oxygen desaturation during upper gastrointestinal endoscopy. *Gastrointest Endosc.* 1992;38(6):684-8.
22. Rozen P, Fireman Z, Gilat T. The causes of hypoxemia in elderly patients during endoscopy. *Gastrointest Endosc.* 1982;28(4):243-6.
23. Kulej-Lyko K, Niewinski P, Tubek S, Ponikowski P. Contribution of Peripheral Chemoreceptors to Exercise Intolerance in Heart Failure. *Front Physiol.* 2022;13:878363.
24. Laffin AE, Kendale SM, Huncke TK. Severity and duration of hypoxemia during outpatient endoscopy in obese patients: a retrospective cohort study. *Can J Anaesth.* 2020;67(9):1182-9.
25. Wani S, Azar R, Hovis CE, Hovis RM, Cote GA, et al. Obesity as a risk factor for sedation-related complications during propofol-mediated sedation for advanced endoscopic procedures. *Gastrointest Endosc.* 2011;74(6):1238-47.

26. Sharma VK, Nguyen CC, Crowell MD, Lieberman DA, de Garmo P, et al. A national study of cardiopulmonary unplanned events after GI endoscopy. *Gastrointest Endosc.* 2007;66(1):27-34.
27. Vargo JJ, Holub JL, Faigel DO, Lieberman DA, Eisen GM. Risk factors for cardiopulmonary events during propofol-mediated upper endoscopy and colonoscopy. *Aliment Pharmacol Ther.* 2006;24(6):955-63.
28. Rodrigues MP, Naves LA, Casulari LA, Silva CA, Araújo RR, et al. Using clinical data to predict sleep hypoxemia in patients with acromegaly. *Arq Neuropsiquiatr.* 2007;65(2a):234-9.
29. Karhu T, Myllymaa S, Nikkonen S, Mazzotti DR, Kulkas A, et al. Diabetes and cardiovascular diseases are associated with the worsening of intermittent hypoxaemia. *J Sleep Res.* 2022;31(1):e13441.
30. Rasche K, Keller T, Hader C, Leidag M, Prinz C. Impact of Obstructive Sleep Apnoea on Type 2 Diabetes and Vice Versa. *Eur Endocrinol.* 2013;9(2):107-9.
31. Walter RE, Beiser A, Givelber RJ, O'Connor GT, Gottlieb DJ. Association between glycemic state and lung function: the Framingham Heart Study. *Am J Respir Crit Care Med.* 2003;167(6):911-6.
32. Lawlor DA, Ebrahim S, Smith GD. Associations of measures of lung function with insulin resistance and Type 2 diabetes: findings from the British Women's Heart and Health Study. *Diabetologia.* 2004;47(2):195-203.
33. Yeh HC, Punjabi NM, Wang NY, Pankow JS, Duncan BB, et al. Cross-sectional and prospective study of lung function in adults with type 2 diabetes: the Atherosclerosis Risk in Communities (ARIC) study. *Diabetes Care.* 2008;31(4):741-6.
34. Tasali E, Mokhlesi B, Van Cauter E. Obstructive sleep apnea and type 2 diabetes: interacting epidemics. *Chest.* 2008;133(2):496-506.
35. Lecube A, Simó R, Pallayova M, Punjabi NM, López-Cano C, et al. Pulmonary Function and Sleep Breathing: Two New Targets for Type 2 Diabetes Care. *Endocr Rev.* 2017;38(6):550-73.
36. Aurora RN, Punjabi NM. Obstructive sleep apnoea and type 2 diabetes mellitus: a bidirectional association. *Lancet Respir Med.* 2013;1(4):329-38.
37. Floras JS. Sleep Apnea and Cardiovascular Disease: An Enigmatic Risk Factor. *Circ Res.* 2018;122(12):1741-64.
38. Tantucci C, Bottini P, Fiorani C, Dottorini ML, Santeusano F, et al. Cerebrovascular reactivity and hypercapnic respiratory drive in diabetic autonomic neuropathy. *J Appl Physiol (1985).* 2001;90(3):889-96.
39. Hung KC, Chang YJ, Chen IW, Soong TC, Ho CN, et al. Efficacy of high flow nasal oxygenation against hypoxemia in sedated patients receiving gastrointestinal endoscopic procedures: A systematic review and meta-analysis. *J Clin Anesth.* 2022;77:110651.
40. Bae JH, Baek YH, Lee JE, Song I, Lee JH, et al. Machine Learning for Detection of Safety Signals From Spontaneous Reporting System Data: Example of Nivolumab and Docetaxel. *Front Pharmacol.* 2020;11:602365.
41. Lin Y, Zhang X, Li L, Wei M, Zhao B, et al. High-flow nasal cannula oxygen therapy and hypoxia during gastroscopy with propofol sedation: a randomized multicenter clinical trial. *Gastrointest Endosc.* 2019;90(4):591-601.

42. Mazzeffi MA, Petrick KM, Magder L, Greenwald BD, Darwin P, et al. High-Flow Nasal Cannula Oxygen in Patients Having Anesthesia for Advanced Esophagogastroduodenoscopy: HIFLOW-ENDO, a Randomized Clinical Trial. *Anesth Analg*. 2021;132(3):743-51.
43. King AB, Alvis BD, Hester D, Taylor S, Higgins M. Randomized trial of a novel double lumen nasopharyngeal catheter versus traditional nasal cannula during total intravenous anesthesia for gastrointestinal procedures. *J Clin Anesth*. 2017;38:52-6.
44. Chhajed PN, Aboyoun C, Malouf MA, Hopkins PM, Plit M, et al. Prophylactic nasopharyngeal tube insertion prevents acute hypoxaemia due to upper-airway obstruction during flexible bronchoscopy. *Intern Med J*. 2003;33(7):317-8.
45. Hillman DR, Platt PR, Eastwood PR. The upper airway during anaesthesia. *Br J Anaesth*. 2003;91(1):31-9.

Figures

Image not available with this version

Figure 1

Flow chart of the study. GIE, gastrointestinal endoscopy

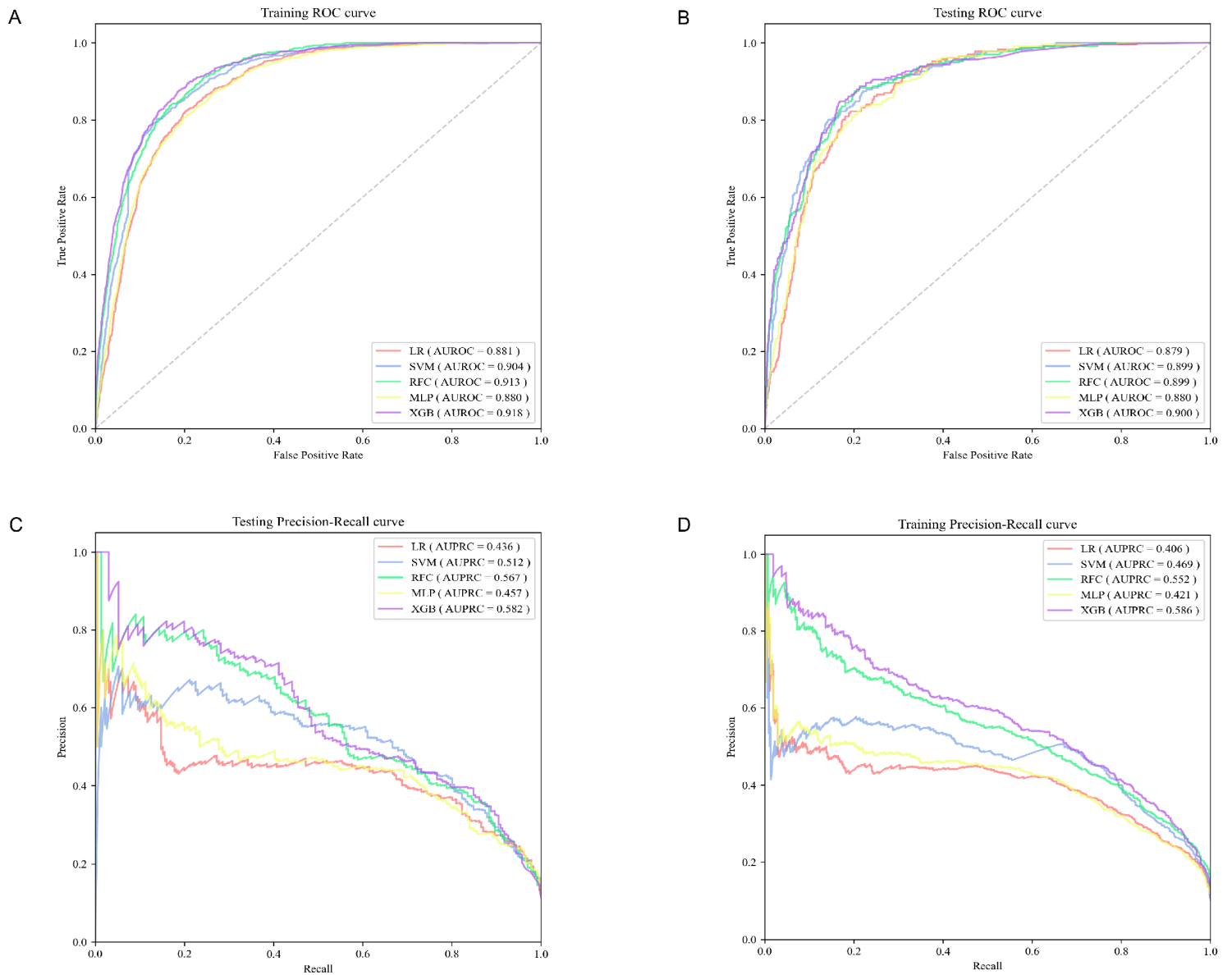


Figure 2

Calibration curve and decision curve analysis (DCA) of the preoperative and holistic models for predicting the risk of hypoxemia in outpatients who received painless gastrointestinal endoscopy (a) The calibration curve of the five preoperativemodels. (b) The DCA of the five preoperative models. (c) The calibration curve of the five holisticmodels. (d) The DCA of the five holistic models. LR, logistic regression; SVM, support vector machines; RFC, random forests classifier; MLP, multi-layer perceptron; XGB, eXtreme Gradient Boosting

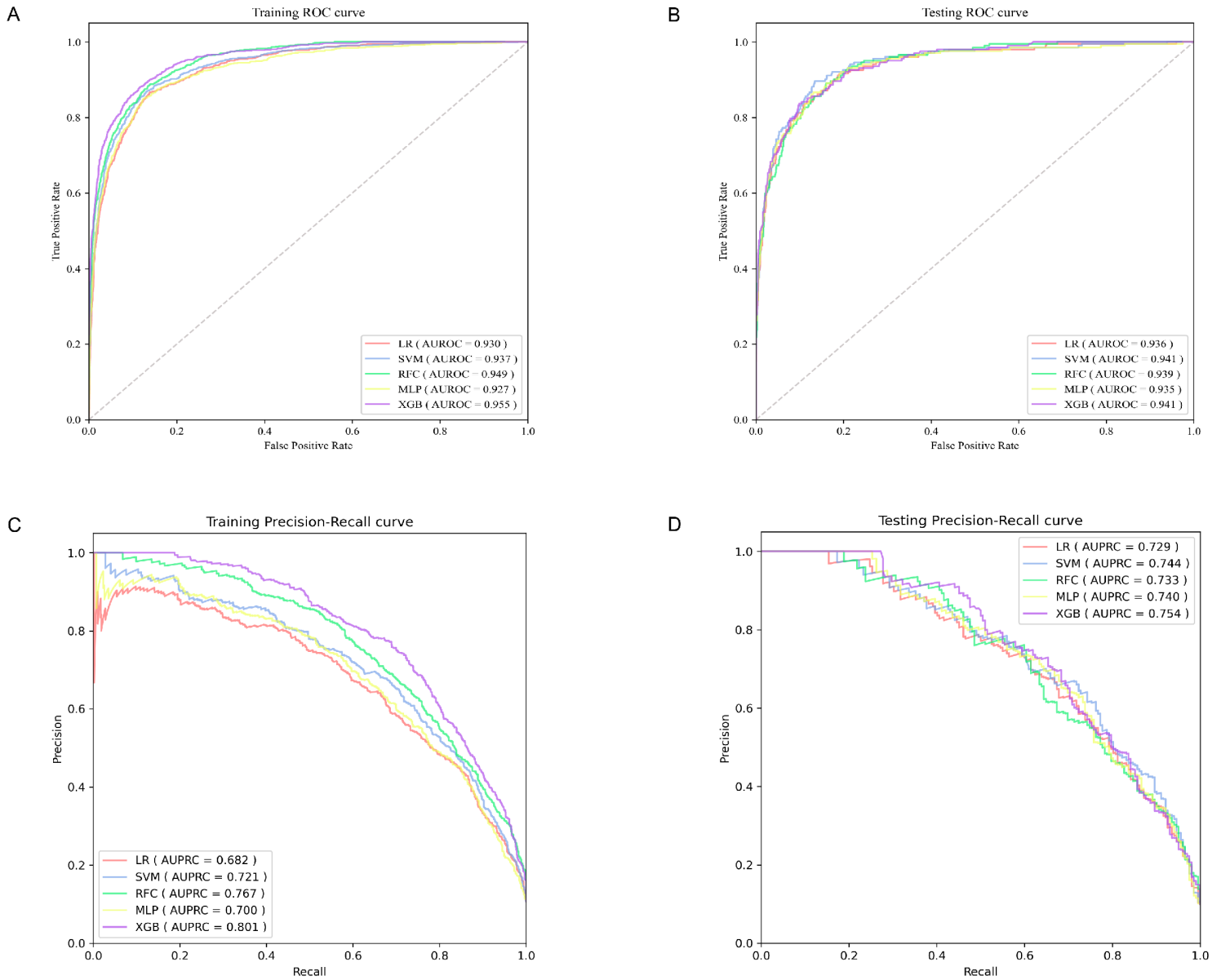


Figure 3

Shapley additive explanations (SHAP) plots of the selected variables in the XGB model (a) The mean absolute SHAP value ranks the relative importance of the variables in preoperative model. (b) The direction of effects for each variable in preoperative model. (c) The mean absolute SHAP value ranks the relative importance of the variables in holistic model. (d) The direction of effects for each variable in holistic model. BMI, body mass index; ASA, American Society of Anesthesiologists class; RHTMD, height-to-thyromental distance

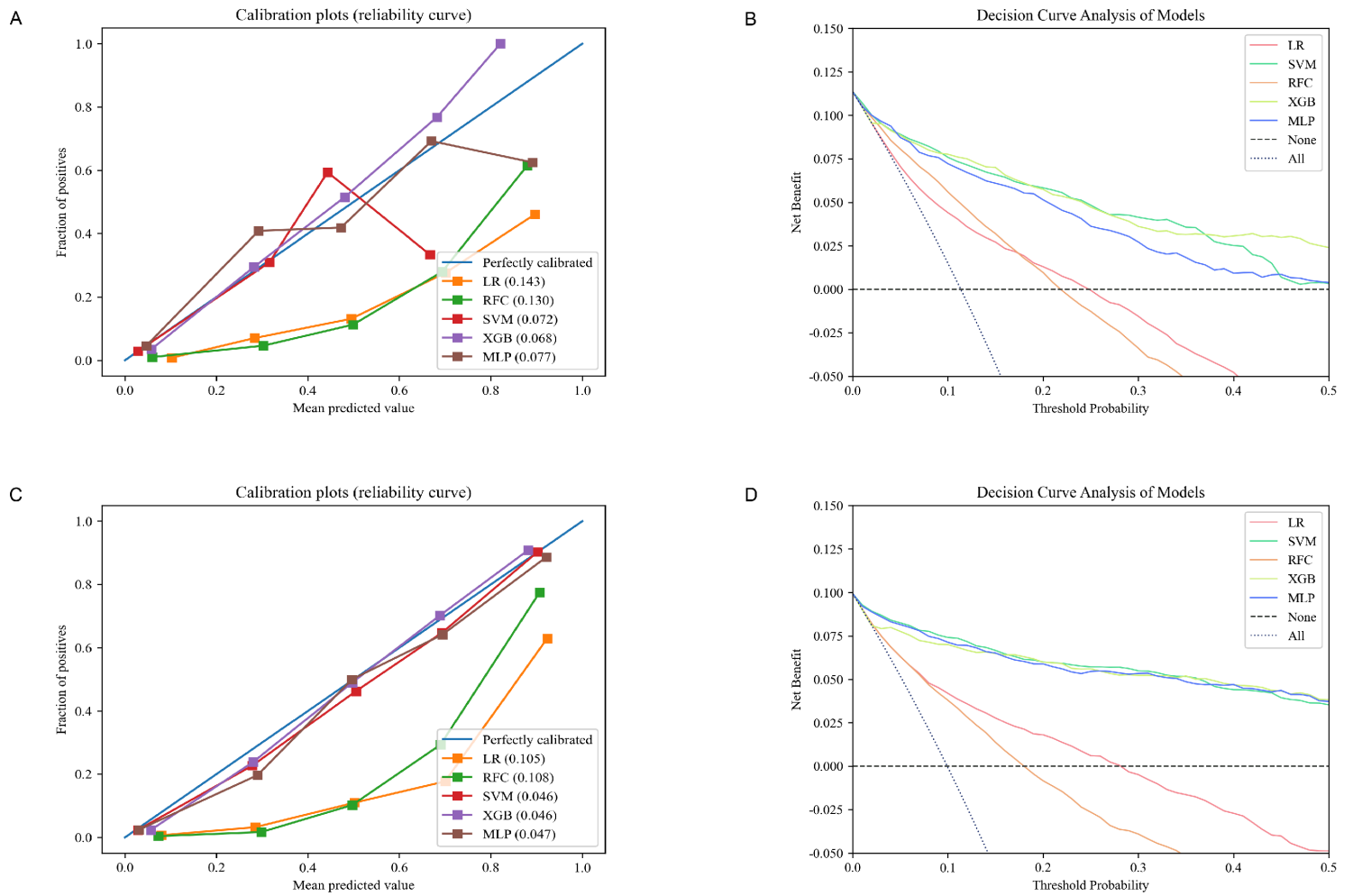


Figure 4

Legend not included with this version.

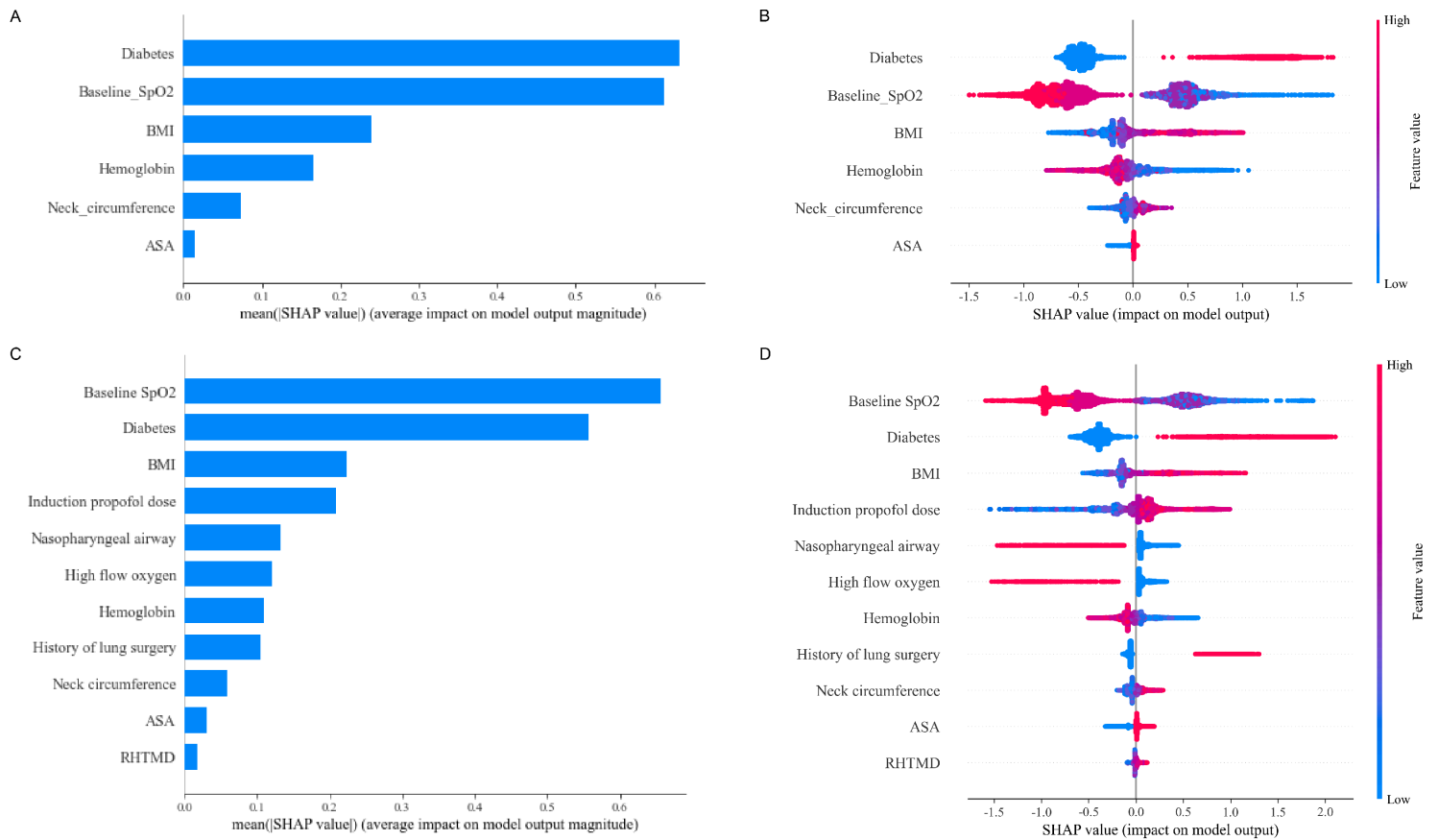


Figure 5

Legend not included with this version.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)
- [SupplementaryTable2.docx](#)
- [SupplementaryTable3.docx](#)