



OPEN Development and validation of a hypoxemia prediction model in middle-aged and elderly outpatients undergoing painless gastroscopy

Leilei Zheng¹, Xinyan Wu¹, Wei Gu⁴, Rui Wang², Jing Wang¹, Hongying He¹, Zhao Wang¹, Bin Yi³✉ & Yi Zhang¹✉

Hypoxemia is a common complication associated with anesthesia in painless gastroscopy. With the aging of the social population, the number of cases of hypoxemia among middle-aged and elderly patients is increasing. However, tools for predicting hypoxemia in middle-aged and elderly patients are lacking. In this study, we investigated the risk factors for hypoxemia in middle-aged and elderly outpatients undergoing painless gastroscopy based on machine learning and constructed a risk prediction model. In this retrospective study, we included the data on 1,348 outpatients undergoing painless gastroscopy. In total, 26 characteristic variables, including demographic information, past medical history, and clinical data of the patients were included, and BorutaShap was used for feature selection. Five machine learning algorithm models, including logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), and light gradient boosting machine (LightGBM), were selected. The best models were selected based on the area under the receiver operating characteristic curve (AUROC). Model feature importance was explained and analyzed using Shapley Additive Explanations (SHAP). The endpoint event of this study was considered to be hypoxemia during the procedure, defined as at least one occurrence of pulse oxygen saturation below 90% without probe misalignment or interference from the beginning of anesthesia induction to the end of painless gastroscopy. In the final cohort of 984 patients, 11% of patients (108/984) experienced hypoxemia during the painless gastroscopy procedure. The AUROCs of the five models were as follows: Logistic Regression (AUROC = 0.893, 95CI: 0.881–0.899), SVM (AUROC = 0.855, 95CI: 0.812–0.884), Random Forest (AUROC = 0.914, 95CI: 0.889–0.924), XGB (AUROC = 0.902, 95CI: 0.865–0.919), and LightGBM (AUROC = 0.891, 95CI: 0.847–0.917). Regarding the explanation of the importance of SHAP features, preoperative variables (baseline SpO₂, body mass index, and micrognathia) and intraoperative variables (operating time of gastroscopy, induction dose of etomidate and propofol mixture, append anesthetic, cough, and repeated pharyngeal irritation) significantly contributed to the model. We identified eight potential risk factors related to the occurrence of hypoxemia in middle-aged and elderly patients undergoing painless gastroscopy, based on machine learning feature engineering. Among the five machine learning algorithms, RF exhibited the best predictive performance in the internal test set and had a certain degree of generalization ability in the external validation set, which indicated that the RF model was more suitable for the data framework of this study. This model was more likely to enhance the accuracy of hypoxemia prediction in middle-aged and elderly patients undergoing painless gastroscopy, and thus, it is suitable for assisting anesthesiologists in clinical decision-making.

Keywords Machine learning, Middle-aged and elderly patients, Painless gastroscopy, Hypoxemia, Prediction model

¹Department of Anesthesiology, Second Affiliated Hospital of Zunyi Medical University, Zunyi 563000, Guizhou, China. ²Department of Anesthesiology, Third Affiliated Hospital of Zunyi Medical University, Zunyi 563000, Guizhou, China. ³Department of Anesthesiology, First Affiliated Hospital of Army Medical University, Chongqing

400038, China. ⁴Department of Anesthesiology, Minhang Hospital of Fudan University, Shanghai 201199, China. ✉email: yibin1974@163.com; zyzy@zmu.edu.cn

Gastrointestinal endoscopy is the most commonly used method to diagnose and treat digestive system diseases. Over 14 million patients in China undergo gastrointestinal endoscopy diagnosis and treatment every year. Painless gastrointestinal endoscopy diagnosis and treatment account for 48.3% of the total number, which is significantly lower than the proportion of patients undergoing painless gastrointestinal endoscopy diagnosis and treatment in the United States (98%) and Germany (82%). Due to the increase in the demand of patients for comfortable medical services, painless and comfortable gastrointestinal endoscopy services gained popularity very fast. By 2030, the number of patients receiving endoscopic diagnosis and treatment services for the digestive tract in China is estimated to reach 51 million¹.

Painless gastrointestinal endoscopy diagnosis and treatment services not only alleviate the pain and discomfort of patients caused by various factors, such as stress, pain, bloating, nausea, and vomiting but also create better diagnostic and treatment conditions for digestive endoscopists. However, they also involve certain risks. Hypoxemia caused by airway obstruction and respiratory depression is a very common complication associated with painless gastroscopy diagnosis and treatment services; hypoxemia has an incidence rate of 1.8–69.0%.² When patients experience severe hypoxemia, gastroscopy diagnosis and treatment services should be suspended and patients should be provided face mask pressure ventilation. Severe cases may even require tracheal intubation. If patients experience prolonged hypoxemia, it can cause serious complications, such as myocardial ischemia, arrhythmia, permanent nerve damage, and even death.³ The endpoint event of this study was hypoxemia during the procedure, defined as at least one occurrence of pulse oxygen saturation below 90% without probe misalignment or interference from the start of anesthesia induction to the end of painless gastroscopy diagnosis and treatment⁴.

The ability to accurately predict complications related to traditional perioperative anesthesia mainly relies on clinical experience, evaluation scales, and examination indicators; however, the method has certain limitations when dealing with complex situations⁵. Along with an increase in the age of the population, the number of middle-aged and elderly patients in hospitals is also increasing. Age is an independent risk factor for hypoxemia.^{6,7} Researchers have investigated various ways to effectively predict and prevent the occurrence of hypoxemia in middle-aged and elderly outpatients undergoing painless gastroscopy, however, reliable risk prediction tools for hypoxemia are lacking.

With the advancement and application of new technologies, such as big data, cloud computing, virtual reality, and intelligent robots, artificial intelligence (AI) is no longer limited to the field of computer technology and has penetrated various fields of social life; AI has become a key driving factor for transforming various fields. Under the impact and influence of the AI wave, the medical field is also undergoing a dramatic transformation. Machine learning (ML), the application core of AI, covers multiple fields, including probability, statistics, and computer science. ML can efficiently perform complex nonlinear data processing, risk prediction, disease diagnosis, and other functions; thus, providing us with new technological means to study various diseases and investigate new research perspectives⁸.

In this study, we investigated and identified the risk factors for hypoxemia in middle-aged and elderly outpatients undergoing painless gastroscopy, based on machine learning. We established a risk prediction model and validated them, to help clinical anesthesiologists identify high-risk patients early and provide relevant interventions.

Method

Research statement involving human participants

Ethical approval for this study was approved by the Ethics Committee of the First Affiliated Hospital of Army Medical University, Chongqing, CHINA (Chairperson Prof Qing Mao) on December 20, 2023 [grant number: (B) KY2023174], and was approved by the Ethics Committee of the Third Affiliated Hospital of Zunyi Medical University, [grant number: (2023) -1-170]. Due to the retrospective nature of the study, the Ethics Committee of the First Affiliated Hospital of Army Medical University, PLA and the Ethics Committee of the Third Affiliated Hospital of Zunyi Medical University waived the need for obtaining informed consent. This study was conducted in strict accordance with the Declaration of Helsinki.

Data source

The clinical data of 984 patients who visited the First Affiliated Hospital of Army Medical University from March 1, 2023 to May 31, 2023 were used for model establishment, with data was accessed on December 22, 2023; and the clinical data of 364 patients who visited the Third Affiliated Hospital of Zunyi Medical University from July 1, 2023 to September 31, 2023 were used for external validation of the model, with data was accessed on October 20, 2023. The inclusion criteria were as follows: outpatient painless gastroscopy patients undergoing elective intravenous general anesthesia; American Society of Anesthesiologists (ASA) class I–III; the age of patients ranged from 50 to 85 years, and individuals could be of either sex. The exclusion criteria were as follows: Patients undergoing complex endoscopic procedures (e.g., cholangiopancreatography, endoscopic ultrasound, endoscopic mucosal resection, submucosal dissection, etc.) and patients undergoing tracheal intubation; patients with potential life-threatening circulatory and respiratory diseases (e.g., uncontrolled severe hypertension, severe arrhythmia, unstable angina, acute respiratory infections, asthma, etc.); patients with liver dysfunction (Child-Pugh grade C or above), acute upper gastrointestinal bleeding with shock, severe anemia, and gastrointestinal obstruction with gastric content retention; patients with allergies to sedative/anesthetic drugs and other serious anesthesia-related risks; data loss greater than 20% or the inability to obtain important data.

Anesthesia management

After senior anesthesiologists confirm that a patient can tolerate general anesthesia and undergo gastroscopy diagnosis and treatment, the patients undergo tripartite verification, and their upper limb venous access is opened. The patients are oriented with their left side on the treatment bed and connected to a bedside electrocardiogram monitoring system. All patients undergoing painless gastroscopy receive nasal cannula oxygen at a rate of 3 L/min for inhalation, followed by a slow infusion of a mixture of etomidate and propofol (1:1 ratio) at 0.2 mL/kg for anesthesia induction. Depending on the depth of sedation, 1–2 mL of the mixture is administered once. Patients with higher anesthesia risks can be administered a slow infusion of opioid drugs (sufentanil 0.1 µg/kg) to assist in sedation. The depth of sedation was evaluated based on the modified Observer Alert/Sedation Score (MOAA/S). When the MOAA/S score was less than 2 points, digestive endoscopy was allowed. After surgery, patients with a MOAA/S score ≥ 4 points were transferred to the anesthesia recovery room. The patient's SpO₂ was continuously monitored during surgery and low-flow (≤ 5 L/min) nasal cannula oxygen was administered when the SpO₂ was between 95% and 100%; when the SpO₂ was between 90% and 94%, high-flow (> 5 L/min) nasal cannula oxygen was administered; when the SpO₂ was $< 90\%$, hypoxemia was recorded, which was the endpoint event of this experiment. Under such low-oxygen conditions, the nasal cannula oxygen flow rate should be increased, the jaw should be raised to open the airway, and if necessary, a simple respirator face mask should be removed from the gastroscope for pressure-assisted ventilation. If the oxygen levels do not increase, tracheal intubation with mechanical ventilation should be implemented.

Data collection

The clinical data of the patients were collected on sex, age, body mass index (BMI), baseline pulse rate, baseline SpO₂, previous underlying medical history (presence of hypertension, diabetes, coronary heart disease, cerebral infarction, chronic obstructive pulmonary disease, and asthma), snoring history, smoking history, drinking history, airway assessment (short neck and micrognathia), electrocardiogram results, ASA class, opioid-assisted sedation, induction dose of etomidate and propofol mixture, repeated pharyngeal irritation, intraoperative coughing, intraoperative motion, the addition of anesthetic drugs, intraoperative biopsy, and operating time of gastroscopy (26 characteristics).

Statistical analysis

All data were statistically analyzed using the SPSS 29.0 software. Normally distributed econometric data were presented as the mean \pm standard deviation ($\bar{X} \pm S$) and analyzed by the independent samples t-test to determine differences between groups. The skewed distribution econometric data were presented as the median (M) and interquartile interval (IQR) and analyzed by the Mann-Whitney U test to determine differences between groups. Count data were presented as count values and percentages, and the differences between groups were determined using χ^2 tests or Fisher's exact test. All statistical tests were two-tailed, and the differences between groups were considered to be statistically significant at $P < 0.05$. Data preprocessing, model development, testing, and validation were performed using the Python software (3.9.13).

Sample size calculation

The sample size for this retrospective cohort study was determined based on the availability of clinical data from two participating hospitals during the specified study periods (March–May 2023 for the internal cohort and July–September 2023 for the external validation cohort). A total of 1375 patients were initially identified, and after applying inclusion and exclusion criteria, 984 patients were retained for model development and 364 patient data from another region were collected for external validation. Given the exploratory nature of machine learning in this clinical context, formal power calculations were not performed. However, the sample size aligns with recommendations for predictive modeling studies, where a minimum of 10–20 events per predictor variable (EPV) is advised to avoid overfitting. With 108 hypoxemia events (11% incidence) and 8 key features selected by BorutaShap, the EPV ratio ($108/8 = 13.5$) falls within this range, supporting model stability. Additionally, stratified sampling (7:3 split) and SMOTE for class imbalance further ensured robustness.

Data preprocessing

The original data set was analyzed by experts to remove significant outliers, and multiple imputation (MI) was used to impute the characteristic variables with a missing degree below 20% of the total population. The classification variables were coded by the one-hot-encoding technique, and the numerical variables were normalized by MinMaxScaler to eliminate the dimensional effect of different data, make different features comparable, and improve the efficiency and accuracy of calculations. After shuffling the pre-processed dataset, stratified sampling was performed, and then, the data were divided into the training set and the test set at a ratio of 7:3 to ensure that the category ratio of the training set to the test set was close to the original data set.

Feature selection

The original data had some noisy or redundant features, which could affect the prediction ability of the model. In this study, BorutaShap was used for feature selection. Based on the random forest feature selection method, the Boruta algorithm determines the importance of features by comparing the importance between original features and randomly generated “shadow features”, which can effectively deal with the correlation and nonlinear relationship between features⁹ Shapley additive explanations¹⁰ (SHAP), based on the concept of Shapley value in game theory, can be used to understand the contribution of each feature to model prediction to provide a more comprehensive explanation. The BorutaShap library combines these two methods for feature selection, provides the explanation of feature importance, and determines the feature variables that significantly affect the

target outcome, to realize feature selection, reduce the risk of model overfitting, and improve the stability and generalization ability of the model.

Imbalanced data processing

In this study, the synthetic minority over-sampling technique¹¹ (SMOTE) was used to solve the problem of data imbalance. This technique is different from the traditional oversampling technique, and it synthesizes new minority samples to balance the gap between different classes instead of directly copying the existing minority samples. This oversampling method reduces the risk of overfitting the model. Compared to traditional oversampling techniques, SMOTE can effectively increase the number of minority samples while avoiding excessive noise, thus improving the performance of the model on an imbalanced dataset. At the same time, we compare the descriptive statistics of the data before and after oversampling with the importance of the variables, and the oversampling can only be justified if the descriptive statistics and/or the importance of the variables remain intact before and after oversampling.

Training and validation

The dataset of the First Affiliated Hospital of Army Medical University was used as the internal cohort (modeling group) for model training and testing, and the dataset of the Third Affiliated Hospital of Zunyi Medical University was used as the external validation cohort (external validation group) for external validation of the model to evaluate the stability and generalization ability of the model. The model performance was evaluated by an exhaustive search for various combinations of the given hyperparameters, and 10-fold cross-validation was used to reduce the bias caused by a single data partition and improve the reliability and generalization ability of model evaluation¹². Grid search and cross-validation were used to select the best combination of hyperparameters for classifier model training and prediction on the test dataset. The accuracy, area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), precision, recall, the F1 score, and the Brier score of the model were calculated. The score and the confusion matrix were used to evaluate the performance of the model.

Interpretation

In the field of machine learning, “black box” models have highly complex structures and parameter configurations, and their internal operating mechanisms are relatively opaque. These facts make it difficult to directly understand their decision-making process or synthesize a reasonable explanation of their prediction results. To increase the interpretability of model features, we introduced Shapley additive explanations¹⁰ (SHAP), which measure the degree to which each feature contributes to model predictions, based on the concept of Shapley value from cooperative game theory. By arranging and combining the features, the Shapley value of each feature was calculated. Then, these features were ranked to determine their importance and influence on model prediction. Features with a higher Shapley value have a greater influence on model prediction. A flow chart of the test (Fig. 1).

Results

Clinical data of patients

In total, 1,375 patients were included in this study; among which, 165 patients treated with painless ultrasound gastroscopy and 226 patients treated with painless gastroscopy were excluded. In the final cohort, 984 patients were included, among which 108 (11%) patients developed hypoxemia during painless gastroscopy. The differences in BMI, baseline SpO₂, induction dose of EP mixture, operating time of gastroscopy, hypertension, coronary disease, COPD, snoring, smoking, drinking, short neck, micrognathia, ASA class, opioids, repeated pharyngeal irritation, motion, cough, append anesthetic, and biopsy between the hypoxemia group and the non-hypoxemia group were significant ($P < 0.05$, Table 1). These results guide the feature selection of multivariate modeling.

The differences in the distribution of clinical indicators between the training set and the internal testing set were not significant ($P > 0.05$, Table 2), which indicated that there was no difference in the distribution of indicators between the two groups.

To accurately evaluate the stability and generalization ability of the model, the data distribution of the training set and the internal test set needs to be consistent, and both sets should have similar feature distribution and category distribution. If the data distribution of the training set and the internal test set is different, the model may overfit the feature or category distribution of the training set, reducing in the performance and stability of the model on the internal test set.

The differences in baseline pulse rate, baseline SpO₂, induction dose of EP mixture, operating time of gastroscopy, hypertension, diabetes, cerebral infarction, snoring, smoking, drinking, opioids, repeated pharyngeal irritation, motion, append anesthetic, and incidence of hypoxemia between the modeling group and the validation group were significant ($P < 0.05$, Table 3).

Continuous numerical variable data distribution

Based on the real-world scenario and the research purpose, appropriate indicators can be selected to describe the distribution characteristics of continuous variable data to better understand the nature and characteristics of the data. The continuous numerical variables measured in this study showed a skewed distribution; therefore, the distribution characteristics of the continuous numerical variables were described by the median (M) and interquartile range (IQR) (Fig. 2).

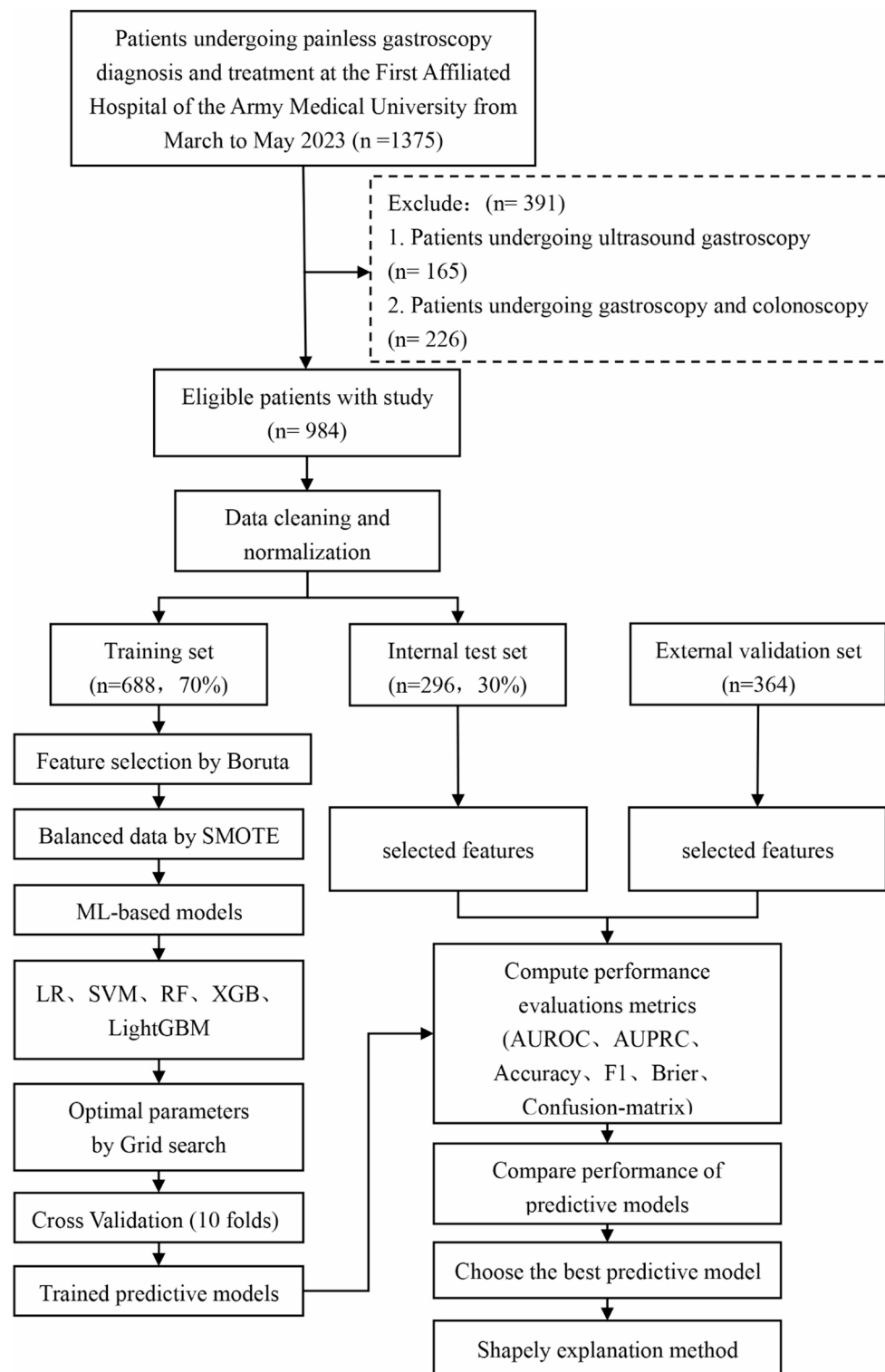


Fig. 1. A flow chart of the test.

BorutaShap feature variable selection

In the training set, BorutaShap was used to perform feature selection on the included variables, to detect the features that significantly affected hypoxemia; these identified features included baseline SpO₂, repeated pharyngeal irritation, operating time of gastroscopy, micrognathia, BMI, induction dose of etomidate and propofol mixture, intraoperative coughing, and append anesthetic (Fig. 3).

Variables	Hypoxemia (n = 108)	Non-hypoxemia (n = 876)	P-value
Age/years	63.0 (56.0, 70.0)	60.0 (56.0, 69.0)	0.258
BMI/kg m ⁻²	25.4 (22.5, 27.7)	22.3 (20.7, 24.5)	0.000▲
PR/min ⁻¹	74.0 (67.0, 81.0)	73.0 (66.0, 81.0)	0.568
SpO ₂ /%	97.0 (95.0, 98.0)	98.0 (97.0, 100.0)	0.000▲
EP Mixture/ml	12.0 (12.0, 14.0)	12.0 (10.0, 13.0)	0.000▲
Operating Time/min	4.0 (3.0, 4.0)	3.0 (2.0, 3.0)	0.000▲
Sex			0.118
Male	62 (57.4)	433 (49.4)	
Female	46 (42.6)	443 (50.6)	
Hypertension			0.013▲
Yes	29 (26.9)	150 (17.1)	
No	79 (73.1)	726 (82.9)	
Diabetes			0.587
Yes	10 (9.3)	68 (7.8)	
No	98 (90.7)	808 (92.2)	
Coronary Disease			0.002▲
Yes	19 (17.6)	73 (8.3)	
No	89 (82.4)	803 (91.7)	
Arrhythmia			0.237
Yes	24 (22.2)	154 (17.6)	
No	84 (77.8)	722 (82.4)	
Cerebral Infarction			0.344
Yes	2 (1.9)	9 (1.1)	
No	106 (98.1)	867 (98.9)	
Asthma			1.000
Yes	2 (1.9)	19 (2.2)	
No	106 (98.1)	857 (97.8)	
COPD			0.004▲
Yes	12 (11.1)	40 (4.6)	
No	96 (88.9)	836 (95.4)	
Snore			0.000▲
Yes	63 (58.3)	239 (27.3)	
No	45 (41.7)	637 (72.7)	
Smoke			0.004▲
Yes	32 (29.6)	157 (17.9)	
No	76 (70.4)	719 (82.1)	
Drink			0.000▲
Yes	36 (33.3)	154 (17.6)	
No	72 (66.7)	722 (82.4)	
Short Neck			0.000▲
Yes	66 (61.1)	152 (17.4)	
No	42 (38.9)	724 (82.6)	
Micrognathia			0.000▲
Yes	45 (41.7)	58 (6.6)	
No	63 (58.3)	818 (93.4)	
ASA Class			0.000▲
I	8 (7.4)	326 (37.2)	
II	67 (62.0)	426 (48.6)	
III	33 (30.6)	124 (14.2)	
Opioids			0.000▲
Yes	50 (46.3)	147 (16.8)	
No	58 (53.7)	729 (83.2)	
Pharyngeal Irritation			0.000▲
Yes	56 (51.9)	93 (10.6)	
No	52 (48.1)	783 (89.4)	
Motion			0.000▲
Continued			

Variables	Hypoxemia (n = 108)	Non-hypoxemia (n = 876)	P-value
Yes	21 (19.4)	51 (5.8)	
No	87 (80.6)	825 (94.2)	
Cough			0.000▲
Yes	34 (31.5)	48 (5.5)	
No	74 (69.5)	828 (94.5)	
Append Anesthetic			0.000▲
Yes	54 (50.0)	118 (13.5)	
No	54 (40.0)	758 (86.5)	
Biopsy			0.000▲
Yes	34 (31.5)	128 (14.6)	
No	74 (68.5)	748 (85.4)	

Table 1. Comparison of clinical data between the hypoxemia group and the non-hypoxemia group [M (P25, P75), cases (%)]. ▲Compared with the control group $P<0.05$.

Model prediction performance and validation results

The feature variables selected by BorutaShap were incorporated into five machine learning algorithms. Next, grid parameter search and 10-fold cross-validation were conducted, and the accuracy, AUROC, AUPRC, precision, recall, F1 score, and Brier score of the models were calculated. The accuracy, AUROC, AUPRC, F1 score, and Brier score of RF in the internal test set showed the best results, and it also performed well in the external validation dataset, which confirmed that RF was the best predictive model in this study and it was more suitable for the dataset (Table 4). The ROC and PRC curves (Fig. 4) and confusion matrix graphs (Fig. 5) were plotted for the five machine learning algorithm test sets using the Python (3.9.13) software. Finally, the best model was selected for the SHAP feature importance analysis (Fig. 6).

ROC curve and PRC curve of classifier model

The ROC curve can be used to determine the performance of the classifier model at different thresholds with TPR as the vertical axis and FPR as the horizontal axis. Values range from 0 to 1, with values closer to 1 indicating better classifier performance. A PRC curve was plotted, with Precision as the vertical axis and Recall as the horizontal axis. The values ranged from 0 to 1, with values closer to 1 indicating better classifier model performance.

Plot of the confusion matrix of the classifier model

A two-dimensional chart was presented to visualize the prediction performance of the classifier model, and the true category was cross-compared with the predicted category, including TP, TN, FP, and FN. Based on the TP, TN, FP, and FN, the classifier model evaluation indices, such as Accuracy, Precision, Recall, and F1 score, were calculated.

Classifier model SHAP interpretability analysis

Feature importance ranking and model interpretability analysis were performed to select the best prediction model, and SHAP was used for model feature importance ranking and interpretability analysis to understand the decision-making process of the model. It is worth noting that random forest impurity or permutation importance metrics are employed to assess the “strength” of the association between the dependent variable and its predictors, while random forest SHAP importance values are used to investigate the “direction” of this association. Based on this, we will derive and focus on important results for random forest impurities or permutations.

Discussion

Hypoxemia is a common anesthesia-related complication in painless gastroscopy diagnosis and treatment services. It has different degrees of effect on the perioperative anesthesia safety of painless gastroscopy patients^{13–15} Published studies lack hypoxemia risk prediction models for middle-aged and elderly outpatients undergoing painless gastroscopy. Considering age as an independent risk factor for hypoxemia during painless gastroscopy,^{6,7} in this study, we used perioperative data of middle-aged and elderly patients undergoing painless gastroscopy in outpatient clinics for modeling. This approach avoided inaccurate parameter estimation caused by directly performing whole population data modeling, enhancing the predictive accuracy of the model applied to the population of middle-aged and elderly patients undergoing painless gastroscopy in outpatient clinics, and making the final prediction model more targeted to effectively predict the occurrence of hypoxemia in middle-aged and elderly patients undergoing painless gastroscopy in outpatient clinics.

In this study, clinical data, such as demographic information, past medical history, surgical anesthesia records, and other relevant data, were collected based on the anesthesia and surgery management system of the hospital. The original dataset was cleaned by removing outliers and imputation of missing values, coding of categorical variables, and standardization of numerical variables. The internal cohort dataset was stratified into the training set and the test set at a ratio of 7:3. BorutaShap and SMOTE were used for feature selection and data imbalance treatment on the training set, and grid search hyperparameters and 10-fold cross-validation

Variables	Training set (n = 688)	Testing set (n = 296)	P-value
Age/years	60.0 (56.0, 69.0)	61.0 (56.0, 70.0)	0.772
BMI/kg m ⁻²	22.7 (20.8, 24.9)	22.5 (20.8, 25.1)	0.867
PR/min ⁻¹	73.0 (66.0, 81.0)	73.0 (66.0, 82.0)	0.597
SpO ₂ /%	98.0 (97.0, 99.0)	98.0 (97.0, 99.0)	0.112
EP Mixture/ml	12.0 (10.0, 13.0)	12.0 (10.0, 13.0)	0.464
Operating Time/min	3.0 (2.0, 3.0)	3.0 (2.0, 3.0)	0.577
Sex			0.412
Male	336 (48.8)	153 (51.7)	
Female	352 (51.2)	143 (48.3)	
Hypertension			0.608
Yes	128 (18.6)	51 (17.2)	
No	560 (81.4)	245 (82.8)	
Diabetes			0.363
Yes	51 (7.4)	27 (9.1)	
No	637 (92.6)	269 (90.9)	
Coronary Disease			0.380
Yes	68 (9.9)	24 (8.1)	
No	620 (90.1)	272 (91.9)	
Arrhythmia			0.793
Yes	123 (17.9)	55 (18.6)	
No	565 (82.1)	241 (81.4)	
Cerebral Infarction			0.742
Yes	7 (1.0)	4 (1.4)	
No	681 (99.0)	292 (98.6)	
Asthma			0.879
Yes	15 (2.2)	6 (2.0)	
No	673 (97.8)	290 (98.0)	
COPD			0.610
Yes	38 (5.5)	14 (4.7)	
No	650 (94.5)	282 (95.3)	
Snore			0.562
Yes	215 (31.3)	87 (29.4)	
No	473 (68.7)	209 (70.6)	
Smoke			0.744
Yes	134 (19.5)	55 (18.6)	
No	554 (80.5)	241 (81.4)	
Drink			0.758
Yes	136 (19.8)	56 (18.9)	
No	552 (80.2)	240 (81.1)	
Short Neck			0.923
Yes	153 (22.2)	65 (22.0)	
No	535 (77.8)	231 (78.0)	
Micrognathia			0.498
Yes	75 (10.9)	28 (9.5)	
No	613 (89.1)	268 (90.5)	
ASA Class			0.340
I	234 (34.0)	100 (33.8)	
II	337 (49.0)	156 (52.7)	
III	117 (17.0)	40 (13.5)	
Opioids			0.207
Yes	145 (21.1)	52 (17.6)	
No	543 (78.9)	244 (82.4)	
Pharyngeal Irritation			0.904
Yes	102 (14.8)	43 (14.5)	
No	586 (85.2)	253 (85.5)	
Motion			0.927
Continued			

Variables	Training set (n = 688)	Testing set (n = 296)	P-value
Yes	50 (7.3)	22 (7.4)	
No	638 (92.7)	274 (92.6)	
Cough			0.276
Yes	53 (7.7)	29 (9.8)	
No	635 (92.3)	267 (90.2)	
Append Anesthetic			0.336
Yes	115 (16.7)	57 (19.3)	
No	573 (83.3)	239 (80.7)	
Biopsy			0.207
Yes	120 (17.4)	42 (14.2)	
No	568 (82.6)	254 (85.8)	
Hypoxemia			0.914
Yes	76 (11.0)	32 (10.8)	
No	612 (89.0)	264 (89.2)	

Table 2. Comparison of clinical data between training set and internal testing set [M (P25, P75), cases (%)].

were used to improve the performance of the model. Five machine learning algorithms, including LR, SVM, RF, XGB, and LightGBM were applied to construct the prediction model of hypoxemia in middle-aged and elderly outpatients undergoing painless gastroscopy.

The results showed that the top eight variables associated with feature importance included the operating time of gastroscopy, baseline SpO₂, BMI, induction dose of etomidate and propofol mixture, append anesthetic, micrognathia, cough, and repeated pharyngeal irritation. Among them, preoperative baseline pulse oxygen saturation was negatively correlated with the occurrence of hypoxemia, which indicated that higher preoperative baseline pulse oxygen saturation values were associated with a lower risk of hypoxemia. The other seven feature variables were positively correlated with the occurrence of hypoxemia.

The operating time of gastroscopy was closely related to the proficiency of the endoscopic physician, the stomach condition of the patients, and whether gastroscopy was suspended during surgery. In our internal cohort, the median duration of painless gastroscopy was 3 min, and the incidence of hypoxemia was 11%. In the external validation cohort, the median duration of painless gastroscopy was 4 min, and the incidence of hypoxemia was 16% ($P < 0.001$); this difference between the cohorts might be related to differences in the proficiency of the endoscopyologists in the two cohorts. Moreover, the gastric conditions of the patients (e.g., gastric polyps) and intraoperative suspension of gastroscopy (e.g., cough and motion caused by insufficient anesthesia depth) increased the dosage of anesthetic and sedative drugs and also increased the duration of gastroscopy. Our findings suggested that the occurrence of hypoxemia was positively correlated with the operating time, which matched the findings of Van¹⁶. Many patients are at risk of hypoxemia during surgical sedation, which suggests that it is positively correlated with the increase in diagnosis and treatment time. However, further prospective studies are needed to confirm this hypothesis.

Preoperative baseline SpO₂ mainly reflects the basal oxygenation status of the patients. Human mitochondrial metabolism relies on sufficient oxygen to synthesize ATP; however, hypoxia leads to a decrease in ATP synthesis, which affects tissue and organ functions¹⁷. Patients with anemia, congenital heart disease, chronic obstructive pulmonary disease, bronchial asthma, and other complications are more likely to have a decrease in SpO₂ than patients who do not have these complications^{18–20}. This is consistent with the results of our study, where we found that the median preoperative basal SpO₂ was 97 in the 108 patients (11%) who experienced hypoxemia in the internal cohort and 98 in the remaining 876 patients (89%) who did not experience hypoxemia ($P < 0.001$). Patients with preoperative basal SpO₂ below 90% were not included in this study. Data on 61 patients with preoperative basal SpO₂ between 90% and 94% were collected in the internal cohort, among which, 26 patients developed hypoxemia, with an incidence of 42.6%. Patients with low preoperative basal SpO₂ may have acute or chronic lung diseases, low tidal volume, and insufficient oxygen reserve^{21,22}, and thus, they should be provided adequate attention. Before anesthesia induction, the oxygen flow through the nasal catheter and the preoxygenation time should be increased to improve the oxygen reserve of patients, to prevent the occurrence of hypoxemia during gastroscopy.

Obesity is an independent risk factor for hypoxemia in patients undergoing painless gastroscopy²³. Compared to patients with a normal BMI range, obese patients are more likely to experience airway obstruction after receiving anesthesia and sedation, which can increase the incidence of sedation-related adverse events and related airway interventions. Among the 984 patients in the internal cohort, 74 patients had a BMI of 28 kg·m⁻² or higher, and hypoxemia occurred in 24 patients (32.4%). Obese patients often experience obstructive sleep apnea (OSA)²⁴, which is a chronic hypoxic state that may affect their lung function and blood oxygen saturation; thus, increasing the risk of hypoxemia during painless gastroscopy.

Anesthetic drugs have an inhibitory effect on the respiratory center and can decrease the respiratory frequency and depth, thus affecting oxygen uptake and transport. Anesthesia induction with propofol alone has disadvantages, such as intravenous pain and high fluctuations in the respiratory and circulatory systems²⁵. In this study, a mixture of etomidate and propofol (1:1) was used to induce anesthesia and achieve a better sedation

Variables	Internal group (<i>n</i> = 984)	External group (<i>n</i> = 364)	<i>P</i> -value
Age/years	61.0 (56.0, 69.0)	63.0 (55.0, 70.0)	0.614
BMI/kg m ⁻²	22.6 (20.8, 24.9)	22.5 (20.8, 24.2)	0.144
PR/min ⁻¹	73.0 (66.0, 81.0)	78.0 (70.0, 84.0)	0.000▲
SpO ₂ /%	98.0 (97.0, 99.0)	97.0 (96.0, 98.0)	0.000▲
EP Mixture/ml	12.0 (10.0, 13.0)	12.0 (12.0, 13.0)	0.000▲
Operating Time/min	3.0 (2.0, 3.0)	4.0 (3.0, 5.0)	0.000▲
Sex			0.176
Male	495 (50.3)	168 (46.2)	
Female	489 (49.7)	196 (53.8)	
Hypertension			0.000▲
Yes	179 (18.2)	126 (34.6)	
No	805 (81.8)	238 (65.4)	
Diabetes			0.000▲
Yes	78 (7.9)	60 (16.5)	
No	906 (92.1)	304 (83.5)	
Coronary Disease			0.151
Yes	92 (9.3)	25 (6.9)	
No	892 (90.7)	339 (93.1)	
Arrhythmia			0.631
Yes	178 (18.1)	70 (19.2)	
No	806 (81.9)	294 (80.8)	
Cerebral Infarction			0.000▲
Yes	11 (1.1)	29 (8.0)	
No	973 (98.9)	335 (92.0)	
Asthma			0.943
Yes	21 (2.1)	8 (2.2)	
No	963 (97.9)	356 (97.8)	
COPD			0.066
Yes	52 (5.3)	29 (8.0)	
No	932 (94.7)	335 (92.0)	
Snore			0.000▲
Yes	302 (30.7)	72 (19.8)	
No	682 (69.3)	292 (80.2)	
Smoke			0.000▲
Yes	189 (19.2)	106 (29.1)	
No	795 (80.8)	258 (70.9)	
Drink			0.002▲
Yes	192 (19.5)	99 (27.2)	
No	792 (80.5)	265 (72.8)	
Short Neck			0.346
Yes	218 (22.2)	72 (19.8)	
No	766 (77.8)	292 (80.2)	
Micrognathia			0.543
Yes	103 (10.5)	34 (9.3)	
No	881 (89.5)	330 (90.7)	
ASA Class			0.572
I	334 (33.9)	113 (31.0)	
II	493 (50.1)	193 (53.0)	
III	157 (16.0)	58 (16.0)	
Opioids			0.000▲
Yes	197 (20.0)	122 (33.5)	
No	787 (80.0)	242 (66.5)	
Pharyngeal Irritation			0.002▲
Yes	145 (14.7)	80 (22.0)	
No	839 (85.3)	284 (78.0)	
Motion			0.000▲
Continued			

Variables	Internal group (n=984)	External group (n=364)	P-value
Yes	72 (7.3)	60 (16.5)	
No	912 (92.7)	304 (83.5)	
Cough			0.789
Yes	82 (8.3)	32 (8.8)	
No	902 (91.7)	332 (91.2)	
Append Anesthetic			0.006▲
Yes	172 (17.5)	88 (24.2)	
No	812 (82.5)	276 (75.8)	
Biopsy			0.544
Yes	162 (16.5)	65 (17.9)	
No	822 (83.5)	299 (82.1)	
Hypoxemia			0.007▲
Yes	108 (11.0)	60 (16.0)	
No	876 (89.0)	304 (84.0)	

Table 3. Comparison of clinical data between modeling group and external validation group [M (P25, P75), cases (%)]. ▲Compared with the control group P<0.05.

effect and patient satisfaction. The results suggested that the increase in the dose of the mixture for anesthesia induction and the addition of intraoperative anesthetic drugs increased the risk of hypoxemia in middle-aged and elderly outpatients undergoing painless gastroscopy. This finding matched the results reported by Hillman²⁶, who found a linear relationship between progressive anesthesia induction with propofol and an increase in upper airway collapse, which can cause airway obstruction, and thus, increase the risk of hypoxemia in patients undergoing gastroscopy.

The lack of complete airway protection measures during painless gastroscopy and the special position during diagnosis and treatment present great challenges to airway management, which increases the risk of airway obstruction and hypoxemia in patients with difficult airways²⁷ (such as micrognathia). The airway condition of the patient should be evaluated before painless gastroscopy to prevent hypoxemia during gastroscopy.

Intraoperative coughing is related to pharyngeal irritation, anesthesia depth, and whether the patient has reflux esophagitis when the endoscopic physician inserts the endoscope^{28,29}. Coughing can cause diaphragmatic and bronchial spasms, affecting normal ventilation and ventilation function, thus increasing the risk of hypoxemia in painless gastroscopy patients.

Compared to previous studies, this study included more characteristic variables, and the evaluation index results of the best model were better. Geng et al.³⁰ used traditional logistic regression methods and showed that age, BMI, and habitual snoring were independent risk factors for hypoxemia. The area under the working characteristic curve of the model subjects was 0.76. The researchers designed a neural network prediction model that included three characteristic variables: BMI, habitual snoring, and neck circumference. The area under the working characteristic curve of the model subjects was 0.80³¹; Li et al.³² used the NoSAS (neck circumference, obesity, snore, age, and sex) questionnaire combined with an improved Mallampati grading to assist clinical anesthesiologists in predicting the occurrence of hypoxemia in painless gastroscopy patients. The sensitivity, specificity, and area under the subject working characteristic curve were 58.3%, 88.4%, and 0.734, respectively. Fang et al.³³ dynamically predicted hypoxemia in patients undergoing painless gastroscopy across all age groups and found that the optimal model was SVM with an AUPRC of 0.650. In our study, the random forest model had an AUROC of 0.914 in the internal test set and 0.844 in the external validation dataset. The AUPRC in the internal test set was 0.721 and the AUPRC result in the external validation set was 0.458; both values were better than those reported in other studies and had good model generalization ability.

This study had certain limitations. (1) Anesthesia was induced using a mixture of etomidate and propofol (1:1), which had a better effect than propofol alone in reducing injection pain, hemodynamic changes, and respiratory depression in patients. It was more suitable for painless gastroscopy and showed good sedative effects^{34,35}. Therefore, we excluded patients who were induced solely by propofol. (2) The external validation dataset was obtained from one hospital in different regions; thus, multiple validation queues from different regions and countries were lacking, which may lead to some obstacles in the promotion and application of the model in other regions and countries. Therefore, before applying our model, more queues and carefully designed multicenter studies need to be performed to verify the generalization ability of the model.

To summarize, the prediction model in this study showed excellent predictive performance, which was achieved by combining various clinical indicators. This model can be used to accurately predict the occurrence of hypoxemia in middle-aged and elderly outpatients undergoing painless gastroscopy and help clinical anesthesiologists identify high-risk patients with hypoxemia during painless gastroscopy in the early stage. Based on the information obtained from the model, anesthesiologists can take active and reasonable preventive intervention measures to reduce the incidence of hypoxemia in middle-aged and elderly outpatients undergoing painless gastroscopy and increase the safety of perioperative anesthesia in such patients.

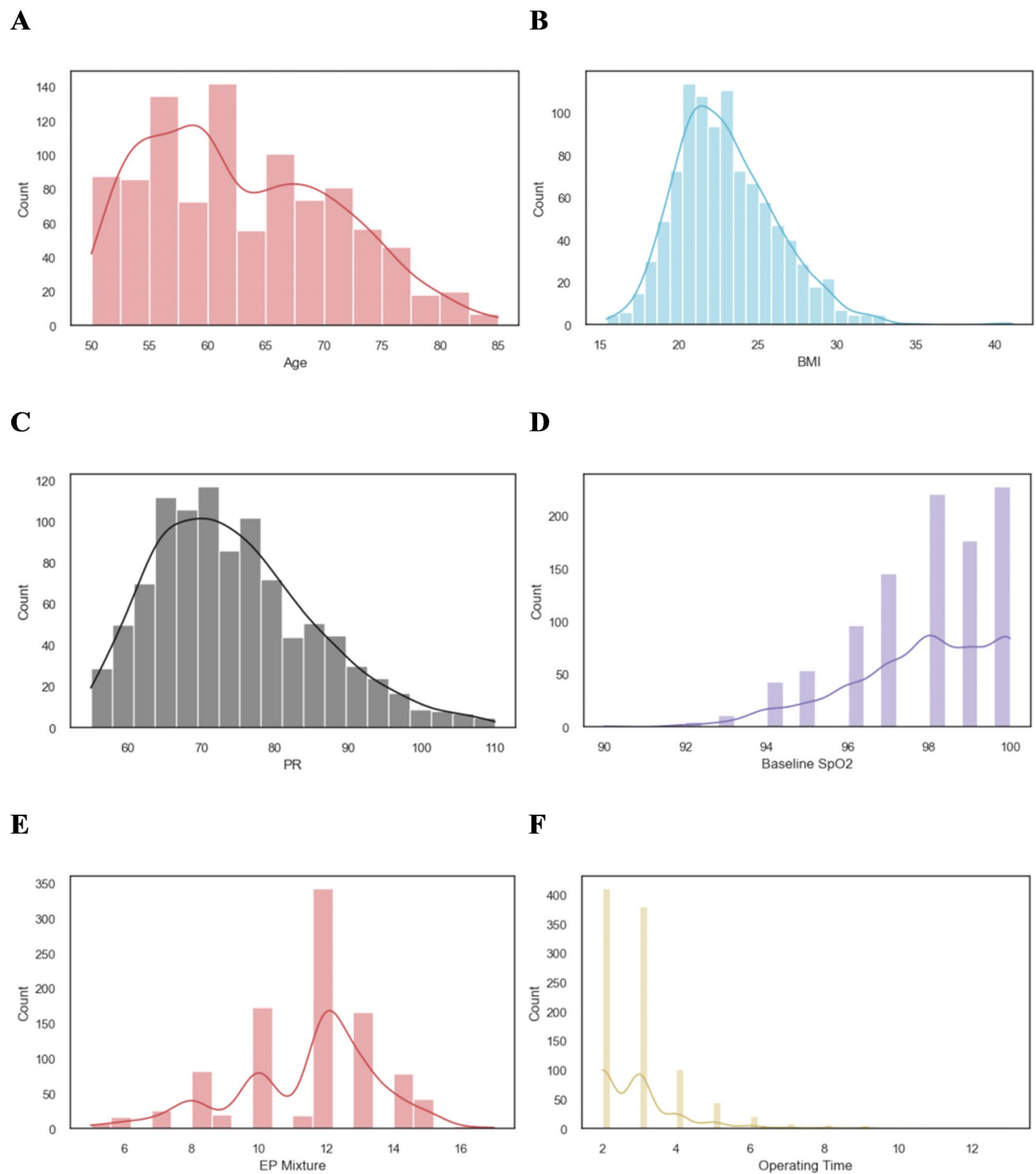


Fig. 2. Distribution of the data on continuous numerical variables. The distribution of (A) age, (B) body mass index (BMI), (C) pulse rate (D) baseline SpO₂, (E) the induction dose of the etomidate and propofol mixture, and (F) the operating time of gastroscopy in the dataset.

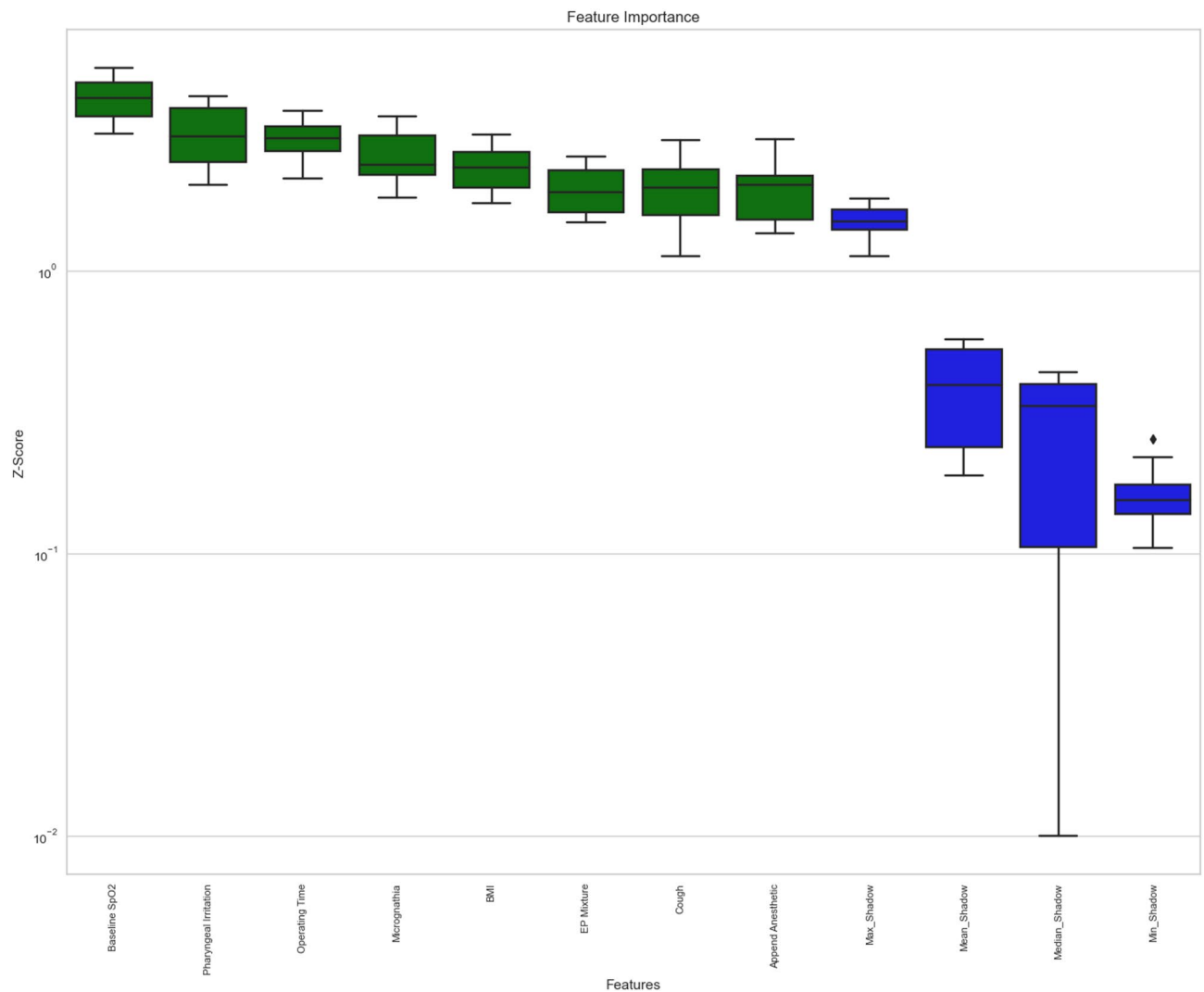


Fig. 3. BorutaShap features variable selection. Max_Shadow: The maximum importance value was calculated to determine the maximum potential impact of the feature. Mean_Shadow: The average importance value was calculated to determine the average potential impact of the feature. Median_Shadow: The median importance value was calculated to determine the typical potential impact of the feature. Min_Shadow: The minimum importance value was calculated to determine the minimum potential impact of the feature.

Model	Group	Accuracy	AUROC, 95 CI	AUPRC	Precision	Recall	F1	Brier
LR	Train	83.252	0.895(0.879,0.915)	0.875	0.833	0.832	0.832	0.128
	Test	80.405	0.893(0.881,0.899)	0.586	0.333	0.781	0.467	0.134
	Validate	75.275	0.832(0.809,0.845)	0.426	0.385	0.833	0.526	0.195
SVM	Train	88.399	0.938(0.931,0.957)	0.914	0.866	0.908	0.887	0.091
	Test	80.405	0.855(0.812,0.884)	0.438	0.303	0.625	0.408	0.113
	Validate	75.275	0.791(0.724,0.817)	0.365	0.383	0.817	0.521	0.196
RF	Train	97.386	0.998(0.998,0.999)	0.998	0.972	0.975	0.974	0.026
	Test	90.878	0.914(0.889,0.924)	0.721	0.561	0.719	0.630	0.072
	Validate	79.945	0.848(0.819,0.860)	0.458	0.434	0.717	0.541	0.130
XGB	Train	94.853	0.988(0.987,0.994)	0.988	0.949	0.948	0.948	0.042
	Test	88.514	0.902(0.865,0.920)	0.702	0.478	0.688	0.564	0.076
	Validate	78.846	0.808(0.745,0.859)	0.440	0.412	0.667	0.510	0.155
LightGBM	Train	95.670	0.991(0.990,0.996)	0.991	0.962	0.951	0.956	0.037
	Test	88.176	0.891(0.847,0.917)	0.652	0.468	0.688	0.557	0.083
	Validate	81.319	0.814(0.711,0.856)	0.470	0.455	0.667	0.541	0.141

Table 4. Comparison of evaluation indexes of five machine learning algorithm models. LR: logistic regression; SVM: support vector machine; RF: random forest; XGBoost: extreme gradient boosting; LightGBM: light gradient boosting machine; Training: training set; Testing: internal testing set; Validation: External validation set; Accuracy: the proportion of correctly predicted samples out of the total samples; AUROC: the area under the receiver's operating characteristic curve; 95 CI: confidence interval; AUPRC, the area under the precision-recall curve; Precision: the proportion of true positive predictions among all positive predictions made; Recall: the proportion of true positive predictions among all actual positive instances; F1 score: the harmonic mean of precision and recall; Brier score: Assess the degree of proximity between the model's risk predictions and the actual observed probabilities. A lower score indicates better model performance, with model output probabilities closer to the true labels.

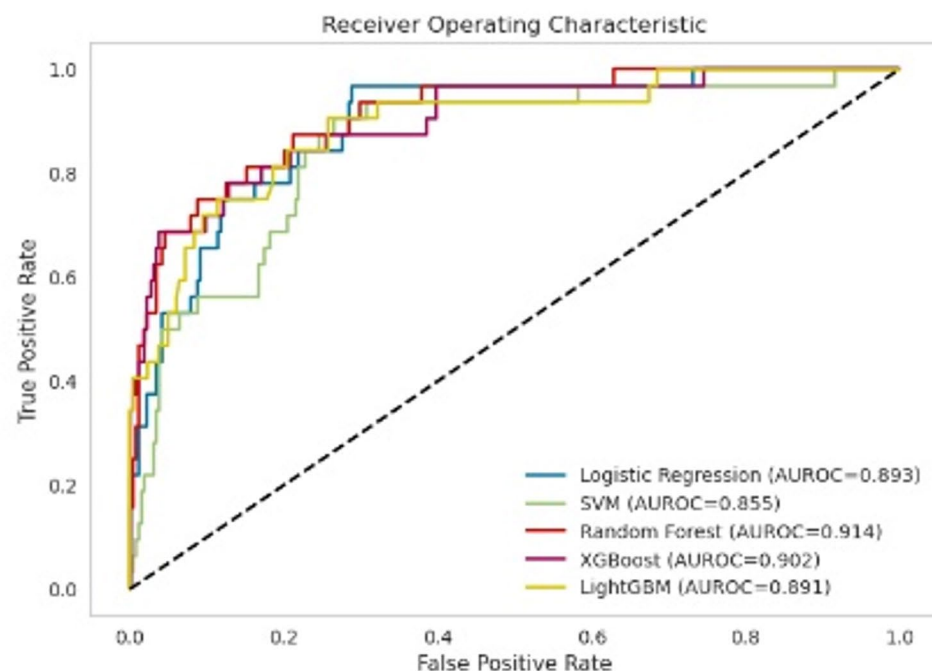
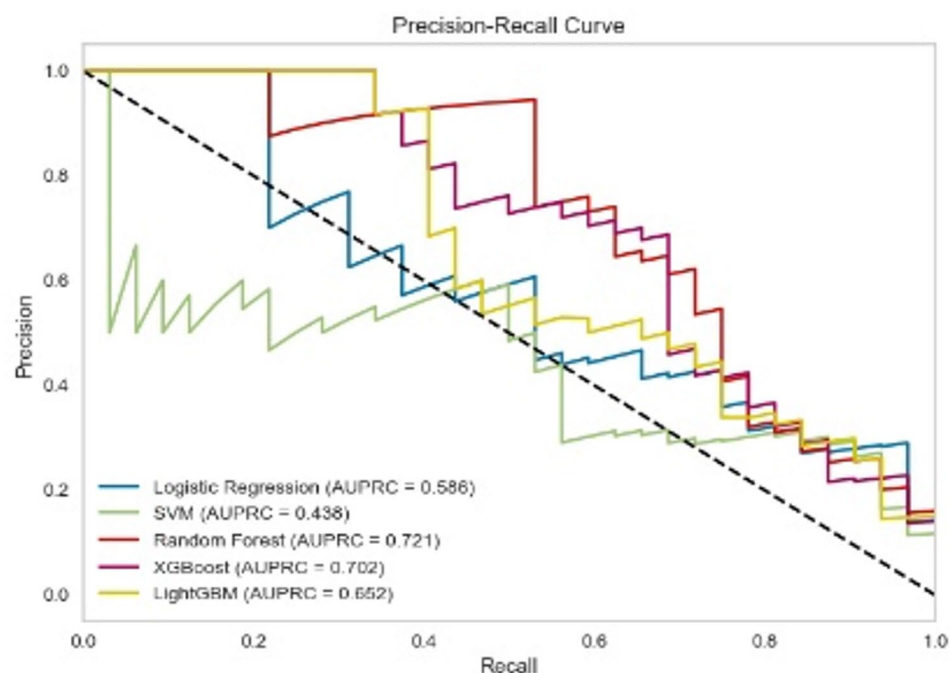
A**B**

Fig. 4. The ROC curve (A) and PRC curve (B) of five models in the internal testing set. AUROC: the area under the receiver operating characteristic curve; AUPRC: the area under the precision-recall curve; LR: logistic regression; SVM: support vector machine; RF: random forest; XGBoost: extreme gradient boosting; LightGBM: light gradient boosting machine.

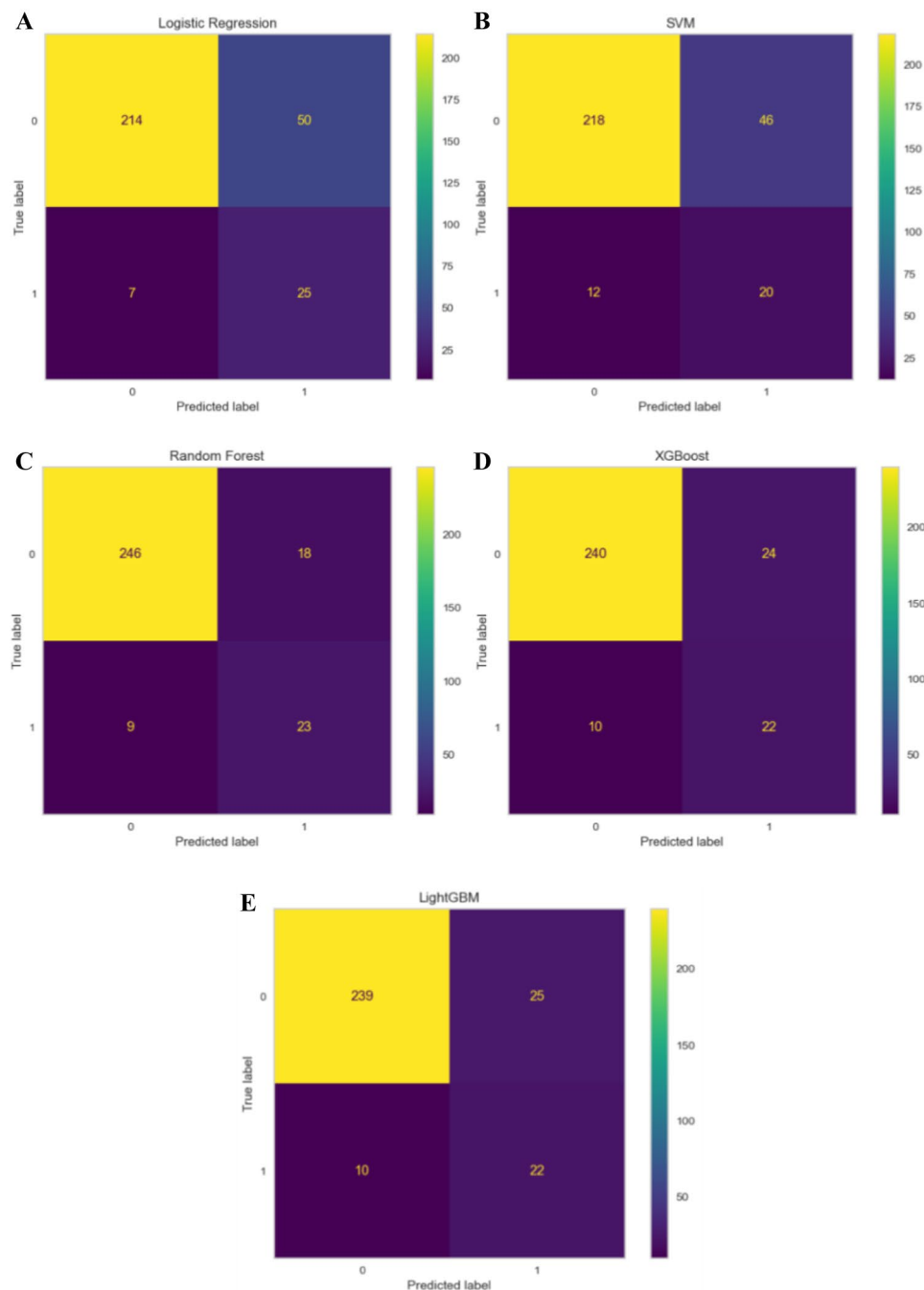


Fig. 5. The confusion matrix of five models in the internal testing set is shown. TP represents the number of samples correctly predicted by the model as positive categories; TN represents the number of samples correctly predicted by the model as negative categories; FP indicates the number of negative class samples incorrectly predicted by the model as positive classes; FN indicates the number of positive category samples incorrectly predicted by the model as negative categories.

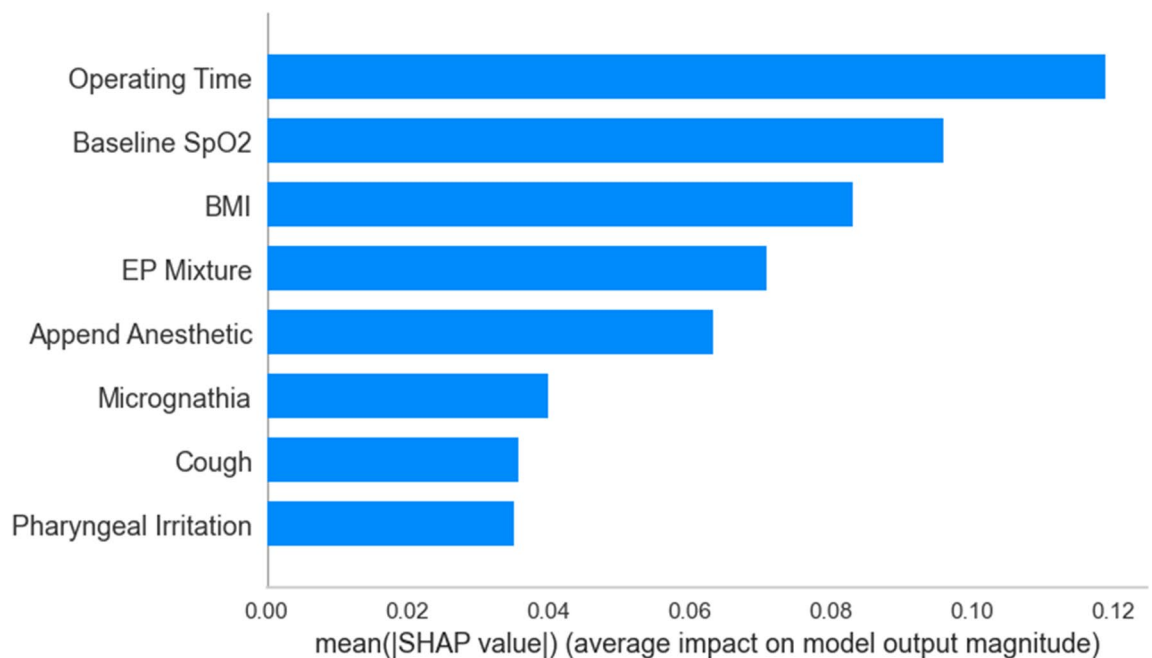
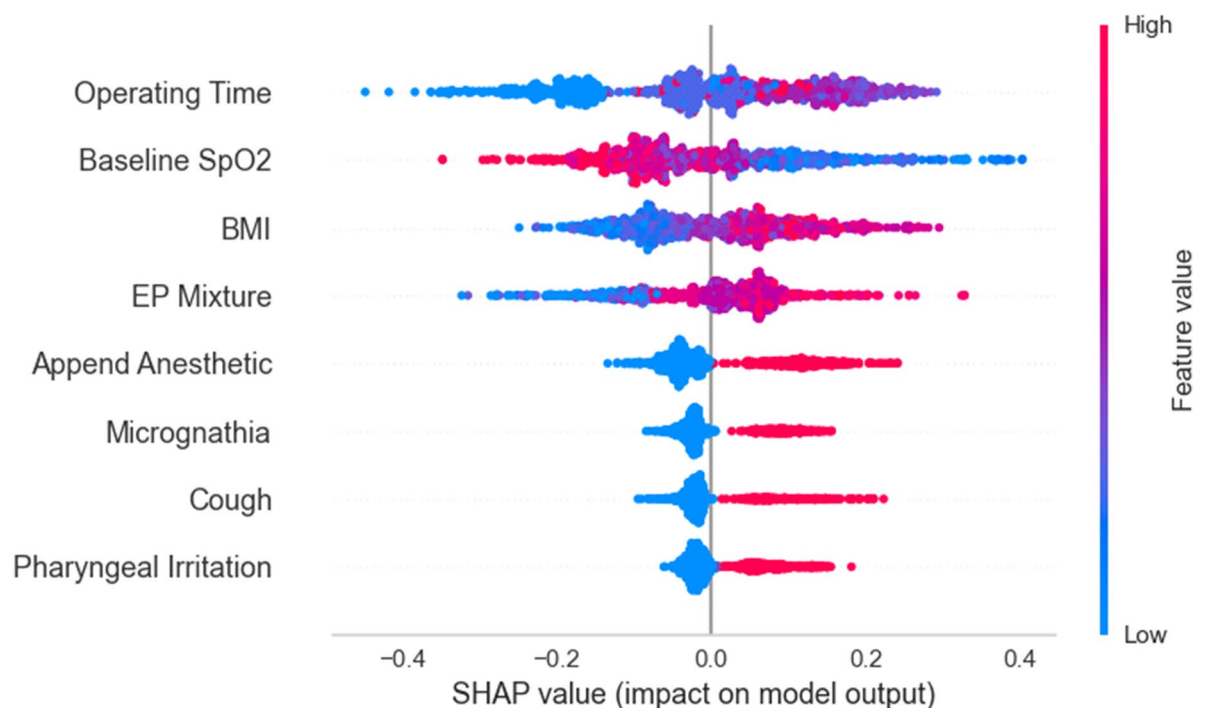
A**B**

Fig. 6. The SHAP summary plot of the RF model. **(A)** The SHAP bar plot for the ranking of the importance of variables. The sorted result was based on the mean absolute SHAP value of each variable in the RF model. **(B)** The SHAP bee swarm plot for the entire distribution of the SHAP values for each variable of the RF model. Each line represents a variable and each dot represents a case. The greater the intensity of the red color of the dot, the higher the value of the variable for the case. In contrast, the greater the intensity of the blue color of the dot, the lower the value of the variable for the case. The abscissa represents the SHAP value; a positive SHAP value helps the model predict the case of developing hypoxemia, and vice versa.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 12 June 2024; Accepted: 14 May 2025

Published online: 23 May 2025

References

1. Zhou, S. et al. National survey on sedation for Gastrointestinal endoscopy in 2758 Chinese hospitals. *Br. J. Anaesth.* **127** (1), 56–64. <https://doi.org/10.1016/j.bja.2021.01.028> (2021).
2. Dumonceau, J. M. 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* **47** (12), 1175–1189. <https://doi.org/10.1055/s-0034-1393414> (2015).
3. Chen, D. X. et al. Comparison of a nasal mask and traditional nasal cannula during intravenous anesthesia for gastroscopy procedures: A randomized controlled trial. *Anesth. Analg.* **134** (3), 615–623. <https://doi.org/10.1213/ane.0000000000005828> (2022).
4. Mason, K. P., Green, S. M. & Piacevoli, Q. Adverse event reporting tool to standardize the reporting and tracking of adverse events during procedural sedation: A consensus document from the world SIVA international sedation task force. *Br. J. Anaesth.* **108** (1), 13–20. <https://doi.org/10.1093/bja/aer407> (2012).
5. Arina, P. et al. Prediction of complications and prognostication in perioperative medicine: A systematic review and PROBAST assessment of machine learning tools. *Anesthesiology* **140** (1), 85–101. <https://doi.org/10.1097/aln.0000000000004764> (2024).
6. Shimizu, H., Homma, Y. & Norii, T. Incidence of adverse events among elderly vs non-elderly patients during procedural sedation and analgesia with Propofol. *Am. J. Emerg. Med.* **44**, 411–414. <https://doi.org/10.1016/j.ajem.2020.04.094> (2021).
7. Travis, A. C., Pievsky, D. & Saltzman, J. R. Endoscopy in the elderly. *Am. J. Gastroenterol.* **107** (10), 1495–1501. <https://doi.org/10.1038/ajg.2012.246> (2012). quiz 1494, 1502.
8. Attia, Z. I. et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *Lancet* **394** (10201), 861–867. [https://doi.org/10.1016/s0140-6736\(19\)31721-0](https://doi.org/10.1016/s0140-6736(19)31721-0) (2019).
9. Silva, I. S. et al. Polycystic ovary syndrome: Clinical and laboratory variables related to new phenotypes using machine-learning models. *J. Endocrinol. Invest.* **45** (3), 497–505. <https://doi.org/10.1007/s40618-021-01672-8> (2022).
10. Dickinson, Q. & Meyer, J. G. Positional SHAP (PoSHAP) for interpretation of machine learning models trained from biological sequences. *PLoS Comput. Biol.* **18** (1), e1009736. <https://doi.org/10.1371/journal.pcbi.1009736> (2022).
11. Xu, Z. et al. A synthetic minority oversampling technique based on Gaussian mixture model filtering for imbalanced data classification. *IEEE Trans. Neural Netw. Learn. Syst.* **35** (3), 3740–3753. <https://doi.org/10.1109/tnnls.2022.3197156> (2022).
12. Adnan, M. et al. Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Comput. Sci.* **8**, e803. <https://doi.org/10.7717/peerj-cs.803> (2022).
13. Mehta, P. P. et al. Capnographic monitoring in routine EGD and colonoscopy with moderate sedation: A prospective, randomized, controlled trial. *Am. J. Gastroenterol.* **111** (3), 395–404. <https://doi.org/10.1038/ajg.2015.437> (2016).
14. Wadhwa, V. et al. Similar risk of cardiopulmonary adverse events between Propofol and traditional anesthesia for Gastrointestinal endoscopy: A systematic review and Meta-analysis. *Clin. Gastroenterol. Hepatol.* **15** (2), 194–206. <https://doi.org/10.1016/j.cgh.2016.07.013> (2017).
15. Wang, S. et al. Bilevel positive airway pressure for gastroscopy with sedation in patients at risk of hypoxemia: A prospective randomized controlled study. *J. Clin. Anesth.* **85**, 111042. <https://doi.org/10.1016/j.jclinane.2022.111042> (2023).
16. van Schaik, E. P. C. et al. Hypoxemia during procedural sedation in adult patients: A retrospective observational study. *Can. J. Anaesth.* **68** (9), 1349–1357. <https://doi.org/10.1007/s12630-021-01992-6> (2021).
17. Hochberg, C. H., Semler, M. W. & Brower, R. G. Oxygen toxicity in critically ill adults. *Am. J. Respir. Crit. Care Med.* **204** (6), 632–641. <https://doi.org/10.1164/rccm.202102-0417CI> (2021).
18. Trzepizur, W. et al. Sleep Apnea-Specific hypoxic burden, symptom subtypes, and risk of cardiovascular events and All-Cause mortality. *Am. J. Respir. Crit. Care Med.* **205** (1), 108–117. <https://doi.org/10.1164/rccm.202105-1274OC> (2022).
19. Dewan, N. A., Nieto, F. J. & Somers, V. K. Intermittent hypoxemia and OSA: Implications for comorbidities. *Chest* **147** (1), 266–274. <https://doi.org/10.1378/chest.14-0500> (2015).
20. Matthay, M. A., Thompson, B. T. & Ware, L. B. The Berlin definition of acute respiratory distress syndrome: Should patients receiving high-flow nasal oxygen be included? *Lancet Respir. Med.* **9** (8), 933–936. [https://doi.org/10.1016/s2213-2600\(21\)00105-3](https://doi.org/10.1016/s2213-2600(21)00105-3) (2021).
21. Santer, P. et al. High-flow nasal oxygen for Gastrointestinal endoscopy improves respiratory safety. *Br. J. Anaesth.* **127** (1), 7–11. <https://doi.org/10.1016/j.bja.2021.03.022> (2021).
22. Hung, K. C. 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.* **77**, 110651. <https://doi.org/10.1016/j.jclinane.2022.110651> (2022).
23. Wani, S. et al. Obesity as a risk factor for sedation-related complications during propofol-mediated sedation for advanced endoscopic procedures. *Gastrointest. Endosc.* **74** (6), 1238–1247. <https://doi.org/10.1016/j.gie.2011.09.006> (2011).
24. Redline, S., Azarbarzin, A. & Peker, Y. Obstructive sleep Apnoea heterogeneity and cardiovascular disease. *Nat. Rev. Cardiol.* **20** (8), 560–573. <https://doi.org/10.1038/s41569-023-00846-6> (2023).
25. Liu, X. et al. The efficacy and safety of remimazolam tosylate versus Etomidate-Propofol in elderly outpatients undergoing colonoscopy: A prospective, randomized, Single-Blind, Non-Inferiority trial. *Drug Des. Dev. Ther.* **15**, 4675–4685. <https://doi.org/10.2147/dddt.S339535> (2021).
26. Hillman, D. R. et al. Evolution of changes in upper airway collapsibility during slow induction of anesthesia with Propofol. *Anesthesiology* **111** (1), 63–71. <https://doi.org/10.1097/ALN.0b013e3181a7ec68> (2009).
27. Leslie, K. et al. Safety of sedation for Gastrointestinal endoscopy in a group of university-affiliated hospitals: A prospective cohort study. *Br. J. Anaesth.* **118** (1), 90–99. <https://doi.org/10.1093/bja/aew393> (2017).
28. Tang, R. et al. Efficacy and safety of sedation with Dexmedetomidine in adults undergoing Gastrointestinal endoscopic procedures: Systematic review and meta-analysis of randomized controlled trials. *Front. Pharmacol.* **14**, 1241714. <https://doi.org/10.3389/fphar.2023.1241714> (2023).
29. Yin, S. et al. Efficacy and tolerability of sufentanil, Dexmedetomidine, or ketamine added to Propofol-based sedation for Gastrointestinal endoscopy in elderly patients: A prospective, randomized, controlled trial. *Clin. Ther.* **41** (9), 1864–1877e1860. <https://doi.org/10.1016/j.clinthera.2019.06.011> (2019).
30. Geng, W. et al. A prediction model for hypoxemia during routine sedation for Gastrointestinal endoscopy. *Clinics (Sao Paulo)*. **73**, e513. <https://doi.org/10.6061/clinics/2018/e513> (2018).
31. Geng, W. et al. An artificial neural network model for prediction of hypoxemia during sedation for Gastrointestinal endoscopy. *J. Int. Med. Res.* **47** (5), 2097–2103. <https://doi.org/10.1177/0300060519834459> (2019).

32. Li, N. et al. Predictive value of NoSAS questionnaire combined with the modified Mallampati grade for hypoxemia during routine sedation for Gastrointestinal endoscopy. *BMC Anesthesiol.* **23** (1), 126. <https://doi.org/10.1186/s12871-023-02075-3> (2023).
33. Fang, Z. et al. Dynamic prediction of hypoxemia risk at different time points based on preoperative and intraoperative features: machine learning applications in outpatients undergoing Esophagogastroduodenoscopy. *Ann. Med.* **55** (1), 1156–1167. <https://doi.org/10.1080/07853890.2023.2187878> (2023).
34. Zhou, X. et al. Etomidate plus Propofol versus Propofol alone for sedation during gastroscopy: A randomized prospective clinical trial. *Surg. Endosc.* **30** (11), 5108–5116. <https://doi.org/10.1007/s00464-016-4861-6> (2016).
35. Wang, C. et al. Physical and chemical compatibility of etomidate and Propofol injectable emulsions. *Pharmacology* **106** (11–12), 644–657. <https://doi.org/10.1159/000519236> (2021).

Acknowledgement

Assistance with the article: we would like to thank Dr Guihua Huang for his assistance with the study.

Author contributions

LL.Z and XY.W wrote the main manuscript text; W.G and R.W prepared Figs. 1, 2, 3, 4, 5 and 6 and J.W and HY.H prepared Tables 1, 2 and 3; Z.W conducts research on overall quality management; B.Y and Y.Z carried out the revision of the manuscript text and proposed the research ideas. All authors reviewed the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China [grant number 82360709]; and the Guizhou High-Level Innovative Talent Training Program “Thousand” Level Talents Program.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-02540-8>.

Correspondence and requests for materials should be addressed to B.Y. or Y.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025