Machine-Learning Model for the Prediction of Hypoxaemia during Endoscopic Retrograde Cholangiopancreatography under Monitored Anaesthesia Care

Huapyong Kang[1,2]*, Bora Lee[3]*, Jung Hyun Jo[4], Hee Seung Lee[4], Jeong Youp Park[4], Seungmin Bang[4], Seung Woo Park[4], Si Young Song[4], Joonhyung Park[5], Hajin Shim[6], Jung Hyun Lee[5], Eunho Yang[5,6], Eun Hwa Kim[7], Kwang Joon Kim[8], Min-Soo Kim[3], and Moon Jae Chung[4]

[1]Division of Gastroenterology, Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine, Incheon;
[2]Department of Medicine, Yonsei University Graduate School, Seoul;
[3]Department of Anesthesiology and Pain Medicine, Anesthesia and Pain Research Institute, Yonsei University College of Medicine, Seoul;
[4]Division of Gastroenterology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul;
[5]Graduate School of AI, Korea Advanced Institute of Science and Technology, Daejeon;
[6]School of Computing, Korea Advanced Institute of Science and Technology, Daejeon;
[7]Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul;
[8]Division of Geriatrics, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea.

Purpose: Hypoxaemia is a significant adverse event during endoscopic retrograde cholangiopancreatography (ERCP) under monitored anaesthesia care (MAC); however, no model has been developed to predict hypoxaemia. We aimed to develop and compare logistic regression (LR) and machine learning (ML) models to predict hypoxaemia during ERCP under MAC.

Materials and Methods: We collected patient data from our institutional ERCP database. The study population was randomly divided into training and test sets (7:3). Models were fit to training data and evaluated on unseen test data. The training set was further split into k-fold (k=5) for tuning hyperparameters, such as feature selection and early stopping. Models were trained over k loops; the i-th fold was set aside as a validation set in the i-th loop. Model performance was measured using area under the curve (AUC).

Results: We identified 6114 cases of ERCP under MAC, with a total hypoxaemia rate of 5.9%. The LR model was established by combining eight variables and had a test AUC of 0.693. The ML and LR models were evaluated on 30 independent data splits. The average test AUC for LR was 0.7230, which improved to 0.7336 by adding eight more variables with an l_1 regularisation-based selection technique and ensembling the LRs and gradient boosting algorithm (GBM). The high-risk group was discriminated using the GBM ensemble model, with a sensitivity and specificity of 63.6% and 72.2%, respectively.

Conclusion: We established GBM ensemble model and LR model for risk prediction, which demonstrated good potential for preventing hypoxaemia during ERCP under MAC.

Key Words: Cholangiopancreatography, endoscopic retrograde; hypoxaemia; machine learning; monitored anaesthesia care; prediction model

INTRODUCTION

Endoscopic retrograde cholangiopancreatography (ERCP) is an essential procedure for pancreato-biliary diseases. The annual rate of ERCP in developed countries is 70–100 per 100000 inhabitants,[1] with more than 500000 and 50000 annual procedures performed in the United States and South Korea, respectively.[2,3]

ERCP requires proper sedation.[4] However, sedation with an intermittent bolus of propofol, as well as the patient's prone position, can increase cardiopulmonary instability.[5-7] Therefore, the latest guidelines recommend anaesthesia provider-administered sedation for ERCP.[8] Respiratory complications necessitate meticulous respiratory monitoring of patients.[9,10] The incidence of hypoxaemia during sedative ERCP is reportedly 10%–28%.[11-14] However, even with pulse oximetry, capnometry, and visual assessment, the ability to reduce hypoxaemia is debatable.[9,15,16]

Several risk factors for sedation-related adverse events (SRAEs) during advanced endoscopic procedures, including high body mass index (BMI), high American Society of Anesthesiologists (ASA) class, advanced age, and sleep apnoea, have been identified.[17-21] A recent prospective randomised controlled trial (RCT) compared SRAEs in two types of anaesthesia provider-administered sedations during ERCP in high-risk patients: general endotracheal anaesthesia (GEA) and monitored anaesthesia care (MAC) without intubation; the authors suggested using GEA for high-risk patients as the incidence of SRAEs, mostly hypoxaemia, was lower in this group.[13] However, GEA for all patients undergoing ERCP is impractical, since GEA is more time-consuming, costly, and medically resource-intensive than MAC.[22,23] Therefore, the selection of GEA or MAC should depend on the precise prediction of hypoxemia risk. A model that can identify patients at an elevated risk of hypoxemia has not been well-established.

Machine learning (ML) approaches have outperformed conventional methods in various tasks, such as logistic regression (LR). Several studies have introduced ML-based clinical prediction models, which have outperformed conventional models.[24-26] In particular, the gradient boosting algorithm (GBM), which learns by boosting multiple weak learners, has delivered superior performance in regression tasks with tabular data.[27,28]

Here, we aimed to develop LR and ML models to predict hypoxaemia during ERCP under MAC and compare their performance.

MATERIALS AND METHODS

Study cohort

We included patients who underwent ERCP at Severance Hospital between May 2012 and September 2017. All procedures were performed by highly experienced endoscopists who have each performed over 5000 ERCPs (S.Y.S., S.W.P., S.B., J.Y.P., and M.J.C.). Patients were identified from our institutional database of ERCPs. Then, we retrospectively abstracted 27 and 32 relevant continuous and categorical variables, respectively, representing patient characteristics, laboratory data, and procedure-related characteristics (Supplementary Table 1, only online). Our exclusion criteria were as follows: 1) age <19 years; 2) conversion to endoscopic procedures other than ERCP; 3) cancelled procedure owing to inadequate sedation; 4) sedation conversion to endoscopist-administered sedation; and 5) ERCP under GEA. Our protocol adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of Severance Hospital (IRB number: 4-2020-0257).

MAC protocol during ERCP

All anaesthesia procedures were performed by senior resident anaesthesiologists with at least 3 years of anaesthesia experience under specialist anaesthesiologists' supervision. Patients were sedated using a standardised protocol of continuous propofol (Fresofol 1% MCT injection, Fresenius Kabi Korea, Seoul, Korea) infusion with intermittent fentanyl administration (fentanyl citrate, Hana Pharm Co., Ltd., Seoul, Korea). Electrocardiography, pulse oximetry, non-invasive arterial blood pressure, and capnometry using a nasal prong were employed as standard monitoring methods. Oxygen was administered through a nasal prong at a 5-L/min flow rate. When blood oxygen saturation (SpO_2) decreased (<90%), the propofol infusion rate was reduced, and airway manoeuvres (e.g., increase oxygen flow, jaw thrust, chin lift, nasopharyngeal airway insertion, bag-mask ventilation) were performed.

Outcome measures and definitions

The primary endpoint was intraprocedural hypoxaemia (SpO_2 <90%) detected by pulse oximetry. Some continuous variables were transformed to categorical variables using cut-off values calculated by the Youden index method or selected with clinical significance. The variables are defined as follows: Habitual snoring-snoring >3 nights/week[29]; same-session endoscopy with sedation before ERCP-any endoscopy (not ERCP) performed under endoscopist-administered sedation immediately before ERCP; and comorbidities-presence of certain groups of International Statistical Classification of Diseases and Related Health Problems-10 codes on the electronic medical records. All ERCPs were categorised to have one most relevant indication after reviewing each case. Among ERCP indications, biliary stricture was defined as biliary stenosis not only due to biliary stone, such as malignant or benign stricture. Indications other than biliary stone or stricture included biliary leakage, pancreatic duct stricture/stone, and ampullary tumor. Detailed definitions of each comorbidity group are shown in Supplementary Table 1 (only online).

Statistical analysis

We randomly divided the study population into training and test sets (7:3 ratio), with the same proportion of patients with ASA class ≥3 in both sets. To compare the ML and LR models, the study population was randomly divided into training and test sets 30 times using the same ratio. The results were evaluated by 30 different data splits, and the performances were measured by computing the area under the curve (AUC). All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA), and R version 3.6.3 (http://www.r-project.org/) with the "rms" package was used to construct the nomogram.

LR model

Univariable LR analysis was conducted to identify the factors associated with hypoxaemia. Multivariable analysis was conducted using variables selected by considering clinical and statistical significances. The hypoxaemia prediction model and nomogram were established using these results, and the Hosmer-Lemeshow test was used to evaluate the fit of the LR model. Model performance was measured by computing the mean value of the AUC generated by 10000 times bootstrap resampling in both sets. A calibration plot was drawn to reflect the agreement between observed outcomes and predicted probabilities.

l_1 regularisation-based feature selection

Additional features to those selected for the LR model by univariable analysis were chosen by fitting populations with whole features to the LR with l_1 regularisation; this added l_1-norm of the model parameters to the original loss of LR.[30] As a result, the model reduced loss function and learned how parameters could be made sparsely non-zero. As model parameters referred to the coefficient of each feature in LR, l_1 regularisation played a role in removing the relatively insignificant features by making their coefficients zero. We singled out the more relevant features that affected the occurrence of hypoxaemia by applying l_1 regularisation to all features and deciding on the number of features to add based on the validation AUC score in the order of the highest coefficient. After the dataset was randomly divided into two sets, the training set was further divided into k folds (k=5). From these folds, k combinations of training and test sets were constructed for ensembling by reserving 1 fold as the test set. For each feature set, k models learned k different training sets, and the validation AUC score of each model was computed with each test set. The averaged k validation scores became the criteria for determining the feature sets.

Gradient boosting ensemble

The GBM first trains one weak learner on the dataset and calculates the residual error; then, another weak learner trains this residual error. By repeating this process and combining weak learners, the GBM can make more accurate predictions. We used Extreme Gradient Boosting (Xgboost), which frequently shows better performance than conventional GBM models.[31] In addition, the ensemble method, which considers predictions of different models through voting or averaging, was adopted to enhance the accuracy of prediction. The ensembled
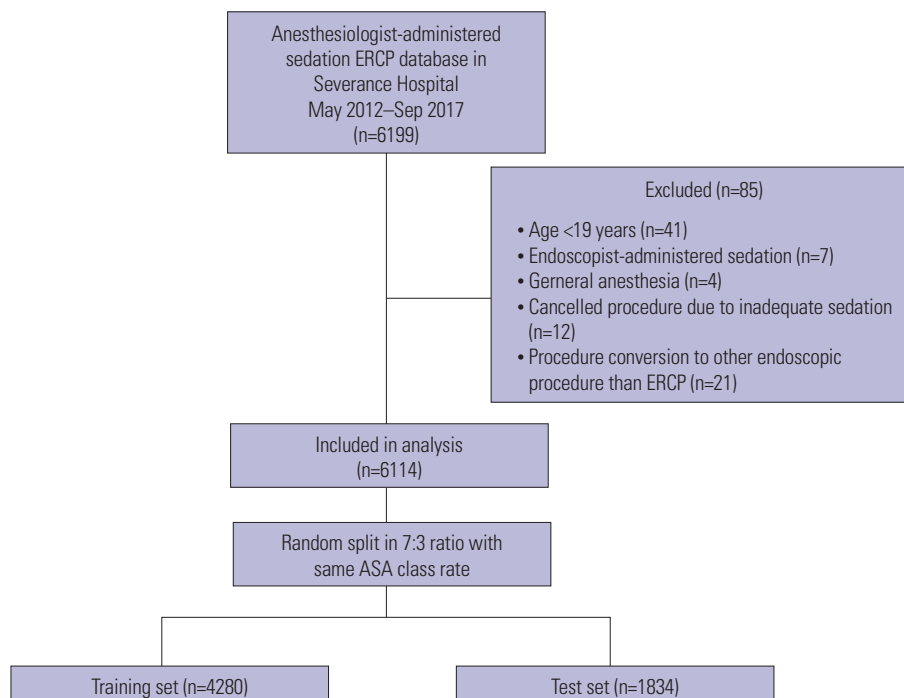

Fig. 1. Flow chart for patient selection and random split. ERCP, endoscopic retrograde cholangiopancreatography; ASA, American Society of Anesthesiologists.

model of Xgboost and LR was trained by k-fold cross validation (k=5) and used to predict hypoxaemia by averaging the prediction logits of each unit ensembled model.

RESULTS

Patients

We identified 6114 cases of ERCP under MAC and divided them into training (n=4280) and test (n=1834) sets (Fig. 1). The total hypoxaemia rate was 5.9% (n=359), and most were successfully managed by adjusting the propofol infusion rate and performing airway manoeuvres. Interruption or premature termination of ERCP due to hypoxemia was required in 110 (1.8%) cases, and emergent endotracheal intubation or conversion to GEA was performed in four of these patients (0.06%). No difference was observed between training and test sets in hypoxaemia rates (5.9% vs. 5.9%, p=0.970) and other baseline characteristics (Table 1).

Table 1. Baseline Characteristics of Training and Test Sets (Summarised)

Variables	Training set (n=4280)	Test set (n=1834)	p value
Age >74 years	933 (21.80)	393 (21.43)	0.7474
Sex, male	2528 (59.07)	1095 (59.71)	0.6406
BMI ≥25 kg/m²	1098 (25.72)	447 (24.51)	0.3186
Current smoker	636 (15.44)	281 (16.07)	0.5460
Never drinker	2175 (51.47)	912 (50.39)	0.4417
ASA classification, III–IV	1549 (36.19)	664 (36.21)	0.9920
Habitual snoring	348 (8.13)	141 (7.69)	0.5587
Nutritional risk: high	477 (12.74)	191 (11.93)	0.4119
History of hypoxaemia during sedation endoscopy	139 (3.25)	45 (2.45)	0.0959
Baseline O₂ administration	92 (2.16)	28 (1.53)	0.1070
Comorbidities			
Malignancy	2163 (50.54)	921 (50.22)	0.8190
Chronic heart disease	390 (9.11)	186 (10.14)	0.2066
Cerebrovascular disease	192 (4.49)	100 (5.45)	0.1044
Chronic respiratory disease	161 (3.76)	77 (4.20)	0.4184
Chronic liver disease (other than cirrhosis)	331 (7.73)	139 (7.58)	0.8353
Liver cirrhosis	246 (5.75)	109 (5.94)	0.7644
ESRD on dialysis	60 (1.40)	23 (1.25)	0.6473
Laboratory test			
Haemoglobin, g/dL	12.1 (10.6–13.4)	12.1 (10.7–13.5)	0.5624
Haematocrit, %	36.3 (31.9–40)	36.3 (31.9–40.1)	0.6421
CRP, mg/dL	22.9 (4.8–68.0)	21.0 (4.5–63.6)	0.4891
Total bilirubin, mg/dL	1.4 (0.6–4.2)	1.3 (0.6–3.9)	0.1398
Creatinine, mg/dL	0.71 (0.58–0.89)	0.73 (0.58–0.89)	0.0676
eGFR ≥30 mL/min/1.73 m²	4166 (97.66)	1790 (97.87)	0.6117
Baseline inotropes administration	64 (1.50)	17 (0.93)	0.0749
Opioids prescription <7 days	496 (11.59)	198 (10.8)	0.3706
Psychotropics prescription <7 days	113 (2.64)	49 (2.67)	0.9439
ERCP indication: biliary stricture	1822 (42.57)	751 (40.95)	0.2393
ERCP indication: biliary stone	1578 (36.87)	703 (38.33)	0.2786
Same-session endoscopy with sedation before ERCP	782 (18.27)	349 (19.03)	0.4840
Propofol used for same-session endoscopy	779 (99.62)	347 (99.43)	0.6470
Propofol dose for same-session endoscopy, mg	60 (50–100)	60 (50–100)	0.8590
Duration of MAC for ERCP, min	20 (15–30)	20 (15–27)	0.3390
Propofol dose used during MAC, mg	130 (100–220)	130 (100–220)	0.9840
Fentanyl dose used during MAC, mcg	75 (50–100)	75 (50–100)	0.4700

BMI, body mass index; ASA, American Society of Anesthesiologists; ESRD, end-stage renal disease; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; ERCP, endoscopic retrograde cholangiopancreatography; MAC, monitored-anaesthesia care.
All categorical variables are presented as n (%). All continuous variables are presented as median (interquartile range).

Table 2. Univariable Logistic Regression Using Training Set (Summarised)

Variables	OR (95% CI)	*p* value
Age >74 years	1.543 (1.164–2.045)	0.0025
Sex, male	0.837 (0.648–1.082)	0.1753
BMI ≥25 kg/m^2	2.497 (1.926–3.236)	<0.0001
Current smoker	0.707 (0.472–1.058)	0.0921
ASA classification		
II vs. I	1.085 (0.703–1.674)	0.7121
III vs. II	1.731 (1.123–2.669)	0.0130
IV vs. III	2.860 (1.427–5.735)	0.0031
Habitual snoring	8.348 (6.267–11.119)	<0.0001
History of hypoxaemia during sedation endoscopy	0.980 (0.474–2.024)	0.9560
Baseline O$_2$ administration	2.493 (1.34–4.638)	0.0039
Comorbidities		
Malignancy	0.518 (0.397–0.676)	<0.0001
Chronic heart disease	1.272 (0.846–1.912)	0.2469
Cerebrovascular disease	1.600 (0.956–2.678)	0.0737
Chronic respiratory disease	1.308 (0.716–2.389)	0.3823
Chronic liver disease (other than cirrhosis)	0.428 (0.218–0.841)	0.0138
Liver cirrhosis	0.524 (0.256–1.073)	0.0774
ESRD on dialysis	3.723 (1.911–7.252)	0.0001
Laboratory test		
Haemoglobin	1.068 (1.002–1.139)	0.0425
Haematocrit	1.022 (1.000–1.045)	0.0540
CRP	1.000 (0.998–1.003)	0.7985
Total bilirubin	0.952 (0.918–0.987)	0.0072
Creatinine	1.235 (1.113–1.371)	<0.0001
eGFR ≥30 mL/min/1.73 m^2	0.340 (0.194–0.598)	0.0002
Baseline inotropes administration	3.840 (2.023–7.291)	<0.0001
Opioids prescription <7 days	0.435 (0.252–0.752)	0.0029
Psychotropics prescription <7 days	0.431 (0.136–1.367)	0.1530
ERCP indication: biliary stricture	0.492 (0.371–0.654)	<0.0001
ERCP indication: biliary stone	1.643 (1.272–2.122)	0.0001
Same-session endoscopy with sedation before ERCP	2.087 (1.577–2.762)	<0.0001
Propofol as sedative for same-session endoscopy	0.771 (0.025–23.765)	0.8819
Propofol dose for same-session endoscopy	1.000 (0.996–1.005)	0.8850
Duration of MAC for ERCP	1.021 (1.013–1.028)	<0.0001
Propofol dose used during MAC	1.001 (0.999–1.002)	0.3660
Fentanyl dose used during MAC	1.006 (1.004–1.008)	<0.0001

OR, odd ratio; CI, confidence interval; BMI, body mass index; ASA, American Society of Anesthesiologists; ESRD, end-stage renal disease; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; ERCP, endoscopic retrograde cholangiopancreatography; MAC, monitored-anaesthesia care.

LR model

Univariable LR analysis of the training set is summarised in Table 2. A combination of variables was selected by considering clinical and statistical significances. Before performing multivariable analysis, patients with missing values of estimated glomerular filtration rate (eGFR) were excluded from both

Table 3. Multivariable Logistic Regression for Selected Variables Using Training Set

Variables	OR (95% CI)	*p* value
Age >74 years	1.443 (1.060–1.964)	0.0198
ASA classification	1.269 (1.045–1.543)	0.0165
Habitual snoring	7.433 (5.514–10.020)	<0.0001
BMI ≥25 kg/m^2	2.070 (1.570–2.728)	<0.0001
ERCP indication: biliary stone	1.539 (1.169–2.027)	0.0021
Baseline inotropes administration	4.123 (2.050–8.293)	<0.0001
Same-session endoscopy with sedation before ERCP	2.191 (1.619–2.966)	<0.0001
eGFR ≥30 mL/min/1.73 m^2	0.389 (0.209–0.723)	0.0029

OR, odd ratio; CI, confidence interval; ASA, American Society of Anesthesiologists; BMI, body mass index; ERCP, endoscopic retrograde cholangiopancreatography; eGFR, estimated glomerular filtration rate.

sets. All selected variables in multivariable analysis, including 1) age >74 years; 2) ASA class; 3) habitual snoring; 4) BMI ≥25 kg/m^2; 5) ERCP indication: biliary stone; 6) baseline administration of inotropes; 7) same-session endoscopy with sedation before ERCP; and 8) eGFR ≥30 mL/min/1.73 m^2, were independently associated with hypoxaemia (Table 3). Based on these results, the risk probability for the prediction model was calculated using the following equation:

Risk probability=exp(LP)/[1+exp(LP)],
where

LP (linear predictor)=-3.5346+0.3667*age>74 years (yes, 1; no, 0)+0.7274*BMI ≥25 kg/m^2 (yes, 1; no, 0)+0.2386*ASA (I, 1; II, 2; III, 3; IV, 4)+2.0060*snoring (yes, 1; no, 0)-0.9450*eGFR ≥30 mL/min/1.73 m^2 (yes, 1; no, 0)+1.4166*inotropes (yes, 1; no, 0)+0.4314*indication: biliary stone (yes, 1; no, 0)+0.7844* same-session endoscopy (yes, 1; no, 0).

The model showed adequate Hosmer-Lemeshow statistics (*p*=0.836). A nomogram was developed with these eight significant variables (Fig. 2A). The sum of the points obtained from each variable was visually matched with the risk probability line. The LR model delivered good prediction performance with an AUC of 0.762 [95% confidence interval (CI): 0.727–0.795] in the training set and 0.670 (95% CI: 0.614–0.723) in the test set (Table 4). The optimal cut-off of the nomogram score for discriminating the high-risk group was 70, which was calculated from the training set by the Youden index, with a sensitivity and specificity of 65.20% and 75.46%, respectively. In the training set, the high-risk group (≥70 points, n=1146) had a significantly higher rate of hypoxaemia compared to the low-risk group (n=3109) (14.22% vs. 2.80%, *p*<0.001). With the same threshold, the sensitivity and specificity of the test set were 49.07% and 75.84%, respectively. Patients in the test set were further divided into two groups with a significantly distinct rate of hypoxaemia (11.28% vs. 4.03%, *p*<0.001). The calibration plot of the training set indicated optimal correlation between
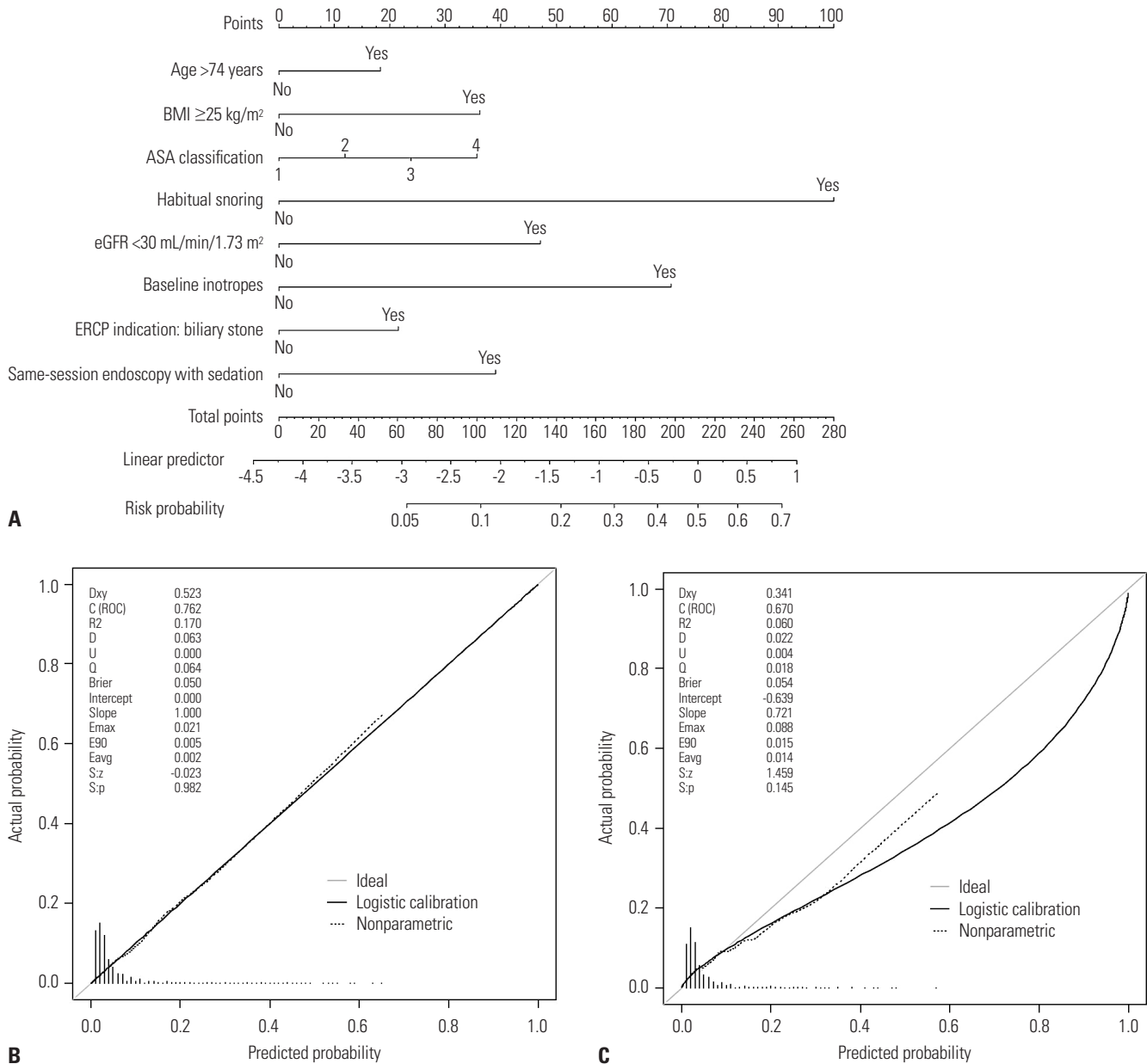
Fig. 2. Nomogram and calibration plots. (A) Developed nomogram of multivariable logistic regression model for the prediction of hypoxemia. Calibration plots in the (B) training set and the (C) test set. BMI, body mass index; ASA, American Society of Anesthesiologists; eGFR, estimated glomerular filtration rate; ERCP, endoscopic retrograde cholangiopancreatography; ROC, receiver operating characteristic.

the predicted and observed probability; however, overall over-estimation was shown in that of the test set (Fig. 2B and C).

Gradient boosting ensemble

Eight features in addition to the eight variables in the LR model were selected using l_1 regularisation: chronic liver disease (other than cirrhosis), chronic heart disease, chronic respiratory disease, propofol as sedative for same-session endoscopy, baseline O_2 administration, current smoker, haematocrit, and total bilirubin. Sixteen features were used to train the Xgboost-LR ensemble model. We noted that experimental model comparison is possible through approximate confidence intervals.

We calculated the confidence intervals from 30 random data splits in both the GBM ensemble and LR models. Table 4 lists the performance parameters and thresholds to discriminate the high-risk group, and Fig. 3 presents the receiver operating characteristic curves of both models in the test set. The test AUC of the GBM ensemble model was 0.7336 (95% CI: 0.7267–0.7404), surpassing the result of the LR model [0.7230 (95% CI: 0.7152–0.7308)]. As a result of measuring the difference in AUC scores between the models for each split, a more direct comparison of the models revealed an average improvement of 0.0106 (95% CI: 0.0074–0.0137). The GBM ensemble model had an enhanced AUC compared to the LR model. The GBM

Table 4. Performance Parameters of the LR Model and the GBM Ensemble Model for Predicting Hypoxaemia

Model	AUC calculation	AUC (95% CI)	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	PPV (95% CI)	NPV (95% CI)
LR								
Training	1 random split, bootstrap resampling	0.762 (0.727–0.795)	Nomogram score 70	65.20 (59.30–71.10)	75.46 (74.12–76.79)	74.85 (73.55–76.16)	14.22 (12.20–16.25)	97.20 (96.62–97.78)
Test		0.670 (0.614–0.723)		49.07 (39.65–58.50)	75.84 (73.82–77.86)	74.26 (72.26–76.26)	11.28 (8.42–14.14)	95.97 (94.92–97.01)
LR								
Training	30 random splits, average value	0.7313 (0.7280–0.7345)	Model output 0.06	57.22 (54.10–60.33)	76.85 (73.90–79.81)	75.70 (73.09–78.30)	14.07 (13.18–14.97)	96.69 (96.55–96.83)
Test		0.7230 (0.7152–0.7308)		54.58 (51.81–57.35)	76.83 (73.84–79.81)	75.52 (72.85–78.19)	13.74 (12.67–14.81)	96.43 (96.27–96.59)
GBM ensemble								
Training	30 random splits, average value	0.7924 (0.7899–0.7948)	Model output 0.09	70.94 (69.92–71.96)	73.44 (72.34–74.53)	73.29 (72.31–74.27)	14.58 (14.22–14.94)	97.62 (97.56–97.68)
Test		0.7336 (0.7267–0.7404)		63.60 (60.13–67.07)	72.21 (69.01–75.42)	71.70 (68.88–74.52)	13.14 (12.33–13.96)	96.97 (96.76–97.19)

AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; LR, logistic regression; GBM, gradient boosting algorithm.
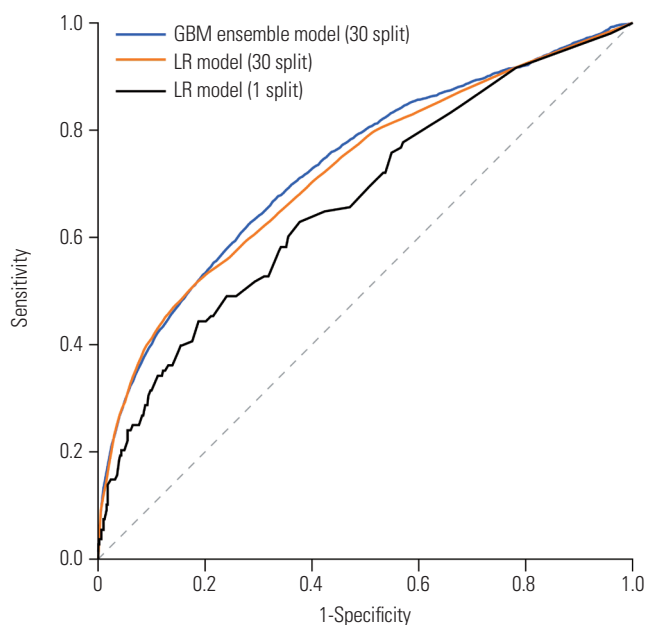

Fig. 3. ROC curves of each model for predicting hypoxia during ERCP in test sets. GBM, gradient boosting algorithm; LR, logistic regression; ROC, receiver operating characteristic; ERCP, endoscopic retrograde cholangiopancreatography.

ensemble model output is scaled between 0 and 1; a value closer to 1 means a higher risk of hypoxaemia. The calculated threshold of the GBM ensemble model output by the Youden index in the training set was 0.09. When this threshold was applied in the test set, the high-risk group showed a significantly higher average rate of hypoxaemia than the low-risk group (13.14±2.32% vs. 3.03±0.61%, $p<0.001$) with a sensitivity and specificity of 63.60% and 72.21%, respectively.

DISCUSSION

We established an LR-based nomogram model and an ML-based model for predicting hypoxemia risk during ERCP under MAC. The LR-based nomogram distinguished the high-risk group for hypoxemia in the training set, but the test set had a less significant AUC than the training set. The GBM ensemble model delivered good prediction performance and risk stratification in the training and test sets, with performance parameters superior to those of the LR-based model.

Various risk factors have been identified for SRAEs during ERCP. In a retrospective study including 650 ERCPs, diagnostic indication and female gender were related with SRAEs in the MAC group.[32] Conversely, a prospective study with 799 advanced endoscopies under MAC reported that male gender was associated with airway modifications.[17] Higher ASA class and BMI are well-known risk factors that were identified in multiple studies; sleep apnoea and emergent ERCP have also been proposed.[18-21] Although anaesthesia provider-administered sedation is increasingly used for ERCP, the decision between MAC and GEA may vary between institutions because of the doctor's preference, institutional policy, and lack of data.[33-35]

A prospective study that evaluated 438 ERCPs reported no difference in hypoxaemia between MAC and GEA.[36] In contrast, another prospective study with 528 ERCPs reported that respiratory events were more common in MAC than those in GEA, but cardiovascular events occurred more frequently in GEA than those in MAC.[18] Patients with high ASA class and BMI were more common in the GEA group of both studies as they were not RCTs, and these factors made the anaesthesiologist choose GEA, which may have resulted in selection bias. A recent RCT comparing ERCP in high-risk patients between MAC and GEA showed higher incidence of SRAEs in the MAC group

than that in the GEA group (51.5% vs. 9.9%, $p<0.001$), including that of hypoxaemia (19.2% vs. 0%, $p<0.001$).[13] These data suggest that GEA should be chosen for ERCP in patients at an elevated risk of SRAEs, which seems more suitable in practice than performing GEA on every patient undergoing ERCP. Providing anaesthesia, particularly GEA, in the endoscopy unit has many hurdles, such as anaesthesiologist's unfamiliarity and resources, including equipment and a post-anaesthesia care unit.[23,35,37] Moreover, conflicting results on whether GEA negatively impacts efficiency metrics in an endoscopy unit have been reported.[4,13]

The aforementioned RCT classified patients as high risk if they had at least one risk factor for SRAEs.[13] However, ERCP has a broad spectrum of procedural difficulties, expected procedure times, and patient conditions. For example, a simple follow-up ERCP for an obese patient may not need general anaesthesia, but a young adult in good physical condition may experience severe hypoxemia under MAC due to a prolonged procedure resulting from difficulty in cannulation. Therefore, a prediction model for a more precise risk stratification may be especially useful.[18] Our risk prediction model for hypoxaemia during ERCP under MAC can help distinguish potential candidates for GEA before the procedure. Among the eight variables in the LR model, advanced age, higher ASA class, habitual snoring, and increased BMI are well-known factors associated with increased SRAEs.[17-21,37] Regarding biliary stone as ERCP indication, we speculate that our institution may have a large proportion of difficult biliary and intrahepatic stones classified as grade 2–3 difficulty that require prolonged procedure times, as we are a large-volume university hospital.[3,38] However, a study from a smaller volume centre showed no difference in predicting procedure failure, even for ERCPs with biliary stones.[20] Same-session endoscopy with sedation before ERCP may require prolonged sedation and procedure times. The eGFR criteria indicated that stage IV or V chronic kidney disease was significantly associated with hypoxaemia, which might be related to the alteration of pharmacokinetics due to the impaired renal excretion of sedative agents.[39] Baseline administration of inotropes indicates cardiovascular insufficiency, mostly as a result of septic shock in a patient with an indication of ERCP. Moreover, sedative agents that inhibit sympathetic activity can cause disorders in blood circulation and oxygen exchange in these patients.[40]

The GBM ensemble model, which showed better prediction performance, was developed with eight additional variables selected by ML, which might indicate increased comorbidity, higher ASA class, prolonged sedation time, and decreased respiratory capacity. However, some of these variables showed no significance in univariate analysis. Although feature selection was performed by ML to construct the most-fitted model, a detailed explanation of how these variables are related to hypoxaemia is not possible due to the nature of ML. Considering the characteristics of the current datasets, GBM was selected among the advanced ML models. GBM is suitable for relatively small tabular datasets,[28] and can effectively approximate nonlinear functions using the model boosting method.[27] A suitable deep-learning architecture to improve the predictive power may be found if more patient datasets become available in the future. Deep-learning models typically require more training data than GBM models and could exhibit enhanced performance by training large-scale data with little performance saturation.

Our prediction models can serve as auxiliary tools for facilitating safety and efficiency and effective tools for communication between anaesthesiologists and endoscopists to select the appropriate type of anaesthesia. The LR-based nomogram is easier to calculate; however, the GBM ensemble model can yield more accurate predictions, although it requires computer-aided calculation. If a patient is distinguished as the high-risk group of hypoxaemia during ERCP by our models, ERCP under GEA, rather than MAC, could be a better option. In addition, institutional experience, patient safety, available resources, and cost should all be considered.[32]

Our study had several limitations. First, our endpoint was hypoxaemia alone, since major differences in hypoxaemia between MAC and GEA were reported in previous RCTs.[13] We focused on hypoxaemia, which can be prevented in GEA with a secured airway. Second, some factors, such as patient position and expected procedure difficulty were not evaluated; evaluating these factors may improve the prediction accuracy. The duration of MAC and the dose of propofol and fentanyl were intentionally left out from the establishment of the model. This was because these variables did not show a strong association with hypoxaemia in univariable analysis; and most importantly, they can only be obtained after the procedure, not before. Third, our total rate of hypoxaemia was 5.9%, which was less than that in previous studies (10%–28%),[11-14] since all MAC processes were implemented by an experienced anaesthesiologist. Immediate airway manoeuvres in response to a detected apnoea or a decrease in SpO_2 may have resulted in our low rate of hypoxaemia. In many hospitals, sedation for ERCP is provided by not only anaesthesiologists but also nurse anaesthetists or endoscopists. Thus, the predicted risk should be interpreted in consideration of each institution's situation. Fourth, since our study was conducted at a single institution, external validation is required. There are two limitations of applying the current dataset to the ML models. First, the size of the dataset is small, so performance variance depends considerably on how the dataset is split into the training and test sets; therefore, additional training data are required to make the ML models more robust. Second, the dataset has an imbalanced label ratio; therefore, techniques, such as oversampling and weighted loss function, should be applied to prevent ML models from biased training.

In conclusion, we established an easily applicable LR-based nomogram with acceptable accuracy and an ML-based GBM ensemble model that showed statistically better performance

than the LR model in predicting the risk of hypoxaemia during ERCP under anaesthesiologist-administered MAC. Our results suggest that the GBM ensemble model has a good potential to prevent hypoxaemia during ERCP under MAC.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Conceptualization: Moon Jae Chung and Min-Soo Kim. Data curation: Huapyong Kang, Bora Lee, and Moon Jae Chung. Formal analysis: Huapyong Kang, Bora Lee, Joonhyung Park, Hajin Shim, Jung Hyun Lee, Eunho Yang, and Eun Hwa Kim. Investigation: Huapyong Kang, Bora Lee, Jung Hyun Jo, Hee Seung Lee, Jeong Youp Park, Seungmin Bang, Seung Woo Park, Si Young Song, Min-Soo Kim, and Moon Jae Chung. Methodology: Eunho Yang, Min-Soo Kim, and Moon Jae Chung. Project administration: Kwang Joon Kim and Moon Jae Chung. Resources: Min-Soo Kim and Moon Jae Chung. Software: Eun Hwa Kim. Supervision: Min-Soo Kim and Moon Jae Chung. Validation: Huapyong Kang, Bora Lee, Hajin Shim, and Eun Hwa Kim. Writing—original draft: Huapyong Kang, Bora Lee, Hajin Shim, and Moon Jae Chung. Writing—review & editing: Jung Hyun Jo, Hee Seung Lee, Jeong Youp Park, Seungmin Bang, Seung Woo Park, Si Young Song, Eun Hwa Kim, Kwang Joon Kim, and Min-Soo Kim. Approval of final manuscript: all authors.

ORCID iDs

Huapyong Kang	https://orcid.org/0000-0003-1790-0809
Bora Lee	https://orcid.org/0000-0002-7699-967X
Jung Hyun Jo	https://orcid.org/0000-0002-2641-8873
Hee Seung Lee	https://orcid.org/0000-0002-2825-3160
Jeong Youp Park	https://orcid.org/0000-0003-0110-8606
Seungmin Bang	https://orcid.org/0000-0001-5209-8351
Seung Woo Park	https://orcid.org/0000-0001-8230-964X
Si Young Song	https://orcid.org/0000-0002-1417-4314
Joonhyung Park	https://orcid.org/0000-0003-4662-0099
Hajin Shim	https://orcid.org/0000-0002-0116-3237
Jung Hyun Lee	https://orcid.org/0000-0002-6580-6289
Eunho Yang	https://orcid.org/0000-0003-2188-0169
Eun Hwa Kim	https://orcid.org/0000-0003-1692-6126
Kwang Joon Kim	https://orcid.org/0000-0002-5554-8255
Min-Soo Kim	https://orcid.org/0000-0001-8760-4568
Moon Jae Chung	https://orcid.org/0000-0002-5920-8549

REFERENCES

1. Hu LH, Xin L, Liao Z, Pan J, Qian W, Wang LW, et al. ERCP development in the largest developing country: a national survey from China in 2013. Gastrointest Endosc 2016;84:659-66.
2. Coelho-Prabhu N, Shah ND, Van Houten H, Kamath PS, Baron TH. Endoscopic retrograde cholangiopancreatography: utilisation and outcomes in a 10-year population-based cohort. BMJ Open 2013;3:e002689.
3. Ahn DW, Han JH, Kim HJ, Kim HK, Son BK, Yi SY, et al. Practice of endoscopic retrograde cholangiopancreatography in Korea: results from a national survey. Korean J Pancreas Biliary Tract 2019;24:21-30.
4. Wadhwa V, Gupta K, Vargo JJ. Monitoring standards in sedation and analgesia: the odyssey of capnography in sedation for gastroenterology procedures. Curr Opin Anaesthesiol 2019;32:453-6.
5. Wehrmann T, Riphaus A. Sedation with propofol for interventional endoscopic procedures: a risk factor analysis. Scand J Gastroenterol 2008;43:368-74.
6. Sudheer PS, Logan SW, Ateleanu B, Hall JE. Haemodynamic effects of the prone position: a comparison of propofol total intravenous and inhalation anaesthesia. Anaesthesia 2006;61:138-41.
7. Edgcombe H, Carter K, Yarrow S. Anaesthesia in the prone position. Br J Anaesth 2008;100:165-83.
8. Early DS, Lightdale JR, Vargo JJ 2nd, Acosta RD, Chandrasekhara V, Chathadi KV, et al. Guidelines for sedation and anesthesia in GI endoscopy. Gastrointest Endosc 2018;87:327-37.
9. Klare P, Reiter J, Meining A, Wagenpfeil S, Kronshage T, Geist C, et al. Capnographic monitoring of midazolam and propofol sedation during ERCP: a randomized controlled study (EndoBreath study). Endoscopy 2016;48:42-50.
10. Gerstenberger PD. Capnography and patient safety for endoscopy. Clin Gastroenterol Hepatol 2010;8:423-5.
11. Yang JF, Farooq P, Zwilling K, Patel D, Siddiqui AA. Efficacy and safety of propofol-mediated sedation for outpatient endoscopic retrograde cholangiopancreatography (ERCP). Dig Dis Sci 2016;61:1686-91.
12. Park CH, Park SW, Hyun B, Lee J, Kae SH, Jang HJ, et al. Efficacy and safety of etomidate-based sedation compared with propofol-based sedation during ERCP in low-risk patients: a double-blind, randomized, noninferiority trial. Gastrointest Endosc 2018;87:174-84.
13. Smith ZL, Mullady DK, Lang GD, Das KK, Hovis RM, Patel RS, et al. A randomized controlled trial evaluating general endotracheal anesthesia versus monitored anesthesia care and the incidence of sedation-related adverse events during ERCP in high-risk patients. Gastrointest Endosc 2019;89:855-62.
14. Hormati A, Aminnejad R, Saeidi M, Ghadir MR, Mohammadbeigi A, Shafiee H. Prevalence of anesthetic and gastrointestinal complications of endoscopic retrograde cholangiopancreatography. Anesth Pain Med 2019;9:e95796.
15. Qadeer MA, Vargo JJ, Dumot JA, Lopez R, Trolli PA, Stevens T, et al. Capnographic monitoring of respiratory activity improves safety of sedation for endoscopic cholangiopancreatography and ultrasonography. Gastroenterology 2009;136:1568-76; quiz 1819-20.
16. Kim SH, Park M, Lee J, Kim E, Choi YS. The addition of capnography to standard monitoring reduces hypoxemic events during gastrointestinal endoscopic sedation: a systematic review and meta-analysis. Ther Clin Risk Manag 2018;14:1605-14.
17. Coté GA, Hovis RM, Ansstas MA, Waldbaum L, Azar RR, Early DS, et al. Incidence of sedation-related complications with propofol use during advanced endoscopic procedures. Clin Gastroenterol Hepatol 2010;8:137-42.
18. Berzin TM, Sanaka S, Barnett SR, Sundar E, Sepe PS, Jakubowski M, et al. A prospective assessment of sedation-related adverse events and patient and endoscopist satisfaction in ERCP with anesthesiologist-administered sedation. Gastrointest Endosc 2011;73:710-7.
19. Wani S, Azar R, Hovis CE, Hovis RM, Cote GA, Hall M, et al. Obesity as a risk factor for sedation-related complications during propofol-mediated sedation for advanced endoscopic procedures. Gastrointest Endosc 2011;74:1238-47.
20. Buxbaum J, Roth N, Motamedi N, Lee T, Leonor P, Salem M, et al. Anesthetist-directed sedation favors success of advanced endoscopic procedures. Am J Gastroenterol 2017;112:290-6.

21. Müller S, Prolla JC, Maguilnik I, Breyer HP. Predictive factors of oxygen desaturation of patients submitted to endoscopic retrograde cholangiopancreatography under conscious sedation. Arq Gastroenterol 2004;41:162-6.

22. Perbtani YB, Summerlee RJ, Yang D, An Q, Suarez A, Williamson JB, et al. Impact of endotracheal intubation on interventional endoscopy unit efficiency metrics at a tertiary academic medical center. Am J Gastroenterol 2016;111:800-7.

23. Martindale SJ. Anaesthetic considerations during endoscopic retrograde cholangiopancreatography. Anaesth Intensive Care 2006; 34:475-80.

24. Allyn J, Allou N, Augustin P, Philip I, Martinet O, Belghiti M, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. PLoS One 2017;12:e0169772.

25. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning–based model for prediction of outcomes in acute stroke. Stroke 2019;50:1263-5.

26. Ming C, Viassolo V, Probst-Hensch N, Chappuis PO, Dinov ID, Katapodi MC. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Res 2019;21:75.

27. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29:1189-232.

28. Chen X, Huang L, Xie D, Zhao Q. EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. Cell Death Dis 2018;9:3.

29. Yeboah J, Redline S, Johnson C, Tracy R, Ouyang P, Blumenthal RS, et al. Association between sleep apnea, snoring, incident cardiovascular events and all-cause mortality in an adult population: MESA. Atherosclerosis 2011;219:963-8.

30. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Methodol 1996;58:267-88.

31. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13-17; San Francisco, CA, USA: KDD; 2016:785-94.

32. Sorser SA, Fan DS, Tommolino EE, Gamara RM, Cox K, Chortkoff B, et al. Complications of ERCP in patients undergoing general anesthesia versus MAC. Dig Dis Sci 2014;59:696-7.

33. Smith ZL, Das KK, Kushnir VM. Anesthesia-administered sedation for endoscopic retrograde cholangiopancreatography: monitored anesthesia care or general endotracheal anesthesia? Curr Opin Anaesthesiol 2019;32:531-7.

34. Smith ZL, Nickel KB, Olsen MA, Vargo JJ, Kushnir VM. Type of sedation and the need for unplanned interventions during ERCP: analysis of the clinical outcomes research initiative national endoscopic database (CORI-NED). Frontline Gastroenterol 2020;11:104-10.

35. Thosani N, Banerjee S. Deep sedation or general anesthesia for ERCP? Dig Dis Sci 2013;58:3061-3.

36. Barnett SR, Berzin T, Sanaka S, Pleskow D, Sawhney M, Chuttani R. Deep sedation without intubation for ERCP is appropriate in healthier, non-obese patients. Dig Dis Sci 2013;58:3287-92.

37. Kuzhively J, Pandit JJ. Anesthesia and airway management for gastrointestinal endoscopic procedures outside the operating room. Curr Opin Anaesthesiol 2019;32:517-22.

38. Baron TH, Petersen BT, Mergener K, Chak A, Cohen J, Deal SE, et al. Quality indicators for endoscopic retrograde cholangiopancreatography. Gastrointest Endosc 2006;63(4 Suppl):S29-34.

39. Triantafillidis JK, Merikas E, Nikolakis D, Papalois AE. Sedation in gastrointestinal endoscopy: current issues. World J Gastroenterol 2013;19:463-81.

40. Hoka S, Yamaura K, Takenaka T, Takahashi S. Propofol-induced increase in vascular capacitance is due to inhibition of sympathetic vasoconstrictive activity. Anesthesiology 1998;89:1495-500.