

RESEARCH

Open Access



# Predicting postoperative pulmonary infection in elderly patients undergoing major surgery: a study based on logistic regression and machine learning models

Jie Liu<sup>1,2†</sup>, Xia Li<sup>1†</sup>, Yanting Wang<sup>1</sup>, Zhenzhen Xu<sup>1</sup>, Yong Lv<sup>1</sup>, Yuyao He<sup>1</sup>, Lu Chen<sup>1</sup>, Yiqi Feng<sup>1</sup>, Guoyang Liu<sup>1</sup>, Yunxiao Bai<sup>1</sup>, Wanli Xie<sup>1</sup> and Qingping Wu<sup>1\*</sup>

## Abstract

**Background** Postoperative pulmonary infection (POI) is strongly associated with a poor prognosis and has a high incidence in elderly patients undergoing major surgery. Machine learning (ML) algorithms are increasingly being used in medicine, but the predictive role of logistic regression (LR) and ML algorithms for POI in high-risk populations remains unclear.

**Methods** We conducted a retrospective cohort study of older adults undergoing major surgery over a period of six years. The included patients were randomly divided into training and validation sets at a ratio of 7:3. The features selected by the least absolute shrinkage and selection operator regression algorithm were used as the input variables of the ML and LR models. The random forest of multiple interpretable methods was used to interpret the ML models.

**Results** Of the 9481 older adults in our study, 951 developed POI. Among the different algorithms, LR performed the best with an AUC of 0.80, whereas the decision tree performed the worst with an AUC of 0.75. Furthermore, the LR model outperformed the other ML models in terms of accuracy (88.22%), specificity (90.29%), precision (44.42%), and F1 score (54.25%). Despite employing four interpretable methods for RF analysis, there existed a certain degree of inconsistency in the results. Finally, to facilitate clinical application, we established a web-friendly version of the nomogram based on the LR algorithm; In addition, patients were divided into three significantly distinct risk intervals in predicting POI.

**Conclusions** Compared with popular ML algorithms, LR was more effective at predicting POI in older patients undergoing major surgery. The constructed nomogram could identify high-risk elderly patients and facilitate perioperative management planning.

**Trial registration** The study was retrospectively registered (NCT06491459).

<sup>†</sup>Jie Liu and Xia Li contributed equally to this work.

\*Correspondence:  
Qingping Wu  
wqp1968@126.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **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/>.

**Keywords** Elderly, Postoperative pulmonary infection, Logistic regression, Machine learning, Prediction model

### Key Summary Point

- Postoperative pulmonary infection (POI) significantly impacts the prognosis of elderly patients undergoing major surgery, and identifying its risk factors can enhance treatment options.
- This study developed a POI prediction model using logistic regression (LR) and machine learning (ML) algorithms, with LR demonstrating the best performance (AUC of 0.80). Additionally, four methods were employed to explain the feature importance in the ML models.
- The nomogram generated from the LR model offers clinicians a practical tool for identifying high-risk elderly patients and optimizing perioperative management strategies.

## Introduction

Postoperative pulmonary infection (POI) is frequently linked to poor prognoses, prolonged hospital stays, and increased healthcare costs [1–3]. The incidence is between 0.9% and 39%, depending on the population and definition of the study [4, 5]. Due to the loss of physiological organ reserves and the increased colonization of gram-negative bacteria in the upper respiratory tract, older people are more likely to develop POI [6]. In a study of cancer surgery, the hazard ratio for the one-year cumulative prevalence of POI in patients over 65 years of age was as high as 3.7-fold [7]. Owing to the characteristics of major surgery, such as long operation time and repeated ischemia and reperfusion, it is easy to cause changes in the immune cells of the body and produce systemic inflammatory reactions, resulting in lung function damage [3, 8]. Major surgery is associated with poor outcomes, and with the aging of the population, the proportion of elderly patients is gradually growing [9–11].

Furthermore, previous studies have shown that changes in circulating inflammatory cells (such as lymphocyte, neutrophil, and monocyte counts) are associated with postoperative complications [12–14]. Circulating inflammatory cell counts are readily available markers that may provide valuable information for predicting the risk of postoperative complications in elderly patients. Although some studies have explored the risk factors for POI [15–17], few studies with large samples focused on POI after major surgery in older adults.

Machine learning (ML) is becoming increasingly prevalent in the medical field, especially for predicting complications [18]. ML is a branch of artificial intelligence that builds models using vast amounts of medical data, leveraging algorithms that do not assume linearity in the relationships between variables. ML has the advantage of revealing deeper correlations in data and is robust to noisy information [19]. ML parses medical information and can be used to build disease occurrence, diagnosis, and prognosis assessment models to aid clinical decision-making [20, 21]. Leading ML algorithms, such as decision

tree (DT), random forest (RF), support vector machine (SVM), gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), and multilayer perceptron (MLP), represent some of the most widely used models in the field [22]. However, some studies have shown that logistic regression (LR) models do not perform worse than ML models and even outperform ML in predicting complications [23–25]. ML faces limitations, including the need for large training datasets and poor interpretability, as the influence of individual variables on model outcomes is often unclear [26]. To our knowledge, although there are studies involving different algorithms, few have investigated the performance of different algorithms in predicting POI in older adults undergoing major surgery.

This study aimed to identify the predictors of POI for perioperative management and to compare the predictive performance of the ML and LR models for POI after major surgery in elderly patients. Furthermore, various interpretable methods were employed to explore the variable importance of ML models.

## Materials and methods

We obtained all data from the electronic database system for elderly patients who underwent surgery at the Wuhan Union Hospital from January 2014 to December 2019. In retrospective enrollment, the inclusion criteria were (1) age  $\geq 65$  years and (2) mechanical ventilation during major surgery. The exclusion criteria were: (1) preoperative tracheal intubation, (2) preoperative pneumonia, (3) organ transplantation, and (4) missing data. According to the literature [27, 28], major surgery is any procedure involving general anesthesia for non-percutaneous, non-endoscopic, and invasive surgery. This study adhered to the principles outlined in the Declaration of Helsinki and was approved by the Ethics Committee of our institution (Ethics Number: 2021–0986).

**Data collection**

We collected appropriate clinical variables based on the literature and clinical experience. Demographic data included age, male sex, body mass index (BMI), history of smoking and alcohol consumption, cancer, chronic obstructive pulmonary disease (COPD), hypertension, coronary artery disease, and diabetes, and laboratory test data included hemoglobin, platelet count, albumin, leukocyte count, lymphocyte count, neutrophil count, monocyte count, blood urea nitrogen (BUN), creatinine, total bilirubin (TBIL), alanine aminotransferase (ALT), and aspartate aminotransferase (AST). Surgery-related factors included type of surgery, emergency surgery, surgical incision, operation time, red blood cell transfusion (RBC transfusion), and colloidal infusion. To ensure the reliability of the data, they were reviewed and

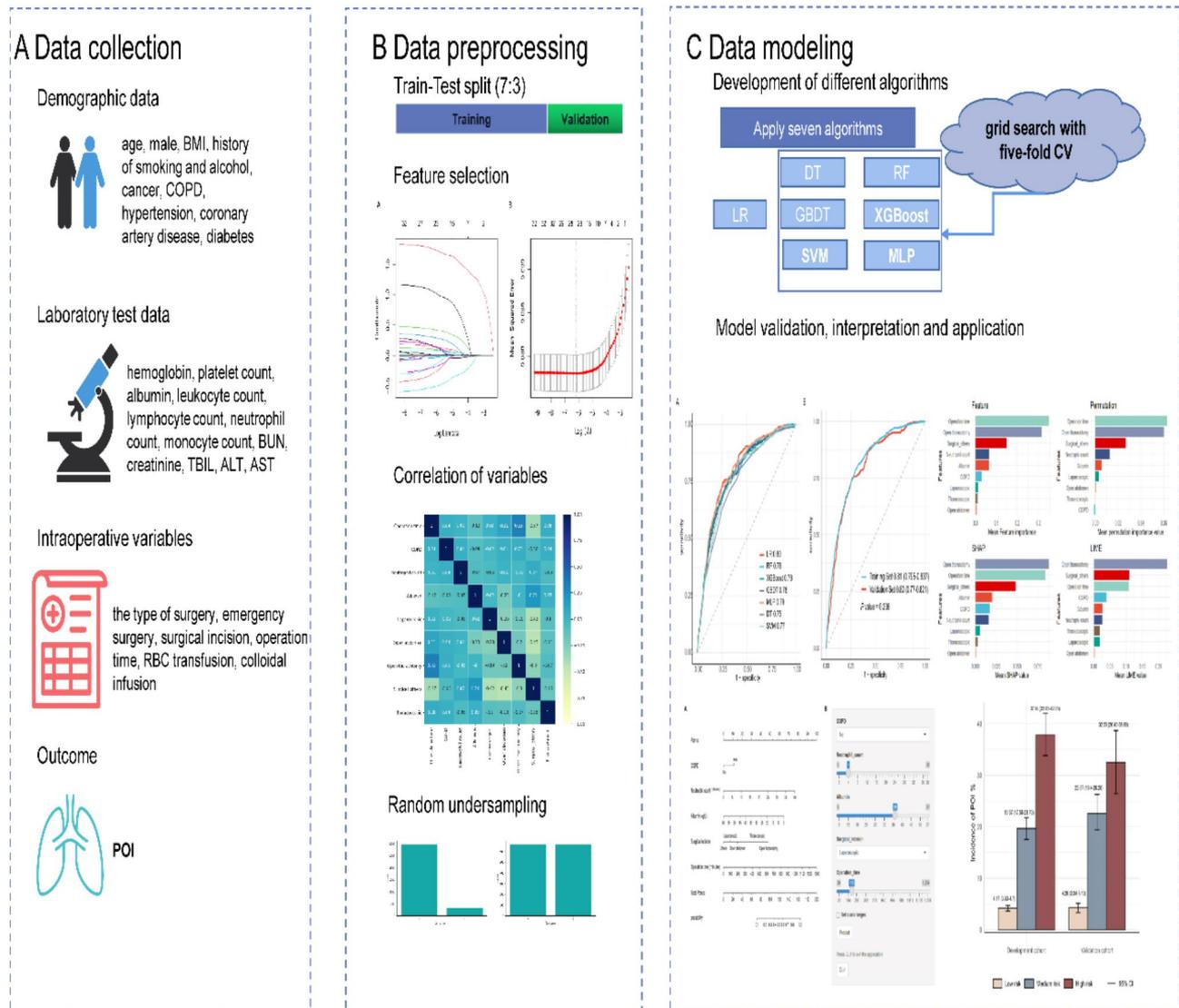
reconfirmed by two researchers, and contradictory data were resolved by the research team after discussion.

**Primary outcome**

The primary outcome was POI incidence during hospitalization. It is defined as new or progressive radiological infiltration accompanied by at least two symptoms: antibiotic treatment, body temperature above 38 °C, increased or decreased white blood cell count, and/or purulent airway secretions [29, 30].

**Data preprocessing**

Figure 1 shows the detailed process of the data preprocessing. We randomly divided the patients into training and validation sets in a ratio of 7:3. To reduce feature dimensionality and mitigate overfitting, we employed the least absolute shrinkage and selection operator (LASSO)



**Fig. 1** The general workflow of this study

regression algorithm to identify relevant risk factors. In the LASSO regression analysis, selected variables were incorporated into the model, and the regularization parameter  $\lambda$  was determined using the minimum standard error method (1-SE rule) through 10-fold cross-validation. A correlation heat map and variance inflation factor (VIF) were used to assess multicollinearity between variables.

Due to the imbalanced distribution of POI in the dataset, random undersampling (RUS) was applied to balance the samples. To prevent data leakage, RUS was performed only on the training set. Dummy coding was applied for unordered multi-categorical variables. For continuous data, we applied standardization to produce data with a mean of 0 and a standard deviation of 1.

### Data modeling

The data modelling process is illustrated in Fig. 1. Two distinct classes of modeling algorithms were considered for the building: the traditional LR and ML algorithms. To obtain more convincing results, we selected the following six popular ML algorithms: DT, RF, SVM, GBDT, XGBoost, and MLP.

To optimize model performance, we tuned hyperparameters using grid search with five-fold cross-validation (GridSearchCV). GridSearchCV evaluates all possible parameter combinations through cross-validation and selects the best set based on performance scores. For cross-validation, the training set was split into five equal parts. In each iteration, one part served as the validation set, while the other four were used for training. This process was repeated five times, and the model's performance was optimized based on the area under the receiver operating characteristic curve (AUC).

The best parameters were applied to the validation set. Model performance was evaluated using AUC and the Brier score, where the Brier score ranges from 0 to 1 (lower values indicate better calibration). Additional metrics, including accuracy, sensitivity, specificity, precision and F1 score were used to compare model performance. The cut-off threshold was determined using clinical experience or the Youden index.

The feature importance attribute with the model, local interpretable model-agnostic explanations (LIME), permutation feature importance, and the Shapley additive explanation (SHAP) method were selected for feature importance analysis. A nomogram was constructed using the LR algorithm. The study population was divided into three groups based on the nomogram and clinical practice. The three groups differed significantly in the predicted risk of POI: low-risk (<10%), intermediate-risk (10-30%), and high-risk (>30%).

### Data-analysis

Continuous variables are presented as means  $\pm$  standard deviations or medians (interquartile ranges, IQR), using independent samples t-tests and non-parametric tests, respectively. Categorical variables were expressed as frequencies and percentages and compared using the chi-squared test or Fisher's exact test.

A P-value < 0.05 (two-sided) was considered statistically significant. The XGBoost algorithm is constructed using the XGBoost package. Other ML algorithms were built using the scikit-learn package. Construction of the ML models was completed using Python (version 3.8.8). The remaining data analyses were performed using R software (version 4.1.1).

## Results

### Baseline characteristics

This study assessed the eligibility of older adults undergoing major surgery at our hospital over a six-year period. Based on the inclusion and exclusion criteria, 9481 patients were analyzed. The clinical characteristics of the patients are summarized in Table 1. A total of 951 patients developed POI, accounting for 10.03% of the population, while 8530 did not. The median age of the patients with POI was 69 years, and 53.78% were male. Compared to non-POI patients, patients with POI were more male, had more comorbidities, lower hemoglobin and albumin levels, higher levels of inflammatory factors, and longer operation time. In addition, the length of postoperative hospital stay and total length of hospitalization were significantly higher in patients who developed POI (16 vs. 10,  $P < 0.001$ ; 25 vs. 16,  $P < 0.001$ ). There were no significant differences between the two cohorts (training and validation sets) for any of the features collected, including the incidence of POI (10.03% vs. 10.02%,  $P = 1$ ).

### Feature selection

Based on the LASSO regression analysis of the training set, five variables were entered into the final model (Fig. 2). These five variables were COPD, neutrophil count, albumin level, surgical incision, and operation time, which were independently associated with POI ( $P < 0.05$ ). To explore multicollinearity between the independent variables, we calculated Spearman's correlation coefficients and visualized them in a heatmap (Fig. 3). The heatmap shows low correlations between the variables. Furthermore, all variance inflation factor values were below 2, confirming minimal multicollinearity. The hyperparameters of the ML models are detailed in Table S1.

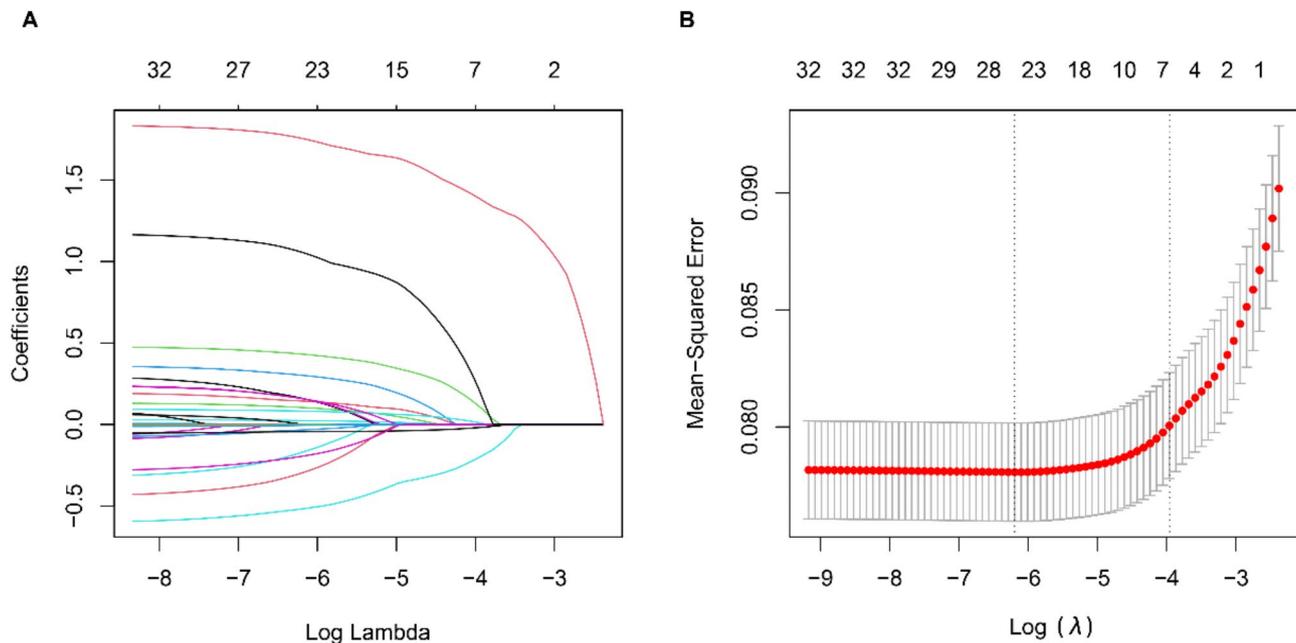
**Table 1** Baseline characteristics of the cohort

Variables	Total (n=9481)	Without POI (n=8530)	With POI (n=951)	P value	Training set (n=6637)	Validation set (n=2844)	P value
<b>Demographics</b>							
Age, years	69 (67, 73)	69 (67, 73)	69 (67, 74)	0.058	69 (67, 73)	69 (67, 73)	0.814
Male, %	5099 (53.78)	4499 (52.74)	600 (63.09)	< 0.001	3586 (54)	1513 (53.2)	0.471
BMI, kg/m <sup>2</sup>	22.89 (20.81, 25.21)	22.89 (20.81, 25.25)	22.66 (20.76, 24.80)	0.046	22.89 (20.83, 25.26)	22.86 (20.76, 25.05)	0.164
Smoking history, %	2224 (23.46)	1899 (22.26)	325 (34.17)	< 0.001	1567 (23.61)	657 (23.1)	0.611
Drinking history, %	1699 (17.92)	1479 (17.34)	220 (23.13)	< 0.001	1197 (18.04)	502 (17.65)	0.676
COPD, %	2220 (23.42)	1899 (22.26)	321 (33.75)	< 0.001	1542 (23.23)	678 (23.84)	0.54
Cancer, %	1278 (13.48)	1092 (12.8)	186 (19.56)	< 0.001	905 (13.64)	373 (13.12)	0.518
Hypertension, %	3559 (37.54)	3171 (37.17)	388 (40.80)	0.031	2510 (37.82)	1049 (36.88)	0.403
Coronary artery disease, %	1198 (12.64)	985 (11.55)	213 (22.4)	< 0.001	841 (12.67)	357 (12.55)	0.9
Diabetes, %	1248 (13.16)	1101 (12.91)	147 (15.46)	0.031	896 (13.5)	352 (12.38)	0.147
<b>Laboratory test</b>							
Hemoglobin count, g/L	125 (112, 135)	125 (113, 135)	122 (109, 134)	< 0.001	125 (112, 135)	125 (113, 135)	0.843
Platelet count, *10 <sup>9</sup> /L	192 (153, 236)	193 (154, 237)	185 (143, 228)	< 0.001	193 (154, 236)	192 (151, 236)	0.146
Albumin, g/L	39.5 (36.4, 42.6)	39.6 (36.6, 42.7)	38.4 (35.3, 41.3)	< 0.001	39.5 (36.3, 42.6)	39.5 (36.5, 42.5)	0.87
Leukocyte count, *10 <sup>9</sup> /L	5.69 (4.68, 6.98)	5.66 (4.67, 6.93)	5.93 (4.85, 7.37)	< 0.001	5.67 (4.67, 6.97)	5.72 (4.7, 7.01)	0.342
Neutrophil count, *10 <sup>9</sup> /L	3.45 (2.64, 4.57)	3.43 (2.62, 4.53)	3.69 (2.84, 4.98)	< 0.001	3.45 (2.64, 4.55)	3.45 (2.65, 4.61)	0.573
Monocyte count, *10 <sup>9</sup> /L	0.41 (0.32, 0.53)	0.41 (0.32, 0.53)	0.45 (0.34, 0.57)	< 0.001	0.41 (0.32, 0.53)	0.41 (0.31, 0.54)	0.798
Blood urea nitrogen, mmol/L	5.52 (4.42, 6.84)	5.5 (4.4, 6.79)	5.81 (4.63, 7.48)	< 0.001	5.51 (4.43, 6.83)	5.53 (4.4, 6.87)	0.799
Creatinine, μmol/L	70.7 (59.9, 83.6)	70.4 (59.6, 83)	74.1 (63.55, 87.85)	< 0.001	70.6 (60, 83.5)	71.1 (59.9, 83.8)	0.716
TBIL, μmol/L	12.1 (9.1, 16.1)	12.1 (9.1, 16.1)	11.6 (8.7, 16.3)	0.09	12 (9.1, 16)	12.1 (9.1, 16.3)	0.572
ALT, U/L	19 (13, 29)	19 (13, 29)	19 (13, 30)	0.827	19 (13, 29)	19 (13, 29)	0.464
AST, U/L	21 (17, 27)	21 (17, 27)	21 (16.5, 28)	0.35	21 (17, 27)	21 (17, 27)	0.403
Serum sodium (mmol/L)	141.5 (139.8, 143.1)	141.6 (139.8, 143.1)	141.2 (139.3, 143)	0.002	141.5 (139.7, 143)	141.6 (139.9, 143.1)	0.313
Serum potassium (mmol/L)	4 (3.77, 4.26)	4 (3.77, 4.25)	4 (3.76, 4.3)	0.406	4 (3.77, 4.25)	4 (3.75, 4.26)	0.958
Serum calcium (mmol/L)	2.22 (2.13, 2.3)	2.22 (2.13, 2.3)	2.2 (2.12, 2.29)	< 0.001	2.22 (2.13, 2.3)	2.22 (2.13, 2.3)	0.996
<b>Intraoperative variables</b>							
Type of surgery				< 0.001			0.718
Otolaryngology	200 (2.1)	196 (2.3)	4 (0.4)		200 (2.1)	143 (2.2)	
Gynecology	304 (3.2)	299 (3.5)	5 (0.5)		304 (3.2)	216 (3.3)	
Breast and thyroid	646 (6.8)	638 (7.5)	8 (0.8)		646 (6.8)	452 (6.8)	
Neurosurgery	396 (4.2)	347 (4.1)	49 (5.2)		396 (4.2)	286 (4.3)	
Digestive	3732 (39.4)	3367 (39.5)	365 (38.4)		3732 (39.4)	2591 (39)	
Vascular	336 (3.5)	325 (3.8)	11 (1.2)		336 (3.5)	231 (3.5)	
Urology	766 (8.1)	747 (8.8)	19 (2)		766 (8.1)	526 (7.9)	
Cardiac	759 (8)	537 (6.3)	222 (23.3)		759 (8)	536 (8.1)	
Thoracic	647 (6.8)	436 (5.1)	211 (22.2)		647 (6.8)	459 (6.9)	
Others	420 (4.4)	414 (4.9)	6 (0.6)		420 (4.4)	312 (4.7)	
Surgical incision, %				< 0.001			0.137
Open thoracotomy	1112 (11.73)	733 (8.59)	379 (39.85)		784 (11.81)	328 (11.53)	
Thorascopic	352 (3.71)	272 (3.19)	80 (8.41)		256 (3.86)	96 (3.38)	
Open abdomen	2208 (23.29)	1976 (23.17)	232 (24.4)		1540 (23.2)	668 (23.49)	
Laparoscopic	2036 (21.47)	1911 (22.4)	125 (13.14)		1383 (20.84)	653 (22.96)	
Others	3773 (39.8)	3638 (42.65)	135 (14.2)		2674 (40.29)	1099 (38.64)	
Emergency surgery, n (%)	275 (2.9)	236 (2.77)	39 (4.1)	0.026	203 (3.06)	72 (2.53)	0.182
Operation time, min	175 (113, 253)	167 (108, 240)	259 (187, 325)	< 0.001	177 (113, 254)	174 (115, 250)	0.811
RBC transfusion, %	1753 (18.49)	1405 (16.47)	348 (36.59)	< 0.001	1225 (18.46)	528 (18.57)	0.924
Colloid infusion				< 0.001			0.349
None	2130 (22.47)	1986 (23.28)	144 (15.14)		1518 (22.87)	612 (21.52)	
0-500 ml	3886 (40.99)	3544 (41.55)	342 (35.96)		2704 (40.74)	1182 (41.56)	
> 500 ml	3465 (36.55)	3000 (35.17)	465 (48.9)		2415 (36.39)	1050 (36.92)	

**Table 1** (continued)

Variables	Total (n=9481)	Without POI (n=8530)	With POI (n=951)	P value	Training set (n=6637)	Validation set (n=2844)	P value
<b>Postoperative variables</b>							
Hospital length of stay, days	17 (12, 23)	16 (11, 22)	25 (18, 35)	<0.001	17 (11, 23)	17 (12, 23)	0.915
Postoperative hospital length of stay, days	10 (7, 14)	10 (7, 14)	16 (11.5, 23)	<0.001	10 (7, 14)	10 (7, 14)	0.834
POI	455 (5.09)	-	-	-	666 (10.03)	285 (10.02)	1

Data are presented as median (interquartile range) or n (%) and compared using Mann-Whitney's test,  $\chi^2$  test, or Fisher's exact test, respectively. BMI, body mass index; COPD, chronic obstructive pulmonary disease; TBIL, total bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; RBC, red blood cell; POI, postoperative pulmonary infection



**Fig. 2** LASSO regression plot. **(A)** Plot of LASSO coefficient paths. **(B)** 10-fold cross-validation curve for the penalty term. LASSO, least absolute shrinkage and selection operator

### Algorithm performance

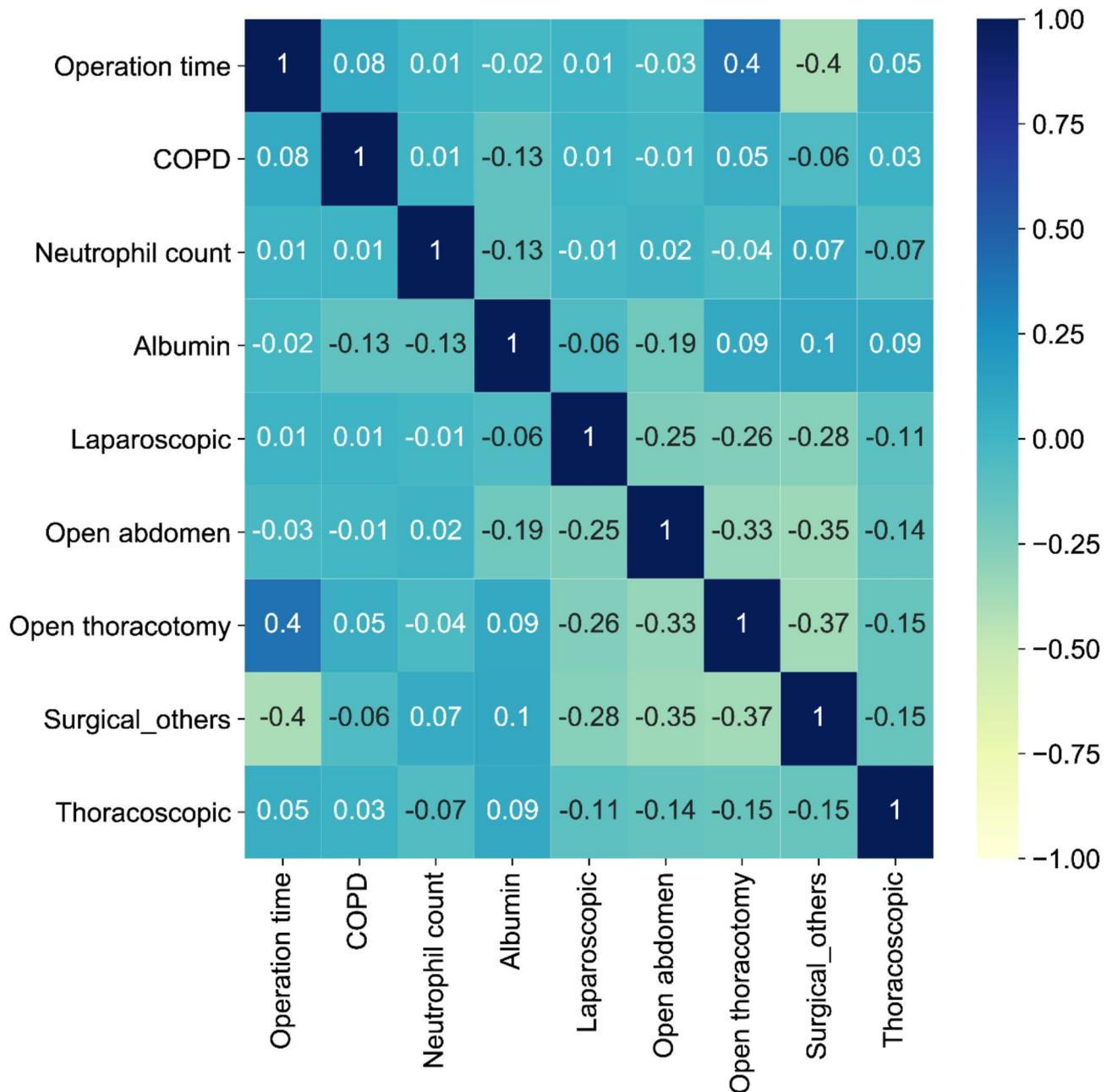
The LR model achieved the highest AUC of 0.80 on the validation set, while the DT model had the lowest AUC of 0.75 (Fig. 4A). Furthermore, the LR model's AUC was 0.81 (95% CI: 0.795–0.837) in the training set (Fig. 4B). The results showed that the AUC of LR significantly improved when compared to DT, RF, SVM, GBDT, and XGBoost (DT vs. LR:  $\Delta$ AUC = -0.045,  $P < 0.001$ ; RF vs. LR:  $\Delta$ AUC = -0.014,  $P = 0.014$ ; GBDT vs. LR:  $\Delta$ AUC = -0.02,  $P = 0.005$ ; XGBoost vs. LR:  $\Delta$ AUC = 0.014,  $P = 0.032$ ; SVM vs. LR:  $\Delta$ AUC = 0.029,  $P = 0.001$ ). Although there was a difference in AUC between LR and MLP, it did not reach statistical significance (MLP vs. LR:  $\Delta$ AUC = -0.002,  $P = 0.56$ ) (Table S2). The Brier scores of the seven algorithms ranged from 0.08 to 0.22, indicating good calibration (LR: 0.081; DT: 0.197; RF: 0.187; XGBoost: 0.186; GBDT: 0.189; SVM: 0.219; MLP: 0.178) (Figure S1). The low Brier score of LR reflects its superior reliability in probability prediction. Calibration curve further confirmed that the LR model outperformed other algorithms

in calibration (Figure S1). Additionally, the LR model outperformed the other ML models in terms of accuracy (88.22%), specificity (90.29%), precision (44.42%), and F1 score (54.25%) (Table S2).

### Algorithm interpretability and application

The four interpretable methods of the RF model showed that operation time, open thoracotomy, albumin level, and neutrophil count were important features (Fig. 5). The rankings of the top five variables in the feature importance and permutation importance methods were generally consistent. However, there were some differences in the rankings between the LIME importance and SHAP methods. Although the top three variables were generally related to surgery, there may be slight variations. In addition, the ranking of COPD fluctuated significantly across various methods.

Based on AUC, specificity, accuracy, precision, F1 score, Brier score, and calibration curve evaluation, we ultimately selected the LR model as the primary model

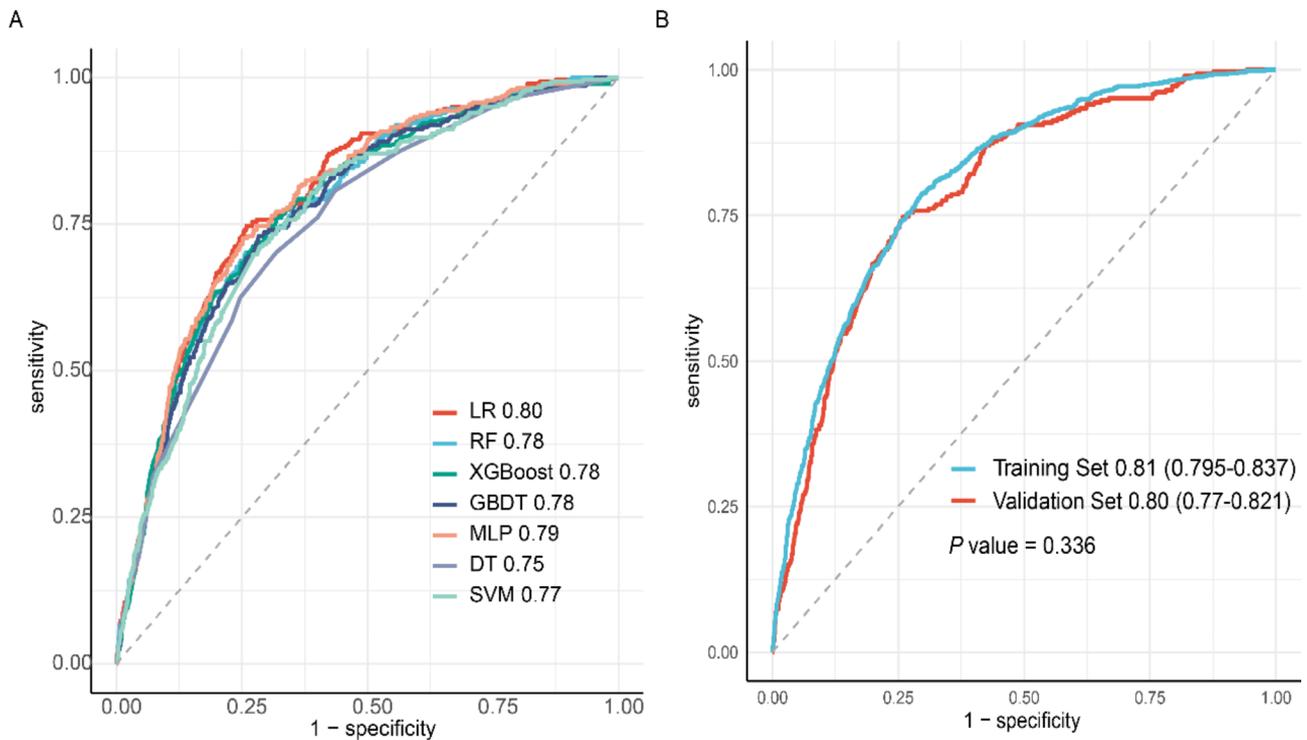


**Fig. 3** The correlation heat map of variables in the final model

and proceeded with its visualization. A web-based version of the nomogram (Fig. 6) is available at <https://liujiel23.shinyapps.io/dynnomapp/>. The proportion of patients stratified by the nomogram was similar in both cohorts, with 71.84%, 19.92%, and 8.24% of patients in the low-, medium-, and high-risk groups (development cohort) and 72.61%, 19.94%, and 7.45% (validation cohort), respectively (Fig. 7). Compared to the low-risk group, the incidence of POI was 4.62-fold and 12.98-fold higher in the medium-risk and high-risk groups of the development cohort, respectively (Table 2).

## Discussion

In this study on POI in older adults undergoing major surgery, we compared the predictive role of multiple models. The results showed that compared with the ML algorithms, the classical LR algorithm had a better prediction effect on POI after major surgery in older patients. We also developed a visual nomogram and performed risk stratification using the LR algorithm. Based on the preoperative and intraoperative variables, the model could help physicians assess the risk of POI earlier and formulate relevant strategies. Our study included



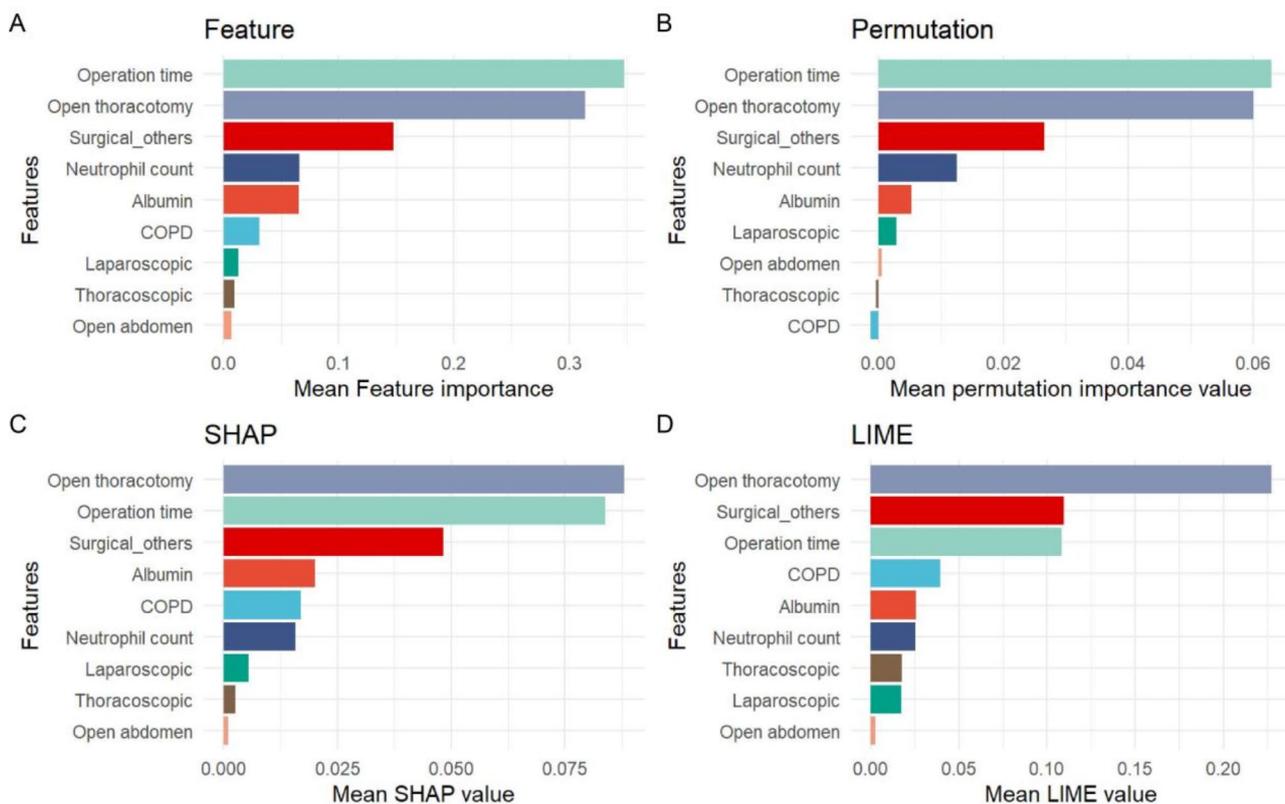
**Fig. 4** ROC curves of different algorithms. **(A)** ROC curves of seven models on the validation set. **(B)** ROC curves of the LR model on the training and validation sets. ROC, receiver operating characteristic; LR, logistic regression; DT, decision tree; RF, random forest; GBDT, gradient boosting decision tree; XGBoost, extreme gradient boosting; MLP, multilayer perceptron; SVM, support vector machine

a large sample size and identified simple predictors. The incidence of POI was 10.03%, consistent with previous literature [31, 32].

Consistent with previous studies, the MLP and LR algorithms performed better [23]. The AUC of the above two models reached more than 0.79, accompanied by high accuracy and specificity, reflecting strong discriminatory ability. Despite progress in ML interpretability, challenges remain due to differing perspectives on explanations. The consistency and variability among the four interpretability methods provided a comprehensive understanding of feature importance. The consistent recognition of features such as operation time, open thoracotomy, albumin level, and neutrophil count underscored their significance in the decision-making process, offering guidance for further analysis and interpretation. However, the variability in feature rankings across methods also reflects challenges in ML development. As ML models grow more complex, explaining their internal mechanisms becomes increasingly difficult [33]. The differences in feature importance rankings stem from the underlying principles of each method. SHAP values are based on game theory, providing a comprehensive explanation of feature contributions in the context of interactions between features [34]. Permutation Importance evaluates feature importance by measuring performance changes when a feature is perturbed, focusing on

individual contributions without considering interactions [35]. The Feature Importance method, based on impurity reduction in tree-based models, may overestimate the importance of high-cardinality features. LIME offers local explanations by approximating the model's behavior around a specific instance but may not fully capture global feature contributions across the entire dataset [36]. These varying approaches can lead to discrepancies in feature importance rankings, reflecting their distinct mechanisms for calculating feature contributions. The interactions between features can lead to different ranking results across methods. We chose LR as the primary model due to its intuitive interpretation of regression coefficients and lower risk of overfitting compared to complex models. Its simplicity and transparency make it suitable for clinical practice, where clear explanations are essential.

In our study, COPD, neutrophil count, albumin level, surgical incision, and operation time were independent risk factors associated with POI. Most studies have confirmed that COPD, operation time, and surgical incision are independently associated with POI [16, 17]. Serum albumin levels reflect the patient's nutritional status and are considered negative for acute-phase protein, which is closely associated with increased postoperative complications and mortality [37]. Low albumin levels may impair immune function and increase infection risk.



**Fig. 5** Global feature importance of the RF model using four interpretability methods, ranked in descending order. **(A)** Ranking of feature importance indicated by Feature Importance plot. The matrix plot depicts displays the mean importance of each feature based on the reduction in impurity or information gain. **(B)** Ranking of feature importance indicated by Permutation Importance plot. The plot showing the mean decrease in model performance when each feature is randomly permuted. **(C)** Ranking of feature importance indicated by SHAP summary plot. The mean SHAP values for each feature represent their contribution to the model's output. **(D)** Ranking of feature importance indicated by LIME plot. The mean LIME values approximate the model's behavior locally using an interpretable surrogate model. RF, random forest; SHAP, Shapley additive explanation; LIME: local interpretable model-agnostic explanations

Moreover, decreased albumin levels may contribute to increased capillary permeability into the extravascular space, causing pleural effusion and pulmonary edema, and increasing the risk of POI [38]. Neutrophils reflect the inflammatory state during disease progression. Neutrophil recruitment and activation are vital for immune defense during pathogenic infections [39].

Some preventive measures can be taken for controllable risk factors in the final model, such as preoperative use of bronchodilators combined with glucocorticoids to treat COPD [40], increasing albumin levels, and improving malnutrition [41]. The LR-based nomogram may aid in risk assessment, stratification, and identifying high-risk patients.

In our study, some factors reported in other studies were not identified as independent risk factors, such as RBC transfusion and urea nitrogen levels [42, 43]. This reflects the great variability of the elderly with major surgery from other study populations, and demonstrates the importance of constructing a POI model for the elderly undergoing major surgery. Additionally, these studies

mainly used the LR algorithm and did not explore the predictive performance of the various ML algorithms.

Compared with previous studies, our study had the following strengths: first, the research population was an elderly group with a high incidence; second, the type of surgery was major surgery, better targeting high-risk groups; third, in addition to using traditional algorithms, we explored model performance using popular ML algorithms; fourth, we applied interpretability techniques to ML algorithms; fifthly, based on the nomogram and predicted probabilities, we stratified the population for better use in clinical practice. In future work, integrating our model into electronic health records or clinical decision support systems could provide clinicians with real-time predictions during preoperative assessments, thereby aiding personalized interventions and optimizing perioperative management.

Our study had some limitations. First, as a single-center retrospective study, it may have underestimated the incidence of POI. However, the large sample size and incidence rate align with literature reports. Second, the



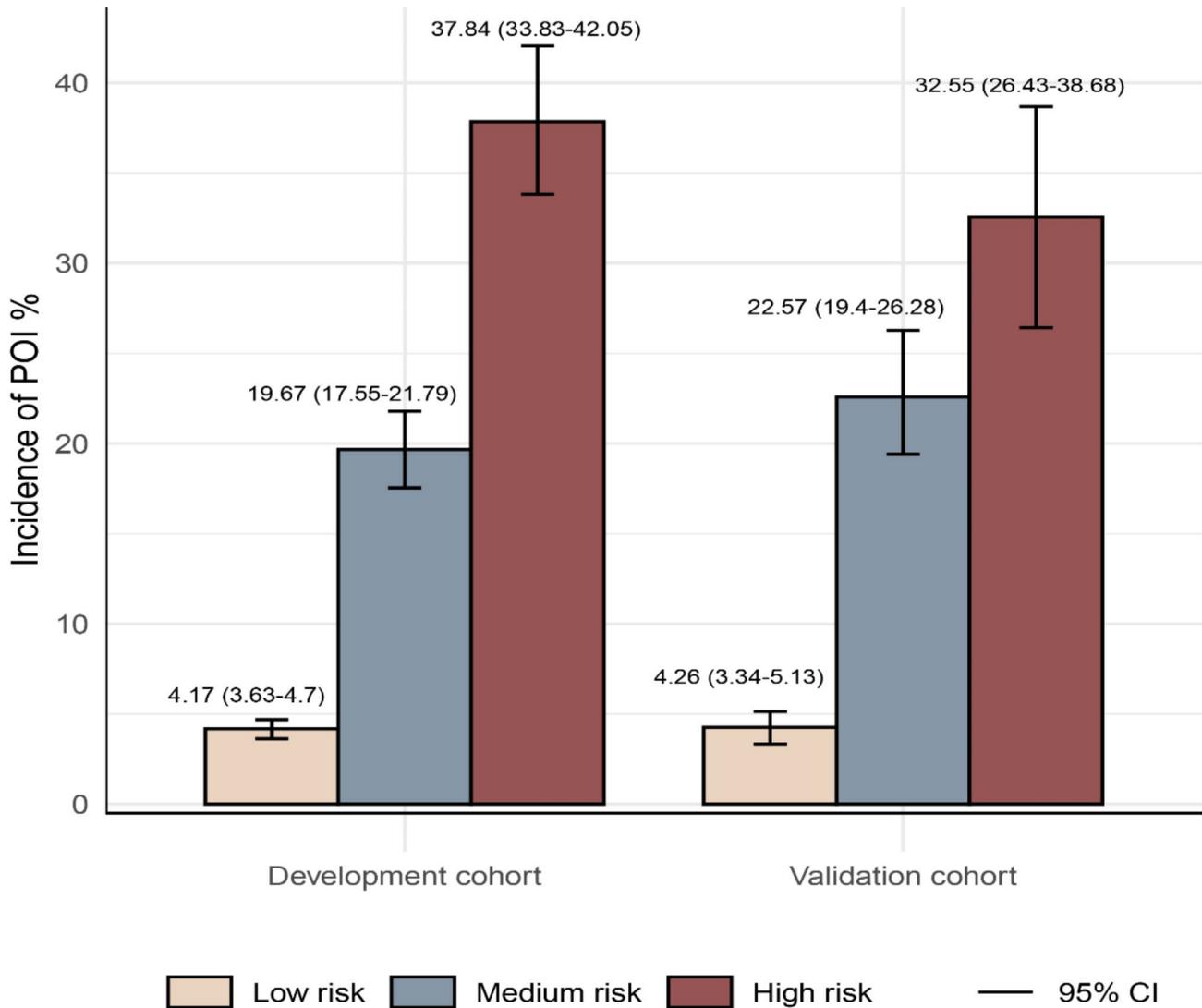
**Fig. 6** The nomogram based on the logistic regression algorithm. **(A)** Building a nomogram in the training cohort by incorporating five parametric parameters; **(B)** Online web server interface for the nomogram

model lacks external validation; therefore, further external validation across multiple centers is needed in the future to assess its generalizability. Third, some parameters related to mechanical ventilation were not collected. However, our model had good discriminative ability. Fourth, the use of LASSO regression for feature selection presents certain limitations. LASSO tends to select a small number of features and, when features are highly correlated, may randomly select one, overlooking others. This could lead to a reduction in model performance, as it may fail to capture all potential variable interactions. Fifth, although the LR model performed well in overall predictions, ML models may exhibit superior predictive power in cases with more features and more complex relationships. Future research will incorporate additional features to further explore the performance of various algorithms.

## Conclusions

In our study, the incidence of POI after major surgery in the elderly patients was 10.03%, which was significantly associated with the length of postoperative hospitalization. This study identified five simple and easy-to-collect

independent risk factors for POI in elderly patients undergoing major surgery and compared the effectiveness of different algorithms for predicting POI. Furthermore, we constructed a visualized nomogram of POI and performed risk stratification to guide the perioperative clinical management of elderly patients undergoing major surgery.



**Fig. 7** The incidence of POI by risk group in the development cohort and validation cohort. POI, postoperative pulmonary infection

**Table 2** Odds for POI by risk group in the development cohort and validation cohort

Group	Development cohort (n = 6637)		Validation cohort (n = 2844)	
	OR (95%CI)	P value	OR (95%CI)	P value
Risk group				
Low risk	Reference		Reference	
Medium risk	5.62 (4.62, 6.84)	< 0.001	6.55 (4.9, 8.76)	< 0.001
High risk	13.98 (11.1, 17.48)	< 0.001	10.84 (7.58, 15.51)	< 0.001

POI, postoperative pulmonary infection; OR, odds ratio; 95%CI, 95% confidence interval

**Abbreviations**

POI	Postoperative pulmonary infection
ML	Machine learning
LR	Logistic regression
AUC	Area under the receiver operating characteristic curve
DT	Decision tree
RF	Random forest
SVM	Support vector machine
GBDT	Gradient boosting decision tree
XGBoost	Extreme gradient boosting

MLP	Multilayer perceptron
BMI	Body mass index
COPD	Chronic obstructive pulmonary disease
BUN	Blood urea nitrogen
TBIL	Total bilirubin
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
RBC	Red blood cell
LASSO	Least absolute shrinkage and selection operator
VIF	Variance inflation factor
RUS	Random undersampling
GridSearchCV	Grid search with cross-validation
LIME	Local interpretable model-agnostic explanations
SHAP	Shapley additive explanation
IQR	Interquartile range

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12890-025-03582-4>.

Supplementary Material 1

## Acknowledgements

We would like to thank all the participants.

## Author contributions

JL and XL designed and analyzed the study. YW, ZX, YH, LC and YL conducted the data collection. YB, YF, GL and WX wrote the manuscript. QW edited the manuscript. All authors reviewed and approved the final version of the manuscript.

## Funding

This study was supported by the National Key Research and Development Program of China (Grant No. 2018YFC2001903), National Natural Science Foundation of China (Grant No. 81873952), and National Natural Science Foundation of China (Grant No. 81901948).

## Data availability

The raw datasets used in the current study can be obtained from the corresponding author upon reasonable request.

## Declarations

### Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Wuhan Union Hospital (Approval No.2021 – 0986). To protect patient privacy, all personally identifiable information was anonymized. Due to the lack of identifiable information and retrospective nature of the study, informed consent was obtained.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Department of Anesthesiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China

<sup>2</sup>Department of Anesthesiology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China

Received: 20 June 2024 / Accepted: 5 March 2025

Published online: 19 March 2025

## References

- Papazian L, Klompas M, Luyt CE. Ventilator-associated pneumonia in adults: a narrative review. *Intensive Care Med.* 2020;46(5):888–906.
- Hui S, Fowler AJ, Cashmore RMJ, Fisher TJ, Schlautmann J, Body S, Lan-Pak-Kee V, Webb M, Kyriakides M, Ng JY, et al. Routine postoperative noninvasive respiratory support and pneumonia after elective surgery: a systematic review and meta-analysis of randomised trials. *Br J Anaesth.* 2022;128(2):363–74.
- Stotts C, Corrales-Medina VF, Rayner KJ. Pneumonia-Induced inflammation, resolution and cardiovascular disease: causes, consequences and clinical opportunities. *Circ Res.* 2023;132(6):751–74.
- Liang S, Zhang X, Hu Y, Yang J, Li K. Association between perioperative chlorhexidine oral care and postoperative pneumonia in non-cardiac surgical patients: A systematic review and meta-analysis. *Surgery.* 2021;170(5):1418–31.
- Redelmeier DA, McAlister FA, Kandel CE, Lu H, Daneman N. Postoperative pneumonia in elderly patients receiving acid suppressants: a retrospective cohort analysis. *BMJ.* 2010;340:c2608.
- Henig O, Kaye KS. Bacterial pneumonia in older adults. *Infect Dis Clin North Am.* 2017;31(4):689–713.
- Jung J, Moon SM, Jang HC, Kang CI, Jun JB, Cho YK, Kang SJ, Seo BJ, Kim YJ, Park SB, et al. Incidence and risk factors of postoperative pneumonia following cancer surgery in adult patients with selected solid cancer: results of Cancer POP study. *Cancer Med.* 2018;7(1):261–9.
- Naito H, Nojima T, Fujisaki N, Tsukahara K, Yamamoto H, Yamada T, Aokage T, Yumoto T, Osako T, Nakao A. Therapeutic strategies for ischemia reperfusion injury in emergency medicine. *Acute Med Surg.* 2020;7(1):e501.
- Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature.* 2018;561(7721):45–56.
- Becher RD, Vander Wyk B, Leo-Summers L, Desai MM, Gill TM. The incidence and cumulative risk of major surgery in older persons in the United States. *Ann Surg.* 2023;277(1):87–92.
- Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-Leitz T, Fu R, Azad T, Chao TE, Berry WR, et al. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *Lancet.* 2015;385(Suppl 2):S11.
- Mungan I, Dicle ÇB, Bektaş Ş, Sari S, Yamanyar S, Çavuş M, Turan S, Bostancı EB. Does the preoperative platelet-to-lymphocyte ratio and neutrophil-to-lymphocyte ratio predict morbidity after gastrectomy for gastric cancer? *Mil Med Res.* 2020;7(1):9.
- Shi H, Sang H, Zhang Z, Chen B, Li L, Liu F, Xia W, Zhou Y, Liu K, Li X, et al. Neutrophil-to-Lymphocyte ratio is associated with clinical outcomes in patients treated with mechanical thrombectomy for posterior circulation large vessel occlusion. *World Neurosurg.* 2024;193:628–35.
- Herrero-Cervera A, Soehnlein O, Kenne E. Neutrophils in chronic inflammatory diseases. *Cell Mol Immunol.* 2022;19(2):177–91.
- Ahn J, Chang JS, Kim JW. Postoperative pneumonia and aspiration pneumonia following elderly hip fractures. *J Nutr Health Aging.* 2022;26(7):732–8.
- Gao YC, Zhang YW, Shi L, Gao W, Li YJ, Chen H, Rui YF. What are risk factors of postoperative pneumonia in geriatric individuals after hip fracture surgery: A systematic review and Meta-Analysis. *Orthop Surg.* 2023;15(1):38–52.
- Russotto V, Sabaté S, Canet J. Development of a prediction model for postoperative pneumonia: A multicentre prospective observational study. *Eur J Anaesthesiol.* 2019;36(2):93–104.
- Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ.* 2023;23(1):689.
- Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, Balicer RD, Feldman B, Wiznitzer A, Segal E. Prediction of gestational diabetes based on nationwide electronic health records. *Nat Med.* 2020;26(1):71–6.
- Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, Kannampallil T, Abraham J. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open.* 2021;4(3):e212240.
- Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, Chihara H, Fukumitsu R, Okawa M, Yamana N, et al. Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning. *Stroke.* 2019;50(9):2379–88.
- Shamout F, Zhu T, Clifton DA. Machine learning for clinical outcome prediction. *IEEE Rev Biomed Eng.* 2021;14:116–26.
- Nusinovici S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, Wong TY, Cheng CY. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol.* 2020;122:56–69.
- Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *Int J Med Inf.* 2021;151:104484.
- Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12–22.
- Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA.* 2017;318(6):517–8.
- Schwarze ML, Barnato AE, Rathouz PJ, Zhao Q, Neuman HB, Winslow ER, Kennedy GD, Hu YY, Dogdion CM, Kwok AC, et al. Development of a list of high-risk operations for patients 65 years and older. *JAMA Surg.* 2015;150(4):325–31.
- Gill TM, Murphy TE, Gahbauer EA, Leo-Summers L, Becher RD. Geriatric vulnerability and the burden of disability after major surgery. *J Am Geriatr Soc.* 2022;70(5):1471–80.
- Bluth T, Serpa Neto A, Schultz MJ, Pelosi P, Gama de Abreu M, Bluth T, Bobek I, Canet JC, Cinnella G, de Baerdemaeker L, et al. Effect of intraoperative high positive End-Expiratory pressure (PEEP) with recruitment maneuvers vs low PEEP on postoperative pulmonary complications in obese patients: A randomized clinical trial. *JAMA.* 2019;321(23):2292–305.

30. Schuijt MTU, Hol L, Nijbroek SG, Ahuja S, van Meenen D, Mazzinari G, Hemmes S, Bluth T, Ball L, Gama-de Abreu M, et al. Associations of dynamic driving pressure and mechanical power with postoperative pulmonary complications-posthoc analysis of two randomised clinical trials in open abdominal surgery. *EClinicalMedicine*. 2022;47:101397.
31. Gupta H, Gupta PK, Schuller D, Fang X, Miller WJ, Modrykamien A, Wichman TO, Morrow LE. Development and validation of a risk calculator for predicting postoperative pneumonia. *Mayo Clin Proc*. 2013;88(11):1241–9.
32. El Solh AA, Bhora M, Pineda L, Dhillon R. Nosocomial pneumonia in elderly patients following cardiac surgery. *Respir Med*. 2006;100(4):729–36.
33. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, Scardapane S, Spinelli I, Mahmud M, Hussain A. Interpreting Black-Box models: A review on explainable artificial intelligence. *Cogn Comput*. 2024;16(1):45–74.
34. Li X, Zhao Y, Zhang D, Kuang L, Huang H, Chen W, Fu X, Wu Y, Li T, Zhang J, et al. Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: findings of the US NHANES from 2003 to 2018. *Chemosphere*. 2023;311(Pt 1):137039.
35. Li W, Huang G, Tang N, Lu P, Jiang L, Lv J, Qin Y, Lin Y, Xu F, Lei D. Effects of heavy metal exposure on hypertension: A machine learning modeling approach. *Chemosphere*. 2023;337:139435.
36. Ribeiro MT, Singh S, Guestrin C. Why Should I Trust You? Explaining the Predictions of Any Classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA*. Association for Computing Machinery 2016: 1135–1144.
37. Eckart A, Struja T, Kutz A, Baumgartner A, Baumgartner T, Zurfluh S, Neeser O, Huber A, Stanga Z, Mueller B, et al. Relationship of nutritional status, inflammation, and serum albumin levels during acute illness: A prospective study. *Am J Med*. 2020;133(6):713–e722717.
38. Hübner M, Mantziari S, Demartines N, Pralong F, Coti-Bertrand P, Schäfer M. Postoperative Albumin Drop Is a Marker for Surgical Stress and a Predictor for Clinical Outcome: A Pilot Study. *Gastroenterol Res Pract* 2016, 2016:8743187.
39. Allie SR, Randall TD. Pulmonary immunity to viruses. *Clin Sci (Lond)*. 2017;131(14):1737–62.
40. Agusti A, Fabbri LM, Singh D, Vestbo J, Celli B, Franssen FME, Rabe KF, Papi A. Inhaled corticosteroids in COPD: friend or foe? *Eur Respir J* 2018, 52(6).
41. Soeters PB, Wolfe RR, Shenkin A. Hypoalbuminemia: pathogenesis and clinical significance. *JPEN J Parenter Enter Nutr*. 2019;43(2):181–93.
42. Russotto V, Sabaté S, Canet J, Network, ftPgtESoACT. Development of a prediction model for postoperative pneumonia: A multicentre prospective observational study. *Eur J Anaesthesiol| EJA*. 2019;36(2):93–104.
43. Arozullah AM, Khuri SF, Henderson WG, Daley J. Development and validation of a multifactorial risk index for predicting postoperative pneumonia after major noncardiac surgery. *Ann Intern Med*. 2001;135(10):847–57.

### Publisher's note

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