Am J Transl Res 2025;17(4):2850-2871
www.ajtr.org /ISSN:1943-8141/AJTR0164000

Original Article

Risk factor identification and prediction of pleural effusion following coronary artery bypass grafting

Caiyun Lu[1,2], Fan Jiang[3], Ling Pan[2], Jingjing Lin[4], Yuanshu Peng[5], Huanzhong Shi[1]

[1]Department of Respiratory and Critical Care Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Chaoyang, Beijing 100020, China; [2]Department of Respiratory and Critical Care Medicine, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning 530011, Guangxi, China; [3]Department of Rehabilitation Medicine, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China; [4]Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China; [5]Department of Chest Surgery, Beijing Chao-Yang Hospital, Capital Medical University, Chaoyang, Beijing 100020, China

Received February 18, 2025; Accepted March 24, 2025; Epub April 15, 2025; Published April 30, 2025

Abstract: Objective: To evaluate the incidence of pleural effusion (PE) following coronary artery bypass grafting (CABG), identify associated risk factors, and develop a validated predictive model for early detection. Methods: A retrospective cohort of 1,979 patients who underwent CABG at Beijing Chaoyang Hospital (Capital Medical University) was randomly divided into training (70%) and validation (30%) sets. Risk factors for PE were identified through univariate analysis, LASSO regression, and multivariate logistic regression. Five machine learning models-nomogram, back-propagation neural network (BPNN), random forest, gradient boosting, and support vector machine-were developed. External validation was performed using data from 289 patients at the First Affiliated Hospital of Guangxi Medical University. Results: PE occurred in 71.0% of patients (1,405/1,979) within 3 days postoperatively. Independent risk factors included body mass index (BMI), carotid artery stenosis, postoperative pneumonia, duration of mechanical ventilation, intraoperative blood loss, operative time, and ejection fraction. Among the models, the BPNN demonstrated the best performance, with area under the curve (AUC) values of 0.828 in the training set and 0.751 in the internal validation set. The AUC for external validation was 0.737, outperforming the other models across all evaluation metrics. Conclusions: This study developed a predictive model for post-CABG pleural effusion with high discriminatory power, providing a useful tool for early risk stratification in clinical settings.

Keywords: Coronary artery bypass grafting, pleural effusion, prediction model, neural network

Introduction

The coronary arteries provide the primary blood supply to the heart. When these vessels become narrowed or occluded, myocardial cells receive insufficient blood and oxygen, resulting in myocardial ischemia and hypoxia. This process is a key driver of coronary artery disease (CAD), which includes coronary heart disease (CHD) and ischemic heart disease (IHD) [1]. Globally, cardiovascular diseases account for over 30% of annual deaths, affecting approximately 17 million people. CAD represents around 40% of these cases, underscoring its significant impact on global health [2].

Coronary artery bypass grafting (CABG) is a highly effective surgical intervention for patients with severe three-vessel disease or critical stenosis of the left main coronary artery [3, 4]. However, postoperative pleural effusion (PE) remains a common complication [5-7]. PE not only prolongs hospital stays but also increases healthcare costs, imposing a substantial burden on patients and healthcare systems [8]. Moreover, it is a leading cause of readmission within one month following CABG [9-11].

Recent model-based studies have achieved promising results in predicting complications after CABG, such as septic shock, thrombocyto-

https://doi.org/10.62347/KGKL5899

penia, liver dysfunction, and new-onset atrial fibrillation [12-14]. However, these studies exhibit several limitations. Many involve relatively small sample sizes, and some focus on a narrow set of clinical indicators, failing to comprehensively assess the factors contributing to complications. Notably, there is a significant lack of research specifically addressing postoperative PE. The few existing studies on PE offer limited and superficial analyses of associated risk factors, and effective predictive models for identifying high-risk patients are lacking.

To address this gap, our study conducted a comprehensive analysis of data collected from CABG patients at a medical center over the past decade. We aimed to identify risk factors associated with the development of postoperative PE and applied machine learning (ML) techniques to develop multiple predictive models, which were then successfully validated in clinical settings.

Materials and methods

Patient selection

We conducted a retrospective analysis of patient data from Beijing Chaoyang Hospital, Capital Medical University. The study included patients who underwent CABG between January 2010 and October 2020. The inclusion criteria were as follows: Patients aged 18 years or older, diagnosed with CHD and those who underwent CABG. Patients were excluded if they met any of the following criteria: Pre-existing PE prior to CABG, history of thoracic or cardiac surgery, other conditions potentially leading to PE (e.g., chylothorax, hemothorax, malignancies), liver or kidney dysfunction, emergency or secondary surgeries, and incomplete data.

To externally validate the model, we collected clinical data from patients who underwent CABG at the First Affiliated Hospital of Guangxi Medical University between 2010 and 2020, applying the same inclusion and exclusion criteria. This study was approved by the Ethics Committee of Beijing Chaoyang Hospital Affiliated to Capital Medical University (No: 2018-Ke-321).

Data extraction

Data were extracted from the medical records of eligible patients. The extracted data included variables relevant to postoperative PE prediction. The comprehensive dataset consisted of the following aspects:

Demographic information: gender, age, height, weight, body mass index (BMI), smoking index, and alcohol consumption.

PE-related indicators: presence or absence of PE, classification of effusion volume, location (unilateral/bilateral), and the first occurrence time of PE.

Disease assessment data: past medical history (hypertension, stroke, diabetes, chronic obstructive pulmonary disease [COPD], hypoalbuminemia, prior percutaneous coronary intervention [PCI], carotid artery stenosis [CAS], unstable angina, myocardial infarction), number of diseased coronary arteries, and ejection fraction (EF).

Surgery-related indicators: operative time, intraoperative blood loss, number of coronary bypass grafts, postoperative complications (e.g., pneumonia, pleural effusion), diagnostic methods for PE, and postoperative laboratory examination indicators (creatine kinase-myocardial band [CK-MB], albumin [ALB], total cholesterol [TC], triglycerides [TG], high-density lipoprotein cholesterol [HDL-C], low-density lipoprotein cholesterol [LDL-C], aspartate transaminase [AST], alanine transaminase [ALT], creatine kinase [CK], lactate dehydrogenase [LDH], gamma-glutamyl transpeptidase [GGT], blood urea nitrogen [BUN], creatinine [Cre], cardiac troponin I [cTnI], white blood cells [WBC], red blood cells [RBC], neutrophils [NEUT], lymphocytes [LYM], neutrophil/lymphocyte ratio [NLR], eosinophils [EOS], hemoglobin [HB], platelets [PLT], and platelets/lymphocyte ratio [PLR]).

Prognostic indicators: length of hospital stay, duration of mechanical ventilation, and time spent in the ICU.

Chest X-ray results obtained three days post-CABG were used to classify patients into two groups: those with PE and those without PE.

The PE group was further stratified into three subgroups based on the volume of pleural effusion: (1) Small Volume Group: pleural fluid volume < 500 mL (characterized by a blunted costophrenic angle). (2) Medium Volume Group: pleural fluid volume 500-1,000 mL (pleural fluid plane near the lung hilum). (3) Large Volume Group: pleural fluid volume > 1,000 mL (pleural effusion plane exceeding the lung hilum) [15].

Outcome measures

The primary objective of this study was to develop and validate a predictive model for postoperative PE following CABG. The predictive performance of the models was evaluated using various metrics, including the area under the receiver operating characteristic (ROC) curve, calibration plots, and decision curve analysis.

Statistical methods

Statistical analysis was performed using SPSS version 27. Descriptive statistics were used to summarize the characteristics of the data. To enhance the clinical interpretation of continuous variables, optimal cut-off values were used to categorize them, with results presented as counts and percentages. Group comparisons were conducted using either the chi-square test or Fisher's exact test, depending on the distribution of the data.

Univariate logistic regression analysis was initially performed to assess the impact of individual variables on the occurrence of postoperative PE. The least absolute shrinkage and selection operator (LASSO) method was then used to identify variables potentially influencing the outcome [16]. Finally, multivariate logistic regression analysis was performed to precisely identify variables for inclusion in the predictive model. A statistical significance level was set at a *p*-value of less than 0.05.

For model construction, the "caret" package in R software (version 4.1.3) was used, with random seeds set to 2023. The enrolled patients were divided into training (70%) and validation (30%) sets. The predictive model was developed using data from the training set. Additionally, 289 patients from the First Affiliated Hospital of Guangxi Medical University were included for external validation of the model. During model development, the "rms", "neuralnet", "randomForest", "gbm", and "e1071" packages in R were used to sequentially develop a nomogram, a back-propagation neural network (BPNN), a random forest (RF), a gradient boosting model (GBM), and a support vector machine (SVM).

Results

General characteristics of CABG patients

Data were collected from 2,398 patients who underwent CABG at Beijing Chaoyang Hospital, Capital Medical University. After excluding 419 participants who did not meet the inclusion criteria, 1,979 patients were ultimately included in the analysis (**Figure 1**). Of these, 1,487 (75.1%) were male and 492 (24.9%) were female (**Figure 2A**). The mean age of the cohort was 63.33 ± 8.80 years, with the highest proportion (42.0%) in the 61 to 70 years age group. A small percentage (0.8%) were aged 30 to 40 years, while 1.3% were over 81 years old (**Figure 2B**).

Among the 1,979 patients, 574 (29%) did not develop postoperative PE within 3 days, while 1,405 (71%) exhibited this complication. The majority (59.9%) presented with left-sided pleural effusion, with 93% showing mild effusion and only 4% having significant volume accumulation (**Figure 2C**). For model development, a random 70% of the dataset (n = 1,385) was allocated to the training set, with the remaining 30% (n = 594) designated for validation. The training cohort consisted of 1,385 patients, including 981 with postoperative PE and 404 controls without PE. These patients were stratified into the PE group and the without PE group.

Comparison of baseline characteristics and serum biochemical markers

A comparative analysis of baseline characteristics and serum biochemical markers was performed between the PE and without PE groups. The baseline demographics and clinical parameters are presented in **Table 1**. Significant differences were identified in multiple clinical parameters, including gender, BMI, hypertension, stroke, CAS, EF, duration in the ICU, mechanical ventilation time, surgical blood loss, and operation time (all P < 0.05). The anal-

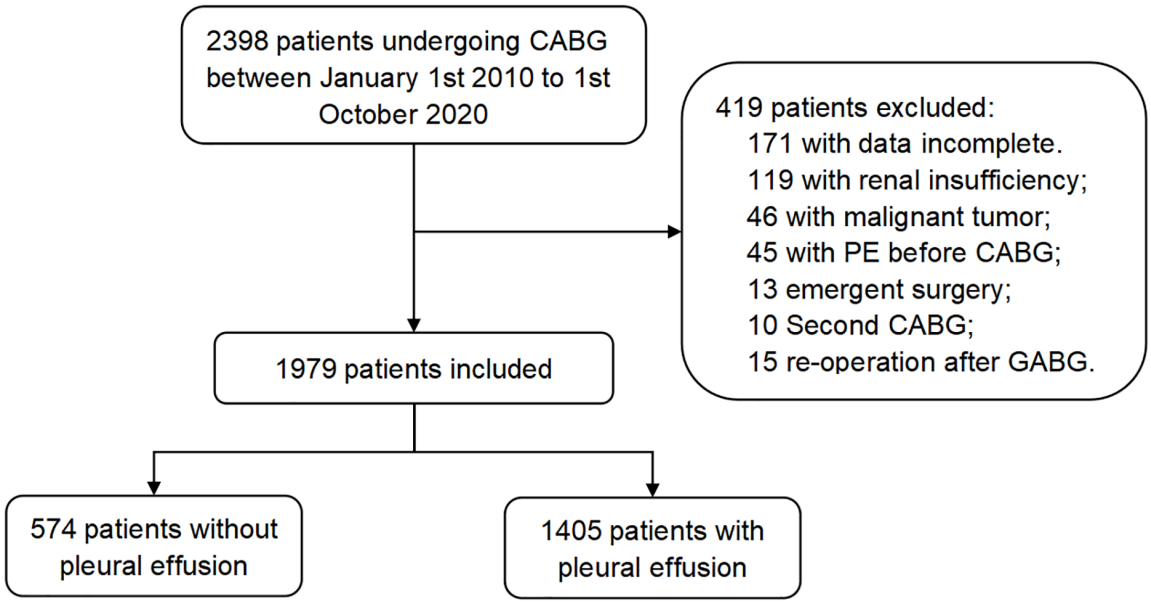


Figure 1. Flow chart of patient inclusion. Abbreviations: PE, pleural effusion; CABG, Coronary artery bypass grafting.

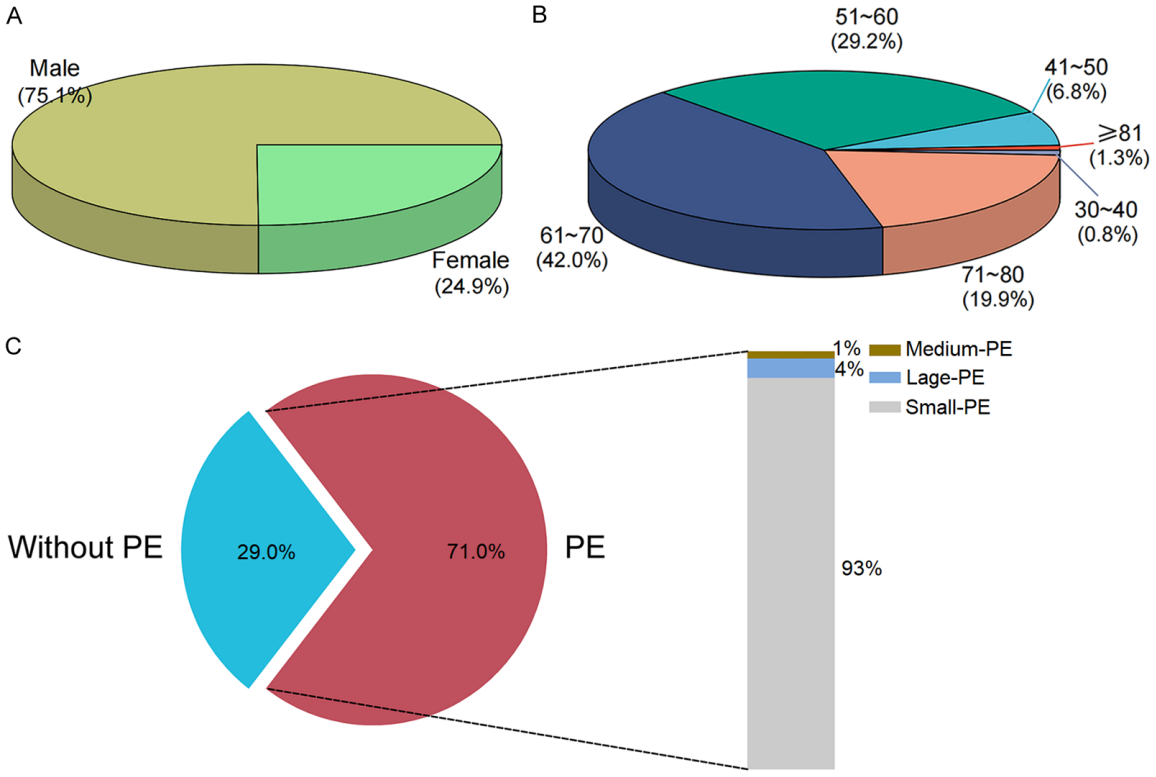


Figure 2. Basic information of patients. A. Gender; B. Age; C. Distribution map of the proportion of 1979 patients in each group. Abbreviation: PE, pleural effusion.

ysis also revealed marked differences in serum levels of HDL-C, AST, ALT, CK, LDH, cTnI, NEUT, NLR, and EOS between the two groups (all $P < 0.05$, **Table 2**).

Table 1. Comparison of baseline characteristics between the PE group and those without-PE

Characteristic	Without PE n = 404	With PE n = 981	χ^2	*P* value
Age, years			2.626	0.105
< 60	145 (35.9)	308 (31.4)		
≥ 60	259 (64.1)	673 (68.6)		
Gender			3.915	0.048
Male	322 (79.7)	733 (74.7)		
Female	82 (20.3)	248 (25.3)		
BMI, kg/m^2			10.831	< 0.001
< 26	260 (64.4)	537 (54.7)		
≥ 26	144 (35.6)	444 (45.3)		
Smoking index			2.982	0.084
< 290	189 (46.8)	509 (51.9)		
≥ 290	215 (53.2)	472 (48.1)		
Alcohol consumption			1.819	0.611
Never	166 (41.1)	422 (43.0)		
Sometimes	156 (38.6)	360 (36.7)		
Former	20 (5.0)	62 (6.3)		
Abuse	62 (15.3)	137 (14.0)		
Medical history				
Hypertension	263 (65.1)	694 (70.7)	4.271	0.039
Dyslipidemia	253 (62.6)	594 (60.6)	0.518	0.472
Diabetes	160 (39.6)	392 (40.0)	0.015	0.902
COPD	10 (2.5)	27 (2.8)	0.084	0.771
Stroke	56 (13.9)	180 (18.3)	4.076	0.043
Hypoproteinemia	194 (48.0)	506 (51.6)	1.451	0.228
Prior PCI	57 (14.1)	139 (14.2)	0.001	0.997
CAS	121 (30.0)	458 (46.7)	32.947	< 0.001
Unstable angina	194 (48.0)	485 (49.4)	0.231	0.631
Myocardial infarction	106 (26.2)	216 (22.0)	2.855	0.091
Postoperative pneumonia	115 (28.5)	340 (34.7)	4.975	0.026
No. of diseased coronary arteries			2.743	0.254
1-2	57 (14.1)	125 (12.7)		
3	318 (78.7)	805 (82.1)		
≥ 4	29 (7.2)	51 (5.2)		
No. of coronary bypasses			0.075	0.963
1-2	148 (36.6)	357 (36.4)		
3	199 (49.3)	480 (48.9)		
≥ 4	57 (14.1)	144 (14.7)		
EF, %			12.006	< 0.001
< 50	43 (10.6)	178803 (18.1)		
≥ 50	361 (89.4)	803 (81.9)		
Time in ICU, day			12.575	< 0.001
< 4	226 (55.9)	446 (45.5)		
≥ 4	178 (44.1)	535 (54.5)		
Use of respirator			1.368	0.713
NO	50 (12.4)	113 (11.5)		
Non-invasive	18 (4.5)	36 (3.7)		
Invasive	308 (76.2)	750 (76.5)		
Invasive and Non-invasive	28 (6.9)	82 (8.3)		

Mechanical ventilation time, day			22.344	< 0.001
< 26	374 (92.6)	812 (82.9)		
≥ 26	30 (7.4)	169 (17.2)		
Amount of bleeding, mL			18.285	< 0.001
< 475	284 (70.3)	569 (58.0)		
≥ 475	120 (29.7)	412 (42.0)		
Time of operation, hour			11.431	< 0.001
< 4	183 (45.3)	349 (35.6)		
≥ 4	221 (54.7)	632 (64.4)		

Abbreviations: PE, Pleural effusion; M, median; BMI, Body Mass Index; COPD, Chronic Obstructive Pulmonary Disease; PCI, Percutaneous Transluminal Coronary Intervention; CAS, Carotid artery stenosis; EF, Preoperative Left Ventricular Ejection Fraction; ICU, Intensive Care Unit. Data are presented as number (%).

Table 2. Comparison of laboratory test results between the PE group and patients without PE

Characteristic	Without PE n = 404	With PE n = 981	χ^2	*P* value
CKMB, U/L			0.907	0.341
< 5.7	221 (54.7)	564 (57.5)		
≥ 5.7	183 (45.3)	417 (42.5)		
ALB, g/L			3.306	0.069
< 20.9	96 (23.8)	280 (28.5)		
≥ 20.9	308 (76.2)	701 (71.5)		
TC, mmol/L			2.023	0.155
< 1.60	184 (45.5)	406 (41.4)		
≥ 1.60	220 (54.5)	575 (58.6)		
TG, mmol/L			0.298	0.585
< 0.59	130 (32.2)	301 (30.7)		
≥ 0.59	274 (67.8)	680 (69.3)		
HDL-C, mmol/L			5.630	0.018
< 0.75	282 (69.8)	745 (75.9)		
≥ 0.75	122 (30.2)	236 (24.1)		
LDL-C, mmol/L			0.924	0.336
< 0.45	99 (24.5)	217 (22.1)		
≥ 0.45	305 (75.5)	764 (77.9)		
AST, U/L			3.941	0.047
< 24	66 (16.3)	206 (21.0)		
≥ 24	338 (83.7)	775 (79.0)		
ALT, U/L			5.327	0.021
< 18	81 (20.0)	254 (25.9)		
≥ 18	323 (80.0)	727 (74.1)		
CK, U/L			4.604	0.032
< 218	56 (13.9)	183 (18.7)		
≥ 218	348 (86.1)	798 (81.3)		
LDH, U/L			5.767	0.016
< 320	368 (91.1)	848 (86.4)		
≥ 320	36 (8.9)	133 (13.6)		
GGT, U/L			1.741	0.187
< 34	279 (69.1)	712 (72.6)		
≥ 34	125 (30.9)	269 (27.4)		

Influencing factors of PE post-CABG

Relying solely on variables identified as statistically significant in multivariate analysis may overlook other crucial factors that could influence the overall interpretation of the data [17]. To address this, we employed the least absolute shrinkage and selection operator (LASSO) regression method due to its advantages in variable selection and model simplification. LASSO regression is particularly effective in high-dimensional datasets, where it can simultaneously perform variable selection and parameter estimation by shrinking less important coefficients to zero. This helps identify a subset of meaningful variables, enhancing model interpretability and reducing overfitting [18].

Using this approach, 20 potentially meaningful variables were identified for further analysis. Due to the coefficient of LDH being zero, this variable was excluded (**Table 3**), narrowing the analysis to 19 variables that could influence postoperative PE (**Figure 3**). These variables included gender, BMI, hypertension, stroke, CAS, postoperative pneumonia, duration in ICU, mechani-

BUN, mmol/L			3.342	0.068
< 5.37	263 (65.1)	587 (69.8)		
≥ 5.37	141 (34.9)	394 (40.2)		
Cre, μmol/L			3.703	0.054
< 103.6	306 (75.7)	693 (70.6)		
≥ 103.6	98 (24.3)	288 (29.4)		
cTnI, ng/L			6.558	0.010
< 24.75	387 (95.8)	902 (91.9)		
≥ 24.75	17 (4.2)	79 (8.1)		
WBC, $*10^9$/L			2.783	0.095
< 10	211 (52.2)	464 (47.3)		
≥ 10	193 (47.8)	517 (52.7)		
RBC, $*10^2$/L			1.503	0.220
< 4	372 (92.1)	921 (93.9)		
≥ 4	32 (7.9)	60 (6.1)		
NEUT, $*10^9$/L			5.696	0.017
< 9.14	251 (62.1)	541 (55.1)		
≥ 9.14	153 (37.9)	440 (44.9)		
LYM, $*10^9$/L			0.938	0.333
< 0.83	140 (34.7)	367 (37.4)		
≥ 0.83	264 (65.3)	614 (62.6)		
NLR			4.270	0.039
< 7.63	200 (49.5)	426 (43.4)		
≥ 7.63	204 (50.5)	555 (56.6)		
EOS, $*10^9$/L			5.194	0.023
< 0.04	143 (35.4)	412 (42.0)		
≥ 0.04	261 (64.6)	569 (58.0)		
HB, g/L			0.091	0.763
< 100	235 (58.2)	562 (57.3)		
≥ 100	169 (41.8)	419 (42.7)		
PLT, $*10^9$/L			2.476	0.116
< 160	254 (62.9)	660 (67.3)		
≥ 160	150 (37.1)	321 (32.7)		
PLR			0.134	0.714
< 150.84	237 (58.7)	565 (57.6)		
≥ 150.84	167 (41.3)	416 (42.4)		

Abbreviations: PE, Pleural effusion; CKMB, creatine kinase myocardial band; ALB, albumin; TC, total cholesterol; TG, triglyceride; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; AST, Aspartate Transaminase; ALT, alanine transaminase; CK, Creatine Kinase; LDH, Lactate Dehydrogenas; GGT, Gamma Glutamyl Transpeptidase; BUN, Blood Urea Nitrogen; Cre, Creatinine; cTnI, Cardiac troponin I; WBC, White Blood Cells; RBC, Red Blood Cells; NEUT, Neutrophil; LYM, Lymphocyte; NLR, Neutrophil/Lymphocyte; EOS, Eosinophil; HB, Hemoglobin; PLT, Platelets; PLR, Platelets/Lymphocyte. Data are presented as number (%).

cal ventilation time, surgical blood loss, operation duration, EF, HDL-C, AST, ALT, CK, cTnI, NEUT, NLR, and EOS.

To further investigate their influence, a multivariate logistic regression analysis was conducted on these 19 variables. The analysis revealed seven independent predictors of postoperative PE in CABG patients: BMI, CAS, postoperative pneumonia, mechanical ventilation duration, surgical blood loss, operation time, and EF (**Figure 4**).

Construction of prediction model for postoperative PE

Based on the seven selected variables, five models were constructed: a nomogram, BPNN, RF, gradient GBM, and SVM. The nomogram model (**Figure 5A**) incorporates seven predictive variables: BMI, CAS, EF, mechanical ventilation time, surgical bleeding volume, surgical duration, and postoperative pneumonia. The β (x - m) term line represents the product of the β coefficient, derived from linear regression using the least squares method, and the deviation of the independent variable x from its mean (m). The β coefficient represents the regression coefficient, x denotes the independent variable, and m signifies the mean value of that variable. The length of this line segment reflects the variable's contribution to the prediction of postoperative PE. The scores of each variable are calculated and summed to determine the total score for the patient. The nomogram model estimates the probability of postoperative PE following CABG by plotting a vertical line from the total score to the probability scale.

The RF model (**Figure 5B**) is an ensemble learning method that combines multiple decision trees to improve predictive performance [19]. Each tree in the forest is trained on a random subset of the data, and the final prediction is made by aggregating the results of all trees.

Table 3. LASSO regression analysis of the coefficients of each variable

Variables	Coefficients
Gender	-0.042
BMI	0.070
Hypertension	0.035
Stroke	0.030
CAS	0.141
Postoperative pneumonia	0.099
Time in ICU	0.035
Mechanical ventilation time	0.096
Amount of bleeding	0.063
Time of operation	0.050
EF	0.073
HDL-C	-0.038
AST	-0.019
ALT	-0.029
CK	-0.033
LDH	0
cTnI	0.041
NEUT	0.019
NLR	0.022
EOS	-0.013

Abbreviations: LASSO, least absolute shrinkage and selection operator; BMI, Body Mass Index; CAS, Carotid artery stenosis; ICU, Intensive Care Unit; EF, Preoperative Left Ventricular Ejection Fraction; HDL-C, high density lipoprotein cholesterol; AST, Aspartate Transaminase; ALT, alanine transaminase; CK, Creatine Kinase; LDH, Lactate Dehydrogenas; cTnI, Cardiac troponin I; NEUT, Neutrophil; NLR, Neutrophil/Lymphocyte; EOS, Eosinophil.

This approach reduces overfitting and enhances the model's ability to generalize to unseen data.

The BPNN model (**Figure 5C** and **5D**) consists of seven input variables and two output results. This artificial neural network employs a supervised learning algorithm to map input data to output predictions through multiple hidden layers [20]. Each input variable contributes to the network's ability to learn complex nonlinear relationships, while the two output results represent the predicted outcomes. The structure of the BPNN enables it to capture intricate patterns in the data, making it particularly effective for tasks requiring high-dimensional input data.

The GBM model (**Figure 5E**) is a powerful ensemble learning method that builds a strong predictive model by sequentially combining multiple weak prediction models, typically decision trees [21]. This approach is based on the principle of gradient descent, where each new tree is trained to correct the errors of the previous ensemble. Specifically, the GBM algorithm minimizes a loss function by iteratively adding trees, with each tree focusing on the residuals (i.e., the differences between predicted and actual values) of the previous iteration. This iterative process ensures that the model gradually improves its predictive accuracy. One of the key features of GBM is its ability to handle complex datasets with nonlinear relationships and interactions between variables. By combining the predictions of multiple trees, the GBM model can capture intricate patterns in the data that may not be apparent in individual models. Additionally, GBM provides feature importance scores, allowing researchers to identify the most influential variables in the dataset.

The SVM model (**Figure 5F**) utilizes a radial kernel function with 846 support vectors and a penalty factor of 1. The radial basis function kernel, commonly used in SVM, maps the input data into a higher-dimensional space to achieve better separation between classes [22]. The 846 support vectors represent the data points that lie closest to the decision boundary and play a crucial role in defining the hyperplane that separates the classes. The penalty factor, set to 1, controls the trade-off between achieving a low training error and maintaining a simple model. This configuration ensures that the SVM model achieves a balance between complexity and generalization, making it robust for classification tasks. Variable importance analysis for the RF, BPNN, and GBM demonstrates that CAS is the most influential variable, followed by the mechanical ventilation time.

Validation of the model

The training and validation set data were utilized to assess and compare the predictive performance of all models. In the training set, the sensitivity and specificity of the models developed to predict postoperative PE following CABG were as follows: Nomogram (0.624, 0.671), BPNN (0.691, 0.797), RF (0.780, 0.641), GBM (0.760, 0.690), and SVM (0.828, 0.616) (**Table 4**). Receiver operating characteristic

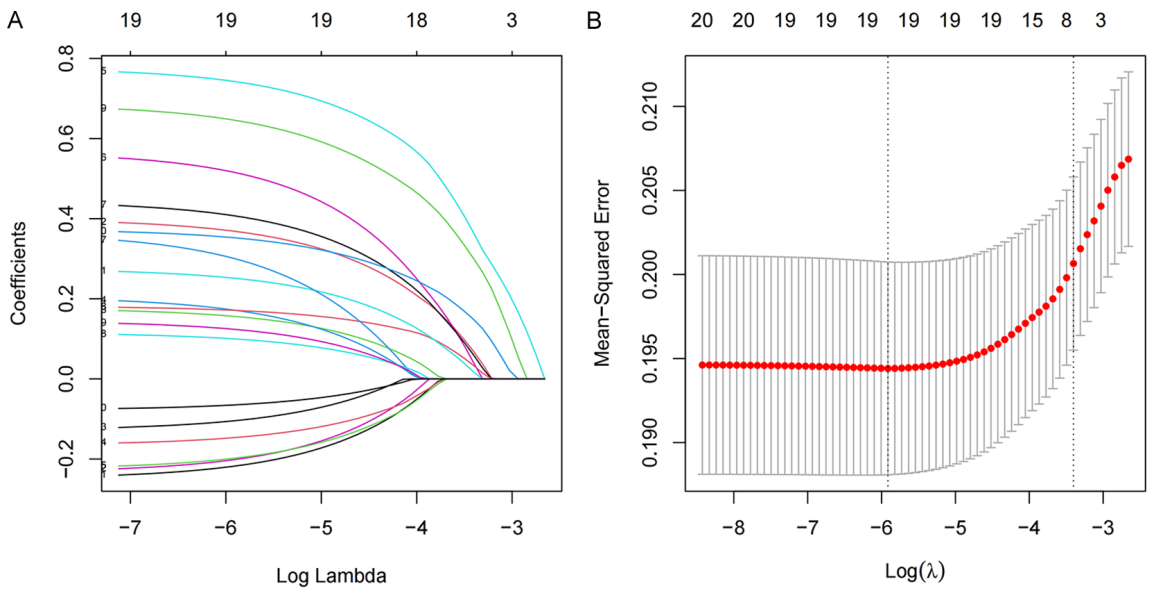


Figure 3. Use LASSO regression to screen potential variables. A. The coefficient curves of 20 overall survival characteristics in the LASSO model. B. The 10-fold cross-validation of the optimal parameter (λ) in the LASSO model was selected by the minimum standard. The relationship curve between the partial likelihood deviation curve and the logarithm (λ). Abbreviation: LASSO, least absolute shrinkage and selection operator.

(ROC) curves were constructed, with sensitivity plotted on the y-axis and (1 - specificity) on the x-axis, to evaluate the predictive accuracy of each model using the area under the curve (AUC). In the test set, only the BPNN achieved an AUC value exceeding 0.800, while the remaining models showed comparable classification capabilities with AUC values around 0.750 (**Figure 6**).

In the validation set, the sensitivity of the nomogram, BPNN, RF, GBM, and SVM for predicting PE were 0.531, 0.693, 0.818, 0.589, and 0.809, respectively. The corresponding specificities were 0.718, 0.729, 0.642, 0.783, and 0.553 (**Table 4**). The AUC values for each model's validation set were 0.667, 0.751, 0.762, 0.745, and 0.708, respectively. Using the calibration function in R software, the validation set was resampled via bootstrapping to obtain calibration curves for each model, effectively mitigating selection bias. Notably, the nomogram and calibration plot for the GBM closely resembled the ideal performance curve. Furthermore, the decision curves for all models surpassed the reference line, as depicted in **Figure 7**. No significant differences were observed between the training and validation sets in terms of data distribution (all $P > 0.05$, **Table 5**).

Validation of the clinical validity of the model

During the internal validation process, it was identified that the BPNN and GBM demonstrated superior predictive capabilities. To comprehensively evaluate their generalization capabilities, robustness, and credibility, data were collected from 289 patients who underwent CABG at the Department of Cardiovascular Surgery, the First Affiliated Hospital of Guangxi Medical University. In the external dataset, the AUC for the BPNN was 0.737 (95% CI: 0.655-0.818), while for the GBM, it was 0.710 (95% CI: 0.631-0.790). Based on Delong's test, no statistically significant difference was found in the AUC between the BPNN and GBM ($Z = 0.563$, $P = 0.573$). Additionally, calibration plots revealed minimal discrepancies in the predictive performance of the models. Nevertheless, the BPNN demonstrated clinical benefits across a broader threshold range (**Figure 8**). It is concluded that the BPNN outperforms other prediction models in all respects, making it the optimal choice for prediction.

Discussion

Pleural effusion following CABG significantly contributes to the rehospitalization rate of patients, as reported in [9-11]. This condition is

Variables	No. of patients	OR(95%CI)	*P* value
Gender			
Female	330	1	
Male	1055	0.779(0.571-1.063)	0.115
BMI (kg/m^2)			
<26	797	1	
≥26	588	1.491(1.154-1.927)	0.002
Hypertension			
No	428	1	
Yes	957	1.193(0.913-1.559)	0.196
Stroke			
No	1149	1	
Yes	236	1.228(0.866-1.741)	0.249
CAS			
No	806	1	
Yes	579	2.175(1.656-2.856)	<0.001
Postoperative pneumonia			
No	930	1	
Yes	455	1.763(1.335-2.329)	<0.001
Time in ICU (day)			
<4	672	1	
≥4	713	1.200(0.928-1.551)	0.165
Mechanical ventilation time (day)			
<26	1186	1	
≥26	199	1.985(1.284-3.070)	0.002
Amount of bleeding (mL)			
<475	853	1	
≥475	532	1.454(1.108-1.907)	0.007
Time of operation (h)			
<4	532	1	
≥4	853	1.317(1.017-1.705)	0.037
EF (%)			
≥50	1164	1	
<50	221	1.558(1.068-2.274)	0.021
HDL-C (mmol/L)			
<0.75	1027	1	
≥0.75	358	0.791(0.598-1.048)	0.102
AST (U/L)			
<24	272	1	
≥24	1113	0.880(0.619-1.250)	0.474
ALT (U/L)			
<18	335	1	
≥18	1050	0.847(0.616-1.165)	0.308
CK (U/L)			
<218	239	1	
≥218	1146	0.798(0.556-1.145)	0.220
cTnI (ng/L)			
<24.75	1289	1	
≥24.75	96	1.442(0.810-2.566)	0.213
NEUT (*10^9/L)			
<9.14	792	1	
≥9.14	593	1.122(0.858-1.468)	0.400
NLR			
<7.63	626	1	
≥7.63	759	1.156(0.873-1.531)	0.311
EOS (*10^9/L)			
<0.04	555	1	
≥0.04	830	0.925(0.698-1.226)	0.588

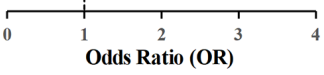
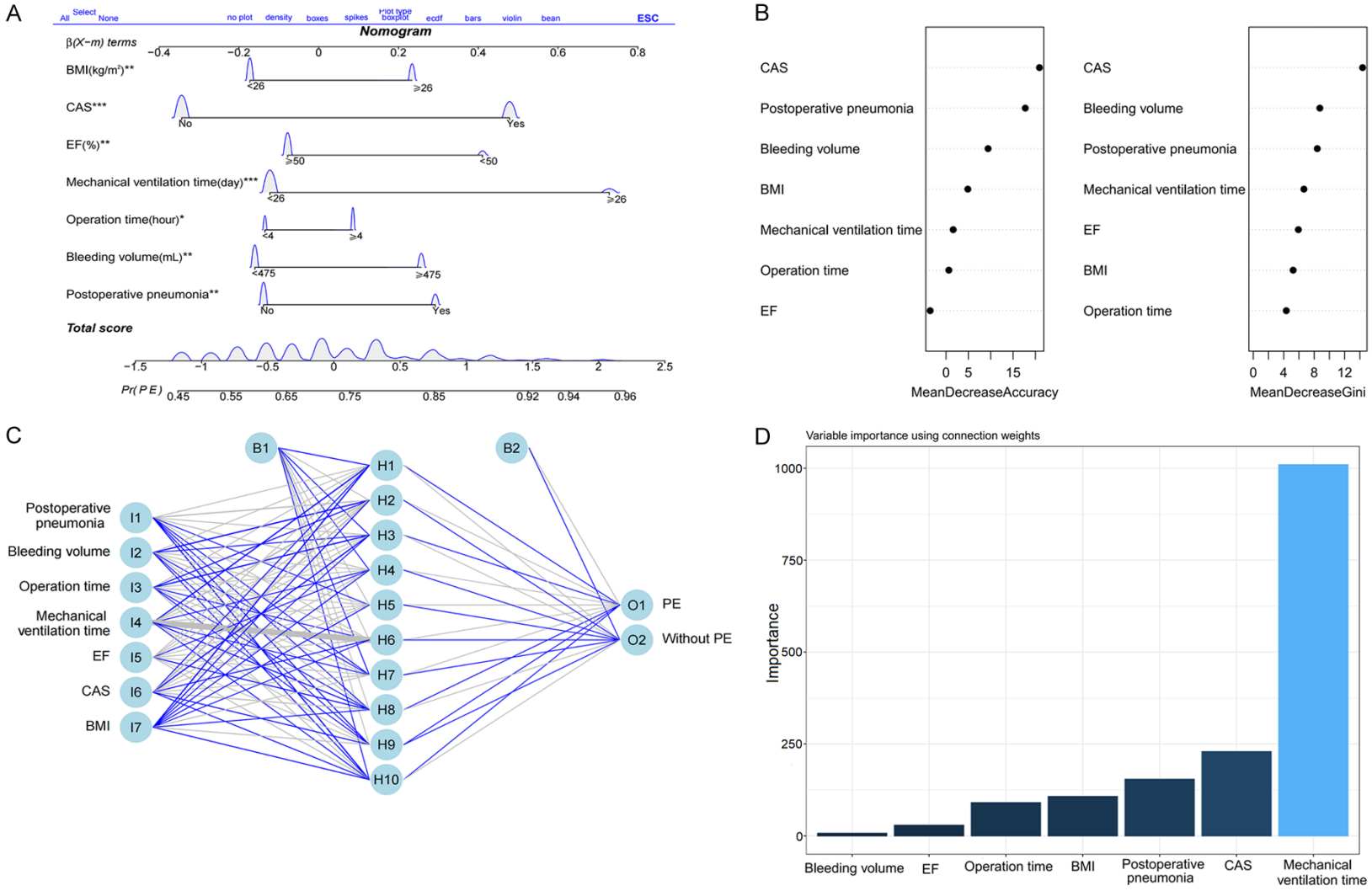
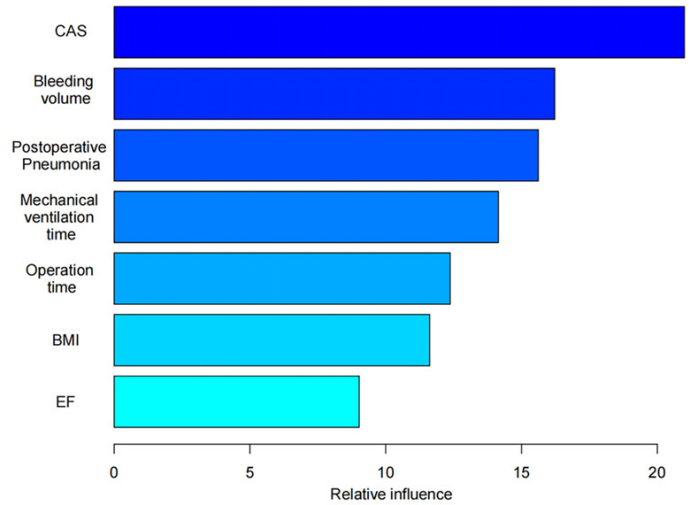


Figure 4. Forest plot of multivariate Logistic regression analysis. Abbreviations: BMI, Body Mass Index; CAS, Carotid artery stenosis; EF, Preoperative Left Ventricular Ejection Fraction; HDL-C, high density lipoprotein cholesterol; ALT, alanine transaminase; CK, Creatine Kinase; LDH, low density lipoprotein cholesterol; cTnI, Cardiac troponin I; NEUT, Neutrophil; NLR, Neutrophil to Lymphocyte rate; EOS, Eosinophil.



E



F

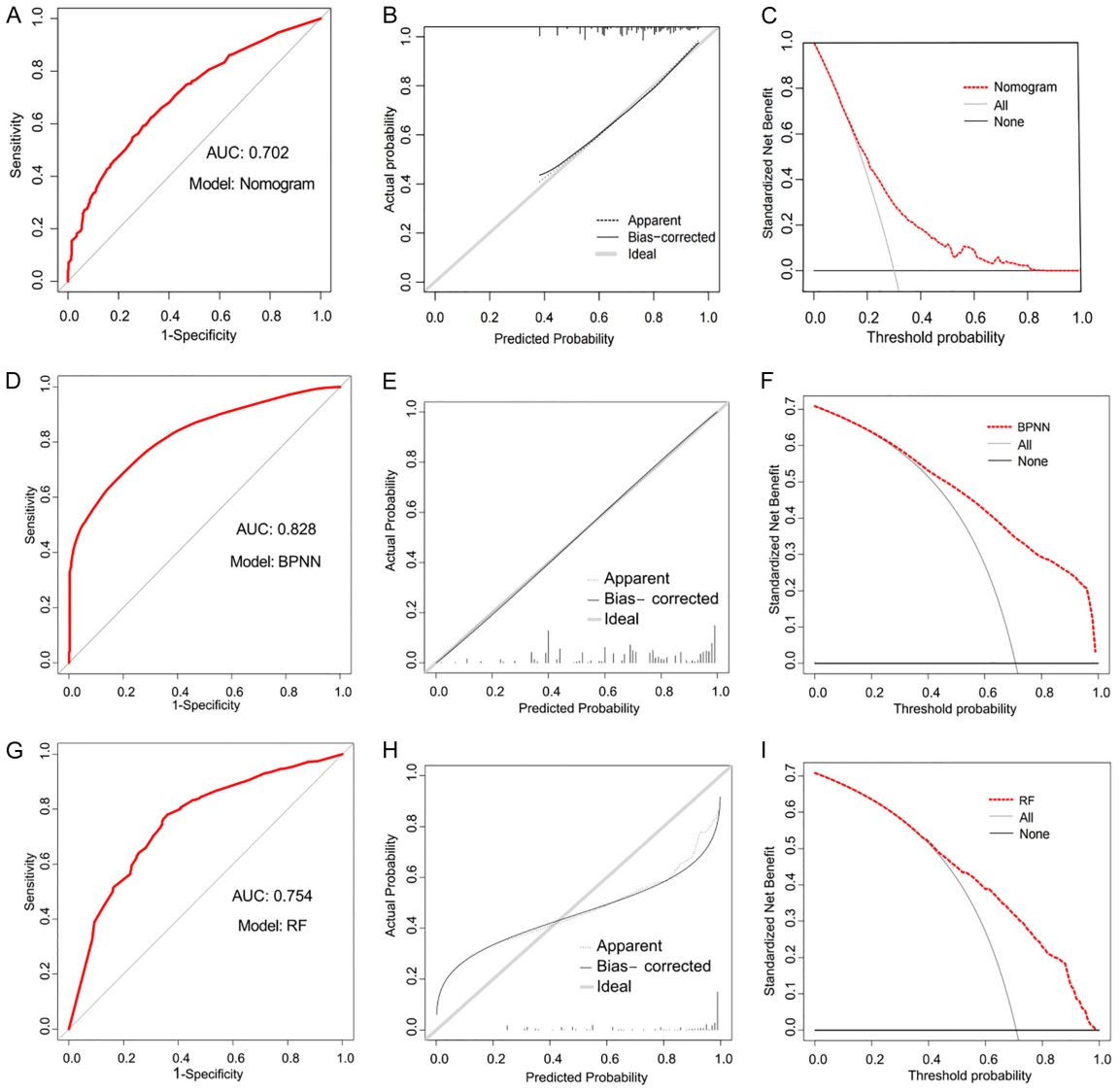
```
> summary(svm_model)

Call:
svm(formula = Group ~ ., data = train, type = "C-classification",
    kernel = "radial")


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  846

 ( 442 404 )


Number of Classes:  2

Levels:
 0 1
```

Figure 5. Model construction of PE post-CABG. A. Nomogram; B. Random forest; C. BP neural network; D. Variable importance ranking in BP neural network; E. Variable importance ranking of gradient boosting model; F. The number of support vectors of support vector machine. Abbreviations: BMI, Body Mass Index; CAS, Carotid artery stenosis; EF, Preoperative Left Ventricular Ejection Fraction.

Table 4. Performance comparison of models

Group	Model	Sensitivity	Specificity	AUC	95% CI
Training set	Nomogram	0.624	0.671	0.702	0.673-0.732
	BPNN	0.691	0.797	0.828	0.806-0.849
	RF	0.780	0.641	0.754	0.725-0.783
	GBM	0.760	0.690	0.790	0.765-0.814
	SVM	0.828	0.616	0.774	0.748-0.800
Validation set	Nomogram	0.531	0.718	0.667	0.620-0.715
	BPNN	0.693	0.729	0.751	0.071-0.792
	RF	0.818	0.642	0.762	0.719-0.804
	GBM	0.589	0.783	0.745	0.702-0.788
	SVM	0.809	0.553	0.708	0.661-0.755

Abbreviations: BPNN, back-propagation neural network; RF, random forest; GBM, gradient boosting model; SVM, support vector machine; AUC, area under the curv; CI, Confidence Interval.



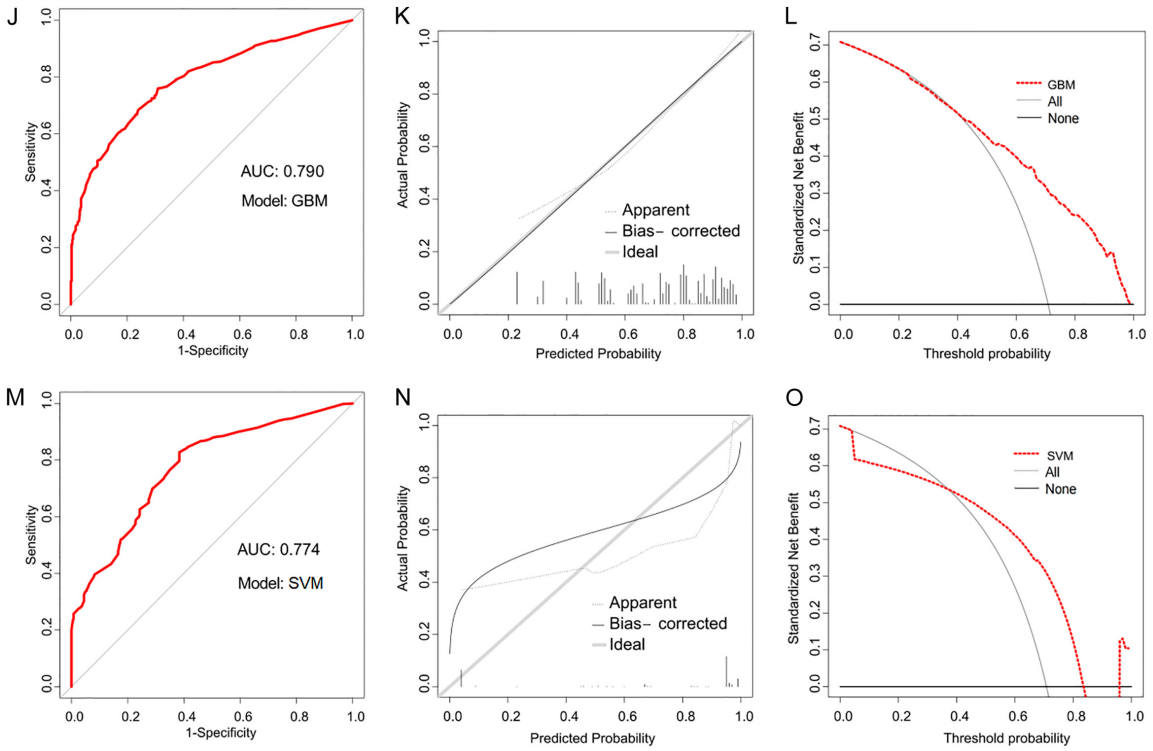
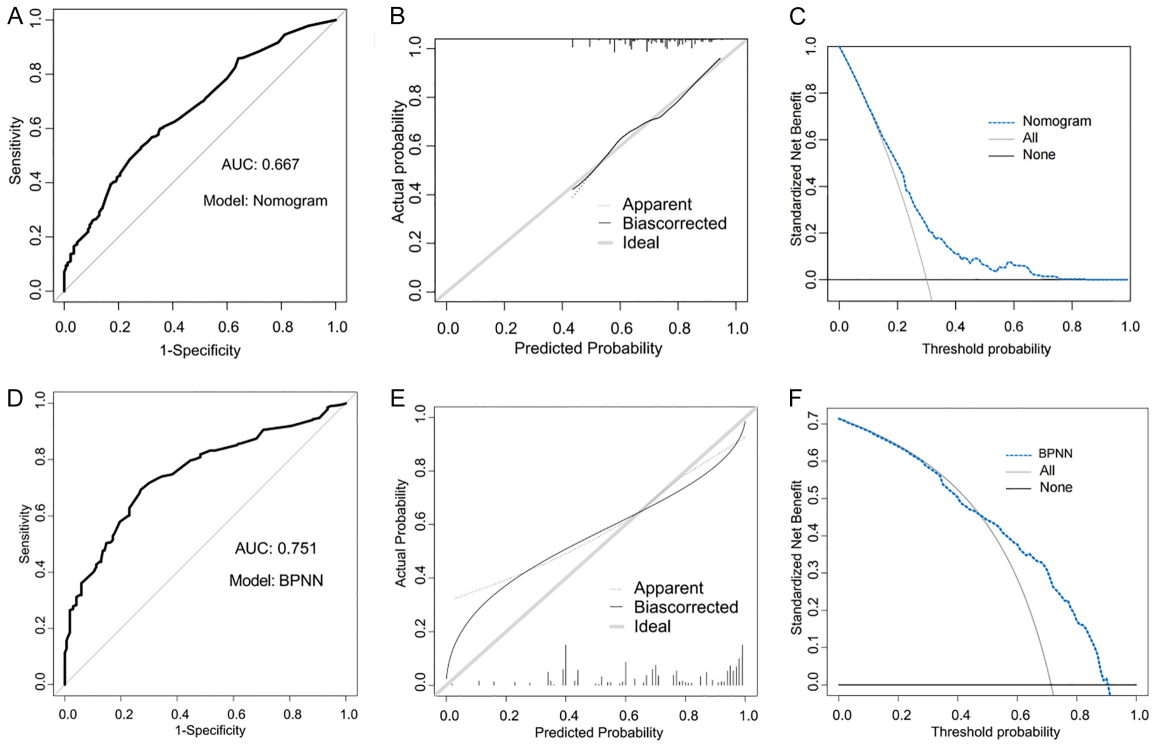


Figure 6. ROC curve, calibration curve and decision curve of the training set of the model. A-C. Nomogram; D-F. BPNN; G-I. RF; J-L. GBM; M-O. SVM. Abbreviations: BPNN, back-propagation neural network; RF, random forest; GBM, gradient boosting model; SVM, support vector machine.



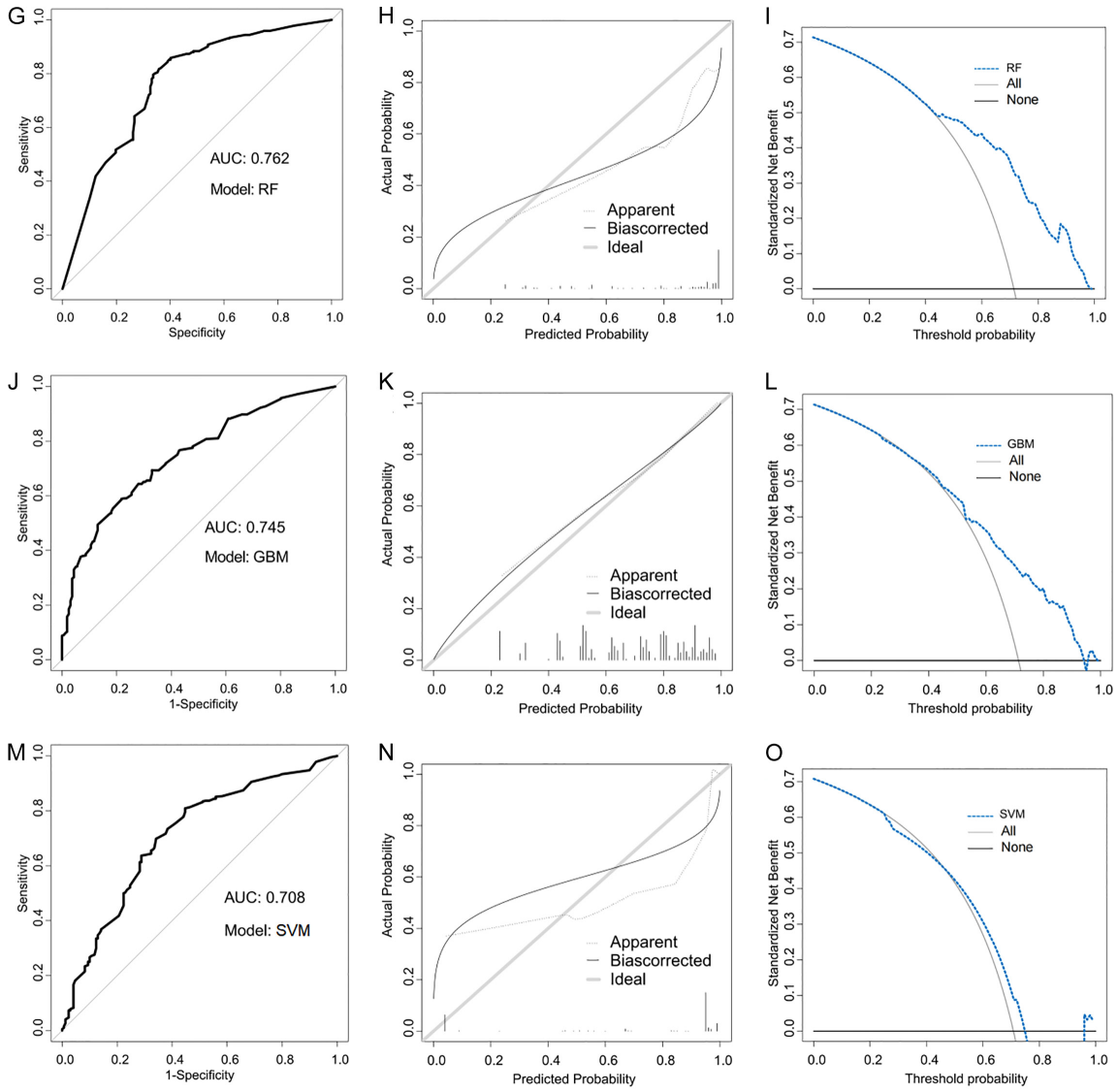


Figure 7. ROC curve, calibration curve and decision curve of the validation set of the model. A-C. Nomogram; D-F. BPNN; G-I. RF; J-L. GBM; M-O. SVM. Abbreviations: BPNN, back-propagation neural network; RF, random forest; GBM, gradient boosting model; SVM, support vector machine.

influenced by a wide range of factors and mechanisms. Historically, efforts to predict and stratify risk based on individual factors have been insufficient to meet clinicians' expectations. Consequently, a risk model that incorporates a composite of various indicators provides a more effective approach for accurate evaluation. In our study, BMI, coronary artery stenosis, postoperative pneumonia, mechanical ventilation time, surgical blood loss, operation duration, and ejection fraction were identified as significant factors associated with the development of postoperative PE following CABG. By focusing on these predictors, five models were developed to estimate the individualized risk of PE following CABG.

In this study, the occurrence of PE in patients three days following CABG was approximately 71% (1,405/1,979). Rezk et al. [23] reported a lower incidence of postoperative PE in patients undergoing CABG, with an incidence of 18% within one to two days post-surgery. Brookes et al. [9] found that approximately 14% (330/409) of patients required thoracic drainage following CABG. The differing incidence rates can be attributed to multiple factors. Firstly, they depend on the study's critical value

Table 5. Comparison of patient data between training set and validation set, n (%)

Characteristic	Total n = 1979	Derivation set n = 1385	Validation set n = 594	χ^2	*P* value
Age, years				0.009	0.925
< 60	646 (32.6)	453 (32.7)	193 (32.5)		
≥ 60	1333 (67.4)	932 (67.3)	401 (67.5)		
Gender				2.643	0.104
Male	1487 (75.1)	1055 (76.2)	432 (72.7)		
Female	492 (24.9)	330 (23.8)	162 (27.3)		
BMI, kg/m^2				0.084	0.771
< 26	1143 (57.8)	797 (57.5)	346 (58.2)		
≥ 26	836 (42.2)	588 (42.5)	248 (41.8)		
Smoking index				0.438	0.508
< 290	1007 (50.9)	698 (50.4)	309 (52.0)		
≥ 290	972 (49.1)	687 (49.6)	285 (48.0)		
Alcohol consumption				4.066	0.254
Never	851 (43.0)	588 (42.5)	263 (44.3)		
Sometimes	748 (37.8)	516 (37.3)	232 (39.1)		
Former	115 (5.8)	82 (5.9)	33 (5.6)		
Abuse	265 (13.4)	199 (14.3)	66 (11.0)		
Medical history					
Hypertension	1364 (68.9)	957 (69.1)	407 (68.5)	0.065	0.799
Dyslipidemia	1225 (61.9)	847 (61.2)	378 (63.6)	1.085	0.298
Diabetes	778 (39.3)	552 (39.3)	226 (38.0)	0.570	0.450
COPD	53 (2.7)	37 (2.7)	16 (2.7)	0.001	0.978
Stroke	328 (16.6)	236 (17.0)	92 (15.5)	0.724	0.395
Hypoproteinemia	985 (4.8)	700 (50.5)	285 (48.0)	1.091	0.296
Prior PCI	300 (15.2)	196 (14.2)	104 (17.5)	3.642	0.056
CAS	839 (42.4)	579 (41.8)	260 (43.8)	0.658	0.417
Unstable angina	957 (48.4)	679 (49.0)	278 (46.8)	0.823	0.364
Myocardial infarction	452 (22.8)	322 (23.2)	130 (21.9)	0.439	0.508
Postoperative pneumonia	628 (31.7)	455 (32.9)	173 (29.1)	2.666	0.103
No. of diseased Coronary arteries					
1-2	277 (14.0)	182 (13.1)	95 (16.0)	3.801	0.150
3	1582 (79.9)	1123 (81.1)	459 (77.3)		
≥ 4	120 (6.1)	80 (5.8)	40 (6.7)		
No. of coronary bypasses				4.992	0.082
1-2	750 (37.9)	505 (36.5)	245 (41.2)		
3	939 (47.4)	679 (49.0)	260 (43.8)		
≥ 4	290 (14.7)	201 (14.5)	89 (15.0)		
EF, %				1.996	0.158
< 50	301 (15.2)	221 (16.0)	80 (13.5)		
≥ 50	1678 (84.8)	1164 (84.0)	514 (86.5)		
Time in ICU, day				0.049	0.826
< 4	957 (48.4)	672 (48.5)	285 (48.0)		
≥ 4	1022 (51.6)	713 (51.5)	309 (52.0)		
Use of respirator				0.819	0.845
NO	228 (11.5)	163 (11.8)	65 (10.9)		
Non-invasive	77 (3.9)	54 (3.9)	23 (3.9)		
Invasive	1522 (76.9)	1058 (76.4)	464 (78.1)		
Invasive and Non-invasive	152 (7.7)	110 (7.9)	42 (7.1)		

Mechanical ventilation time, day				1.276	0.259
< 26	1706 (86.2)	1186 (85.6)	520 (87.5)		
≥ 26	273 (13.8)	199 (14.4)	74 (12.5)		
Amount of bleeding, mL				0.003	0.953
< 475	1218 (61.5)	853 (61.6)	365 (61.4)		
≥ 475	761 (38.5)	532 (38.4)	229 (38.6)		
Time of operation, hour				2.798	0.094
< 4	784 (39.6)	532 (38.4)	252 (42.4)		
≥ 4	1195 (604)	853 (61.6)	342 (57.6)		
CKMB, U/L				0.258	0.611
< 5.7	1129 (57.0)	785 (56.7)	344 (57.9)		
≥ 5.7	850 (43.0)	600 (43.3)	250 (42.1)		
ALB, g/L				0.083	0.773
< 20.9	541 (27.3)	376 (27.1)	165 (27.8)		
≥ 20.9	1438 (72.7)	1009 (72.9)	429 (72.2)		
TC, mmol/L				2.807	0.094
< 1.60	819 (41.4)	590 (42.6)	229 (38.6)		
≥ 1.60	1160 (58.6)	795 (57.4)	365 (61.4)		
TG, mmol/L				0.803	0.370
< 0.59	628 (31.7)	431 (31.1)	197 (33.2)		
≥ 0.59	1351 (68.3)	954 (68.9)	397 (66.8)		
HDL-C, mmol/L				1.632	0.201
< 0.75	1451 (73.3)	1027 (74.2)	424 (71.4)		
≥ 0.75	528 (26.7)	358 (25.8)	170 (28.6)		
LDL-C, mmol/L				0.754	0.385
< 0.45	441 (22.3)	316 (22.8)	125 (21.0)		
≥ 0.45	1538 (77.7)	1069 (77.2)	469 (79.0)		
AST, U/L				0.709	0.400
< 24	379 (19.2)	272 (19.6)	107 (18.0)		
≥ 24	1600 (80.8)	1113 (80.4)	487 (82.0)		
ALT, U/L				0.119	0.730
< 18	483 (24.4)	335 (24.2)	148 (24.9)		
≥ 18	1496 (75.6)	1050 (75.8)	446 (75.1)		
CK, U/L				1.329	0.249
< 218	329 (16.6)	239 (17.3)	90 (15.2)		
≥ 218	1650 (83.4)	1146 (82.7)	504 (84.8)		
LDH, U/L				3.176	0.075
< 320	1754 (88.6)	1216 (87.8)	538 (90.6)		
≥ 320	225 (11.4)	169 (12.2)	56 (9.4)		
GGT, U/L				0.053	0.819
< 34	1413 (71.4)	991 (71.6)	422 (71.0)		
≥ 34	566 (28.6)	394 (28.4)	172 (29.0)		
BUN, mmol/L				3.009	0.083
< 5.37	1239 (62.6)	850 (61.4)	389 (65.5)		
≥ 5.37	740 (37.4)	535 (38.6)	205 (34.5)		
Cre, μmol/L				3.058	0.080
< 103.6	1450 (73.3)	999 (72.1)	451 (75.9)		
≥ 103.6	529 (26.7)	386 (27.9)	143 (24.1)		

cTnI, ng/L				2.467	0.116
< 24.75	1853 (93.6)	1289 (93.1)	564 (94.9)		
≥ 24.75	126 (6.4)	96 (6.9)	30 (5.1)		
WBC, $*10^9$/L				0.340	0.560
< 10	956 (48.3)	675 (48.7)	281 (47.3)		
≥ 10	1023 (51.7)	710 (51.3)	313 (52.7)		
RBC, $*10^2$/L				0.233	0.630
< 4	1844 (93.2)	1293 (93.4)	551 (92.8)		
≥ 4	135 (6.8)	92 (6.6)	43 (7.2)		
NEUT, $*10^9$/L				1.626	0.202
< 9.14	1150 (58.1)	792 (57.2)	358 (60.3)		
≥ 9.14	829 (41.9)	593 (42.8)	236 (39.7)		
LYM, $*10^9$/L				0.455	0.500
< 0.83	715 (36.1)	507 (36.6)	208 (35.0)		
≥ 0.83	1264 (63.9)	878 (63.4)	386 (65.0)		
NLR				1.000	0.317
< 7.63	909 (45.9)	626 (45.2)	283 (48.2)		
≥ 7.63	1070 (54.1)	759 (54.8)	311 (51.8)		
EOS, $*10^9$/L				5.433	0.020
< 0.04	760 (38.4)	555 (40.1)	205 (34.5)		
≥ 0.04	1219 (61.6)	830 (59.9)	389 (65.5)		
HB, g/L				3.703	0.054
< 100	1111 (56.1)	797 (57.5)	314 (52.9)		
≥ 100	868 (43.9)	588 (42.5)	280 (47.1)		
PLT, $*10^9$/L				2.070	0.150
< 160	1286 (65.0)	914 (66.0)	372 (62.6)		
≥ 160	693 (35.0)	471 (34.0)	222 (37.4)		
PLR				0.703	0.402
< 150.84	1158 (58.5)	802 (57.9)	356 (59.9)		
≥ 150.84	821 (41.5)	583 (42.1)	238 (40.1)		

Abbreviations: PE, Pleural effusion; M, median; BMI, Body Mass Index; COPD, Chronic Obstructive Pulmonary Disease; PCI, Percutaneous Transluminal Coronary Intervention; CAS, Carotid artery stenosis; EF, Preoperative Left Ventricular Ejection Fraction; ICU, Intensive Care Unit; CKMB, creatine kinase myocardial band; ALB, albumin; TC, total cholesterol; TG, triglyceride; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; AST, Aspartate Transaminase; ALT, alanine transaminase; CK, Creatine Kinase; LDH, Lactate Dehydrogenas; GGT, Gamma Glutamyl Transpeptidase; BUN, Blood Urea Nitrogen; Cre, Creatinine; cTnI, Cardiac troponin I; WBC, White Blood Cells; RBC, Red Blood Cells; NEUT, Neutrophil; LYM, Lymphocyte; NLR, Neutrophil/Lymphocyte; EOS, Eosinophil; HB, Hemoglobin; PLT, Platelets; PLR, Platelets/Lymphocyte. Data are presented as number (%).

and hard endpoint, such as whether drainage is considered a key factor. Studies focusing on smaller amounts of effusion tend to report higher incidence rates due to sensitive detection methods. Pleural effusion volume varies significantly among individuals. A small volume may be absorbed spontaneously, while a large volume requires intervention. Studies focused on hard endpoints, like drainage, and only including severe cases may result in a lower incidence rate. Furthermore, different studies define meaningful effusion volume differently. Additionally, postoperative effusion changes over time, and the timing of evaluation can influence the results. The high proportion of patients with small-volume effusion in this study may contribute to the higher statistical incidence rate.

Among numerous factors, being overweight and obesity are recognized as common risk factors for coronary artery disease [23] and significant contributors to mortality following CABG [24]. A BMI greater than 40 kg/m^2 is

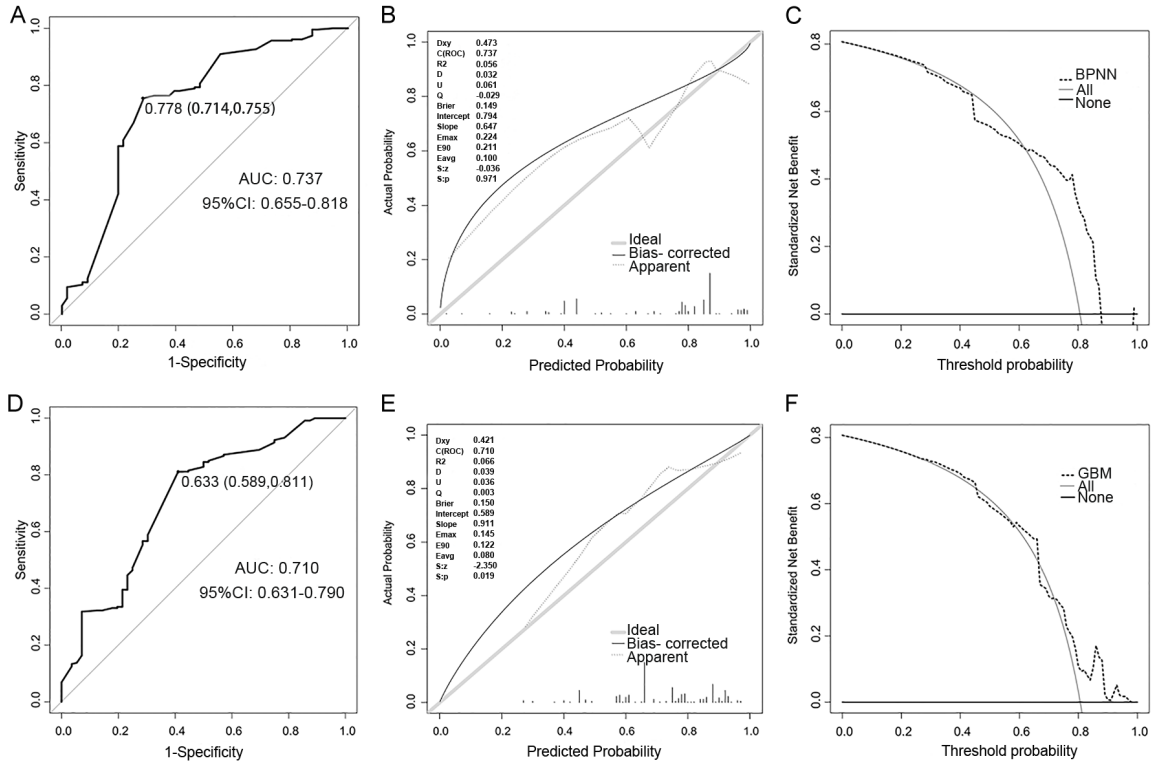


Figure 8. External validation of BPNN and GBM. A and D. ROC curve; B and E. Calibration curve; C and F. Decision curve. Abbreviations: BPNN, back-propagation neural network; GBM, gradient boosting model.

associated with an increased risk of deep and organ-space surgical site infections following CABG [25]. These infections, in turn, can elevate the probability of thoracic infections. Obesity can also impair the body's specific and non-specific immune responses, exacerbating bacterial infections and predisposing individuals to chest infections [26], which can lead to postoperative PE. Pneumonia is another common complication following CABG [27]. Our findings demonstrate a significant association between postoperative pneumonia and the risk of developing postoperative PE. Furthermore, in patients admitted to the ICU, prolonged mechanical ventilation is associated with an increased risk of developing PE [28]. This finding is consistent with the research by Akhtar et al. [7], which indicated that patients requiring pleural effusion drainage tend to have longer intubation times. Prompt removal of ventilators in CABG patients can effectively improve cardiopulmonary function, minimize the risk of pulmonary embolism (PE), shorten ICU and hospital stays, and subsequently reduce healthcare costs [29].

Our observations demonstrate that the risk of developing postoperative PE increases by 1.349-fold when the surgical duration exceeds four hours, suggesting that prolonged operation time contributes to the occurrence of postoperative PE. During CABG, extended exposure to inflammatory cytokines in the lungs and alveoli has been reported to cause swelling, reduced lung compliance and expiratory volume, and lung injury [30]. The longer the surgical procedure lasts, the longer the chest is exposed to air, which may cause trauma or ischemia-reperfusion injury to the heart and lung tissue. This leads to an elevated inflammatory response and increased vascular permeability, thereby raising the likelihood of postoperative chest infection. The EF plays a crucial role in predicting the risk of postoperative PE. In the European System for Cardiac Operative Risk Evaluation (Euro SCORE II) [31], the EF is considered a vital factor for comprehensive surgical risk assessment.

Due to the substantial risks associated with perioperative cardiac surgery, several risk assessment systems have been developed to

quantify the probability of mortality and postoperative complications following CABG. These include Euro SCORE [31], the Society of Cardiothoracic and Vascular Anesthesia Score (Sino SCORE) [32], and the SSII-CABG model [33]. Euro SCORE and Sino SCORE predict the likelihood of death within 30 days post-surgery, while the SSII-CABG model mainly focuses on long-term survival outcomes, particularly over a four-year period.

Although numerous predictive models exist that utilize various factors to assess the risk of postoperative PE, most studies focus primarily on predicting PE after tumor resection, such as liver cancer [34], and lung cancer [35]. Additionally, many studies are dedicated to developing diagnostic models to distinguish between benign and malignant pleural effusion [36-39]. However, the sample sizes of these models tend to be limited, and the variables they incorporate mainly consist of preoperative indicators. A significant gap remains in the development of a comprehensive risk assessment framework specifically designed for postoperative PE following cardiac surgery. Consequently, we believe it is crucial to develop a risk model capable of accurately assessing the probability of postoperative PE in patients undergoing CABG.

By utilizing machine learning algorithms, five predictive models for postoperative PE were successfully developed in our research. These models incorporated a diverse set of predictors, including surgery-related factors, preoperative considerations, and postoperative multidimensional variables. Specifically, BMI, coronary artery stenosis, postoperative pneumonia, duration of mechanical ventilation, surgical blood loss, operation duration, and EF were identified as critical predictors in our models. In terms of prediction performance, our analysis indicated that the BPNN significantly outperformed the other four models, namely the nomogram, RF, GBM, and SVM. Consequently, BPNN was chosen as the preferred predictive model. The ability of BPNN to effectively approximate complex nonlinear relationships makes it particularly suitable for handling problems with intricate internal mechanisms [40].

However, it is important to note the limitations of BPNN. One major drawback is its poor interpretability. Unlike a nomogram, which presents a visual and intuitive approach to understanding the relationship between variables and the predicted outcome, the internal workings of BPNN are often regarded as a “black box”. It is challenging to precisely explain how each input variable contributes to the final prediction, which could limit its application in fields where interpretability is crucial, such as clinical decision-making, where doctors need to understand the reasoning behind a prediction. Therefore, further research is needed to explore ways to improve its interpretability, such as developing novel methods for feature importance analysis or visualization techniques, to enhance its applicability in clinical settings.

Our study has some limitations that need to be acknowledged. The primary aim of this study was to analyze the risk factors for pleural effusion after CABG and construct a predictive model to reduce its incidence. However, most of the patients included in the study had a small amount of pleural effusion. Even if these patients developed PE after surgery, it often resolved spontaneously without requiring special intervention. Therefore, the clinical value of predicting this type of situation is limited. Future studies could focus on patients with severe pleural effusion for in-depth analysis. In addition, this study only included the available risk factors and lacked detailed information on surgical techniques used (such as open and closed pleural procedures) and postoperative care plans (such as extubation strategies or physical therapy programs). This may result in an incomplete assessment of the risk of PE.

In conclusion, our study has demonstrated that pleural effusion is a prevalent complication after coronary artery bypass grafting, with a relatively high incidence rate. To address this issue, we have effectively developed a prediction model for PE with excellent predictive performance. This model is invaluable for the early identification of high-risk patients for PE.

Disclosure of conflict of interest

None.

Address correspondence to: Huanzhong Shi, Department of Respiratory and Critical Care Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Chaoyang, Beijing 100020, China. Tel: +86-0771-2188007; E-mail: shihuanzhong@sina.com

References

[1] Forrest IS, Petrazzini BO, Duffy A, Park JK, Marquez-Luna C, Jordan DM, Rocheleau G, Cho JH, Rosenson RS, Narula J, Nadkarni GN and Do R. Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. Lancet 2023; 401: 215-225.

[2] Cheraghi M and Pooria A. A review: nanofibrous scaffold in possible prevention and treatment of coronary artery disease. Biotechnol Appl Biochem 2019; 66: 478-483.

[3] Kusu-Orkar TE, Kermali M, Oguamanam N, Bithas C and Harky A. Coronary artery bypass grafting: factors affecting outcomes. J Card Surg 2020; 35: 3503-3511.

[4] Krittanawong C, Rizwan A, Khawaja M, Newman N, Escobar J, Virk HUH, Alam M, Al-Azzam F, Yong CM and Jneid H. The current state of coronary revascularization: coronary artery bypass graft surgery versus percutaneous coronary interventions. Curr Cardiol Rep 2024; 26: 919-933.

[5] Lusquinhos J, Tavares M and Abelha F. Postoperative pulmonary complications and perioperative strategies: a systematic review. Cureus 2023; 15: e38786.

[6] Montrief T, Koyfman A and Long B. Coronary artery bypass graft surgery complications: a review for emergency clinicians. Am J Emerg Med 2018; 36: 2289-2297.

[7] Akhtar M, Razick DI, Saeed A, Baig O, Kamran R, Ansari U, Sajid Z and Rahman JE. Complications and outcomes of the nuss procedure in adult patients: a systematic review. Cureus 2023; 15: e35204.

[8] Nakamura E, Sofue T, Higashitani M, Saiki K, Yamamoto T, Shiga T, Aoki Y, Shiraishi A, Kunisho Y, Onishi K, Kato A and Minamino T. A case of a peritoneal dialysis patient with left pleuroperitoneal communication caused by a pericardial defect after coronary artery bypass surgery. CEN Case Rep 2024; 13: 457-462.

[9] Brookes JDL, Williams M, Mathew M, Yan T and Bannon P. Pleural effusion post coronary artery bypass surgery: associations and complications. J Thorac Dis 2021; 13: 1083-1089.

[10] Jain A, Devarajan A, Assallum H, Malekan R, Lanier GM and Epelbaum O. Characteristics of early pleural effusions after orthotopic heart transplantation: comparison with coronary artery bypass graft surgery. Pleura Peritoneum 2021; 6: 161-165.

[11] Feng TR, White RS, Gaber-Baylis LK, Turnbull ZA and Rong LQ. Coronary artery bypass graft readmission rates and risk factors - a retrospective cohort study. Int J Surg 2018; 54: 7-17.

[12] Zhong Z, Yuan X, Liu S, Yang Y and Liu F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. Sci Rep 2021; 11: 3384.

[13] Ma H, Chen D, Lv W, Liao Q, Li J, Zhu Q, Zhang Y, Deng L, Liu X, Wu Q, Liu X and Yang Q. Performance of an AI prediction tool for new-onset atrial fibrillation after coronary artery bypass grafting. EClinicalMedicine 2025; 81: 103131.

[14] Zhang H, Qiao H, Yang B, Lu Y, Bai T, Xue J and Liu Y. Development and validation of a diagnostic model based on left atrial diameter to predict postoperative atrial fibrillation after off-pump coronary artery bypass grafting. J Thorac Dis 2023; 15: 3708-3725.

[15] Roberts ME, Rahman NM, Maskell NA, Bibby AC, Blyth KG, Corcoran JP, Edey A, Evison M, de Fonseka D, Hallifax R, Harden S, Lawrie I, Lim E, McCracken DJ, Mercer R, Mishra EK, Nicholson AG, Noorzad F, Opstad K, Parsonage M, Stanton AE and Walker S; BTS Pleural Guideline Development Group. British Thoracic Society Guideline for pleural disease. Thorax 2023; 78 Suppl 3: s1-s42.

[16] Buch G, Schulz A, Schmidtmann I, Strauch K and Wild PS. A systematic review and evaluation of statistical methods for group variable selection. Stat Med 2023; 42: 331-352.

[17] Vasquez MM, Hu C, Roe DJ, Halonen M and Guerra S. Measurement error correction in the least absolute shrinkage and selection operator model when validation data are available. Stat Methods Med Res 2019; 28: 670-680.

[18] Sun T, Liu J, Yuan H, Li X and Yan H. Construction of a risk prediction model for lung infection after chemotherapy in lung cancer patients based on the machine learning algorithm. Front Oncol 2024; 14: 1403392.

[19] Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, Yu W and Yan J. Study of cardiovascular disease prediction model based on random forest in eastern China. Sci Rep 2020; 10: 5245.

[20] Jan Ben S, Dorner M, Gunther MP, von Kanel R and Euler S. Proof of concept: predicting distress in cancer patients using back propagation neural network (BPNN). Heliyon 2023; 9: e18328.

[21] Wang W, Sheng R, Liao S, Wu Z, Wang L, Liu C, Yang C and Jiang R. LightGBM is an effective predictive model for postoperative complications in gastric cancer: a study integrating radiomics with ensemble learning. J Imaging Inform Med 2024; 37: 3034-3048.

[22] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y and Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics 2018; 15: 41-51.

[23] Rezk ME, Elgazzar MA, Abo Youssef SM, Emeraa AS, Elkafoury AE and Moussa HH. Open versus closed pleura internal mammary artery harvesting and early pulmonary function after coronary artery bypass grafting. Heart Lung Circ 2020; 29: 1412-1417.
[24] Pencina MJ, Navar AM, Wojdyla D, Sanchez RJ, Khan I, Elassal J, D'Agostino RB Sr, Peterson ED and Sniderman AD. Quantifying importance of major risk factors for coronary heart disease. Circulation 2019; 139: 1603-1611.
[25] Lv M, Gao F, Liu B, Pandey P, Feng Y, Wang Y, Zhang Y and Li Z. The effects of obesity on mortality following coronary artery bypass graft surgery: a retrospective study from a single center in China. Med Sci Monit 2021; 27: e929912.
[26] Bajpai G and Nahrendorf M. Infectious and lifestyle modifiers of immunity and host resilience. Immunity 2021; 54: 1110-1122.
[27] Bartlett B, Sanfilippo FM, Lee S, Ludewick H, Waterer G, Rajwani A, Bharat C, Ihdayhid AR, Corrales-Medina V and Dwivedi G. The risk of adverse cardiac events following pneumonia in patients with coronary artery disease. Ann Am Thorac Soc 2025; [Epub ahead of print].
[28] Luo Q, Zhao W, Su Z, Liu Y, Jia Y, Zhang L, Wang H, Li Y, Wu X, Li S and Yan F. Risk factors for prolonged pleural effusion following total cavopulmonary connection surgery: 9 years' experience at Fuwai Hospital. Front Pediatr 2019; 7: 456.
[29] Pooria A, Pourya A and Gheini A. Postoperative complications associated with coronary artery bypass graft surgery and their therapeutic interventions. Future Cardiol 2020; 16: 481-496.
[30] Guo MH, Toubar O, Issa H, Glineur D, Ponnambalam M, Vo TX, Rahmouni K, Chong AY and Ruel M. Long-term survival, cardiovascular, and functional outcomes after minimally invasive coronary artery bypass grafting in 566 patients. J Thorac Cardiovasc Surg 2024; 168: 1080-1088, e2.
[31] Halvorsen S, Mehilli J, Cassese S, Hall TS, Abdelhamid M, Barbato E, De Hert S, de Laval I, Geisler T, Hinterbuchner L, Ibanez B, Lenarczyk R, Mansmann UR, McGreavy P, Mueller C, Muneretto C, Niessner A, Potpara TS, Ristic A, Sade LE, Schirmer H, Schupke S, Sillesen H, Skulstad H, Torracca L, Tutarel O, Van Der Meer P, Wojakowski W and Zacharowski K; ESC Scientific Document Group. 2022 ESC Guidelines on cardiovascular assessment and management of patients undergoing non-cardiac surgery. Eur Heart J 2022; 43: 3826-3924.
[32] Peng K, McIlroy DR, Bollen BA, Billings FT 4th, Zarbock A, Popescu WM, Fox AA, Shore-Lesserson L, Zhou S, Geube MA, Ji F, Bhatia M, Schwann NM, Shaw AD and Liu H. Society of cardiovascular anesthesiologists clinical practice update for management of acute kidney injury associated with cardiac surgery. Anesth Analg 2022; 135: 744-756.
[33] Yuksel Y, Yildiz C and Kose S. Assessment of predictive value of syntax-ii score for adverse cardiac events and clinical outcomes in patients with acute coronary syndrome. Angiology 2024; 75: 754-763.
[34] Nitta H, Mitsuura C, Shiraishi Y, Miyata T, Shimizu K, Harada K, Karashima R, Masuda T, Matsumoto K, Okino T, Yamashita YI, Baba H and Takamori H. Predictive model for postoperative pleural effusion after hepatectomy. Ann Gastroenterol Surg 2020; 5: 373-380.
[35] Li R, Qiu J, Qu C, Ma Z, Wang K, Zhang Y, Yue W and Tian H. Comparison of perioperative outcomes with or without routine chest tube drainage after video-assisted thoracoscopic pulmonary resection: a systematic review and meta-analysis. Front Oncol 2022; 12: 915020.
[36] Wang S, Tian S, Li Y, Zhan N, Guo Y, Liu Y, Xu J, Ma Y, Zhang S, Song S, Geng W, Xia H, Ma P, Wang X, Liao T, Duan Y, Jin Y and Dong W. Development and validation of a novel scoring system developed from a nomogram to identify malignant pleural effusion. EBioMedicine 2020; 58: 102924.
[37] Wang S, Tan X, Li P, Fan Q, Xia H, Tian S, Pan F, Zhan N, Yu R, Zhang L, Duan Y, Xu J, Ma Y, Chen W, Li Y, Zhao Z, Liu C, Bao Q, Yang L and Jin Y. Differentiation of malignant from benign pleural effusions based on artificial intelligence. Thorax 2023; 78: 376-382.
[38] Pan Y, Bai W, Chen J, Mao Y, Qian X, Xu K, Tang S, Zhang J, Chen C, Chen J and Hu X. Diagnosing malignant pleural effusion using clinical and analytical parameters. J Clin Lab Anal 2019; 33: e22689.
[39] Ferreiro L, Gude F, Toubes ME, Lama A, Suarez-Antelo J, San-Jose E, Gonzalez-Barcala FJ, Golpe A, Alvarez-Dobano JM, Rabade C, Rodriguez-Nunez N, Diaz-Louzao C and Valdes L. Predictive models of malignant transudative pleural effusions. J Thorac Dis 2017; 9: 106-116.
[40] Akella A and Akella S. Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution. Future Sci OA 2021; 7: FSO698.