Construction of a Risk Prediction Model for Hospital-Acquired Pulmonary Embolism in Hospitalized Patients

Lengchen Hou, PhD[1],[*], Longjun Hu, PhD[1],[*],
Wenxue Gao, MD[1], Wenbo Sheng, PhD[2],
Zedong Hao, PhD[2], Yiwei Chen, MPH[2],
and Jiyu Li, PhD[1] (iD)

Abstract

The purpose of this study is to establish a novel pulmonary embolism (PE) risk prediction model based on machine learning (ML) methods and to evaluate the predictive performance of the model and the contribution of variables to the predictive performance. We conducted a retrospective study at the Shanghai Tenth People's Hospital and collected the clinical data of in-patients that received pulmonary computed tomography imaging between January 1, 2014 and December 31, 2018. We trained several ML models, including logistic regression (LR), support vector machine (SVM), random forest (RF), and gradient boosting decision tree (GBDT), compared the models with representative baseline algorithms, and investigated their predictability and feature interpretation. A total of 3619 patients were included in the study. We discovered that the GBDT model demonstrated the best prediction with an area under the curve value of 0.799, whereas those of the RF, LR, and SVM models were 0.791, 0.716, and 0.743, respectively. The sensibilities of the GBDT, LR, RF, and SVM models were 63.9%, 68.1%, 71.5%, and 75%, respectively; the specificities were 81.1%, 66.1, 72.7%, and 65.1%, respectively; and the accuracies were 77.8%, 66.5%, 72.5%, and 67%, respectively. We discovered that the maximum D-dimer level contributed the most to the outcome prediction, followed by the extreme growth rate of the plasma fibrinogen level, in-hospital duration, and extreme growth rate of the D-dimer level. The study demonstrates the superiority of the GBDT model in predicting the risk of PE in hospitalized patients. However, in order to be applied in clinical practice and provide support for clinical decision-making, the predictive performance of the model needs to be prospectively verified.

Introduction

Venous thromboembolism (VTE) is a disorder that includes deep venous thrombosis (DVT) and pulmonary embolism (PE). The incidence of VTE is increasing annually. According to relevant data, the total incidence of VTE in the Asian population is 17 to 70 per 100,000 people, thus making it the third most common vascular disease.[1,2] PE, which is a leading cause of cardiovascular death,[3] occurs when an embolus breaks off a thrombus in a vein and occludes blood vessels of the pulmonary artery.[4] PE affects approximately 112 patients per 100 000 annually in the United States;[5] furthermore, the incidence of PE increased from 23 per 100 000 in 1993 to more than 65 per 100,000 in 2012.[6] A study estimated the overall incidence of postpartum PE among Taiwanese women who had undergone Cesarean sections to be 0.154 per 1000.[7]

[1] Shanghai Tenth People's Hospital, Shanghai, China
[2] Shanghai Synyi Medical Technology Co., Ltd, Shanghai, China

*As co-first authors, the two authors have an equally important contribution to this research.

Corresponding Author:
Jiyu Li, Shanghai Tenth People's Hospital, 301 Middle Yanchang Road, Jiang'an District, 200072, Shanghai, China.
Email: leejiyu@sina.com

The symptoms of PE are generally detected incidentally; in many cases, patients with PE die without presenting symptoms.[8] An estimated 60 000 to 100 000 people died from PE in the United States in 2015.[9] In the same study, the all-cause mortality rate of PE patients was 11.4% during the first 2 weeks after diagnosis and 17.4% at 3 months, where 45.1% of deaths were ascribed to PE. Furthermore, 7.9% of patients had recurrent PE within 3 months of presentation of symptoms. Among these patients, the mortality rate was 33.7% at 2 weeks and 46.8% at 3 months.[10] Furthermore, PE survivors can develop post-PE syndrome and have a higher risk of recurrent PE and dysfunction.[11,12]

Early diagnosis and appropriate management can reduce mortality and morbidity in PE patients.[13] However, the diagnosis and screening of PE is challenging owing to the nonspecific nature of the classic presenting symptoms, and highly specialized expertise, such as that typically possessed by nuclear radiologists, is required. Therefore, it is crucial to establish a prevention concept for PE in hospitalized patients and assess the risk of PE. Currently, the VTE screening scale typically used in hospitals is Caprini; however, the accuracy of the Caprini scale is poor, rendering it ineffective for accurately screening high-risk VTE patients. Moreover, the current screening scales require manual scoring, which is inconvenient and makes performing large-scale screenings difficult; moreover, key data may be lost. Therefore, a VTE risk prediction tool that is easy to use and demonstrates a better predictive performance should be developed. Risk prediction models use variables (covariates) to estimate the absolute probability or risk that a certain outcome will occur or will occur within a certain time period in a patient with a certain predictor variable.[14]

The implementation of individualized VTE risk assessment based on machine learning (ML) can significantly reduce the risk of VTE in in-hospital patients. It was discovered in a previous study that ML in the form of logistic regression (LR) and artificial neural networks contributed positively to the risk factor analysis of complications after spine surgery, including VTE.[15] Furthermore, the ML approach may offer clinical benefits for VTE risk stratification in cancer outpatients undergoing chemotherapy. Using the ML approach to devise a risk assessment model may also assist clinicians at critical stages of clinical decision-making.[16,17] The study shows that ML-based risk prediction models can identify high-risk patients with greater accuracy compared with previously developed VTE scoring models. In addition, the ML model can be used even when the operator does not possess expertise regarding the relevant risk factors.[18]

The purpose of this study is to establish a novel PE risk prediction model based on ML methods, as well as to evaluate the predictive performance of the model and the contribution of variables to the predictive performance.

Materials and Methods

Study Design and Patients

The study was approved by the Ethics Committee of Shanghai Tenth People's Hospital (SHSY-IEC-4.1/19-140/01).

We conducted a retrospective study at the Shanghai Tenth People's Hospital using clinical inpatient data. All inpatients that received pulmonary computed tomography (CT) imaging from January 1, 2014 to December 31, 2018 in the Shanghai Tenth People's Hospital were included. The other inclusion criteria were as follows: admitted patients older than 18 years, hospital length of stay of at least 2 days, and patients suspected of PE. The exclusion criteria included patients with a history of PE before admission and those who had previously consumed antithrombotic drugs (antiplatelet, anticoagulant, thrombolytic, and defibrillation drugs) for prophylaxis. We retrieved and parsed the texts from CT verification reports for PE outcomes via computer programming, followed by manual reconfirmation by physicians. Finally, the cohort comprised 3619 samples, of which 629 were positive, and 2990 were negative.

Diagnostic Criteria of PE

PE-positive diagnoses include those where a filling defect is identified in the central, segmental, or subsegmental pulmonary arteries. PE-negative diagnoses refer to those where evidence of PE is not found.[19] Inconsistent conclusion of test results was adjudicated by 5 experienced respiratory physicians and radiologists.

Data Preprocess and Feature Construction

The outcome, inclusion, and exclusion conditions and potential risk factors were obtained from clinical EHR databases: inpatient records, diagnoses, drug orders, imaging checks, laboratory tests, vital signs and nursing records, and operation and anesthesia records. Based on a system provided by Synyi Medical Technology (Shanghai), which aims at the EHR system integration and data governance, we constructed a feature engineering pipeline to collect and process related data before PE examination. This procedure is also known as extract–transform–load (ETL).

According to ETL, all features can be categorized into 3 types (static features, dynamic features, and derived features). Some static features are retrievable from original databases directly, such as demographic characteristics, life and social behaviors, and admission diagnoses. Meanwhile, some other features are dynamic and recorded as a time series, such as vital signs and laboratory tests. For specific features, we set up a group of time windows before the outcomes (PE examination) and then aggregated all records within these time windows to derive statistics as features, including the maximal, minimal, and mean values, as well as the average change rates. The settings of the time windows depended on data fields and were advised by physicians and medical knowledge experts. For instance, blood cell tests were windowed within 1, 3, and 7 d, whereas enzyme and antibody tests were windowed as early as 1 month and 1 year. Furthermore, we included some derived variables such as the body mass index, compound symptoms, and drug histories, including anticoagulant usage.

Feature Evaluation and Filtering

Thousands of variables can be obtained through our ETL pipeline. However, some features might have high missing value rates that are inconsistent with PE, or they might exhibit strong mutual collinearity with each other, which may yield unstable ML results. For instance, certain uncommon surgery history belongs to the former case, whereas the latter case includes test values within different time windows for the same laboratory items. As an example, red blood cell count could be expanded to "maximum red blood cell count within 3 days" and "minimum mean red blood cell volume within 2 days." Therefore, it is critical to perform feature selection and filtering to improve the robustness, interpretability, and generalization performance of ML prediction.

We used the scorecard method,[20] a popular technique used in credit scoring, to filter less significant variables. We calculated the information value (IV) for a categorical variable based on the weight of evidence of each level, whereas numerical variables were grouped into bins automatically using a decision tree algorithm before the scorecard measurement. A high IV score represents a strong predictive power, and a feature is preserved only if its IV scores exceed a fixed threshold. It is noteworthy that variables with high missing rates need not be managed specifically as they can be implicitly removed during the scorecard procedure.

To address multicollinearity among variables, we considered the correlation coefficients (CCs) of feature pairs as exclusion criteria. For computation, only features with significant IV were included as candidates. Among all feature pairs with absolute CC values exceeding 0.6, the feature with a smaller IV score was removed.

Machine Learning Models

To assess the predictability of the ML models, the cohort was split randomly into training and testing sets in proportions of 80% (485 positive cases and 2382 negative cases) and 20% (144 positive cases and 608 negative cases), respectively. The training set was used to train the models, including the optimization of the model architectures and parameters, whereas the testing set was used to assess the performances of the prediction models and model overfitting.

Five-fold cross-validation was employed in the training set to estimate the prediction errors. At each cross-validation round, one-fold data were treated as the validation set and the remaining data were used for the parameter fitting of the ML models. Subsequently, the results of all rounds were combined to obtain an overall model evaluation result. Specifically, we computed the area under the curve (AUC) of the receiver operating characteristics and confusion matrices, as well as derived the accuracy, f1 score, precision, sensitivity, and specificity.

We adopted the gradient boosting decision tree (GBDT) algorithm to build a risk prediction model. As a tree boosting method, the GBDT has been shown to outperform many classification tasks and has been used widely in recent years owing to its efficiency, accuracy, and interpretability. In addition, we investigated other representative algorithms for a baseline comparison: LR, support vector machine (SVM) with radial basis function kernels, and random forest (RF).

The configuration variables of the ML models, also known as hyperparameters, such as the tree number, maximal tree depth, and learning rate of the GBDT, as well as the kernel coefficients of the SVM, comprise another important issue affecting the prediction power of the model. We used the average AUC value from cross-validation as the model evaluation for one configuration of the hyperparameters. Searching for the best hyperparameters for maximizing the AUC value is referred to a black box optimization problem, which has been solved using the simulated annealing method, a practical stochastic optimization technique that is more effective compared with alternatives such as grid searching or random searching.

The ML component of this project was developed in Python. The LR, SVM, and RF models used were based on the Scikit-learn package, whereas the GBDT was based on XGBoost,[21] a widely used method to realize the GBDT. The hyperparameters of all the models were optimized using Hyperopt[22] with the default configuration of simulated annealing.

Outcome Indicators

To evaluate the predictive efficacy of the ML-based PE risk prediction models for PE risk in hospitals, we evaluated the AUC, sensitivity, specificity, and accuracy of the risk prediction models.

Statistical Analysis

The patients' demographics and features included in the final ML model were summarized descriptively. Continuous variables were expressed as N, mean ± SD, and median (Q1, Q3), whereas categorical variables were presented as frequency (%). In addition, the distribution balance between the training and testing groups was evaluated via statistical tests, and $P < .05$ was considered statistically significant. Statistical analyses were conducted using SPSS version 22.0 (SPSS Inc., Chicago, IL).

Results

A total of 3619 patients were included in this study, of which 2867 and 752 patients were assigned to the training and testing sets, respectively. The training set comprised 485 PE-positive patients (age, 70.5 ± 12.8 years old; male, 48.66%) and 2382 PE-negative patients (age, 65.6 ± 14.1 years old; male, 46.94%), whereas the testing set comprised 144 PE-positive patients (age, 69.1 ± 15 years old; male, 41.67%) and 608 PE-negative patients (age, 65.2 ± 14.7 years old; male, 46.55%) (Table 1).

The following variables were selected in the study (Table 2): in-hospital duration, maximum neutrophil count within 2 weeks, maximum serum albumin level within 3 d, minimum plasma fibrinogen level within 1 d, extreme growth rate of plasma fibrinogen level within 2 weeks, plasma prothrombin

Table 1. Demographic Characteristics of In-Hospital Patients.

	Training (n = 2867)		Testing (n = 752)	
	PE positive (n = 485)	PE negative (n = 2382)	PE positive (n = 144)	PE negative (n = 608)
Age				
N	485	2382	144	608
Mean ± SD	70.5 ± 12.8	65.6 ± 14.1	69.1 ± 15	65.2 ± 14.7
Median (Q1, Q3)	73 (63, 80)	66 (58, 77)	70 (61.5, 81)	66 (58, 76.5)
Gender				
Male, n (%)	236 (48.66%)	1118 (46.94%)	60 (41.67%)	283 (46.55%)
Female, n (%)	249 (51.34%)	1264 (53.06%)	84 (58.33%)	325 (53.45%)

Abbreviation: PE, pulmonary embolism.

time average growth rate within 2 weeks, minimum mean red blood cell volume within 2 d, last thrombin time within 1 week, extreme growth rate of urea nitrogen level within 2 weeks, maximum red blood cell count within 3 d, maximum D-dimer level within 2 weeks, extreme growth rate of D-dimer level within 2 weeks, maximum C-reactive protein level within 2 weeks, extreme growth rate of C-reactive protein level within 2 weeks, any primary care within 1 month, and base excess level.

Figure 1 shows the curves of the receiver operating characteristics of the 4 predictors, namely LR, SVM, RF, and GBDT, on the test dataset. The detailed results of the prediction performance are presented in Table 3, including the values of sensitivity, specificity, accuracy, and AUC with 95% confident intervals. The classification threshold was specified by maximizing the F1 value. Based on the AUC for predicting the PE risks of different ML models, we discovered that the GBDT model demonstrated the best prediction with an AUC value of 0.799, whereas the RF model (AUC 0.791) was comparable, yet slightly weaker than the GBDT model. In contrast, the results of LR and SVM decreased significantly, yielding AUC values of 0.716 and 0.743, respectively. The sensibilities of these risk prediction models (GBDT, LR, RF, and SVM) were 63.9%, 68.1%, 71.5%, and 75%, respectively; the specificities were 81.1%, 66.1, 72.7%, and 65.1%, respectively; and the accuracies were 77.8%, 66.5%, 72.5%, and 67%, respectively (Table 3).

We further analyzed the feature contribution based on the GBDT. The feature importance is defined as the frequency of a feature used in the ensemble of decision trees in the GBDT, which is proportional to its effect on the overall model performance. As shown in Figure 2, the maximum D-dimer level (D_dimer_max) contributed the most to the outcome prediction of the ML models, followed by the extreme growth rate of the plasma fibrinogen level (plasma_fibrinogen_rate), inhospital_duration, and extreme growth rate of the D_dimer level (D_dimer_rate).

Discussion

In this study, we established a novel PE risk prediction model based on ML and evaluated its efficacy in terms of AUC, sensitivity, specificity, and accuracy. We discovered that the GBDT risk prediction model demonstrated the best predictive efficacy. The AUC for predicting PE using the GBDT risk

prediction model was 0.799 (95% CI: 0.762-0.837), whereas the sensibility, specificity, and accuracy were 63.9%, 81.1%, and 77.8%, respectively.

The prediction of PE risk is crucial for the prevention and treatment of PE. Previous studies have used different types of predictive models or risk scoring tools to estimate the risk of PE in hospitalized patients. Miniati's research found that the LR model can be used to estimate the risk of PE before obtaining definitive test results.[23] Several studies have confirmed that the Wells rule, modified Wells rule, simplified Wells rule, Geneva score, revised Geneva score, and simplified revised Geneva score can all be used for PE risk prediction, and the efficiency of these models was 43% to 48%.[24,25] The AUC of Wells score and the revised Geneva score for the PE risk prediction of outpatients ≥65 years old are 0.632 and 0.610, respectively.[26] The Wells score and the revised Geneva score seem to have no value in predicting PE in pregnant and postpartum populations.[27] In clinical practice, the Caprini score is often used to predict the risk of VTE, but its use still has certain limitations. The study found that the Caprini VTE risk assessment model can effectively predict the VTE risk of critically ill surgical patients.[28] A Caprini score ≥11 can identify high-risk surgical patients who need more effective prevention programs.[29] However, a retrospective clinical study found that, despite the linear relationship between Caprini RAM and the risk of VTE, Caprini RAM was unable to identify a subset of medical patients who would benefit from pharmacologic prophylaxis.[30] The study found that the Caprini model has no significant correlation with the PE risk of DVT patients or the PE risk with significant hemodynamics.[31]

Numerous risk factors of PE exist, in which both inherited and acquired risk factors increase the likelihood of VTE and PE, including factor V Leiden, prothrombin gene mutation (G20210-A), antithrombin deficiency, protein C deficiency, protein S deficiency, acute and chronic medical illness, trauma, surgery, malignancy and related factors, peripartum state estrogen therapy, aging, and obesity.[32,33] Feature selection is a perennial yet challenging issue, and a universal solution has not been discovered. In this study, we adopted a greedy-like strategy for feature selection in the preprocessing stage. Although the strategy might only seek for a suboptimal feature subset, it is reasonable in practice considering its computational feasibility and prediction accuracy. In our study,

Table 2. Characteristics Included in the GBDT Model.

	Training (n = 2867)		Testing (n = 752)	
	PE positive (n = 485)	PE negative (n = 2382)	PE positive (n = 144)	PE negative (n = 608)
In-hospital duration				
N	485	2382	144	608
Mean ± SD	14.7 ± 10.6	11.3 ± 13	16.3 ± 13.2	10.9 ± 8.3
Median (Q1, Q3)	12 (9, 17)	9 (6, 13)	14 (9, 20)	8 (6, 13)
Maximum neutrophil count within 2 weeks				
N	473	2311	142	587
Mean ± SD	7.53 ± 3.91	6.46 ± 4.27	7.63 ± 4.17	6.3 ± 4.09
Median (Q1, Q3)	6.76 (4.74, 9.36)	5.08 (3.53, 8.21)	6.99 (4.6, 9.69)	5.01 (3.4, 7.82)
Maximum serum albumin level within 3 days				
N	224	1188	74	281
Mean ± SD	37.48 ± 6.66	38.88 ± 6.89	36.2 ± 4.98	39.12 ± 7.7
Median (Q1, Q3)	37 (33, 40)	39 (35, 43)	36.5 (34, 39)	39 (35, 43)
Minimum plasma fibrinogen level within 1 day				
N	232	759	69	196
Mean ± SD	3.49 ± 1.24	3.63 ± 1.49	3.49 ± 1.12	3.67 ± 1.47
Median (Q1, Q3)	3.27 (2.69, 4.23)	3.24 (2.5, 4.54)	3.3 (2.77, 4.08)	3.31 (2.58, 4.79)
Extreme growth rate of plasma fibrinogen level within 2 weeks				
N	216	558	62	145
Mean ± SD	0.007 ± 0.03	0 ± 0.027	0.004 ± 0.028	−0.716 ± 8.65
Median (Q1, Q3)	0.004 (−0.007, 0.017)	−0.002 (−0.014, 0.01)	0.002 (−0.01, 0.01)	−0.003 (−0.011, 0.009)
Plasma prothrombin time average growth rate within 2 days				
N	312	1277	97	320
Mean ± SD	−0.018 ± 0.062	−0.017 ± 0.075	−0.026 ± 0.057	−0.02 ± 0.078
Median (Q1, Q3)	−0.008 (−0.037, 0.013)	−0.004 (−0.038, 0.022)	−0.012 (−0.041, 0.004)	−0.004 (−0.042, 0.024)
Plasma prothrombin time average growth rate within 2 weeks				
N	465	2239	140	570
Mean ± SD	−0.01 ± 0.04	−0.01 ± 0.06	−0.02 ± 0.05	−0.01 ± 0.07
Median (Q1, Q3)	0 (−0.02, 0.02)	0 (−0.03, 0.03)	−0.01 (−0.03, 0.01)	0 (−0.03, 0.03)
Minimum mean red blood cell volume within 2 days				
N	322	1377	103	341
Mean ± SD	90.39 ± 5.75	90.35 ± 5.84	91.34 ± 7.32	90.13 ± 6.3
Median (Q1, Q3)	90.75 (87.9, 93.6)	90.6 (87.3, 93.6)	91.6 (88.4, 94.6)	90.7 (87, 93.6)
Last thrombin time within 1 week				
N	441	2110	134	537
Mean ± SD	19.44 ± 3.23	20.45 ± 6.56	19.53 ± 3.47	20.75 ± 8.6
Median (Q1, Q3)	18.8 (17.5, 20.7)	19.7 (18.3, 21.3)	18.9 (17.7, 20.4)	19.7 (18.3, 21.5)
Extreme growth rate of urea nitrogen level within 2 weeks				
N	242	748	75	188
Mean ± SD	−0.007 ± 0.074	−0.001 ± 0.069	−0.014 ± 0.056	0.003 ± 0.077
Median (Q1, Q3)	−0.011 (−0.04, 0.019)	−0.008 (−0.029, 0.013)	−0.014 (−0.048, 0.015)	−0.007 (−0.031, 0.024)
Maximum red blood cell count within 3 days				
N	382	1705	120	426
Mean ± SD	4.04 ± 0.68	4.13 ± 0.67	4.03 ± 0.63	4.16 ± 0.69
Median (Q1, Q3)	4.09 (3.63, 4.49)	4.17 (3.74, 4.58)	4.06 (3.65, 4.42)	4.19 (3.72, 4.64)
Maximum D-dimer level within 2 weeks				
N	472	2271	142	572
Mean ± SD	31.42 ± 460.01	3.92 ± 7.42	9.15 ± 9.82	4.42 ± 9.71
Median (Q1, Q3)	5.82 (2.38, 12.51)	1.31 (0.32, 4.26)	6.77 (2.86, 10.62)	1.33 (0.31, 4.14)
Extreme growth rate of D-dimer level within 2 weeks				
N	294	916	86	236
Mean ± SD	0.729 ± 12.14	−0.009 ± 0.194	0.036 ± 0.205	−0.05 ± 0.362
Median (Q1, Q3)	0.018 (−0.018, 0.077)	0 (−0.019, 0.02)	0.015 (−0.029, 0.076)	−0.001 (−0.03, 0.018)
Maximum C-reactive protein level within 2 weeks				
N	433	1995	134	523
Mean ± SD	54.52 ± 57.31	47.36 ± 60.67	58.09 ± 55.12	46.86 ± 58.36
Median (Q1, Q3)	31.52 (7.87, 87.14)	12.7 (3.4, 81)	41.99 (10.5, 99.9)	13.5 (3.4, 79.51)

Table 2. (continued)

	Training (n = 2867)		Testing (n = 752)	
	PE positive (n = 485)	PE negative (n = 2382)	PE positive (n = 144)	PE negative (n = 608)
Extreme growth rate of C-reactive protein level within 2 weeks				
N	245	798	78	203
Mean ± SD	0.411 ± 1.433	0.215 ± 1.48	0.27 ± 1.006	0.06 ± 1.17
Median (Q1, Q3)	0.022 (−0.255, 0.512)	−0.019 (−0.392, 0.405)	0.138 (−0.302, 0.684)	−0.021 (−0.398, 0.381)
Any primary care within 1 month				
Yes	243 (50.1%)	761 (31.95%)	80 (55.56%)	212 (34.87%)
No	242 (49.9%)	1621 (68.05%)	64 (44.44%)	396 (65.13%)
Base excess level				
Low	24 (8.28%)	127 (10.73%)	8 (8.42%)	23 (7.8%)
Normal	171 (58.97%)	743 (62.75%)	51 (53.68%)	195 (66.1%)
High	95 (32.76%)	314 (26.52%)	36 (37.89%)	77 (26.1%)

Abbreviations: PE, pulmonary embolism; GBDT, gradient boosting decision tree.

we collected the variables of a specific time window and then determined variables suitable for the PE risk prediction model through ML analysis (the 17 variables above).

In recent years, many studies have confirmed that plasma D-dimer (the degradation product of cross-linked fibrin) can be used as a diagnostic tool for PE.[34,35]

The C-reactive protein levels were significantly higher in patients with PE and concomitant pneumonia,[36] which points to the fact that the C-reactive protein is a risk factor for PE. A previous study indicated that PE can be excluded in the standard C-reactive protein test, which is either performed alone or combined with other assessments.[37] Moreover, it was discovered that the C-reactive protein was associated with right ventricular dysfunction, which is a predictor of PE and may be a promising biomarker for PE risk stratification.[38]

The activity of the coagulation system, which is directly related to thrombosis, increases in high-risk PE patients.[39] A study revealed that increased plasma fibrinogen levels are associated with an increased risk of PE in combination with DVT.[40] In addition, preoperatively high fibrinogen and low plasminogen
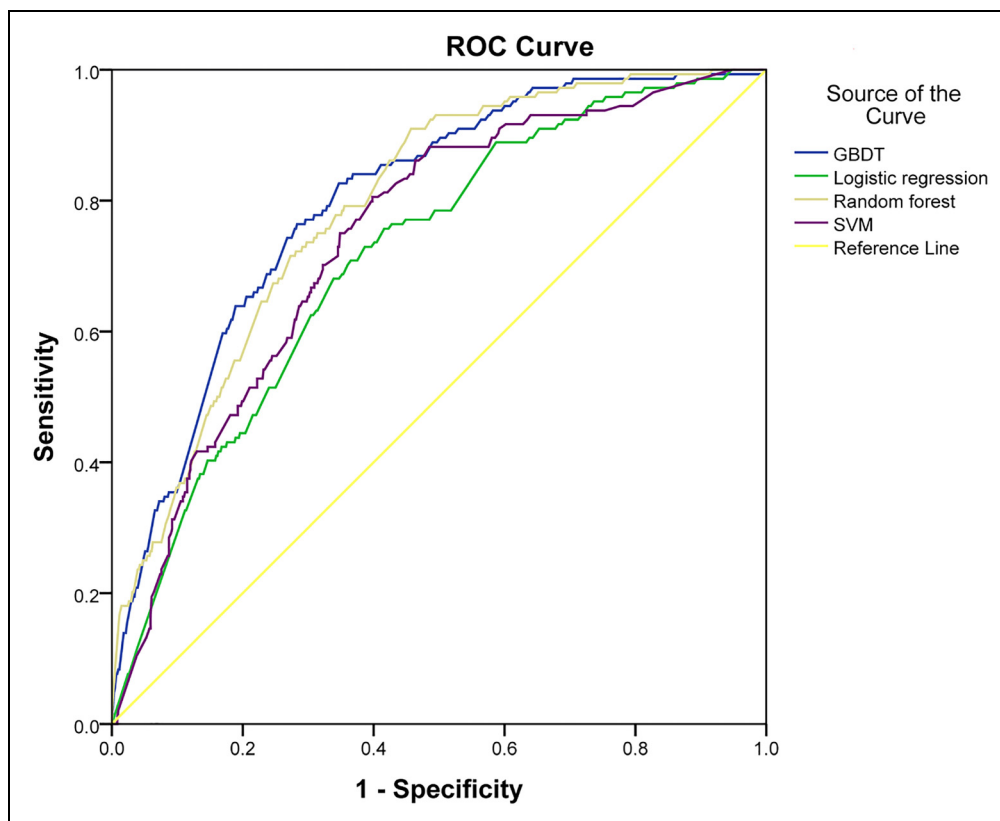

Figure 1. Receiver operating curves for the prediction of pulmonary embolism (PE) risk of different machine learning models (validation set).
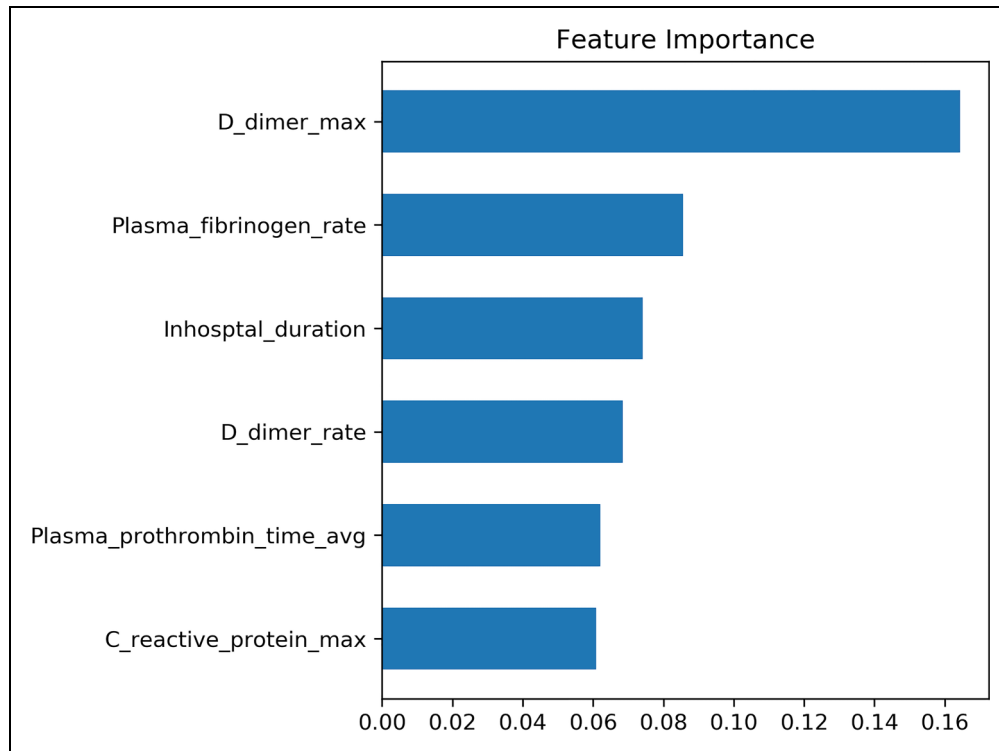
Figure 2. Importance of the top 10 risk factors in the prediction model of machine learning: average scores of each feature among the overall gbtree models after cross-validation and downsampling.
Abbreviations: _dimer_max, maximum D-dimer level; plasma_fibrinogen_rate, extreme growth rate of plasma fibrinogen level; D_dimer_rate, extreme growth rate of D-dimer level; plasma_prothrombin_time_avg, plasma prothrombin time average growth rate; C_reactive_protein_max, maximum C-reactive protein level.

levels are associated with poor long-term outcomes after pulmonary endarterectomy in patients with chronic thromboembolic pulmonary hypertension.[41]

The effect of albumin and arterial blood gas on the risk of PE has also been confirmed in several studies. An experimental study revealed that ischemic-modified albumin levels may contribute to PE.[42] Inconsistent findings regarding the effect of PE diagnosis were obtained from arterial blood gas value analysis. In a retrospective study, Cvitanic and Marino[43] discovered that both the arterial blood carbon dioxide partial pressure ($PaCO_2$) and alveolar–arterial oxygen partial pressure difference (P[A-a]O_2) can be used as a basis for the exclusion of APE.

In our study, multiple time window detection indicators were used as model variables. Each indicator exhibits a time trend characteristic. Using the ML method, variables with a better correlation with the outcome (PE) were selected, thereby facilitating PE risk prediction. To the best of our knowledge, the proposed PE risk prediction model is the first risk prediction model to use variables within multiple time windows. The variables used in the risk prediction model in previous studies were variables at fixed time points. Furthermore, the detection of variables at a certain time point was affected by multiple factors, and the detected values differed from the original level. Furthermore, we estimated the frequency of the feature used in the ensemble of decision trees in the GBDT; D_dimer_max contributed the most to the outcome prediction of the ML models, followed by the plasma_fibrinogen_rate, inhospital_duration, and D_dimer_rate. Based on the detected values of the variables at different times, the risk of APE in hospitalized patients can be estimated through the model, such that clinicians can monitor the risk of APE in patients on a regular basis.

Table 3. Predictive Efficacy Analysis (Verification Set) of Different Machine Learning Models for PE.

Model	AUC (95% CI)	Sensibility	Specificity	Accuracy	F1
GBDT	0.799 (0.762, 0.837)	63.9%	81.1%	77.8%	0.524
Logistic regression	0.716 (0.672, 0.761)	68.1%	66.1%	66.5%	0.438
Random forest	0.791 (0.753, 0.828)	71.5%	72.7%	72.5%	0.499
SVM	0.743 (0.701, 0.785)	75%	65.1%	67%	0.466

Abbreviations: PE, pulmonary embolism; GBDT, gradient boosting decision tree; SVM, support vector machine.

However, this study has several limitations. First, this is a retrospective study. In addition, we have included patients only from a single center, and thus, the application of our findings in other populations and institutions needs to be further verified.

Conclusion

The GBDT model exhibited the best performance in terms of PE risk prediction in our study. However, before it is applied in clinical practice to provide support for clinical decision-making, the predictive performance of the model needs to be prospectively verified.

ORCID iD

Jiyu Li https://orcid.org/0000-0002-4851-0312

References

1. Zakai NA, McClure LA. Racial differences in venous thromboembolism. *Journal of Thrombosis and Haemostasis: JTH*. 2011;9(10):1877-1882. doi:10.1111/j.1538-7836.2011.04443.x
2. Cushman M. Epidemiology and risk factors for venous thrombosis. *Semin Hematol*. 2007;44(2):62-69. doi:10.1053/j.seminhematol. 2007.02.004
3. Cohen AT, Agnelli G, Anderson FA, et al. Venous thromboembolism (VTE) in Europe. The number of VTE events and associated morbidity and mortality. *Thromb Haemostasis*. 2007;98(4):756-764.
4. Huisman MV, Barco S, Cannegieter SC, et al. Pulmonary embolism. *Nature Reviews Disease Primers*. 2018;7(4):18028. doi:10. 1038/nrdp.2018.28
5. Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. *Arch Intern Med*. 2011;171(9):831-837. doi:10.1001/archinternmed. 2011.178
6. Smith SB, Geske JB, Kathuria P, et al. Analysis of national trends in admissions for pulmonary embolism. *Chest*. 2016;150(1): 35-45. doi:10.1016/j.chest.2016.02.638
7. Wang HC, Tsai PS, Li KY, Fan YC, Huang CJ. Perioperative risk factors for postpartum pulmonary embolism in Taiwanese cesarean section women. *Asian J Anesthesiol*. 2017;55(2):35-40. doi:10.1016/j.aja.2017.05.002
8. Bougouin W, Marijon E, Planquette B, et al. Factors associated With pulmonary embolism-related sudden cardiac arrest. *Circulation*. 2016;134(25):2125-2127. doi:10.1161/circulationaha. 116.024746
9. Prevention. CfDCa. National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention. Venous Thromboembolism (Blood Clots). Data and Statistics on Venous Thromboembolism. Atlanta: CDC. wwwcdcgov/ncbddd/dvt/datahtml. 2015
10. Goldhaber SZ, Visani L, De Rosa M. Acute pulmonary embolism: clinical outcomes in the international cooperative pulmonary embolism registry (ICOPER). *Lancet (London, England)*. 1999; 353(9162):1386-1389. doi:10.1016/s0140-6736(98)07534-5
11. Klok FA, van der Hulle T, den Exter PL, Lankeit M, Huisman MV, Konstantinides S. The post-PE syndrome: a new concept for chronic complications of pulmonary embolism. *Blood Rev*. 2014;28(6):221-226. doi:10.1016/j.blre.2014.07.003
12. Klok FA, Barco S. Follow-up after acute pulmonary embolism. *Hamostaseologie*. 2018;38(1):22-32. doi:10.5482/hamo-17-06-0020
13. Konstantinides SV, Torbicki A, Agnelli G, et al. 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J*. 2014;35(43):3033-3069. 69a-69k. doi:10.1093/ eurheartj/ehu283.
14. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer; 2009.
15. Kim JS, Merrill RK, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine*. 2018;43(12):853-860. doi:10.1097/brs.0000000000002442
16. Ferroni P, Zanzotto FM, Scarpato N, Riondino S, Guadagni F. Validation of a machine learning approach for venous thromboembolism risk prediction in oncology. *Dis Markers*. 2017;2017:8781379. doi:10.1155/2017/8781379
17. Ferroni P, Zanzotto FM, Scarpato N, et al. Risk assessment for venous thromboembolism in chemotherapy-treated ambulatory cancer patients. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*. 2017;37(2): 234-242. doi:10.1177/0272989x16662654
18. Kawaler E, Cobian A, Peissig P, Cross D, Yale S, Craven M. Learning to predict post-hospitalization VTE risk from EHR data. AMIA annual symposium proceedings. *AMIA Symposium*. 2012;2012:436-445.
19. Girardi AM, Bettiol RS, Garcia TS, et al. Wells and Geneva scores are not reliable predictors of pulmonary embolism in critically ill patients: a retrospective study. *J Intensive Care Med*. 2020;35(10): 1112–1117. doi:10.1177/0885066618816280
20. Siddiqi N. Credit risk scorecards :developing and implementing intelligent credit scoring2005.

21. Chen T, Guestrin C, eds. XGBoost: A Scalable Tree Boosting System. Acm Sigkdd International Conference on Knowledge Discovery & Data Mining; 2016.

22. Bergstra J, Yamins D, Cox DD. Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms. *Python in Science Conference.* 2013.

23. Miniati M, Monti S, Bottai M. A structured clinical model for predicting the probability of pulmonary embolism. *Am J Med.* 2003;114(3):173-179. doi:10.1016/s0002-9343(02)01478-x

24. Hendriksen JM, Geersing GJ, Lucassen WA, et al. Diagnostic prediction models for suspected pulmonary embolism: systematic review and independent external validation in primary care. *Br Med J.* 2015;351:h4438. doi:10.1136/bmj.h4438

25. Ceriani E, Combescure C, Le Gal G, et al. Clinical prediction rules for pulmonary embolism: a systematic review and meta-analysis. *Journal of Thrombosis and Haemostasis : JTH.* 2010;8(5):957-970. doi:10.1111/j.1538-7836.2010.03801.x

26. Coelho J, Divernet-Queriaud M, Roy P-M, Penaloza A, Le Gal G, Trinh-Duc A. Comparison of the wells score and the revised Geneva score as a tool to predict pulmonary embolism in outpatients over age 65. *Thromb Res.* 2020;196:120-126.

27. Touhami O, Marzouk SB, Bennasr L, et al. Are the wells score and the revised geneva score valuable for the diagnosis of pulmonary embolism in pregnancy? *European Journal of Obstetrics & Gynecology and Reproductive Biology.* 2018;221:166-171.

28. Obi AT, Pannucci CJ, Nackashi A, et al. Validation of the Caprini venous thromboembolism risk assessment model in critically Ill surgical patients. *JAMA Surg.* 2015;150(10):941-948. doi:10.1001/jamasurg.2015.1841

29. Lobastov K, Barinov V, Schastlivtsev I, Laberko L, Rodoman G, Boyarintsev V. Validation of the Caprini risk assessment model for venous thromboembolism in high-risk surgical patients in the background of standard prophylaxis. *Journal of Vascular Surgery Venous and Lymphatic Disorders.* 2016;4(2):153-160. doi:10.1016/j.jvsv.2015.09.004

30. Grant PJ, Greene MT, Chopra V, Bernstein SJ, Hofer TP, Flanders SA. Assessing the Caprini score for risk assessment of venous thromboembolism in hospitalized medical patients. *Am J Med.* 2016;129(5):528-535. doi:10.1016/j.amjmed.2015.10.027

31. Huynh N, Fares WH, Brownson K, et al. Risk factors for presence and severity of pulmonary embolism in patients with deep venous thrombosis. *Journal of Vascular Surgery Venous and Lymphatic Disorders.* 2018;6(1):7-12. doi:10.1016/j.jvsv.2017.08.015

32. Turetz M, Sideris AT, Friedman OA, Triphathi N, Horowitz JM. Epidemiology, pathophysiology, and natural history of pulmonary embolism. *Semin Intervent Radiol.* 2018;35(2):92-98. doi:10.1055/s-0038-1642036

33. Doherty S. Pulmonary embolism An update. *Aust Fam Physician.* 2017;46(11):816-820.

34. Bounameaux H, de Moerloose P, Perrier A, Miron MJ. D-dimer testing in suspected venous thromboembolism: an update. *QJM: Monthly Journal of the Association of Physicians.* 1997;90(7):437-442. doi:10.1093/qjmed/90.7.437

35. Miron MJ, Perrier A, Bounameaux H, et al. Contribution of non-invasive evaluation to the diagnosis of pulmonary embolism in hospitalized patients. *Eur Respir J.* 1999;13(6):1365-1370.

36. Cha SI, Choi KJ, Shin KM, et al. Clinical characteristics of pulmonary embolism with concomitant pneumonia. *Blood Coagulation & Fibrinolysis : an International Journal in Haemostasis and Thrombosis.* 2016;27(3):281-286. doi:10.1097/mbc.0000000000000411

37. Steeghs N, Goekoop RJ, Niessen RW, Jonkers GJ, Dik H, Huisman MV. C-reactive protein and D-dimer with clinical probability score in the exclusion of pulmonary embolism. *Br J Haematol.* 2005;130-(4):614-619. doi:10.1111/j.1365-2141.2005.05652.x

38. Abul Y, Karakurt S, Ozben B, Toprak A, Celikel T. C-reactive protein in acute pulmonary embolism. *Journal of Investigative Medicine: The Official Publication of the American Federation for Clinical Research.* 2011;59(1):8-14.

39. Lehnert P, Johansson PI, Ostrowski SR, et al. Coagulopathy in patients with acute pulmonary embolism: a pilot study of whole blood coagulation and markers of endothelial damage. *Scand J Clin Lab Invest.* 2017;77(1):19-26. doi:10.1080/00365513.2016.1239130

40. Klovaite J, Nordestgaard BG, Tybjaerg-Hansen A, Benn M. Elevated fibrinogen levels are associated with risk of pulmonary embolism, but not with deep venous thrombosis. *Am J Respir Crit Care Med.* 2013;187(3):286-293. doi:10.1164/rccm.201207-1232OC

41. Kato F, Tanabe N, Ishida K, et al. Coagulation-fibrinolysis system and postoperative outcomes of patients with chronic thromboembolic pulmonary hypertension. *Circulation Journal: Official Journal of the Japanese Circulation Society.* 2016;80(4):970-979. doi:10.1253/circj.CJ-15-1208

42. Turedi S, Patan T, Gunduz A, et al. Ischemia-modified albumin in the diagnosis of pulmonary embolism: an experimental study. *Am J Emerg Med.* 2009;27(6):635-640. doi:10.1016/j.ajem.2008.05.002

43. Cvitanic O, Marino PL. Improved use of arterial blood gas analysis in suspected pulmonary embolism. *Chest.* 1989;95(1):48-51. doi:10.1378/chest.95.1.48