Am J Transl Res 2024;16(12):7438-7447
www.ajtr.org /ISSN:1943-8141/AJTR0160289

Original Article

Efficacy of three predictive models for deep vein thrombosis in patients with lumbar disc herniation

Shuai Yang, Qingfeng Guo, Yaqing Xing, Erjun Liu, Fugang Zhao, Weiling Zhang

Department of Traditional Chinese Medicine, The First Hospital of Hebei Medical University, Shijiazhuang 050091, Hebei, China

Received September 2, 2024; Accepted November 14, 2024; Epub December 15, 2024; Published December 30, 2024

Abstract: Objective: To develop predictive models for assessing deep vein thrombosis (DVT) risk among lumbar disc herniation (LDH) patients and evaluate their performances. Methods: A retrospective study was conducted on 798 LDH patients treated at the First Hospital of Hebei Medical University from January 2017 to December 2023. The patients were divided into a training set (n = 558) and a test set (n = 240) using computer-generated random numbers in a ratio of 7:3. Patients without DVT in the training set were categorized as the non-DVT group (n = 463), while those diagnosed with DVT were the DVT group (n = 95). Univariate analysis was performed to compare clinical data between the two groups. Data with statistical significance were used for the development of a Logistic regression model, Gradient boosting model, and Random Forest model. Model performance was evaluated through receiver operating characteristic (ROC) curve analysis and calibration curve assessment. Results: In the training set, univariate analysis revealed significant differences in age, platelets (PLT), cholesterol (TC), triglycerides (TG), glycated hemoglobin (HbAlc), D-dimer (D-D), fibrinogen (FIB), activated partial thromboplastin time (APTT), prothrombin time (PT), and thrombin time (TT) between the non-DVT group and the DVT group (all $P<0.05$). Predictive models were constructed based on these indicators. The areas under the ROC curves (AUCs) in the training set were as follows (in descending order): Random Forest model (0.978) > Gradient boosting model (0.943) > Logistic regression model (0.919). In the test set, the AUCs were: Random Forest model (0.952) > Gradient boosting model (0.941) > Logistic regression model (0.908). The DeLong test indicated that the AUC of the Random Forest model in the training set was significantly higher than that of the Logistic regression model ($P<0.05$); however, no significant difference was observed between the other two models. Calibration curves demonstrated that the predictive probabilities from all three models closely aligned with actual DVT incidence in both sets. Conclusion: The Logistic regression model, Gradient boosting model, and Random Forest model constructed in this study exhibit good predictive value for the occurrence of DVT in LDH patients, aiding in the optimization of clinical management of clinical management. Among them, the Random Forest model performed the best of the three.

Keywords: Lumbar disc herniation, deep vein thrombosis, risk factor, Logistic regression model, Gradient boosting model, Random Forest model

Introduction

Lumbar disc herniation (LDH) is a prevalent condition in orthopedic spine disorders, primarily stemming from degenerative alterations in the nucleus pulposus, fibrous ring, and cartilage plate of the lumbar intervertebral disc. These alterations lead to the protrusion of nucleus pulposus through the ruptured fibrous ring, compressing and irritating the nerve roots and cauda equina. This, in turn, causes lumbar pain and radicular pain in the lower limbs [1]. The progression of LDH is often prolonged, with a high recurrence rate. Throughout this extended course, patients not only experience a marked decline in their quality of life but also may have restricted physical activities due to pain, resulting in prolonged bed rest. These lifestyle alterations inevitably diminish blood circulation to the lower extremities, contributing to venous stasis in the leg veins [2, 3]. Such hemodynamic changes create an environment conducive to the development of deep vein thrombosis (DVT), a thrombotic condition that insidiously manifests within the deep venous system, predominantly affecting the lower limbs. Notably, early symptoms of DVT are often subtle and challenging for patients to recognize;

https://doi.org/10.62347/TWTG6803

thus, they are frequently overlooked [4]. If left untreated or inadequately managed, thrombus can dislodge and travel to the pulmonary artery, leading to fatal pulmonary embolism or precipitating severe complications such as myocardial infarction and even sudden cardiac death - thereby posing significant threats to patient safety [5, 6]. Early and accurate prediction of DVT occurrence holds paramount significance for improving the prognosis among LDH patients. However, there is limited information on the incidence rates and the risk factors for DVT in this patient population. Therefore, this study aims to identify the key risk factors for DVT in LDH patients and develop predictive models based on these factors. The objective is to enhance clinical capabilities for early detection of DVT risk in LDH patients, thereby facilitating more proactive preventive measures and reducing the incidence of DVT.

Materials and methods

Patients

A retrospective cohort study was conducted on 798 patients diagnosed with LDH who underwent treatment at The First Hospital of Hebei Medical University between January 2017 and December 2023. The participants were stratified into a training cohort (n = 558) and a test cohort (n = 240) using computer-generated randomization in a ratio of 7:3. The training cohort was employed for model validation. This study was approved by the Ethics Committee of The First Hospital of Hebei Medical University.

Inclusion criteria: (1) Clinically confirmed diagnosis of primary LDH; (2) Age >18 years; (3) Receipt of conservative management; (4) Availability of complete requisite data. Exclusion criteria: (1) Presence of other musculoskeletal conditions; (2) History of DVT; (3) Long-term usage of anticoagulant or antiplatelet medications; (4) Significant cardiac, hepatic, or renal impairment; (5) Severe cardiovascular, cerebrovascular, malignant neoplasms, coagulation abnormalities, or infectious ailments; (6) Pregnancy or lactation; (7) Loss to follow-up.

DVT diagnosis

A color Doppler ultrasound device operating at a probe frequency of 9.0 MHz was used for diagnosis. The patient was positioned in a supine posture with the lower limbs slightly extended and externally rotated, ensuring relaxation and encouraging regular deep breathing. Intermittent compression was applied to assess the femoral vein, superficial femoral vein, deep femoral vein, popliteal vein, posterior tibial vein, anterior tibial vein, peroneal vein, and calf muscle interosseous venous network for the presence of any thrombus indicative of DVT.

Data collection

A comprehensive review of the literature on DVT risk factors informed the data collection process. The hospital's electronic information system was employed to collect the following data: age, gender, body mass index (BMI), smoking history, underlying diseases (hypertension, diabetes, hyperlipidemia), hospital stay, and laboratory tests measured within 24 hours of admission [red blood cells (RBC), white blood cells (WBC), platelets (PLT), hemoglobin (Hb), cholesterol (TC), triglycerides (TG), fasting blood glucose (FBG), glycated hemoglobin (HbAlc), D-dimer (D-D), fibrinogen (FIB), activated partial thromboplastin time (APTT), prothrombin time (PT), and thrombin time (TT)].

Statistical methods

Statistical analysis was performed using SPSS 26.0. Qualitative data were presented as frequency and percentage [n (%)], and the Chi-square test was applied for comparison. Following univariate analysis, Logistic regression model, Gradient boosting model, and Random Forest model were developed using RStudio software. The discriminative ability and calibration of these models were evaluated by receiver operating characteristic (ROC) curve and calibration curve analyses. Comparisons of the area under the ROC curve (AUC) were conducted using the DeLong test. A *P*-value of less than 0.05 was considered a significant difference.

Results

Univariate analysis of clinical data in the training set and test set

In the training set, 463 patients without DVT were included in the non-DVT group, while 95 patients with DVT were included in the DVT group. In the DVT group, the proportion of patients with age >52 years, PLT>306 × 10^9/L,

Table 1. Univariate analysis of clinical data [n (%)]

Data	Training set				Test set			
	Non-DVT group (n = 463)	DVT group (n = 95)	χ^2	*P*	Non-DVT group (n = 193)	DVT group (n = 47)	χ^2	*P*
Age			29.130	<0.001			42.969	<0.001
≤52 years	299 (64.58)	33 (37.74)			134 (69.43)	8 (17.02)		
>52 years	164 (35.42)	62 (65.26)			59 (30.57)	39 (82.98)		
Gender			1.102	0.294			6.067	0.014
Female	142 (30.67)	24 (25.26)			59 (30.57)	6 (12.77)		
Male	321 (69.33)	71 (74.74)			134 (69.43)	41 (87.23)		
BMI			2.683	0.101			7.349	0.007
≤23.1 kg/m^2	265 (57.24)	63 (66.32)			106 (54.92)	36 (76.60)		
>23.1 kg/m^2	198 (42.76)	32 (33.68)			87 (45.08)	11 (23.40)		
Smoking history	162 (34.99)	37 (38.95)	0.538	0.463	70 (36.27)	25 (53.19)	4.526	0.033
Hypertension	147 (31.75)	26 (27.37)	0.707	0.400	56 (29.02)	14 (29.79)	0.135	0.713
Diabetes	83 (17.93)	13 (13.68)	0.996	0.318	32 (16.24)	5 (10.64)	1.023	0.312
Hyperlipidemia	96 (20.73)	18 (18.95)	0.155	0.694	51 (26.42)	5 (10.64)	5.265	0.022
Length of stay			1.060	0.303			0.013	0.909
≤8 days	126 (27.21)	21 (22.11)			55 (28.50)	13 (27.66)		
>8 days	337 (72.79)	74 (77.89)			138 (71.50)	34 (72.34)		
RBC			0.337	0.562			3.558	0.059
≤6.2 × 10^{12}/L	210 (45.36)	40 (42.11)			91 (47.15)	15 (31.91)		
>6.2 × 10^{12}/L	253 (54.64)	55 (57.89)			102 (52.85)	32 (68.09)		
WBC			2.928	0.087			0.003	0.958
≤8.3 × 10^9/L	320 (69.11)	74 (77.89)			143 (74.09)	35 (74.47)		
>8.3 × 10^9/L	143 (30.89)	21 (22.11)			50 (25.91)	12 (25.53)		
PLT			43.130	<0.001			10.327	0.001
≤306 × 10^9/L	329 (71.06)	34 (35.79)			138 (71.50)	22 (46.81)		
>306 × 10^9/L	134 (28.94)	61 (64.21)			55 (28.50)	25 (53.19)		
Hb			1.173	0.279			0.366	0.545
≤131 g/L	262 (56.59)	48 (50.53)			108 (55.96)	24 (51.06)		
>131 g/L	201 (43.41)	47 (49.47)			85 (44.04)	23 (48.94)		
TC			6.819	0.009			0.111	0.739
≤5.6 mmol/L	275 (59.40)	70 (73.68)			116 (60.10)	27 (57.45)		
>5.6 mmol/L	188 (40.60)	25 (26.32)			77 (39.90)	20 (42.55)		
TG			134.140	<0.001			38.927	<0.001
≤1.0 mmol/L	376 (81.21)	21 (22.11)			146 (75.65)	13 (27.66)		
>1.0 mmol/L	87 (18.79)	74 (77.89)			47 (24.35)	34 (72.34)		
FBG			2.176	0.140			1.803	0.179
≤4.5 mmol/L	121 (26.13)	18 (18.95)			51 (26.42)	8 (17.02)		
>4.5 mmol/L	342 (73.87)	77 (81.05)			142 (73.58)	39 (82.98)		
HbAlc			22.116	<0.001			16.281	<0.001
≤6.5%	372 (80.35)	55 (57.89)			154 (79.79)	24 (51.06)		
>6.5%	91 (19.65)	40 (42.11)			39 (20.21)	23 (48.94)		
D-D			221.946	<0.001			108.087	<0.001
≤0.4 mg/L	409 (88.34)	16 (16.84)			175 (90.67)	9 (19.15)		
>0.4 mg/L	54 (11.66)	79 (83.16)			18 (9.33)	38 (80.85)		
FIB			121.750	<0.001			20.423	<0.001
≤4.5 g/L	351 (75.81)	16 (16.84)			141 (73.06)	18 (38.30)		
>4.5 g/L	112 (24.19)	79 (83.16)			52 (26.94)	29 (61.70)		
APTT			60.940	<0.001			14.970	<0.001
≤30 s	201 (43.41)	83 (87.37)			83 (43.01)	35 (74.47)		
>30 s	262 (56.59)	12 (12.63)			110 (56.99)	12 (25.53)		

PT			269.848	<0.001			103.901	<0.001
≤9 s	39 (8.42)	80 (84.21)			16 (8.29)	36 (76.60)		
>9 s	424 (91.58)	15 (15.79)			177 (91.71)	11 (23.40)		
TT			40.825	<0.001			16.268	<0.001
≤14 s	240 (51.84)	83 (87.37)			93 (48.19)	38 (80.85)		
>14 s	223 (48.16)	12 (12.63)			100 (51.82)	9 (19.15)		

BMI: body mass index; RBC: red blood cell count; WBC: white blood cell count; PLT: platelet; Hb: hemoglobin; TC: cholesterol; TG: triglycerides; FBG: fasting blood glucose; HbAlc: glycated hemoglobin; D-D: D-dimer; FIB: fibrinogen; APTT: activated partial thromboplastin time; PT: prothrombin time; TT: thrombin time.

Table 2. Assignment of independent variables

Independent variable	Assignment
Age	0 = ≤52 years; 1 = >52 years
PLT	0 = ≤306 × 10^9/L; 1 = >306 × 10^9/L
TC	0 = ≤5.6 mmol/L; 1 = >5.6 mmol/L
TG	0 = ≤1.0 mmol/L; 1 = >1.0 mmol/L
HbAlc	0 = ≤6.5%; 1 = >6.5%
D-D	0 = ≤0.4 mg/L; 1 = >0.4 mg/L
FIB	0 = ≤4.5 g/L; 1 = >4.5 g/L
APTT	0 = ≤30 s; 1 = >30 s
PT	0 = ≤9 s; 1 = >9 s
TT	0 = ≤14 s; 1 = >14 s

PLT: platelets; TC: cholesterol; TG: triglycerides; HbAlc: glycated hemoglobin; D-D: D-dimer; FIB: fibrinogen; APTT: activated partial thromboplastin time; PT: prothrombin time; TT: thrombin time.

TC≤5.6 mmol/L, TG>1.0 mmol/L, HbAlc>6.5%, D-D>0.4 mg/L, FIB>4.5 g/L, APTT≤30 s, PT≤9 s, and TT≤14 s were significantly higher than in the non-DVT group (all P<0.05). In the test set, 193 patients were included in the non-DVT group and 47 patients were included in the DVT group. In the DVT group, the proportion of patients with age >52 years, male sex, BMI≤23.1 kg/m^2, smoking history, combined hyperlipidemia, PLT>306 × 10^9/L, TG>1.0 mmol/L, HbAlc>6.5%, D-D>0.4 mg/L, FIB>4.5 g/L, APTT≤30 s, PT≤9 s, and TT≤14 s were significantly higher than in the non-DVT group (all P<0.05). Details are shown in **Table 1**.

Construction of the predictive models

The occurrence of DVT was considered as the dependent variable (assigned values: 0 = absent, 1 = present). Age, PLT, TC, TG, HbAlc, D-D, FIB, APTT, PT, and TT were included as independent variables in the logistic regression analysis. The assignments are detailed in **Table 2**. The results indicated that TG, D-D, FIB, APTT, PT, and TT were significant independent factors influencing the occurrence of DVT in patients, as presented in **Table 3**. Subsequently, a Logistic regression model was developed with the specific formula: logit (P) = -2.550 + 2.433 × TG + 2.924 × D-D + 2.644 × FIB - 3.107 × APTT - 4.553 × PT - 1.570 × TT. A line chart was generated to visualize the model performance, depicted in **Figure 1**.

For the Gradient boosting model, age, PLT, TC, TG, HbAlc, D-D, FIB, APTT, PT, and TT were used as independent variables. The model was optimized with the following parameters: shrinkage = 0.01, cv.folds = 10, n.trees = 2000, interaction.depth = 1. The optimal generalization error was achieved when n.trees = 1280. Furthermore, the model determined the relative importance of each independent variable, ranked in descending order as: PT, D-D, TG, FIB, APTT, PLT, TT, HbAlc, age, and TC, as illustrated in **Figure 2**.

For the Random Forest model, age, PLT, TC, TG, HbAlc, D-D, FIB, APTT, PT, and TT were used as independent variables. The model parameters were set to ntree = 500, mtry = 3, with stable error performance. The mean decrease Gini values of the variables were ranked in descending order as follows: PT, D-D, FIB, TG, APTT, PLT, TT, age, HbAlc, and TC, as depicted in **Figure 3**.

Validation of performance of predictive models

In the training set, AUCs were ranked in descending order as follows: the Random Forest model (0.978, 95% CI: 0.963-0.992) > the Gradient boosting model (0.943, 95% CI: 0.904-0.982) > the Logistic regression model (0.919, 95% CI: 0.866-0.952). Similarly, in the test set, AUCs followed a similar pattern with

Table 3. Logistic regression analysis

Independent variable	*B*	*SE*	*Wald*	*P*	*OR (95% CI)*
Age	0.588	0.609	0.934	0.334	1.801 (0.546-5.941)
PLT	0.906	0.633	2.045	0.153	2.474 (0.715-8.561)
TC	-0.074	0.605	0.015	0.903	0.929 (0.284-3.042)
TG	2.433	0.618	15.493	<0.001	11.397 (3.393-38.282)
HbAlc	1.071	0.668	2.572	0.109	2.920 (0.788-10.815)
D-D	2.924	0.661	19.574	<0.001	18.613 (5.097-67.976)
FIB	2.644	0.681	15.072	<0.001	14.071 (3.703-53.465)
APTT	-3.107	0.852	13.307	<0.001	0.045 (0.008-0.237)
PT	-4.553	0.768	35.122	<0.001	0.011 (0.002-0.047)
TT	-1.570	0.702	5.000	0.025	0.208 (0.053-0.824)
Constant	-2.550	0.813	9.830	0.002	-

PLT: platelets; TC: cholesterol; TG: triglycerides; HbAlc: glycated hemoglobin; D-D: D-dimer; FIB: fibrinogen; APTT: activated partial thromboplastin time; PT: prothrombin time; TT: thrombin time.

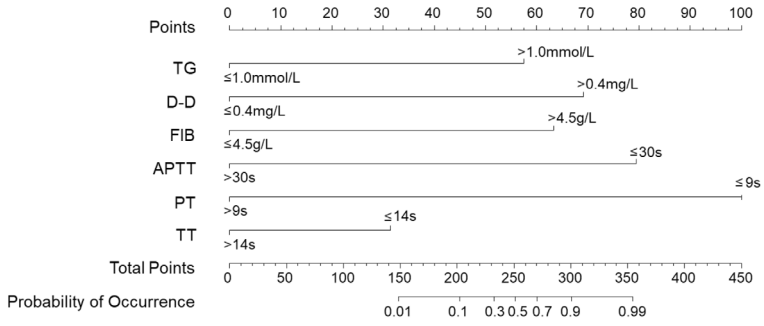

Figure 1. Line chart. TG: triglycerides; D-D: D-dimer; FIB: fibrinogen; APTT: activated partial thromboplastin time; PT: prothrombin time; TT: thrombin time.

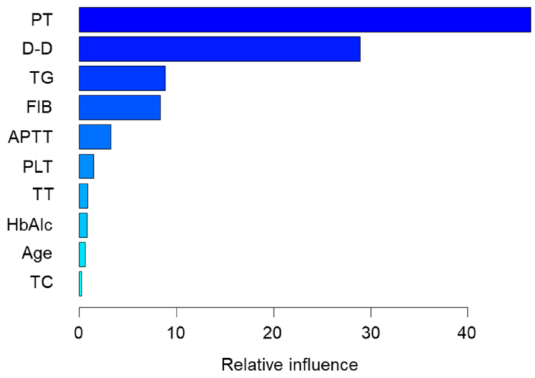

Figure 2. Relative importance of variables in the Gradient boosting model. PT: prothrombin time; D-D: D-dimer; TG: triglycerides; FIB: fibrinogen; APTT: activated partial thromboplastin time; PLT: platelets; TT: thrombin time; HbAlc: glycated hemoglobin; TC: cholesterol.

the Random Forest model (0.952, 95% CI: 0.928-0.977) > Gradient boosting model (0.941, 95% CI: 914-967) > the Logistic regression model (0.908, 95% CI: 0.856-0.960), as illustrated in **Figure 4**.

The DeLong test indicated that the AUC of the Random Forest model in the training set was significantly greater than that of the Logistic regression model ($P<0.05$); however, no significant difference was observed between the other two models, as detailed in **Table 4**. The calibration curves demonstrated a close alignment between the predicted probability of DVT incidence and the actual occurrence across both training and test sets for all three models, as illustrated in **Figure 5**.

Discussion

DVT can pose an immediate threat to a patient's life [7], making prompt diagnosis and treatment crucial. While venography is considered as the 'gold standard' for DVT diagnosis due to its high precision in visualizing deep veins using contrast-enhanced X-rays [8, 9], its invasive nature limits clinical use due to the potential renal burden from contrast agents. Presently, clinical DVT diagnosis relies heavily on imaging modalities like color Doppler ultrasound. However, diagnostic delays may occur in cases lacking specific symptoms or when patients cannot undergo these tests [10]. Therefore, it is essential to identify the DVT risk factors in patients with LDH and to develop a predictive model.

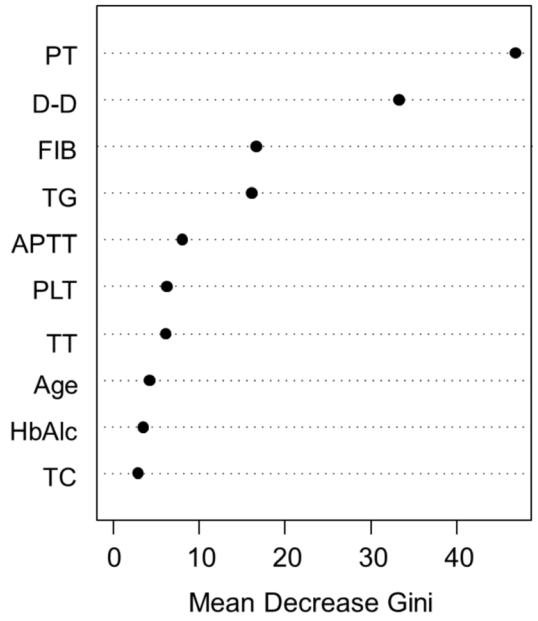

Figure 3. Relative importance of variables in the Random Forest model. PT: prothrombin time; D-D: D-dimer; FIB: fibrinogen; TG: triglycerides; APTT: activated partial thromboplastin time; PLT: platelets; TT: thrombin time; HbAlc: glycated hemoglobin; TC: cholesterol.

The development of DVT is widely believed to be associated with blood hypercoagulability, endothelial cell injury, and reduced blood flow. Several contributing factors have been identified [11]. Previous research [12-14] has demonstrated that age, gender, hypertension, alcohol consumption, and smoking are significant risk factors for the occurrence of DVT. Based on these factors, clinicians can identify high-risk populations for DVT and initiate early preventive measures. Univariate analysis in this study revealed significant differences in age, PLT, TC, TG, HbAlc, D-D, FIB, APTT, PT, and TT between the non-DVT group and the DVT group in the training set. Logistic regression analysis further identified TG, D-D, FIB, APTT, PT, and TT as independent influencing factors for DVT among patients. This can be attributed to the fact that elderly patients generally exhibit higher blood viscosity, which, in combination with vascular endothelial ageing, predisposes them to hypercoagulability and an increased risk of thrombosis [15]. Elevated TC and TG levels can reduce blood flow velocity, further increasing the risk of thrombosis. These lipid abnormalities can also damage endothelial cells, leading to inflammatory responses that promote clot formation. Additionally, elevated TC levels are linked to atherosclerosis, where plaque deposition on arterial walls can rupture, causing platelets to aggregate and clot formation [16, 17]. HbAlc is as a crucial marker for glucose control, with higher levels indicating poorer glucose control. Even in a non-diabetic population, an increase in glycemic index is associated with an elevated risk of blood clotting [18]. D-D is commonly utilized to assess the extent of thrombosis and fibrinolysis. Elevated D-D levels indicate ongoing clot breakdown, as fibrinolysin degrades fibrin in clots [19]. However, various diseases, including sepsis, can also elevate D-D levels [20]. FIB, APTT, PT, and TT reflect body's coagulation function, with abnormal values suggesting impaired coagulation [21]. Elevated FIB levels may suggest increased blood viscosity and heightened tendency for blood clotting, raising the risk of DVT [22]. Therefore, it is imperative to implement comprehensive preventive measures targeting the identified risk factors. Regular monitoring of the patient's TG, D-D, FIB, APTT, PT, and TT levels is essential for early detection of DVT. Additionally, vigilant observation of changes in the patient's condition is essential for promptly addressing any potential complications.

In the era of precision medicine, disease prediction models play a crucial role in managing disease risk. This study developed three predictive models to assess the risk of DVT in LDH patients based on identified risk factors. Logistic regression, a widely utilized statistical method in medical research, is part of the general linear regression family and is particularly effective in handling binary or multi-class dependent variables. In this study, the Logistic regression model exhibited AUCs of 0.919 and 0.908 for the training and test sets, respectively. In recent years, driven by the rapid advancement of information technology and artificial intelligence, there has been a growing trend towards constructing risk prediction models using machine learning methods. Gradient boosting, a machine learning technique utilized for regression and classification tasks, constructs a robust learner through iterative training and refinement of weak learners to minimize the loss function [23]. In this study, the constructed Gradient boosting model achieved AUC values of 0.943 and 0.941 on the training set and test set, respectively. Random Forest model, an ensemble learning method that en-

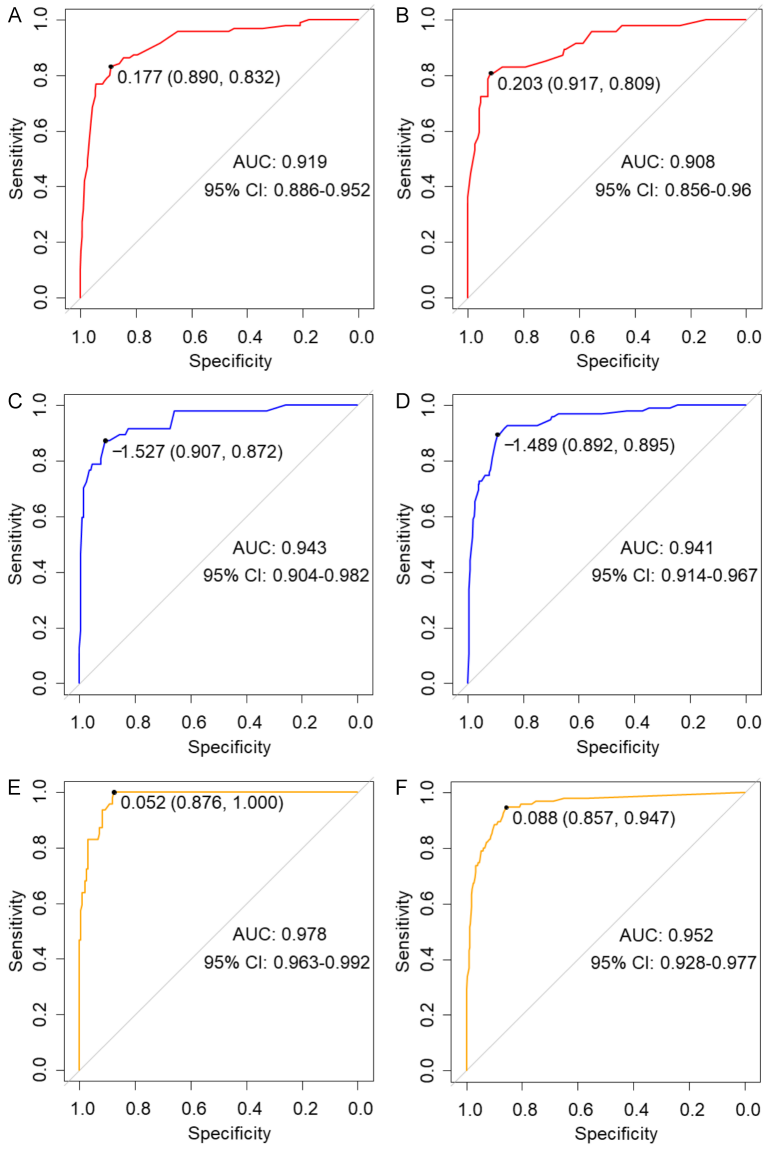

Figure 4. ROC curve analysis for the performance of three predictive models. A: Logistic regression model in training set; B: Logistic regression model in test set; C: Gradient boosting model in training set; D: Gradient boosting model in test set; E: Random Forest model in training set; F: Random Forest model in test set.

hances model accuracy and stability by constructing multiple decision trees and aggregating their predictions [24]. The AUC values for the training set and test set of the Random Forest model developed in this study were 0.978 and 0.952, respectively. All three models demonstrated strong discriminative ability, with the AUC ranking as follows: Random Forest model > Gradient boosting model > Logistic regression model. The DeLong test revealed a significantly higher AUC for the Random Forest model compared to the Logistic regression model in the training set. This difference may be attributed to the more stringent assumptions of the Logistic regression model, including independence of observed variables, absence of interaction effects, and linear relationship between logit (P) and independent variables. In contrast, the Random Forest model imposes fewer assumptions, is less sensitive to multicollinearity and less susceptible to overfitting. Additionally, it provides insight into feature importance. All three models exhibited high precision. Considering the current results, the application of Random Forest models should be prioritized for prediction and decision-making in practical contexts, while integrating Gradient boosting models for a more comprehensive analysis. Furthermore, it is crucial to acknowledge the limitations inherent in Logistic regression models and consider more flexible modeling approaches.

While this investigation has yielded valuable findings, several limitations must be acknowledged. First, the data were sourced exclusively from clinical records at a single medical facility, which may introduce selection and information biases. Furthermore, the retrospective nature limits the ability to establish causal relationships; thus, only correlational outcomes can be inferred. Additionally, using univariate analysis for variable selection in model construction may have led to the omission of relevant factors that could influence DVT occurrence. Consequently, future research should use multicenter prospective cohort studies with larger sample sizes to explore more variables such as genetic predispositions, lifestyle patterns, and medication use so as to enhance prognostic precision.

Table 4. Comparison of AUC for predictive models

	Training set		Test set	
	Z	*P*	*Z*	*P*
Logistic regression model VS Gradient boosting model	0.914	0.361	1.101	0.271
Gradient boosting model VS Random Forest model	1.652	0.099	0.598	0.550
Logistic regression model VS Random Forest model	3.209	0.001	1.468	0.142

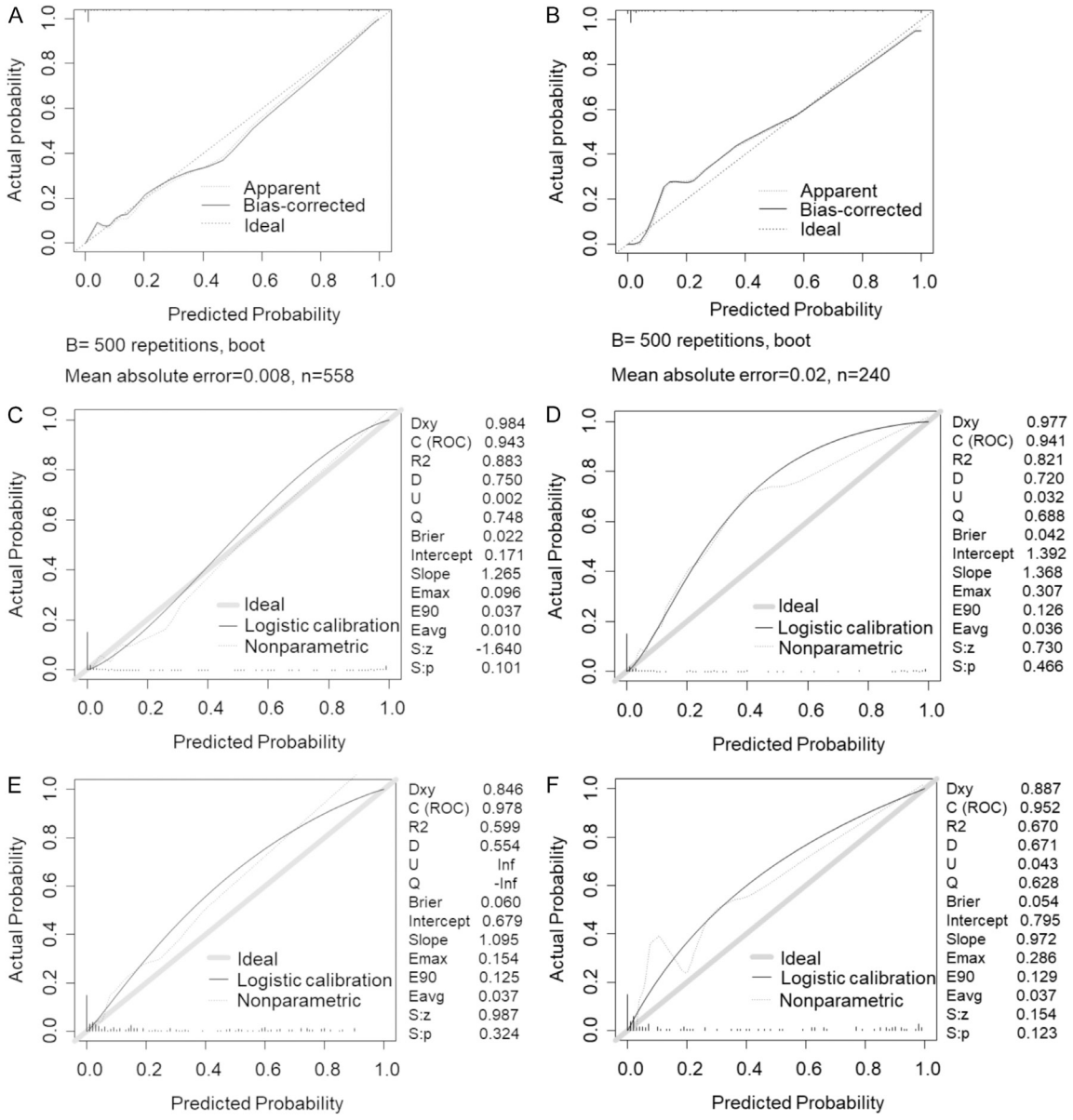

Figure 5. Calibration curves for three predictive models. A: Logistic regression model in training set; B: Logistic regression model in test set; C: Gradient boosting model in training set; D: Gradient boosting model in test set; E: Random Forest model in training set; F: Random Forest model in test set.

In summary, the Logistic regression model, Gradient boosting model, and Random Forest model constructed in this study all demonstrated good predictive value for the occurrence of DVT in LDH patients, offering potential benefits for optimizing clinical management. Among

them, the Random Forest model exhibited the best performance and should be prioritized for use in clinical settings.

Acknowledgements

This work was supported by Sci-Tech Plan Project of Hebei Province Administration of Traditional Chinese Medicine (2021130).

Disclosure of conflict of interest

None.

Address correspondence to: Weiling Zhang, Department of Traditional Chinese Medicine, The First Hospital of Hebei Medical University, No. 89 Donggang Road, Shijiazhuang 050091, Hebei, China. Tel: +86-0311-87157527; E-mail: 18633888386@163.com

References

[1] Yu P, Mao F, Chen J, Ma X, Dai Y, Liu G, Dai F and Liu J. Characteristics and mechanisms of resorption in lumbar disc herniation. Arthritis Res Ther 2022; 24: 205.

[2] Basindwah S, AlHamzah M, Balsharaf F, AlRajhi B, Sewaralthahab S, Altoijry A, Altuwaijri T and Al-Habib A. Prevalence of asymptomatic deep vein thrombosis in preoperative state of spine surgeries. World Neurosurg 2023; 170: e737-e743.

[3] Xu J, Ding X, Wu J, Zhou X, Jin K, Yan M, Ma J, Wu X, Ye J and Mo W. A randomized controlled study for the treatment of middle-aged and old-aged lumbar disc herniation by Shis spine balance manipulation combined with bone and muscle guidance. Medicine (Baltimore) 2020; 99: e23812.

[4] Kafeza M, Shalhoub J, Salooja N, Bingham L, Spagou K and Davies AH. A systematic review of clinical prediction scores for deep vein thrombosis. Phlebology 2017; 32: 516-531.

[5] He SY, Zhang P, Qin HJ, Jiang N and Yu B. Incidence and risk factors of preoperative deep venous thrombosis following hip fracture: a retrospective analysis of 293 consecutive patients. Eur J Trauma Emerg Surg 2022; 48: 3141-3147.

[6] Skeik N, Smith JE, Jensen JD, Nowariak ME, Manunga JM and Mirza AK. Literature review of distal deep vein thrombosis. J Vasc Surg Venous Lymphat Disord 2021; 9: 1062-1070, e6.

[7] Braithwaite I, Healy B, Cameron L, Weatherall M and Beasley R. Venous thromboembolism risk associated with protracted work- and computer-related seated immobility: a case-control study. JRSM Open 2016; 7: 2054270416632670.

[8] Sato T, Yoshimura N, Horii Y, Yamazaki M, Sato K, Kumagai K and Ishikawa H. Low tube voltage computed tomography venography for patients with deep vein thrombosis of the lower extremities - a comparison with venous ultrasonography. Circ J 2021; 85: 369-376.

[9] Tanoue S, Nakaura T, Iyama Y, Iyama A, Nagayama Y, Yoshida M and Yamashita Y. Diagnostic performance of dual-layer computed tomography for deep vein thrombosis in indirect computed tomography venography. Circ J 2020; 84: 636-641.

[10] Sturm L, Bettinger D, Klinger C, Krauss T, Engel H, Huber JP, Schmidt A, Caca K, Thimme R and Schultheiss M. Validation of color Doppler ultrasound and computed tomography in the radiologic assessment of non-malignant acute splanchnic vein thrombosis. PLoS One 2021; 16: e0261499.

[11] de Gouveia RH, Gulczynski J, Canzonieri V and Nesi G; History of Pathology Working Group of the European Society of Pathology. Rudolf Virchow: 200th birth anniversary. Virchows Arch 2021; 479: 1063-1065.

[12] Meng H, Zhu Y, Zhang J, Li J, Zhao K, Zhang Y and Chen W. Incidence and risk factor for preoperative deep vein thrombosis (DVT) in isolated calcaneal fracture, a prospective cohort study. Foot Ankle Surg 2021; 27: 510-514.

[13] Zhu Y, Chen W, Li J, Zhao K, Zhang J, Meng H, Zhang Y and Zhang Q. Incidence and locations of preoperative deep venous thrombosis (DVT) of lower extremity following tibial plateau fractures: a prospective cohort study. J Orthop Surg Res 2021; 16: 113.

[14] Ma J, Qin J, Hu J, Shang M, Zhou Y, Liang N, Zhang Y and Zhu Y. Incidence and hematological biomarkers associated with preoperative deep venous thrombosis following foot fractures. Foot Ankle Int 2020; 41: 1563-1570.

[15] He SY, Zhang P, Qin HJ, Jiang N and Yu B. Incidence and risk factors of preoperative deep venous thrombosis following hip fracture: a retrospective analysis of 293 consecutive patients. Eur J Trauma Emerg Surg 2022; 48: 3141-3147.

[16] Huang Y, Ge H, Wang X and Zhang X. Association between blood lipid levels and lower extremity deep venous thrombosis: a population-based cohort study. Clin Appl Thromb Hemost 2022; 28: 10760296221121282.

[17] Kruger PC, Eikelboom JW, Douketis JD and Hankey GJ. Deep vein thrombosis: update on diagnosis and management. Med J Aust 2019; 210: 516-524.

[18] Guadagni F, Riondino S, Formica V, Del Monte G, Morelli AM, Lucchetti J, Spila A, D'Alessandro R, Della-Morte D, Ferroni P and Roselli M. Clinical significance of glycemic parameters on ve-

nous thromboembolism risk prediction in gastrointestinal cancer. World J Gastroenterol 2017; 23: 5187-5195.
[19] Chopard R, Albertsen IE and Piazza G. Diagnosis and treatment of lower extremity venous thromboembolism: a review. JAMA 2020; 324: 1765-1776.
[20] Bhatt M, Braun C, Patel P, Patel P, Begum H, Wiercioch W, Varghese J, Wooldridge D, Alturkmani HJ, Thomas M, Baig M, Bahaj W, Khatib R, Kehar R, Ponnapureddy R, Sethi A, Mustafa A, Nieuwlaat R, Lim W, Bates SM, Lang E, Le Gal G, Righini M, Husainat NM, Kalot MA, Al Jabiri YN, Schünemann HJ and Mustafa RA. Diagnosis of deep vein thrombosis of the lower extremity: a systematic review and meta-analysis of test accuracy. Blood Adv 2020; 4: 1250-1264.
[21] Hopp MT, Ugurlar D, Pezeshkpoor B, Biswas A, Ramoji A, Neugebauer U, Oldenburg J and Imhof D. In-depth structure-function profiling of the complex formation between clotting factor VIII and heme. Thromb Res 2024; 237: 184-195.
[22] Fang X, Shen Y, Wang M, Dai L, Shi L, Zhang F, Chen C and Yuan Y. Predictive value of Caprini risk assessment model, D-dimer, and fibrinogen levels on lower extremity deep vein thrombosis in patients with spontaneous intracerebral hemorrhage. Front Neurol 2024; 15: 1370029.
[23] Griesbach C, Säfken B and Waldmann E. Gradient boosting for linear mixed models. Int J Biostat 2021; 17: 317-329.
[24] Hu J and Szymczak S. A review on longitudinal data analysis with random forest. Brief Bioinform 2023; 24: bbad002.