Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning-Based Model for Predicting Preoperative Deep Vein Thrombosis in Patients with Peri-ankle Fractures

Meihui Zhao
Baoding First Central Hospital
Li Zhou
Chengde Medical University
Shengxun Huang
Chengde Medical University
Jiale Dai
Chengde Medical University
Yuanyuan Liu
18617789626@163.com

Baoding First Central Hospital
Xiangjuan Li
Baoding First Central Hospital

Article

Keywords: Peri-ankle Fracture, Deep Vein Thrombosis、Machine Learning、Prediction Model

Posted Date: March 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8087004/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Machine Learning-Based Model for Predicting Preoperative Deep Vein Thrombosis in Patients with Peri-ankle Fractures

Meihui Zhao[1],Li Zhou [2],Shengxun Huang[2],Jiale Dai [2],Yuanyuan Liu[1]*,Xiangjuan Li[1]

1 Baoding First Central Hospital, Baoding, Hebei, China

2 Chengde Medical University, Chengde, Hebei, China

Corresponding author:Yuanyuan Liu 18617789626@163.com

Funding:Baoding City Key Laboratory of Mechanisms of Perioperative Orthopedic Complications and Integrated Chinese-Western Medicine Interventions

Abstract：Objective: To investigate the risk factors for preoperative deep vein thrombosis (DVT) in patients with peri-ankle fractures and to develop a machine learning-based prediction model for preoperative DVT risk.**Methods:**A retrospective study was conducted involving 1000 patients with peri-ankle fractures. Predictors were selected using Lasso regression, and the dataset was randomly split into a training set and a validation set at a 7:3 ratio. Four models—Logistic Regression, Decision Tree, Random Forest, and XGBoost—were constructed and internally validated using the Bootstrap method. Model performance was compared using metrics including the area under the ROC curve (AUC), calibration curves, decision curve analysis, and the F1 score. The optimal model was interpreted using the SHAP method.**Results:**Four predictors were identified: age, BMI, preoperative waiting time, and D-dimer level. The XGBoost model demonstrated the best performance, with an AUC of 0.827 in the validation set. SHAP analysis confirmed that higher values of these four features contributed positively to the model's predictions, aligning with clinical knowledge and ensuring a transparent and credible decision-making process.**Conclusion:**The XGBoost-based prediction model exhibits favorable performance and interpretability. It delineates core risk factors to guide targeted nursing interventions and holds potential for translation into a clinical decision-support tool, offering a novel strategy for the precise prevention of preoperative DVT.

Keywords：Peri-ankle Fracture；Deep Vein Thrombosis；Machine Learning；Prediction Model

Introduction

Peri-ankle fractures represent one of the most common types of fractures, accounting for approximately 10.2% of all fractures[1]. The cornerstone of treatment lies in the restoration of anatomical alignment and functional stability of the joint, with surgical intervention required in approximately 23.4% to 83% of patients[2-4]. However, due to the thin subcutaneous tissue around the ankle, progressive post-

traumatic soft tissue swelling often peaks within 24 to 72 hours, frequently necessitating postponement of surgery until the swelling has subsided—typically indicated by the presence of the skin wrinkle sign—which generally occurs between 3 to 14 days after injury[5,6]. Such delays in surgical management significantly increase both the complexity of patient care and associated risks. Prolonged immobilization and restricted mobility not only extend the duration of hospitalization but also place patients at an elevated risk of complications such as deep vein thrombosis (DVT) and pressure ulcers.

DVT represents a critical preoperative risk factor in these patients. Trauma alone can induce a hypercoagulable state within 24 hours, while pain and immobilization following the fracture further exacerbate venous stasis and endothelial injury. This combination creates a synergistic effect, placing patients with ankle fractures in a high-risk category for DVT[7,8]. Research indicates that the average time from fracture occurrence to DVT diagnosis is merely 4.2 days[9]， yet the preoperative waiting period often extends to 7-10 days[10]. More alarmingly, each day of surgical delay is associated with a 9.3% increase in DVT risk[11]， with the overall reported preoperative incidence ranging from 2.1% to 37.5%[12,13]. ince a dislodged thrombus can lead to life-threatening pulmonary embolism, preoperative DVT prevention has become a paramount priority in clinical nursing practice.

Currently, generalized risk assessment tools such as the Caprini score are widely used in clinical practice; however, they lack specificity for warning of DVT risk in the particular population of peri-ankle fracture patients. Moreover, early symptoms of DVT can often be masked by the trauma response, posing significant challenges to accurate prevention [14]. In recent years, machine learning (ML) techniques have demonstrated remarkable potential in medical risk prediction by integrating multifactorial clinical data to develop high-accuracy predictive models. Therefore, this study aims to develop a preoperative DVT risk prediction model specifically for patients with peri-ankle fractures using machine learning algorithms. The goal is to enable early identification of DVT and facilitate individualized, precise prophylaxis, thereby offering a novel approach to enhancing perioperative nursing safety.

1. Subjects and Methods

1.1 Study Population

This retrospective study enrolled patients with peri-ankle fractures who underwent surgical treatment in the Department of Orthopedics at a tertiary Grade A hospital in Baoding City between January 2022 and December 2024.

The inclusion criteria were as follows: (1) radiologically confirmed diagnosis of peri-ankle fracture; (2) age ≥18 years; and (3) acute fracture (time from injury to presentation <21 days). Patients meeting

any of the following criteria were excluded: (1) open fracture; (2) concomitant fractures in other anatomical regions; (3) associated vascular or neurological injuries; (4) pre-existing deep vein thrombosis (DVT) prior to admission; (5) severe hepatic or renal dysfunction, or diagnosis of malignant tumor; or (6) pregnancy.

The minimum sample size was determined to be 652 participants based on the methodology described by Riley et al.[15]To account for a potential 10%–20% rate of excluded or non-evaluable cases, a minimum of 782 subjects were targeted. The study ultimately included 1000 participants, who were randomly allocated into a training set (n=700) and a validation set (n=300) at a ratio of 7:3. This study was approved by the Medical Ethics Committee of Baoding First Central Hospital (Approval No.: Quick [2024]241). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects and/or their legal guardian(s).

1.2 Methods

1.2.1 Study Instrument

Based on a comprehensive literature review and expert panel discussions, a Checklist for Risk Factors of Preoperative DVT in Patients with Peri-Ankle Fractures was developed. This checklist encompassed the following categories:(1) Demographic characteristics: sex, age, body mass index (BMI), preoperative waiting time, and time from injury to hospital admission;(2) Clinical and disease-related characteristics: diabetes, hypertension, cardiovascular disease, cerebrovascular disease, smoking history, alcohol use history, fracture type, and injury mechanism;(3) Laboratory parameters: D-dimer, hemoglobin, prothrombin time (PT), activated partial thromboplastin time (APTT), platelet count, hematocrit, platelet distribution width (PDW), albumin, C-reactive protein (CRP), and fasting blood glucose.

1.2.2 Data Collection and Quality Control

Two researchers, who had received standardized training, independently extracted data from the hospital's electronic medical record system using the aforementioned checklist. Following data collection, a dual-entry verification and cross-validation process was performed to ensure accuracy. Data preprocessing included removal of duplicate entries, along with logical checks and imputation of missing values for key variables to maintain data integrity. The study design adhered to the TRIPOD statement[16].

1.2.3 Statistical Analysis

Data analyses were conducted using SPSS 26.0 and R 4.3. Continuous variables were compared between groups using the t-test or Mann–Whitney U test, as appropriate, while categorical variables were compared using the χ^2 test or Fisher's exact test. For prediction modeling, least absolute shrinkage and selection operator (Lasso) regression was first applied to screen predictors from all candidate variables. The dataset was then randomly split into training and validation sets in a 7:3 ratio. Based on the selected variables, four models were constructed: logistic regression, decision tree, random forest, and XGBoost. Internal validation was performed using the bootstrap method with 1000 resamples. Model performance was comprehensively evaluated using the area under the receiver operating characteristic curve, calibration curves, decision curve analysis, and the F1 score. The optimal model was further interpreted using SHAP (SHapley Additive exPlanations). A two-sided p-value < 0.05 was considered statistically significant.

2. Results

2.1 Baseline Characteristics of the Study Population

A total of 1000 eligible patients were randomly allocated into a training set (n = 700) and a validation set (n = 300) at a 7:3 ratio. The incidence of DVT was 14.9% (104/700) in the training set and

16.0% (48/300) in the validation set. Comparison of baseline characteristics between the two sets revealed a statistically significant difference only in albumin levels ($P < 0.05$), with all other variables showing no significant differences ($P > 0.05$), indicating that the randomization process was generally successful. Although albumin levels differed between the groups, this variable was not selected by the LASSO regression as a final predictor; therefore, its influence on subsequent model development is likely minimal. Detailed baseline characteristics are presented in Table 1.

Table 1. Baseline Characteristics of the Study Participants in the Training and Validation Sets [Mean ± SD / n (%) / Median (IQR)]

Variable	**Group**	**Train Set (n=700)**	**Valiadation Set (n=300)**	**Test Statistic**	***P-Value***
Sex	Male	440（62.86）	181（60.33）	0.568[b]	0.451
	Female	260（37.14）	119（39.67）		
Age(years)		46.84±14.53	47.59±14.27	-0.753[a]	0.452
BMI（kg/m^2）	＜24.0	487（69.57）	210（70.00）	0.018[b]	0.892
	≥24.0	213（30.43）	90（30.00）		
Time from Injury to Admission（h）		7.00（4.00，24.00）	6.00（3.00，24.00）	-1.406[c]	0.160
Preoperative Waiting Time（d）		9.11±3.71	8.91±3.61	0.794[a]	0.428
Type of Injury	High-energy	395（56.43）	151（50.33）	3.147[b]	0.076
	Low-energy	305（43.57）	149（49.67）		
Fracture Classification	Unimalleolar Fracture	298（42.57）	127（42.33）	5.859[b]	0.663
	Bimalleolar Fracture	18（2.57）	10（3.33）		
	Trimalleolar Fracture	60（8.57）	18（6.00）		
	Calcaneal Fracture	147（21.00）	66（22.00）		
	Talus Fracture	9（1.29）	1（0.33）		
	Metatarsal Fracture	48（6.86）	21（7.00）		
	Distal Tibiofibular Fracture	48（6.86）	24（8.00）		
	Lower Tibiofibular Fracture	57（8.14）	23（7.67）		
	Pilon Fracture	15（2.14）	10（3.33）		

Hypertension	Yes	139（19.86）	72（24.00）	2.165[b]	
	No	561（80.14）	228（76.00）		0.141
Diabetes	Yes	53（7.57）	27（9.00）	0.582[b]	
	No	647（92.43）	273（91.00）		0.445
Cardiovascular	Yes	16（2.29）	10（3.33）	0.910[b]	
Disease	No	684（97.71）	290（96.67）		0.340
Cerebrovascular	Yes	8（1.14）	8（2.67）	3.097[b]	
Disease	No	692（98.86）	292（97.33）		0.078
Smoking	Yes	192（27.43）	78（26.00）	0.217[b]	
	No	508（72.57）	222（74.00）		0.641
Alcohol Use	Yes	158（22.57）	68（22.67）	0.001[b]	
	No	542（77.43）	232（77.33）		0.974
D-dimer	＜500	221（31.57）	86（28.67）	0.833[b]	
（μg/L）	≥500	479（68.43）	214（71.33）		0.361
Prothrombin	＜10	12（1.71）	6（2.00）	4.456[b]	
Time（s）	10-14	680（97.14）	285（95.00）		
	＞14	8（1.14）	9（3.00）		0.108
Activated Partial	＜23	14（2.00）	3（1.00）	1.257[b]	
Thromboplastin	23-37	679（97.00）	294（98.00）		
Time（s）	＞37	7（1.00）	3（1.00）		0.533
Hemoglobin	＜110	59（8.43）	35（11.67）	3.909[b]	
（g/L）	110-150	529（75.57）	210（70.00）		
	＞150	112（16.00）	55（18.33）		0.142
Hematocrit	＜40	321（45.86）	137（45.67）	1.007[b]	
（%）	40-50	363（51.86）	159（53.00）		0.605
	＞50	16（2.29）	4（1.33）		
Platelet Count	＜100	4（0.57）	2（0.67）	0.071[b]	
（10^9/L）	100-300	615（87.86）	262（87.33）		0.965
	＞300	81（11.57）	36（12.00）		
Platelet	＜9	30（4.29）	13（4.33）	0.024[b]	
Distribution Width	9-17	657（93.86）	281（93.67）		0.988
（fL）	＞17	13（1.86）	6（2.00）		

Albumin（g/L）	＜35	19（2.71）	21（7.00）	10.046[b]	
	35-50	671（95.86）	275（91.67）		0.007
	＞50	10（1.43）	4（1.33）		
C-Reactive Protein（mg/L）	＜10	406（58.00）	165（55.00）	0.772[b]	0.380
	≥10	294（42.00）	135（45.00）		
Fasting Blood Glucose（＞6.1mmol/L）	Yes	190（27.14）	77（25.67）	0.234[b]	
	No	510（72.86）	223（74.33）		0.629

Note: a, b, and c represent the t-value, χ^2 value, and Z-value, respectively.

2.2 Identification of Predictors for Preoperative DVT in Patients with Peri-ankle Fractures

2.2.1 Variable Selection Using LASSO Regression

The least absolute shrinkage and selection operator (LASSO) regression method, which incorporates L1 regularization, was employed to automatically select relevant predictors and exclude non-contributory variables, thereby reducing the risk of overfitting. The optimal regularization parameter (λ) was determined via cross-validation. At λ-1SE = 0.048, the model achieved a favorable balance between goodness-of-fit and simplicity, retaining only a limited set of predictors. Four variables were ultimately selected: age, body mass index (BMI), preoperative waiting time, and D-dimer level. These variables were subsequently included in the multivariable analysis. The variable selection path and cross-validation curve are shown in Figures 1 and 2, respectively.

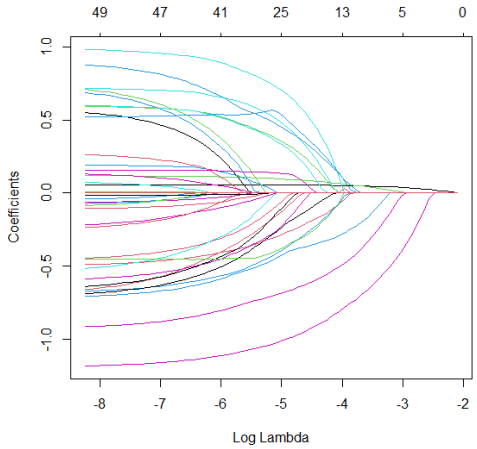


Figure 1. Coefficient Path Plot

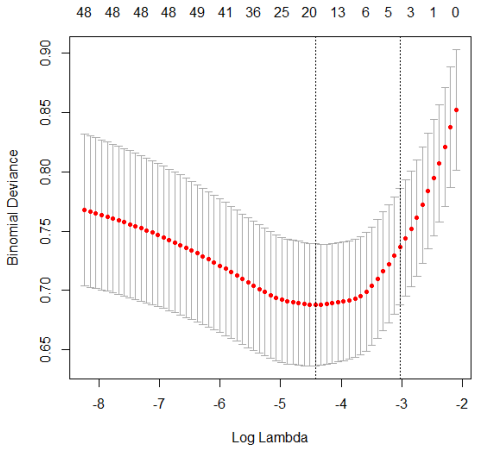


Figure 2. Cross-Validation Plot

2.2.2 Multivariable Analysis of Preoperative DVT in Patients with Peri-Ankle Fractures

Multivariable analysis was performed with the occurrence of DVT as the dependent variable and the four predictors selected by LASSO regression as independent variables. Detailed value assignments for these variables are provided in Table 2. The results identified the following as independent risk factors for preoperative DVT ($P < 0.05$): age, BMI ≥ 24 kg/m², preoperative waiting time, and D-dimer ≥ 500 μg/L. Complete results of the multivariable analysis are presented in Table 2.

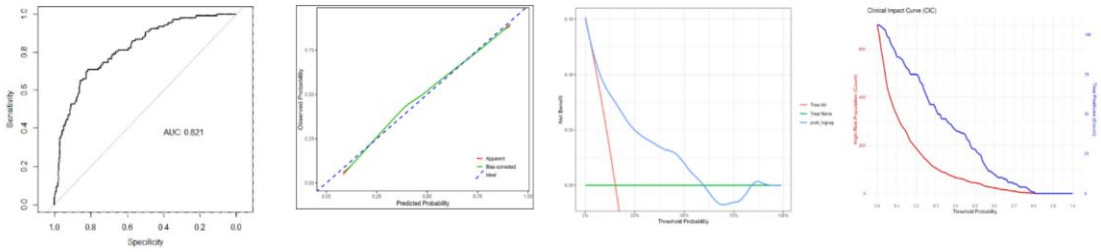
Table 2. Multivariable Analysis of Preoperative DVT in Patients with Peri-Ankle Fractures

Variable	**Regression Coefficient**	**Standard Error**	**Waldχ^2 Value**	***P-value***	**Odds Ratio (OR)**	**95%*CI***	
						Lower	**Upper**
Constant	-7.362	0.692	113.149	＜0.001	0.001		

Age	0.066	0.010	44.935	＜0.001	1.069	1.048	1.090
BMI≥24kg/m^2	1.056	0.239	19.546	＜0.001	2.875	1.800	4.591
Preoperative Waiting Time	0.098	0.032	9.655	0.002	1.103	1.037	1.174
D-dimer≥500μg/L	1.016	0.350	8.413	0.004	2.763	1.390	5.490

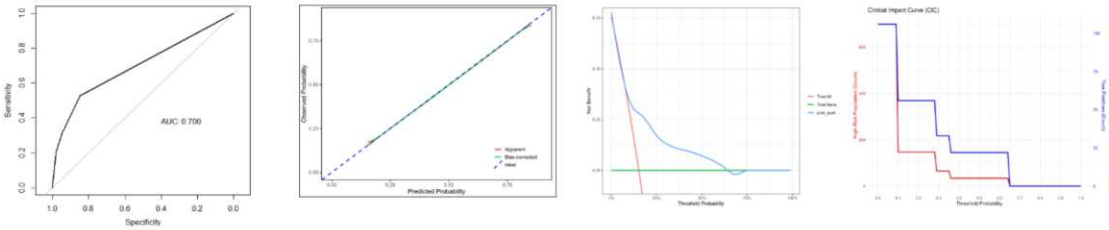
2.3 Development of a Prediction Model for Preoperative DVT in Patients with Peri-Ankle Fractures

Using the predictors identified through LASSO regression and multivariable analysis, four distinct prediction models were constructed: logistic regression, decision tree, XGBoost, and random forest. The predictive performance of each model on the training set was evaluated by generating receiver operating characteristic (ROC), calibration, decision curve analysis (DCA), and clinical impact curve (CIC). Quantitative metrics including the area under the curve (AUC), precision, accuracy, recall, and F1-score were calculated. Detailed results are presented in Figures 3–6 and Table 3.



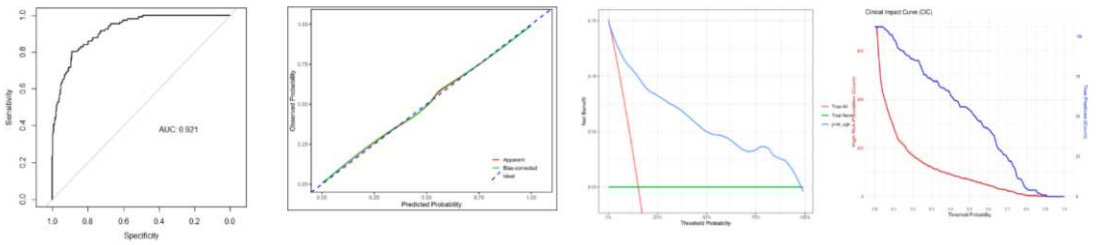
Logistic Regression Model ROC Curve | Logistic Regression Model Calibration Curve | Logistic Regression Model DCA Curve | Logistic Regression Model CIC Curve

Figure 3. Evaluation Metrics for the Logistic Regression Model on the Training Set



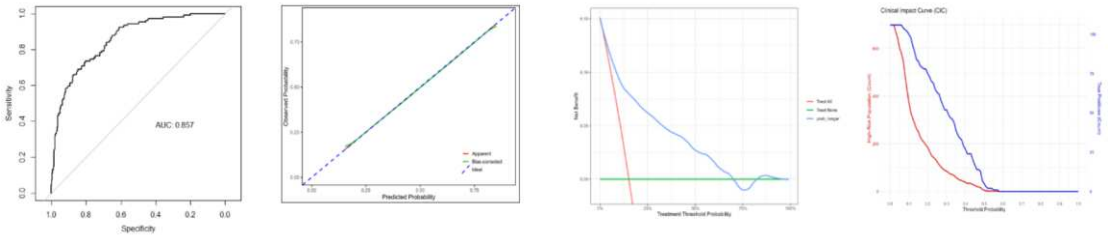
Decision Tree Model ROC Curve | Decision Tree Model Calibration Curve | Decision Tree Model DCA Curve | Decision Tree Model CIC Curve

Figure 4. Evaluation Metrics for the Decision Tree Model on the Training Set



XGBoost Model ROC Curve | XGBoost Model Calibration Curve | XGBoost Model DCA Curve | XGBoost Model CIC Curve

Figure 5. Evaluation Metrics for the XGBoost Model on the Training Set



Random Forest Model ROC Curve | Random Forest Model Calibration Curve | Random Forest Model DCA Curve | Random Forest Model CIC Curve

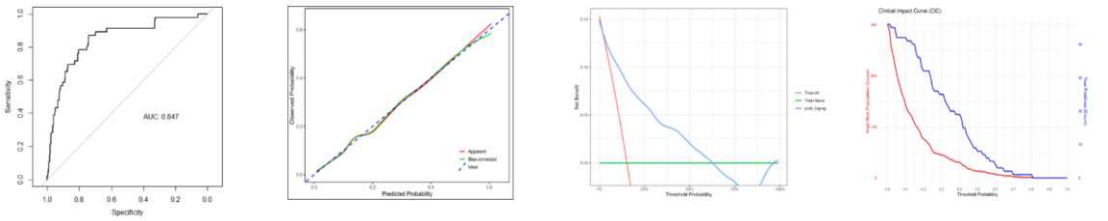
Figure 6. Evaluation Metrics for the Random Forest Model on the Training Set

Table 3. Comparison of Predictive Performance Among the Four Models on the Training Set

Model		AUC	Accuracy	Precision	Recall	F1 Score
Logistic Regression	Training set	0.821	0.864	0.881	0.971	0.924
Decision Tree	Training set	0.700	0.863	0.874	0.980	0.924
XGboost	Training set	0.921	0.904	0.917	0.975	0.945
Random Forest	Training set	0.857	0.856	0.854	1	0.922

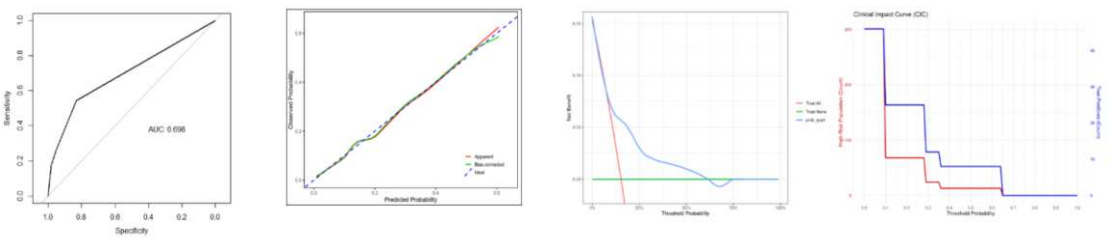
2.4 Internal Validation of the Prediction Models for Preoperative DVT in Patients with Peri-Ankle Fractures

Internal validation of the four machine learning prediction models was performed using the bootstrap method with 1,000 resamples to assess their predictive performance on the validation set. The results of this validation, including ROC curves, calibration plots, and decision curve analysis for each model, are detailed in Figures 7–10 and Table 4.



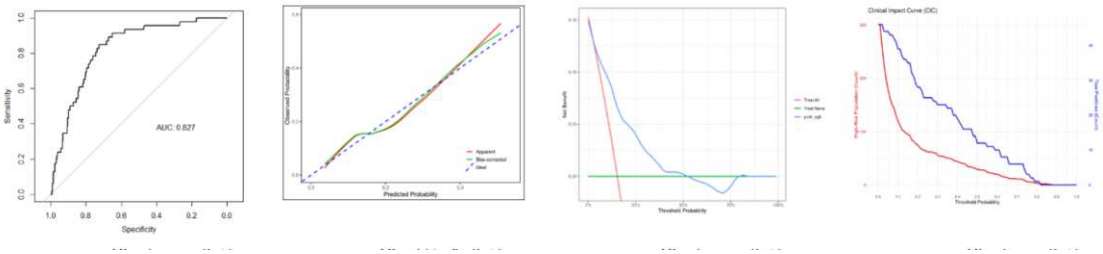
Logistic Regression Model ROC Curve | Logistic Regression Model Calibration Curve | Logistic Regression Model DCA Curve | Logistic Regression Model CIC Curve

Figure 7. Evaluation Metrics for the Logistic Regression Model on the Validation Set



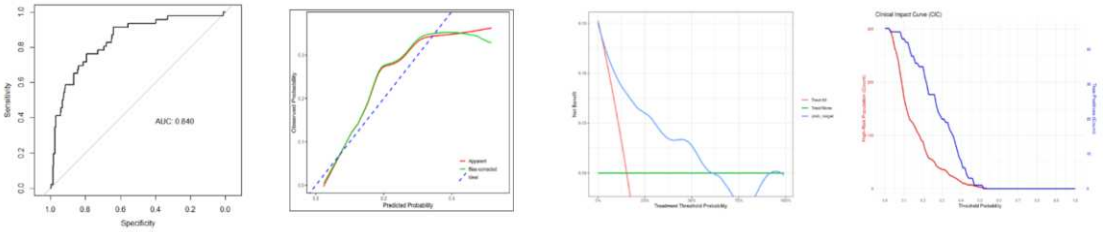
Decision Tree Model ROC Curve | Decision Tree Model Calibration Curve | Decision Tree Model DCA Curve | Decision Tree Model CIC Curve

Figure 8. Evaluation Metrics for the Decision Tree Model on the Validation Set



XGBoost Model ROC Curve | XGBoost Model Calibration Curve | XGBoost Model DCA Curve | XGBoost Model CIC Curve

Figure 9. Evaluation Metrics for the XGBoost Model on the Validation Set



Random Forest Model ROC Curve | Random Forest Model Calibration Curve | Random Forest Model DCA Curve | Random Forest Model CIC Curve

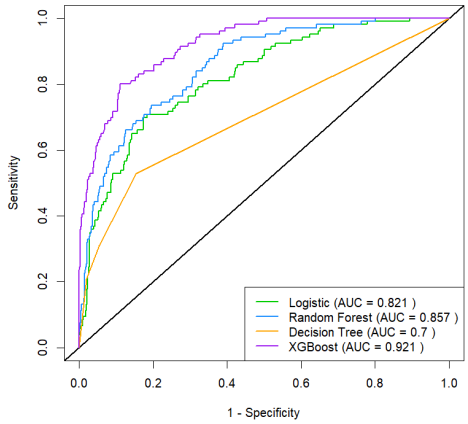
Figure 10. Evaluation Metrics for the Random Forest Model on the Validation Set

Table 4. Comparison of Predictive Performance Among the Four Models on the Validation Set

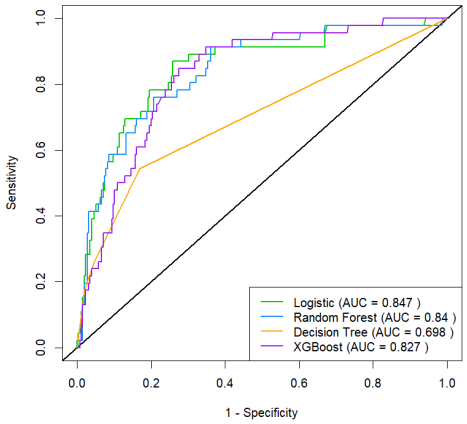
Model		AUC	Accuracy	Precision	Recall	F1 Score
Logistic Regression	Validation set	0.847	0.857	0.868	0.980	0.921
Decision Tree	Validation set	0.698	0.857	0.868	0.980	0.921
XGboost	Validation set	0.827	0.831	0.875	0.933	0.903
Random Forest	Validation set	0.840	0.844	0.849	0.922	0.915

2.5 Comprehensive Comparison of Predictive Performance Across the Four Machine Learning Models

Based on the overall performance in both the training and validation sets, the XGBoost model demonstrated strong fitting capability and enhanced predictive accuracy, while the Logistic Regression model exhibited superior stability. The Random Forest model ranked next in predictive performance, whereas the Decision Tree model underperformed compared to the other three. A comprehensive comparison is detailed in Figure 11 and Table 5.



ROC Curves of the Four Models in the Training Set



ROC Curves of the Four Models in the Validation Set

Figure 11. Comprehensive Comparison of ROC Curves for the Four Models

Table 5. Comprehensive Comparison of Predictive Performance Across the Four Models in Training and Validation Sets

Model		AUC	Accuracy	Precision	Recall	F1 Score
Logistic Regression	Training set	0.821	0.864	0.881	0.971	0.924
	Validation set	0.847	0.857	0.868	0.980	0.921
Decision Tree	Training set	0.700	0.863	0.874	0.980	0.924
	Validation set	0.698	0.857	0.868	0.980	0.921
XGboost	Training set	0.921	0.904	0.917	0.975	0.945
	Validation set	0.827	0.831	0.875	0.933	0.903
Random Forest	Training set	0.857	0.856	0.854	1	0.922
	Validation set	0.840	0.844	0.849	0.922	0.915

2.6 Interpretability Analysis of the XGBoost Model

To elucidate the decision-making logic of the optimal model, we performed an interpretability analysis on the XGBoost model using SHapley Additive exPlanations (SHAP). The features were ranked by their importance in influencing the prediction outcome in the following order: age, BMI, preoperative waiting time, and D-dimer (Figure 12). The SHAP summary plot further revealed that higher values of these four features corresponded to increased SHAP values, indicating a greater predicted risk of DVT. This finding aligns with clinical understanding and the results of the multivariable analysis, confirming the rationality and interpretability of the model's decision-making process (Figure 13).

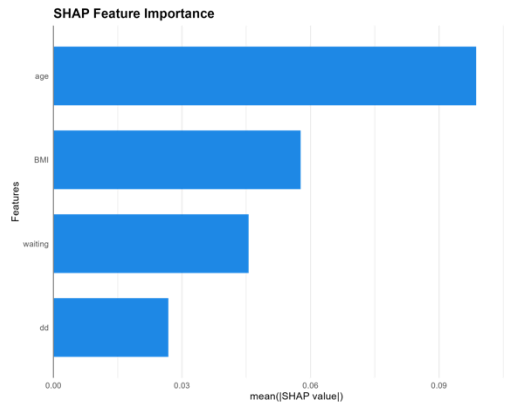


Figure 12. Feature Importance Ranking of the XGBoost Model

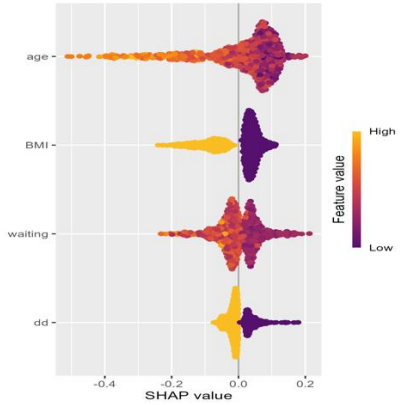


Figure 13. SHAP Summary Plot of the XGBoost Model

3. Discussion

3.1 Incidence of Preoperative DVT in Patients with Peri-Ankle Fractures

This study found a preoperative DVT incidence of 15.3% in patients with peri-ankle fractures. This rate falls within the range of 6.4% to 37.5% reported in previous literature[12,17-19]. The observed variation may be attributed to differences in study populations, inclusion criteria, and DVT detection protocols. Notably, the relatively high detection rate in our study is likely associated with the systematic use of Doppler ultrasonography for screening, suggesting that active surveillance facilitates the identification of a greater number of asymptomatic DVT cases. Despite the variability in reported incidence, our findings consistently affirm that this patient population faces a substantial risk of preoperative DVT. Given that DVT is a preventable yet potentially fatal complication, these findings underscore the importance of implementing systematic and individualized preoperative risk assessment in patients with peri-ankle fractures, rather than relying solely on empirical or generalized approaches.

3.2 Risk Factors for Preoperative DVT in Patients with Peri-Ankle Fractures

The four predictors identified in this study are not only statistically significant but also reflect clear pathophysiological mechanisms and hold substantial clinical relevance for targeted interventions.

Advanced age was confirmed as an independent risk factor for DVT, with each additional year of age associated with a 6.9% increase in risk (OR = 1.069). This correlation aligns with age-related endothelial dysfunction, hemodynamic abnormalities, and a progressively hypercoagulable state[20,21]. Consequently, nursing assessments should explicitly recognize advanced age as a core risk dimension, warranting more intensive prophylaxis in elderly patients.

A BMI ≥ 24 kg/m² was associated with a 2.875-fold increase in DVT risk, underscoring overweight/obesity as a major modifiable factor[22,23]. This is attributable to the chronic inflammatory state, dyslipidemia, and prothrombotic tendency commonly observed in obesity[24,25]. These findings highlight a clear target for nursing intervention: integrating weight management strategies and targeted health education into perioperative care plans, particularly for obese individuals.

Preoperative waiting time emerged as a critical, systems-related risk factor. Each additional day of delay increased DVT risk by 10.3% (OR = 1.103), directly illustrating a dose–response relationship between immobilization duration and venous stasis[13,22,26]. This insight shifts focus beyond patient-specific factors to healthcare processes, emphasizing the vital role nurses can play in coordinating and advocating for optimized clinical pathways to mitigate patient risk.

A D-dimer level ≥ 500 μg/L, serving as a sensitive laboratory marker, increased DVT risk by 2.763-fold. Although limited in specificity, it provided an objective early warning within our specific study population[9,27]. In nursing practice, elevated D-dimer levels should trigger enhanced monitoring and prophylactic measures, even in the absence of typical clinical symptoms.

In summary, these four factors—spanning patient baseline status (age, BMI), care process efficiency (preoperative waiting time), and objective laboratory metrics (D-dimer)—collectively establish a comprehensive and readily assessable risk prediction framework. This model provides a solid foundation for implementing precise, evidence-based nursing interventions.

3.3 Comprehensive Comparison of the Four Machine Learning Models

The present study demonstrated that the XGBoost model achieved the best overall predictive performance in the validation set (AUC = 0.827). Its strength lies in effectively capturing complex, non-linear relationships among variables, which aligns well with the multifactorial and interactive pathogenesis of DVT. Consistent with our results, Wei et al.[28]also reported superior performance of XGBoost in predicting DVT risk in patients with lower limb fractures (AUC = 0.979). Compared to traditional assessment tools such as the Caprini score, the XGBoost model maintains high accuracy while offering enhanced clinical applicability.

3.4 Model Interpretability Analysis Based on SHAP

This study employed SHAP analysis to successfully transform the "black-box" model into an interpretable decision-support tool[29]. The feature importance ranking identified age as the primary risk factor, necessitating that nursing staff assign the highest priority to risk assessment in elderly patients. BMI and preoperative waiting time, as modifiable factors, indicate that nursing practice should reinforce health education for overweight or obese patients and work collaboratively with the healthcare team to optimize clinical pathways and minimize preoperative delays. D-dimer, serving as an objective laboratory marker, provides a crucial quantitative basis for nursing assessment; its elevation should trigger more frequent monitoring and proactively suggest vascular ultrasound examinations. In summary, SHAP analysis not only validates the model's reliability but also translates the four predictors into clear nursing assessment priorities and intervention targets, establishing a solid foundation for developing a precise and efficient preoperative DVT prevention protocol.

4. Conclusion

This study identified age, BMI, preoperative waiting time, and D-dimer level as independent predictors for preoperative DVT in patients with peri-ankle fractures. The XGBoost model developed based on these predictors demonstrated optimal performance, with satisfactory discrimination and calibration. Through SHAP analysis, the model achieved transparent interpretation of its decision-making process, thereby enhancing its credibility and potential acceptability in clinical nursing practice. The proposed model shows promise for development into a decision-support tool to facilitate precise preoperative DVT risk assessment in this patient population. Future efforts should focus on validating its generalizability using multi-center external data, integrating the model into nursing information systems, and conducting interventional studies to evaluate its clinical utility.

Author Contributions

Meihui Zhao: Conceptualization, Data Curation, Writing-Original Draft Preparation;Li Zhou, Shengxun Huang, Jiale Dai: Data Curation, Validation;Yuanyuan Liu and Xiangjuan Li: Supervision, Writing - Review & Editing.

Data Availability

The datasets generated and/or analysed during the current study are not publicly available due to patient privacy concerns but are available from the corresponding author on reasonable request.

References

[1] KANG H J, LEE J W, KWON Y M,et al.Epidemiology of Ankle Fractures in Korea: A Nationwide Population-Based Study[J].J Korean Med Sci,2022,37(38):e288.

[2] FENELON C, GALBRAITH J G, FAHEY T,et al.The Operative Treatment of Ankle Fractures: A 10-Year Retrospective Study of 1529 Patients[J].J Foot Ankle Surg,2021,60(4):663-668.

[3] VANDERKARR M F, RUPPENKAMP J W, VANDERKARR M,et al.Incidence, costs and post-operative complications following ankle fracture - A US claims database analysis[J]. BMC Musculoskelet Disord,2022,23(1):1129.

[4] HAPPONEN V, KRÖGER H, KUISMIN M,et al.Ankle fractures in Finland: 118,929 operatively treated between 1987 and 2019[J].Acta Orthop,2022,93:327-333.

[5] LI T, SUN Z, CHAI Y, et al. Expert consensus on optimizing diagnosis and treatment strategies for ankle fractures within the ERAS framework[J]. Chinese Journal of Bone and Joint Surgery,,2019,12(01):3-12.

[6] WANG Y, LI T, MI M, et al. Effectiveness of ERAS-based nursing interventions on preoperative limb edema reduction in patients undergoing elective ankle fracture surgery[J]. Chinese Journal of Bone and Joint Surgery,2023,16(02):169-174.

[7] LIN Q, YANG M, HOU Z. Guidelines for prevention of perioperative venous thromboembolism in Chinese trauma orthopedics patients (2021),2021,23(3):185-192.

[8] LIMEIRA THOMSON K, POOL E, KERRAY P.Risk assessment and thromboprophylaxis in adult patients with lower-limb immobilisation[J].Emergency Nurse,2024,32(1):28-33.

[9] MA J, QIN J, HU J,et al.Incidence and Hematological Biomarkers Associated With Preoperative Deep Venous Thrombosis Following Foot Fractures[J].Foot Ankle Int,2020,41(12):1563-1570.

[10] LIU Y, ZHOU X, GUO D, et al. Clinical efficacy of blood-letting puncture combined with red light therapy in managing local swelling and pain before surgery for peri-ankle fractures[J]. Modern Journal of Integrated Traditional Chinese and Western Medicine,2022,31(06):799-802.

[11] CONG Y, HUANG H, ZHANG B F,et al.The Correlation Between Preoperative Lower Extremity Deep Vein Thrombosis (DVT) and the Time from Injury to Surgery (TFITS)[J]. Clinical and Applied Thrombosis/Hemostasis, 2022,28:10760296221108961.

[12] LI W, WANG D. Risk factors for deep vein thrombosis following ankle fractures and the monitoring value of D-dimer levels[J]. Trauma and Critical Care Medicine,2018,6(2):81-83.

[13] SUN N, YANG F, ZHU S. Analysis of risk factors for preoperative deep vein thrombosis in ankle fracture patients[J]. Shandong Medical Journal,2015,55(28):82-83.

[14] KOSKI E, MURPHY J.AI in Healthcare[J].Stud Health Technol Inform,2021,284:295-299.

[15] RILEY R D, ENSOR J, SNELL K I E,et al.Calculating the sample size required for developing a clinical prediction model[J].Bmj,2020,368:m441.

[16] COLLINS G S, MOONS K G M, DHIMAN P,et al.TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods[J].Bmj,2024,385:e078378.

[17] SUO H, WANG N, SUO M, et al. Analysis of risk factors for preoperative deep vein thr ombosis in patients with ankle fractures[J]. Chinese Journal of Bone and Joint Injuries,202 0,35(08):857-859.

[18] GAO Z, ZHAO K, JIN L,et al.Combination of neutrophil to lymphocyte ratio, platelet to lymphocyte ratio with plasma D-dimer level to improve the diagnosis of deep venous thro mbosis (DVT) following ankle fracture[J].Journal of Orthopaedic Surgery and Research, 20 23,18(1):362.

[19] LUO Z, CHEN W, LI Y,et al.Preoperative incidence and locations of deep venous thromb osis (DVT) of lower extremity following ankle fractures[J].Scientific Reports,2020,10(1):10 266.

[20] DING K, WANG H, JIA Y,et al.Incidence and risk factors associated with preoperative d eep venous thrombosis in the young and middle-aged patients after hip fracture[J].J Ortho p Surg Res,2022,17(1):15.

[21] BLOOM S I, ISLAM M T, LESNIEWSKI L A,et al.Mechanisms and consequences of en dothelial cell senescence[J].Nat Rev Cardiol,2023,20(1):38-51.

[22] LU Z, CAO Z, GAO G, et al. Utility of acoustic radiation force impulse imaging in asse ssing acute deep vein thrombosis risk during preoperative waiting period for calcaneal frac tures[J]. Chinese Journal of Tissue Engineering Research,2017,21(03):423-427.

[23] NASSOUR N, AKHBARI B, RANGANATHAN N,et al.Correlation Between Statin Use a nd Symptomatic Venous Thromboembolism Incidence in Patients With Ankle Fracture: A Machine Learning Approach[J].Foot & ankle specialist,2023,19386400231207692-.

[24] LIU X, PANG P, LUO Z,et al.Prevalence and risk factors for proximal deep vein thromb osis at admission in patients with traumatic fractures: a multicenter retrospective study[J].F ront Cardiovasc Med,2024,11:1372268.

[25] CONSTANTINESCU A E, BULL C J, GOUDSWAARD L J,et al.A phenome-wide appro ach to identify causal risk factors for deep vein thrombosis[J].BMC Med Genomics,2023,1 6(1):284.

[26] MENG H, ZHU Y, ZHANG J,et al.Incidence and risk factor for preoperative deep vein t hrombosis (DVT) in isolated calcaneal fracture, a prospective cohort study[J].Foot and An kle Surgery,2021,27(5):510-514.

[27] ZHU Y, MENG H, MA J,et al.Prevalence of Preoperative Lower Extremity Deep Vein T hrombosis in Bilateral Calcaneal Fractures[J].J Foot Ankle Surg,2021,60(5):950-955.

[28] WEI C, WANG J, YU P,et al.Comparison of different machine learning classification mo dels for predicting deep vein thrombosis in lower extremity fractures[J].Sci Rep,2024,14(1): 6901.

[29] XU Q, YU B, ZHOU P, et al. Development of an interpretable prediction model for dee p vein thrombosis in total hip arthroplasty patients using machine learning and SHAP ana lysis[J]. Chinese Journal of Hospital Statistics,2024,31(01):11-8+24.