


Ten-Year Multicenter Retrospective Study Utilizing Machine Learning Algorithms to Identify Patients at High Risk of Venous Thromboembolism After Radical Gastrectomy

Yuan Liu ^{*}, Chen Song^{*}, Zhiqiang Tian, Wei Shen

Department of General Surgery, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, People's Republic of China

^{*}These authors contributed equally to this work

Correspondence: Wei Shen, Department of General Surgery, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, People's Republic of China, Tel +86 13385110723, Email shenweij@s@outlook.com

Purpose: This study aims to construct a machine learning model that can recognize preoperative, intraoperative, and postoperative high-risk indicators and predict the onset of venous thromboembolism (VTE) in patients.

Patients and Methods: A total of 1239 patients diagnosed with gastric cancer were enrolled in this retrospective study, among whom 107 patients developed VTE after surgery. We collected 42 characteristic variables of gastric cancer patients from the database of Wuxi People's Hospital and Wuxi Second People's Hospital between 2010 and 2020, including patients' demographic characteristics, chronic medical history, laboratory test characteristics, surgical information, and patients' postoperative conditions. Four machine learning algorithms, namely, extreme gradient boosting (XGBoost), random forest (RF), support vector machine (SVM), and k-nearest neighbor (KNN), were employed to develop predictive models. We also utilized Shapley additive explanation (SHAP) for model interpretation and evaluated the models using k-fold cross-validation, receiver operating characteristic (ROC) curves, calibration curves, decision curve analysis (DCA), and external validation metrics.

Results: The XGBoost algorithm demonstrated superior performance compared to the other three prediction models. The area under the curve (AUC) value for XGBoost was 0.989 in the training set and 0.912 in the validation set, indicating high prediction accuracy. Furthermore, the AUC value of the external validation set was 0.85, signifying good extrapolation of the XGBoost prediction model. The results of SHAP analysis revealed that several factors, including higher body mass index (BMI), history of adjuvant radiotherapy and chemotherapy, T-stage of the tumor, lymph node metastasis, central venous catheter use, high intraoperative bleeding, and long operative time, were significantly associated with postoperative VTE.

Conclusion: The machine learning algorithm XGBoost derived from this study enables the development of a predictive model for postoperative VTE in patients after radical gastrectomy, thereby assisting clinicians in making informed clinical decisions.

Keywords: gastric neoplasms, gastrectomy, venous thromboembolism, risk factors, machine learning, prediction model

Introduction

The incidence of gastric cancer, which ranks as the second highest among all malignant tumors, is increasing yearly due to changes in lifestyle and dietary habits.¹ Patients are frequently diagnosed in advanced stages, resulting in a poor prognosis.² Surgery remains a primary approach to treating gastric cancer.³ Advances in minimally invasive surgical techniques, such as laparoscopic and robotic surgery, have reduced postoperative recurrence rates and minimized patient trauma, thereby enhancing survival and quality of life.^{4,5} However, given the intricate anatomy of the stomach and the distribution of surrounding lymph nodes, as well as the high technical demands of radical gastrectomy, life-threatening complications may occur, including anastomotic leakage (AL) and venous thromboembolism (VTE).^{6,7} VTE is a serious

complication in cancer patients, causing significant noncancer-related mortality.⁸ Although more common in patients with malignancies such as pancreatic, brain, lung, and ovarian cancers, postoperative patients with gastric cancer are also at risk for this complication.⁹ VTE can have serious consequences, including respiratory distress, heart failure, and death, if the thrombus travels to the lungs. In the lower limbs, VTE can cause edema, pain, and even skin ulceration if it affects the deep veins.¹⁰ This complication prolongs hospitalization and imposes a significant financial burden on patients and their families. Prophylactic treatment can effectively reduce the risk of patient death by 50% in high-risk patients.^{11,12} Accurate prediction of postoperative VTE in patients undergoing radical gastrectomy and identification of high-risk patients is therefore crucial.

Surgeons often rely on their previous clinical experience to assess the risk of VTE in surgical patients, but this method is often unreliable due to the subjective nature of the surgeon's experience and its temporal limitations. Ten years ago, some researchers utilized parametric regression methods to predict VTE in surgical patients preoperatively, which were relatively accurate. However, due to the complex relationships between clinical characteristic variables and the multifactorial nature of postoperative complications, using regression models cannot solve the problem of predicting disease. Moreover, when the number of cases for variables in the regression model is small, the predicted results can be highly variable, which limits the usefulness of the regression model.^{13–15} In recent years, artificial intelligence (AI) has rapidly advanced in the medical field, with machine learning being a major branch of AI that offers more stable model building and accurate predictions, making it popular among clinicians for clinical prediction and other applications.^{16,17} Here, we analyzed the clinical information of gastric cancer patients and used machine learning algorithms to develop a prediction model for VTE after radical gastrectomy. This model can identify high-risk patients for VTE without relying on conventional imaging examinations such as abdominal computed tomography (CT), which can reduce medical costs and assist clinicians in providing timely and precise individualized treatment plans.

Materials and Methods

Study Subjects

In this study, we utilized clinical data obtained from the databases of two medical institutions, namely, Wuxi People's Hospital affiliated with Nanjing Medical University and Wuxi Second People's Hospital. Case inclusion criteria were as follows: (1) patients aged 18–80 years with gastric cancer; (2) patients undergoing open radical gastrectomy or laparoscopic-assisted radical gastrectomy; and (3) the surgical team consisted of senior surgeons with the ability to independently perform radical gastrectomy. Case exclusion criteria were as follows: (1) patients with other malignant tumors; (2) patient with a previous diagnosis of VTE; (3) patients diagnosed with hemophilia, thrombocytopenic purpura, vascular purpura, and other disorders characterized by a tendency towards bleeding or abnormal bleeding; (4) patients diagnosed with hematologic conditions such as vascular purpura, platelets, and coagulation factors; (5) patients diagnosed with vital organ conditions such as liver and kidney diseases that are intolerant to surgical interventions; (6) patients on anticoagulant medications such as warfarin and heparin, anti-platelet medications such as aspirin and clopidogrel, and non-steroidal anti-inflammatory drugs such as ibuprofen during the perioperative period; and (7) patients with missing cases, clinical data, or visits. All patients in the study were followed up for at least 3 years after surgery. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Wuxi People's Hospital and Wuxi Second People's Hospital, with approval number KY22085.

Study Design and Data Collection

Clinical data on gastric cancer patients between 2010 and 2020 were obtained from the databases of Wuxi People's Hospital and Wuxi Second People's Hospital. The data included 42 preoperative variables (within 24 h before the day of surgery), intraoperative variables, and postoperative variables (occurring 48 h after the initial surgery). Preoperative variables collected included patient demographic characteristics (gender, age, history of smoking, alcohol abuse, and body mass index), basic clinical characteristics (American Society of Anesthesiologists score, nutrition risk screening 2002 score, history of surgery, adjuvant chemotherapy history, adjuvant radiotherapy history, and use of central venous catheters), basic medical history (anemia, ileus, ulcerative colitis, Crohn's disease, diabetes mellitus, hypertension,

chronic obstructive pulmonary disease, hyperlipidemia, and coronary artery disease), laboratory tests (albumin, carcinoembryonic antigen and carbohydrate antigen 19–9), and tumor characteristics (T-stage, N-stage, peripheral nerve invasion, tumor size, and tumor number). The intraoperative variables collected included the type of surgery, surgical approach, number of lymph nodes dissected, duration of surgery, intraoperative bleeding, intraoperative blood transfusion, intraoperative percutaneous arterial oxygen saturation status, abdominal drainage, whether the patient experienced intraoperative tachycardia, and whether the surgery was an emergency procedure. Postoperative variables collected included laboratory test indices (neutrophil to lymphocyte ratio, procalcitonin, C-reactive protein, and serum amyloid A). The outcome variable of the study was postoperative VTE.

Development and Evaluation of Predictive Models

In the present study, statistical analysis was carried out using SPSS software and R software. The following steps were taken for the construction and evaluation of the clinical prediction models:¹ Data preprocessing. The dataset used in this study comprised patients diagnosed with gastric cancer at Wuxi People's Hospital between January 2010 and January 2020. This dataset was employed for building the predictive model. The establishment set was randomly divided into a training set (70%) and a test set (30%). A separate dataset of gastric cancer patients from Wuxi Second People's Hospital during the same period was used as an external validation set to evaluate the model's performance.² Univariate and multivariate regression analyses were conducted on the model establishment set's data. The chi-square test was utilized to compare the differences between two groups for categorical variables, while a *t*-test was applied for continuous variables that followed a normal distribution. For continuous variables that did not conform to a normal distribution, the rank sum test was employed. A *p* value of less than 0.05 was considered statistically significant. Logistic regression analysis was performed for variables that were significant in the univariate analysis to obtain independent influences on postoperative VTE. Four models, extreme gradient boosting (XGBoost), random forest (RF), support vector machine (SVM), and k-nearest neighbor algorithm (KNN), were used to score the importance of each factor and rank them according to the importance of the weight of the influencing factor. The variables that ranked in the top ten in all four model rankings and were meaningful in univariate and multivariate analyses were selected.³ Evaluate and build prediction models. The filtered clinical variables were incorporated into four machine learning algorithms, including SVM, RF, XGBoost, and KNN. The four models were evaluated using three criteria: discrimination, calibration, and clinical utility, with the best performing model being selected for further analysis. A receiver operating characteristic (ROC) curve was used to calculate the area under the curve (AUC) value and gauge the model's predictive ability. The calibration curve was plotted to assess the agreement between the predicted and actual results, and decision curve analysis (DCA) was performed to determine the benefit to patients from interventional treatment. Internal validation was conducted through a k-fold cross-validation methodology.⁴ External validation of the best model. The generalizability and predictive efficiency of the model was assessed by applying it to an external validation set and plotting ROC curves.⁵ Model interpretation. The contribution of each feature in the sample to the prediction was obtained through Shapley value-based Shapley additive explanation (SHAP). The ranking of risk factors' importance was depicted through the SHAP summary plot, and the prediction results of individual samples were analyzed and interpreted through the SHAP force plot.

Results

Clinical Information of the Patients

In this study, we included a total of 1239 patients diagnosed with gastric cancer, among whom 107 patients (8.64%) developed postoperative VTE (Table 1 and Figure 1). The original data presented in the study are included in Table S1.

Screening for Risk Factors for Postoperative VTE

The results of univariate and multivariate analyses showed that body mass index (BMI), history of adjuvant radiotherapy, history of adjuvant chemotherapy, albumin (ALB), use of central venous catheters (CVCs), history of hypertension, emergency surgery, duration of surgery, intraoperative bleeding, T-stage, N-stage, and neutrophil to lymphocyte ratio

Table 1 Characteristic Distribution of Data in the Establishment Set and External Validation Set

Variables		Overall (n=1239)	Establishment Set (n=873)	External Validation Set (n=366)	P-value
Sex	Female	728(58.757)	577(66.094)	151(41.257)	<0.001
	Male	511(41.243)	296(33.906)	215(58.743)	
Age	<65	831(67.070)	687(78.694)	144(39.344)	<0.001
	≥65	408(32.930)	186(21.306)	222(60.656)	
BMI	<25 kg/m ²	909(73.366)	680(77.892)	229(62.568)	<0.001
	≥25 kg/m ²	330(26.634)	193(22.108)	137(37.432)	
ASA	<3	638(51.493)	412(47.194)	226(61.749)	<0.001
	≥3	601(48.507)	461(52.806)	140(38.251)	
Smoking history	No	676(54.560)	438(50.172)	238(65.027)	<0.001
	Yes	563(45.440)	435(49.828)	128(34.973)	
Drinking history	No	630(50.847)	399(45.704)	231(63.115)	<0.001
	Yes	609(49.153)	474(54.296)	135(36.885)	
Surgical history	No	931(75.141)	661(75.716)	270(73.770)	0.47
	Yes	308(24.859)	212(24.284)	96(26.230)	
Adjuvant Chemotherapy	No	883(71.267)	645(73.883)	238(65.027)	0.002
Adjuvant Radiotherapy	Yes	356(28.733)	228(26.117)	128(34.973)	
ALB	No	863(69.653)	587(67.239)	276(75.410)	0.004
	Yes	376(30.347)	286(32.761)	90(24.590)	
NRS2002 score	≥30 g/L	746(60.210)	592(67.812)	154(42.077)	<0.001
	<30 g/L	493(39.790)	281(32.188)	212(57.923)	
CEA level	<3	859(69.330)	592(67.812)	267(72.951)	0.074
	≥3	380(30.670)	281(32.188)	99(27.049)	
CA19-9 level	<5 ng/mL	878(70.864)	634(72.623)	244(66.667)	0.035
	≥5 ng/mL	361(29.136)	239(27.377)	122(33.333)	
Anemia	<37 U/mL	898(72.478)	645(73.883)	253(69.126)	0.087
	≥37 U/mL	341(27.522)	228(26.117)	113(30.874)	
Ileus	No	904(72.962)	612(70.103)	292(79.781)	<0.001
	Yes	335(27.038)	261(29.897)	74(20.219)	
CHD	No	902(72.801)	602(68.958)	300(81.967)	<0.001
	Yes	337(27.199)	271(31.042)	66(18.033)	
COPD	No	985(79.500)	695(79.611)	290(79.235)	0.881
	Yes	254(20.500)	178(20.389)	76(20.765)	
Diabetes	No	996(80.387)	702(80.412)	294(80.328)	0.973
	Yes	243(19.613)	171(19.588)	72(19.672)	
Hyperlipidemia	No	876(70.702)	575(65.865)	301(82.240)	<0.001
	Yes	363(29.298)	298(34.135)	65(17.760)	
Hypertension	No	905(73.043)	611(69.989)	294(80.328)	<0.001
	Yes	334(26.957)	262(30.011)	72(19.672)	
Ulcerative colitis	No	999(80.630)	695(79.611)	304(83.060)	0.161
	Yes	240(19.370)	178(20.389)	62(16.940)	
Crohn's Disease	No	988(79.742)	689(78.923)	299(81.694)	0.268
	Yes	251(20.258)	184(21.077)	67(18.306)	
Surgery type	No	1017(82.082)	705(80.756)	312(85.246)	0.06
	Yes	222(17.918)	168(19.244)	54(14.754)	
Surgery type	Laparoscopic surgery	608(49.072)	438(50.172)	170(46.448)	0.232
	Open surgery	631(50.928)	435(49.828)	196(53.552)	

(Continued)

Table 1 (Continued).

Variables		Overall (n=1239)	Establishment Set (n=873)	External Validation Set (n=366)	P-value
Surgical procedure	Proximal gastrectomy	402(32.446)	287(32.875)	115(31.421)	0.689
	Distal gastrectomy	429(34.625)	305(34.937)	124(33.880)	
	Total gastrectomy	408(32.930)	281(32.188)	127(34.699)	
Emergency surgery	No	854(68.927)	587(67.239)	267(72.951)	0.047
	Yes	385(31.073)	286(32.761)	99(27.049)	
Surgery time	<270 min	883(71.267)	628(71.936)	255(69.672)	0.422
	≥270 min	356(28.733)	245(28.064)	111(30.328)	
CVCs	No	870(70.218)	637(72.967)	233(63.661)	0.001
	Yes	369(29.782)	236(27.033)	133(36.339)	
Bleeding volume	<100 mL	951(76.755)	692(79.267)	259(70.765)	0.001
	≥100 mL	288(23.245)	181(20.733)	107(29.235)	
Blood Transfusion	No	897(72.397)	596(68.270)	301(82.240)	<0.001
	Yes	342(27.603)	277(31.730)	65(17.760)	
Lymph node dissection	<25	917(74.011)	662(75.830)	255(69.672)	0.024
	≥25	322(25.989)	211(24.170)	111(30.328)	
SpO ₂	≥90%	922(74.415)	630(72.165)	292(79.781)	0.005
	<90%	317(25.585)	243(27.835)	74(20.219)	
Tachycardia	No	984(79.419)	712(81.558)	272(74.317)	0.004
	Yes	255(20.581)	161(18.442)	94(25.683)	
Abdominal drainage	No	924(74.576)	717(82.131)	207(56.557)	<0.001
	Yes	315(25.424)	156(17.869)	159(43.443)	
T-stage	T1~T2	898(72.478)	619(70.905)	279(76.230)	0.056
	T3~T4	341(27.522)	254(29.095)	87(23.770)	
N-stage	N0	921(74.334)	640(73.310)	281(76.776)	0.203
	N1~N3	318(25.666)	233(26.690)	85(23.224)	
PNI	No	1082(87.328)	766(87.743)	316(86.339)	0.498
	Yes	157(12.672)	107(12.257)	50(13.661)	
Tumor number	<2	998(80.549)	717(82.131)	281(76.776)	0.03
	≥2	241(19.451)	156(17.869)	85(23.224)	
Tumor size	<5 cm	836(67.474)	582(66.667)	254(69.399)	0.349
	≥5 cm	403(32.526)	291(33.333)	112(30.601)	
PCT level	<0.05 ng/mL	970(78.289)	715(81.901)	255(69.672)	<0.001
	≥0.05 ng/mL	269(21.711)	158(18.099)	111(30.328)	
CRP level	<10 mg/l	941(75.948)	683(78.236)	258(70.492)	0.004
	≥10 mg/l	298(24.052)	190(21.764)	108(29.508)	
SAA level	<10 mg/l	885(71.429)	675(77.320)	210(57.377)	<0.001
	≥10 mg/l	354(28.571)	198(22.680)	156(42.623)	
NLR	<3	867(69.976)	598(68.499)	269(73.497)	0.08
	≥3	372(30.024)	275(31.501)	97(26.503)	

Abbreviations: BMI, body mass index; ASA, The American Society of Anesthesiologists; ALB, albumin; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 19-9; PCT, procalcitonin; CRP, C-reactive protein; SAA, serum amyloid A; NRS2002, nutrition risk screening 2002; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; PNI, peripheral nerve invasion; SPO₂, percutaneous arterial oxygen saturation; NLR, neutrophil to lymphocyte ratio; CVCs, central venous catheters.

(NLR) were independent influencing factors for VTE ($P < 0.05$) (Table 2). XGBoost, RF, SVM, and KNN algorithms screened for risk factors influencing postoperative VTE included BMI, history of adjuvant radiotherapy, history of adjuvant chemotherapy, use of CVC, duration of surgery, intraoperative bleeding, T-stage, N-stage, and NLR (Figure 2A–D). After comprehensive analysis, the risk factors included in this prediction model included BMI ≥ 25 kg/m², history of adjuvant radiotherapy,

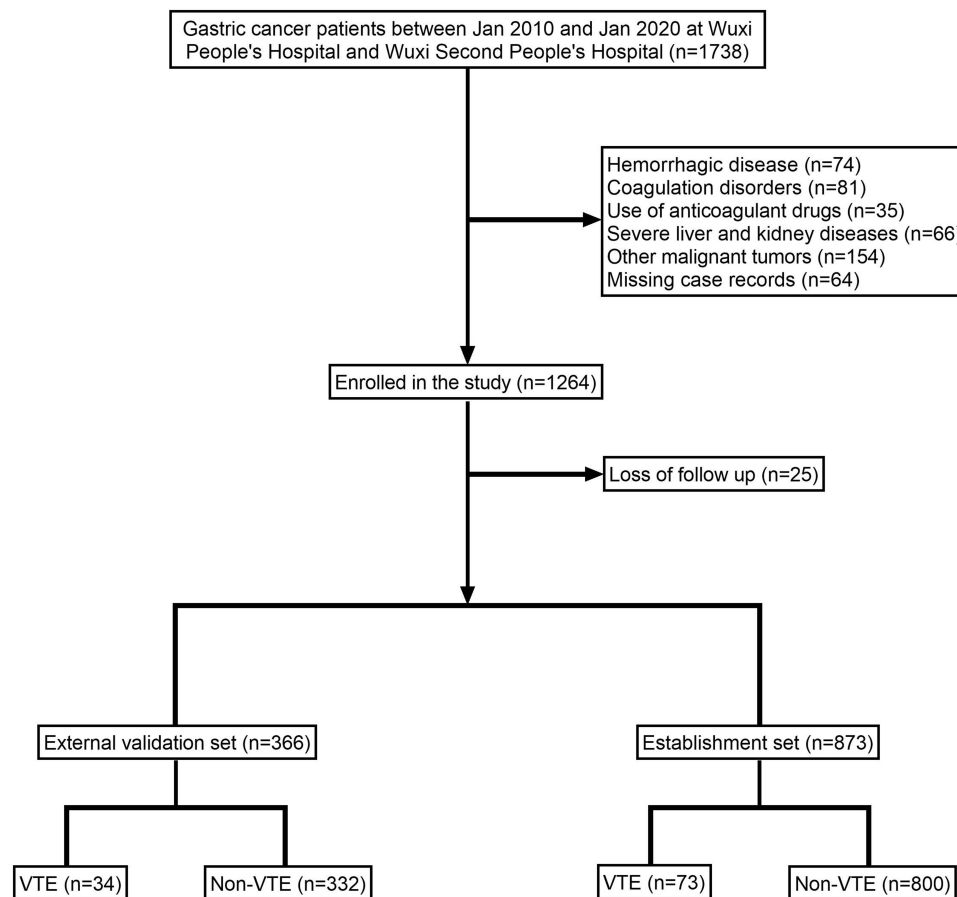


Figure 1 Flow diagram of patients included in the study.

Abbreviation: VTE, venous thromboembolism.

history of adjuvant chemotherapy, use of CVC, surgery time ≥ 270 min, intraoperative bleeding ≥ 100 mL, T-stage, tumor lymph node metastasis, and NLR ≥ 3 .

Model Building and Evaluation

The ROC curve results show that XGBoost has an AUC value as high as 0.989 in the training set; the AUC value in the validation set is 0.912, which is the best performance among the four models (Table 3). The Brier score is a significant measure of the accuracy of model prediction, and a small Brier score for all four models signifies an accurate prediction. In this study, the slight deviation of the calibration curves of all four models from the diagonal line suggests their excellent calibration performance, and their predicted probabilities can be relied upon for making decisions. The net benefit curves of all four models showed an “S” shaped pattern, indicating an increase in net benefit with an increase in the probability threshold, followed by a decline after reaching the highest point. The XGBoost model showed the high net benefit among all models, and the optimal threshold point was located within the probability range of the actual disease state (Figure 3A–D). The generalization ability of the four models was compared using k-fold cross-validation. In this study, the test set comprised N=262 cases (30.01%), and the remaining samples were used for 10-fold cross-validation. The XGBoost algorithm exhibited an AUC value of 0.9036 ± 0.0566 in the validation set and an AUC value of 0.9212 in the test set, with an accuracy of 0.8855 (Figure 4A–C). The RF algorithm displayed an AUC value of 0.8820 ± 0.0511 in the validation set and an AUC value of 0.8485 in the test set, with an accuracy of 0.7786. The SVM algorithm showed an AUC value of 0.8788 ± 0.0704 in the validation set and an AUC value of 0.8788 in the test set, with an accuracy of 0.8702. Finally, the KNN algorithm exhibited an AUC value of 0.8263 ± 0.1192 in the validation set and an AUC value of 0.8065 in the test set, with an accuracy of 0.9046. After a comprehensive comparison, the XGBoost algorithm was selected to construct the model in this study.

Table 2 Univariate and Multivariate Analyses of Variables Related to Postoperative VTE

Variables		Univariate Analysis			Multivariate Analysis		
		OR	95% CI	P-value	OR	95% CI	P-value
Sex	Female	Reference			Reference		
	Male	3.89	[2.388,6.336]	<0.001	1.354	[0.581,3.176]	0.483
Age	<65	Reference			Reference		
	≥65	3.35	[2.066,5.432]	<0.001	0.981	[0.416,2.271]	0.965
BMI	<25 kg/m ²	Reference			Reference		
	≥25 kg/m ²	6.975	[4.265,11.407]	<0.001	3.01	[1.335,6.87]	0.008
ASA	<3	Reference					
	≥3	0.91	[0.570,1.452]	0.691			
Smoking history	No	Reference					
	Yes	0.925	[0.579,1.477]	0.744			
Drinking history	No	Reference					
	Yes	1.354	[0.839,2.185]	0.215			
Surgical history	No	Reference					
	Yes	0.947	[0.545,1.644]	0.846			
Adjuvant Chemotherapy	No	Reference			Reference		
	Yes	0.496	[0.263,0.936]	0.03	0.314	[0.115,0.789]	0.018
Adjuvant Radiotherapy	No	Reference			Reference		
	Yes	4.128	[2.532,6.728]	<0.001	3.301	[1.573,7.202]	0.002
ALB	≥30 g/L	Reference			Reference		
	<30 g/L	9.982	[5.634,17.686]	<0.001	3.874	[1.61,9.707]	0.003
NRS2002 score	<3	Reference					
	≥3	1.152	[0.705,1.883]	0.572			
CEA level	<5 ng/mL	Reference					
	≥5 ng/mL	0.564	[0.310,1.027]	0.061			
CA19-9 level	<37 U/mL	Reference					
	≥37 U/mL	0.662	[0.369,1.189]	0.168			
Anemia	No	Reference					
	Yes	1.141	[0.691,1.882]	0.606			
Ileus	No	Reference			Reference		
	Yes	3.151	[1.959,5.068]	<0.001	1.488	[0.688,3.158]	0.304
CHD	No	Reference					
	Yes	1.417	[0.827,2.427]	0.205			
COPD	No	Reference					
	Yes	1.282	[0.735,2.237]	0.381			
Diabetes	No	Reference			Reference		
	Yes	3.018	[1.873,4.864]	<0.001	0.997	[0.421,2.325]	0.994
Hyperlipidemia	No	Reference					
	Yes	0.637	[0.364,1.115]	0.114			
Hypertension	No	Reference			Reference		
	Yes	9.048	[5.482,14.935]	<0.001	2.817	[1.267,6.277]	0.011
Ulcerative colitis	No	Reference					
	Yes	0.67	[0.354,1.268]	0.219			
Crohn's Disease	No	Reference					
	Yes	0.926	[0.506,1.697]	0.804			
Surgery type	Laparoscopic surgery	Reference					
	Open surgery	0.979	[0.613,1.563]	0.93			
Surgical procedure	Proximal gastrectomy	Reference					
	Distal gastrectomy	1.403	[0.774,2.541]	0.264			
	Total gastrectomy	1.477	[0.812,2.690]	0.202			

(Continued)

Table 2 (Continued).

Variables		Univariate Analysis			Multivariate Analysis		
		OR	95% CI	P-value	OR	95% CI	P-value
Emergency surgery	No	Reference			Reference		
	Yes	5.336	[3.218,8.849]	<0.001	2.983	[1.43,6.396]	0.004
Surgery time	<270 min	Reference			Reference		
	≥270 min	10.75	[6.185,18.682]	<0.001	2.72	[1.204,6.34]	0.018
CVCs	No	Reference			Reference		
	Yes	6.073	[3.698,9.974]	<0.001	3.52	[1.643,7.687]	0.001
Bleeding volume	<100 mL	Reference			Reference		
	≥100 mL	5.672	[3.494,9.207]	<0.001	2.561	[1.196,5.48]	0.015
Blood transfusion	No	Reference			Reference		
	Yes	2.702	[1.684,4.336]	<0.001	1.973	[0.845,4.669]	0.118
Lymph node dissection	<25	Reference					
	≥25	1.111	[0.651,1.897]	0.699			
SpO ₂	≥90%	Reference					
	<90%	1.275	[0.772,2.105]	0.343			
Tachycardia	No	Reference					
	Yes	1.077	[0.596,1.947]	0.806			
Abdominal drainage	No	Reference			Reference		
	Yes	2.278	[1.358,3.820]	0.002	1.939	[0.857,4.341]	0.108
T-stage	T1~T2	Reference			Reference		
	T3~T4	3.721	[2.307,6.001]	<0.001	3.061	[1.449,6.622]	0.004
N-stage	N0	Reference			Reference		
	N1~N3	4.548	[2.809,7.363]	<0.001	2.303	[1.093,4.845]	0.027
PNI	No	Reference					
	Yes	0.943	[0.456,1.950]	0.873			
Tumor number	<2	Reference					
	≥2	0.748	[0.386,1.452]	0.391			
Tumor size	<5 cm	Reference			Reference		
	≥5 cm	4.006	[2.459,6.528]	<0.001	1.023	[0.448,2.329]	0.957
PCT level	<0.05 ng/mL	Reference					
	≥0.05 ng/mL	1.006	[0.549,1.845]	0.984			
CRP level	<10 mg/l	Reference			Reference		
	≥10 mg/l	1.965	[1.189,3.247]	0.008	1.266	[0.535,2.901]	0.583
SAA level	<10 mg/l	Reference					
	≥10 mg/l	0.885	[0.498,1.572]	0.677			
NLR	<3	Reference			Reference		
	≥3	4.414	[2.706,7.199]	<0.001	3.926	[1.895,8.42]	<0.001

Abbreviations: OR, odds ratio; CI, confidence interval; BMI, body mass index; ASA, The American Society of Anesthesiologists; ALB, albumin; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 19-9; PCT, procalcitonin; CRP, C-reactive protein; SAA, serum amyloid A; NRS2002, nutrition risk screening 2002; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; PNI, peripheral nerve invasion; SPO₂, percutaneous arterial oxygen saturation; NLR, neutrophil to lymphocyte ratio; CVCs, central venous catheters.

Model External Validation

The external validation set yielded an AUC value of 0.85, suggesting that the disease prediction model was characterized by a high degree of accuracy (Figure 4D).

Model Explanation

The SHAP summary plot results demonstrated that the risk factors for VTE after radical gastrectomy were prioritized as follows: surgery time ≥ 270 min, usage of CVC, intraoperative bleeding ≥ 100 mL, history of adjuvant radiotherapy, NLR ≥ 3, gastric cancer of T3 and T4, history of adjuvant chemotherapy, tumor lymph node metastasis, and BMI ≥ 25 kg/m²

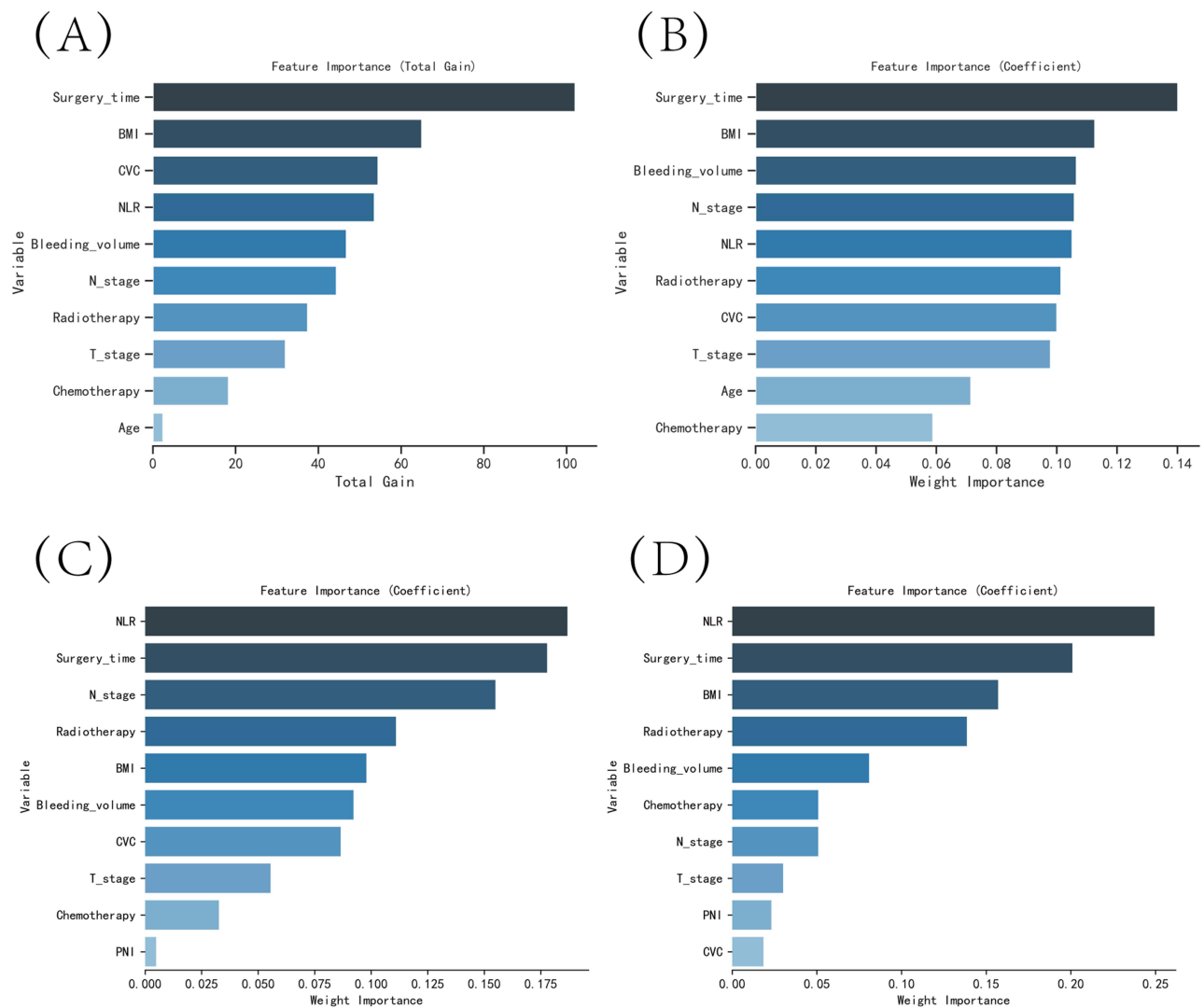


Figure 2 The variable influence factor ranking plots of the four models. **(A)** Variable importance ranking diagram of the XGBoost model. **(B)** Variable importance ranking diagram of the RF model. **(C)** Variable importance ranking diagram of the SVM model. **(D)** Variable importance ranking diagram of the KNN model.

Abbreviations: BMI, body mass index; ASA, PNI, peripheral nerve invasion; NLR, neutrophil to lymphocyte ratio; CVC, central venous catheter.

(Figure 5). The SHAP force plot presents the results of predictive analysis for three patients with VTE using the study model. For patient one, the model predicted disease occurrence with a probability of 0.029, with increased probability associated with gastric cancer of T3 and T4, use of CVC, and surgery time ≥ 270 min and decreased probability associated with a history of adjuvant chemotherapy. For patient two, the model predicted disease occurrence with a probability of 0.374, with increased probability associated with gastric cancer of T3 and T4, use of CVC, and surgery time ≥ 270 min. For patient three, the model predicted a disease occurrence probability of 0.748, with increased probability associated with NLR ≥ 3 , use of CVC, gastric cancer at T3 and T4, and surgery time ≥ 270 min (Figure 6A–C).

Discussion

In this study, we developed risk prediction models using four machine learning algorithms. Despite each algorithm having unique characteristics and potential for various scenarios, we evaluated and determined that the XGBoost algorithm demonstrated the highest accuracy, better stability, and superior generalization ability. Compared to the RF algorithm, the XGBoost algorithm allows for easy assessment of the feature contribution to the model by ranking the importance of the features, providing more intuitive output of the model's conclusions. Additionally, the XGBoost

Table 3 Evaluation of the Four Models

		AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 Score (95% CI)	Cutoff (95% CI)	Kappa (95% CI)
KNN	Training set	0.971 (0.955–0.988)	0.949(0.945–0.953)	0.988(0.983–0.993)	0.913(0.905–0.921)	0.820(0.798–0.841)	0.200(0.200–0.200)	0.681(0.656–0.705)
	Validation set	0.852 (0.746–0.958)	0.921(0.912–0.931)	0.773(0.718–0.829)	0.906(0.889–0.923)	0.676(0.617–0.734)	0.200(0.200–0.200)	0.547(0.501–0.593)
XGBoost	Training set	0.989 (0.982–0.996)	0.944(0.937–0.951)	0.978(0.969–0.987)	0.939(0.931–0.948)	0.752(0.726–0.779)	0.157(0.130–0.184)	0.653(0.613–0.693)
	Validation set	0.912 (0.848–0.975)	0.901(0.886–0.915)	0.898(0.853–0.943)	0.837(0.793–0.880)	0.636(0.601–0.670)	0.157(0.130–0.184)	0.490(0.434–0.545)
RF	Training set	0.913 (0.880–0.946)	0.845(0.822–0.868)	0.855(0.815–0.895)	0.842(0.815–0.869)	0.488(0.455–0.521)	0.125(0.106–0.145)	0.402(0.365–0.439)
	Validation set	0.904 (0.844–0.965)	0.843(0.816–0.869)	0.879(0.825–0.932)	0.829(0.786–0.873)	0.514(0.468–0.559)	0.125(0.106–0.145)	0.369(0.332–0.405)
SVM	Training set	0.985 (0.974–0.997)	0.976(0.969–0.982)	0.901(0.888–0.915)	0.982(0.974–0.990)	0.871(0.842–0.899)	0.154(0.116–0.191)	0.887(0.875–0.899)
	Validation set	0.884 (0.808–0.960)	0.888(0.870–0.906)	0.874(0.820–0.928)	0.794(0.749–0.838)	0.594(0.537–0.652)	0.154(0.116–0.191)	0.430(0.398–0.462)

Abbreviations: AUC, area under the curve; RF, random forest; XGBoost, extreme gradient boosting; SVM, support vector machine; KNN, k-nearest neighbor algorithm; CI, confidence interval.

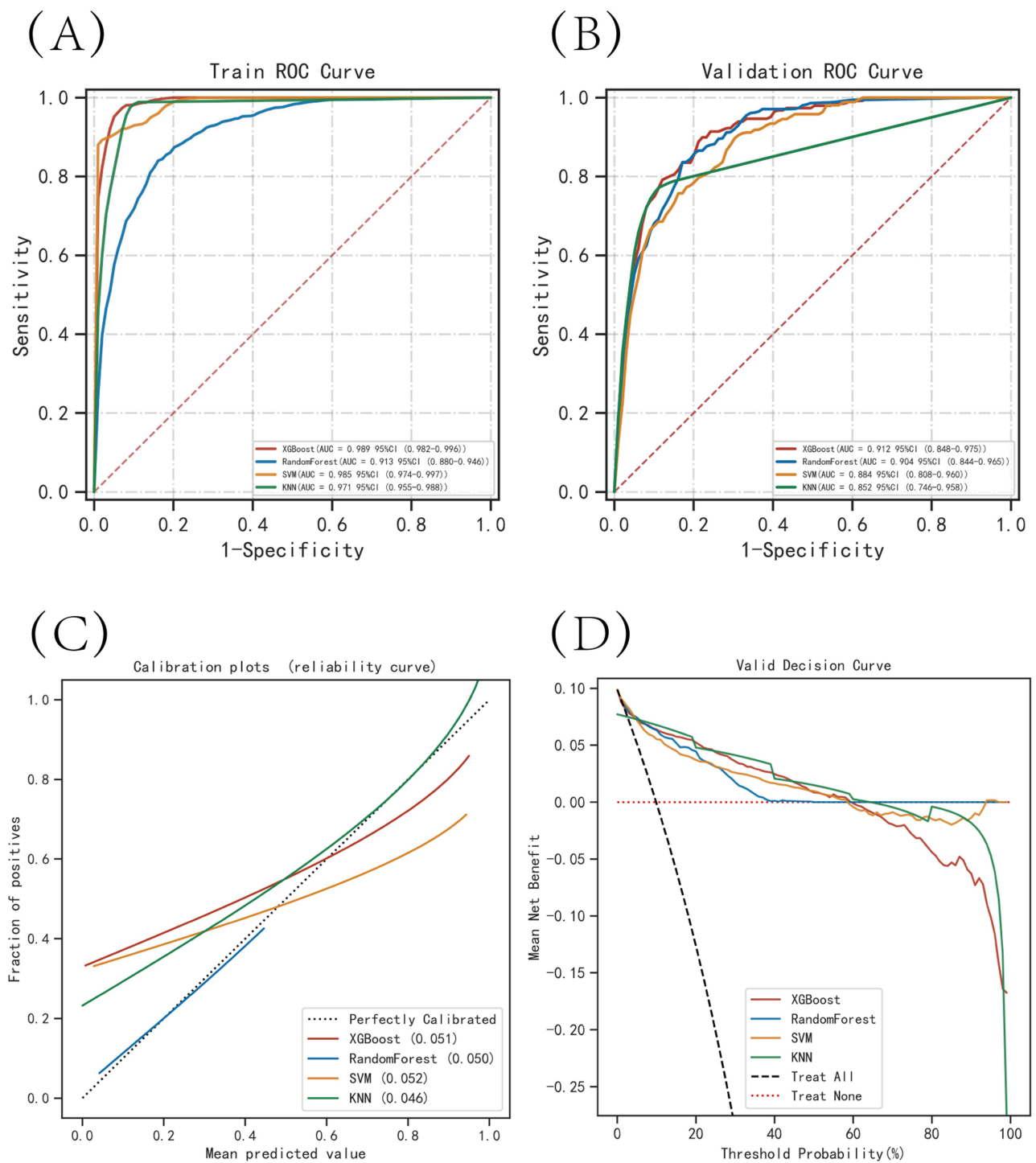


Figure 3 Evaluation of the four models for predicting VTE. **(A)** ROC curves for the training set of the four models. **(B)** ROC curves for the validation set of the four models. **(C)** Calibration plots of the four models. The 45-degree dashed line in each plot represents the ideal correspondence between the predicted (x-axis) and observed (y-axis) probabilities of complications. The closer the distance between the two curves, the higher the predictive accuracy. **(D)** DCA curves of the four models. The point of intersection between the red curve and the “All” curve represents the baseline or starting point, while the point of intersection between the red curve and the “None” curve indicates the decision node where the corresponding patients may derive benefit.

Abbreviations: XGBoost, extreme gradient boosting; SVM, support vector machine; KNN, k-nearest neighbor algorithm.

algorithm offers more options for parameter tuning, which enhances the model’s performance control.¹⁸ The SVM and KNN algorithms have high model complexity, potentially hindering performance on large datasets.¹⁹ Conversely, the XGBoost algorithm has lower model complexity and is better suited for multidimensional studies, reducing

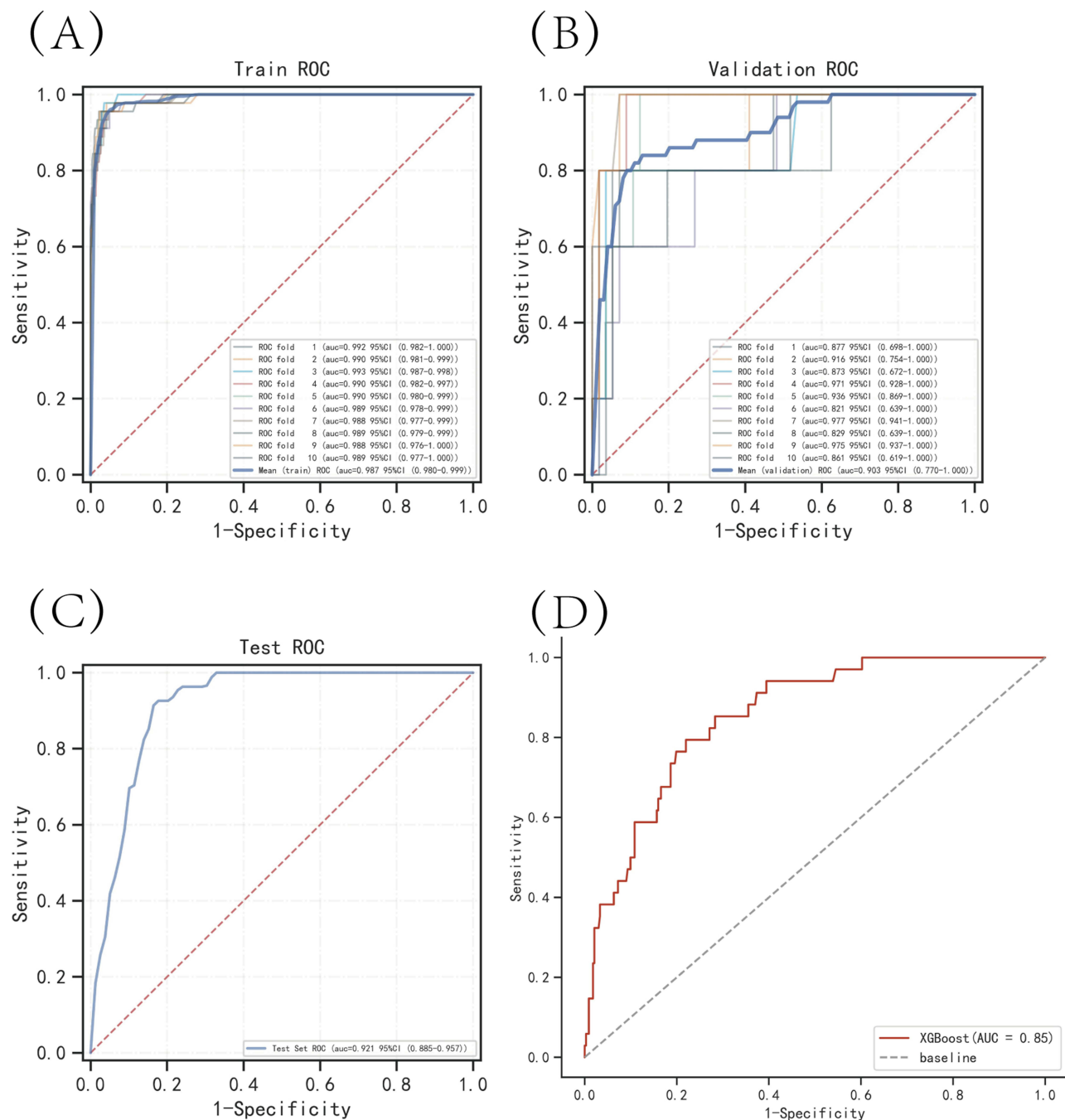


Figure 4 Internal validation of the XGBoost model. **(A)** ROC curve of the XGBoost model for the training set. **(B)** ROC curve of the XGBoost model for the validation set. **(C)** ROC curve of the XGBoost model for the test set. **(D)** External validation of the XGBoost model.

Abbreviations: AUC, area under the curve; XGBoost, extreme gradient boosting.

computational effort and training time. After comprehensive comparison of the four algorithms, we chose the XGBoost algorithm to construct a model predicting postoperative VTE formation in patients in this study.

Several studies^{20,21} have demonstrated the effectiveness of machine learning algorithms in clinical diagnosis and prognosis, as well as their ability to accurately predict adverse outcomes in disease progression compared to traditional diagnostic methods. Machine learning algorithms were also employed in the development of the prediction model in this study. This model can aid clinical decision-makers in accurately identifying patients at high risk for postoperative VTE, enabling patients to avoid unnecessary tests and reducing the financial and physical burdens of diagnostic procedures.

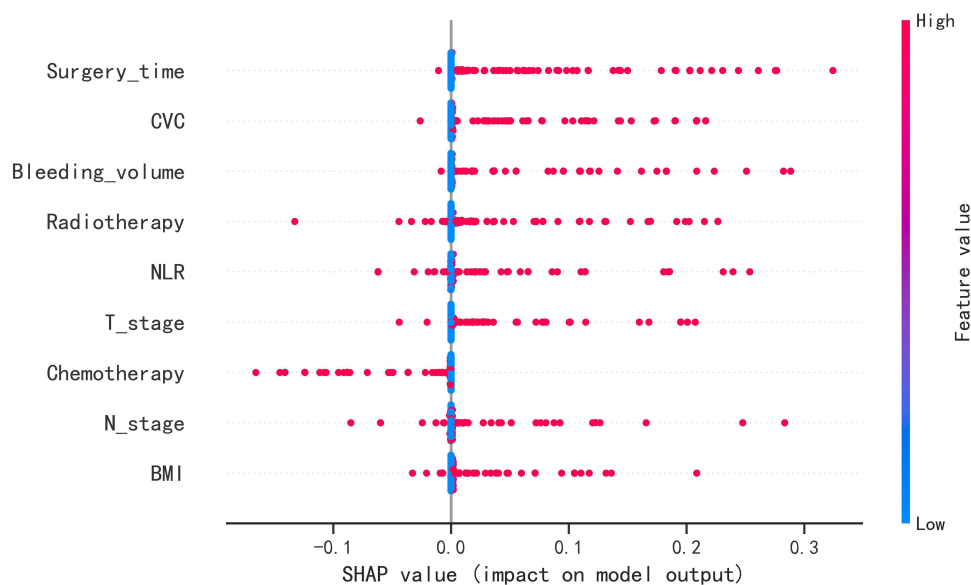


Figure 5 SHAP summary plot. The risk factors are ranked on the y-axis according to their significance, which is determined by the mean of their absolute Shapley values. The higher the risk factor appears on the plot, the more crucial it is for the model.

Abbreviations: BMI, body mass index; ASA, NLR, neutrophil to lymphocyte ratio; CVC, central venous catheter.

Importantly, the model considers risk factors that are often overlooked by clinicians, such as NLR and BMI, and thus may improve patient survival rates. Previous studies^{22,23} have focused on hematological indicators such as platelets and fibrinogen to assess the risk of postoperative thrombosis, but with poor accuracy. This study expands the list of risk factors, with the hope of drawing attention to other high-risk patients. The current study visualized the model by SHAP analysis and showed that VTE was strongly associated with BMI, history of adjuvant radiotherapy, history of adjuvant chemotherapy, use of CVC, duration of surgery, intraoperative bleeding, T-stage, N-stage, and NLR.

Obesity is associated with increased surface tension on the abdominal wall, which impedes instrument movement during surgery and compresses the operating space in the abdominal cavity, requiring greater surgical skill. In addition, due to looser vascular tissue in obese patients, tissue resection during surgery is more likely to cause injury and affect blood supply to the trauma and intestine, increasing the risk of postoperative VTE. A recent study by Jeong²⁴ confirmed the strong association between high BMI and surgical trauma and thrombosis in tumor patients after surgery. Obesity is also associated with increased inflammation and impaired endothelial cell function, and the added pressure on veins further increases the risk of postoperative VTE.

A retrospective investigation conducted in the United States, which encompassed a total of 399 participants,^{25,26} discovered that the risk of venous thrombosis was six times higher among patients who underwent radiotherapy in comparison with those who did not. The results of the study suggest that radiotherapy is a significant factor that influences venous thrombosis, corroborating the present study's findings. Radiotherapy exerts a potent effect in impeding tumor growth and killing tumor cells; nevertheless, it also generates severe detrimental effects.^{27,28} High-energy particles interact with tissues during radiotherapy, resulting in the production of free radicals that cause harm to cell membranes and other cellular structures, including endothelial cells.^{29,30} The damage to endothelial cells expedites the release of pro-inflammatory molecules and triggers the activation of peripheral platelets and leukocytes, leading to thrombosis. Furthermore, impairment of the natural anticoagulant properties of blood vessels can arise due to damage to endothelial cells, amplifying the likelihood of coagulation.³¹ Conversely, radiation therapy can also provoke localized limb swelling, rendering the veins stiff and inelastic, which, in turn, results in venous blood stasis. Curiously, the converse physiological transformations were observed in patients undergoing chemotherapy. While some researchers contend that chemotherapy impairs estrogen-producing cells, diminishing the influence of estrogen on the coagulation system and thus making patients more vulnerable to venous thrombosis,³² chemotherapy is exceedingly efficacious in inhibiting coagulation

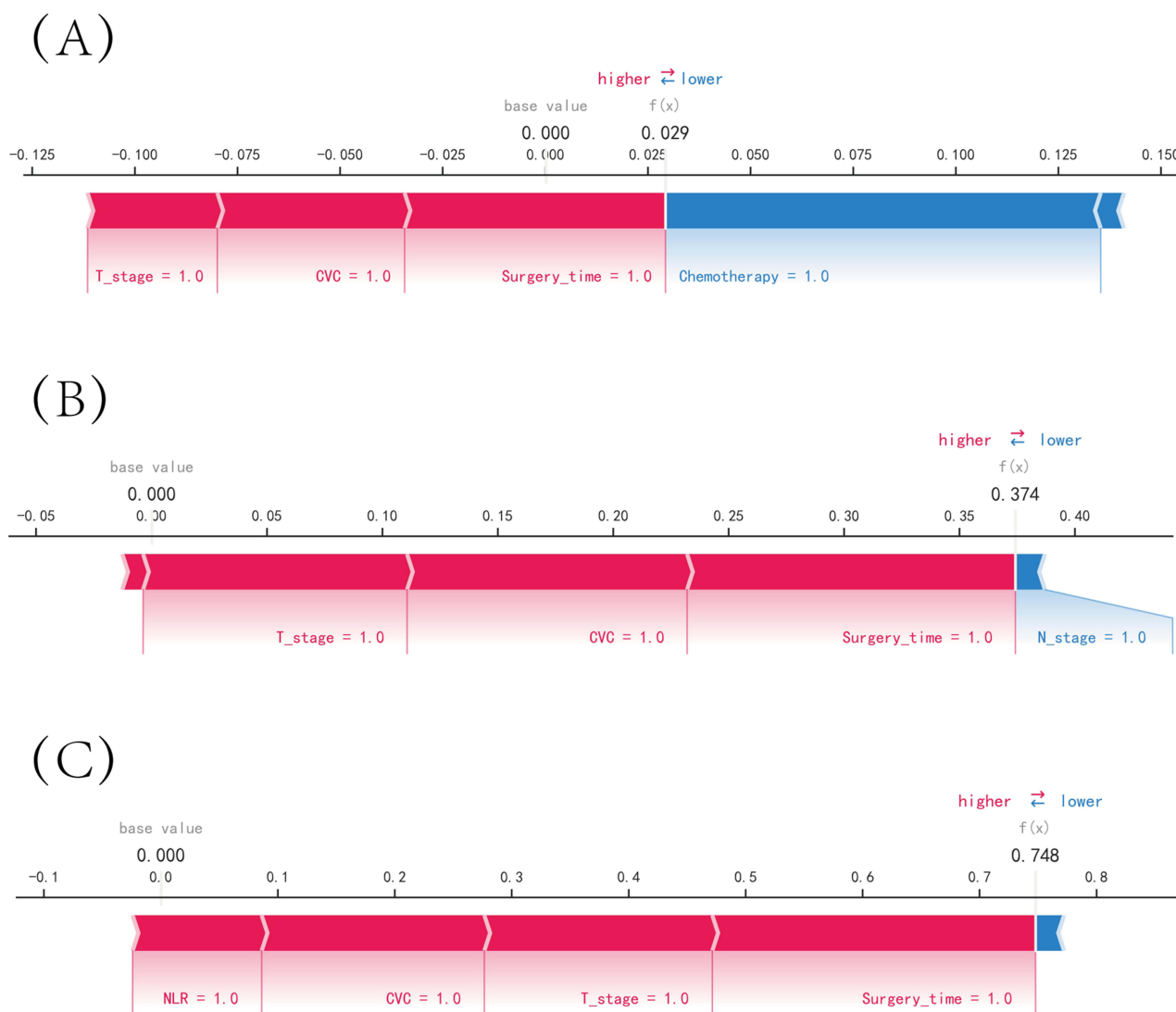


Figure 6 SHAP force plot. The explanatory variables are ordered along the horizontal axis based on the absolute value of their impact, with blue representing features that negatively affect disease prediction, as indicated by a decrease in SHAP values, and red representing features that positively affect disease prediction, as indicated by an increase in SHAP values. **(A)** Predictive Analysis of Patient I. **(B)** Predictive Analysis of Patient II. **(C)** Predictive Analysis of Patient III.

Abbreviations: NLR, neutrophil to lymphocyte ratio; CVC, central venous catheter.

factors such as thrombin and fibrinogen, natural anticoagulants such as protein C and protein S, and also reduces platelet adhesion. These effects play a crucial role in curtailing the incidence of postoperative venous thrombosis.³³

This study has also identified that, in addition to the impact of radiotherapy on thrombogenesis, surgery heightens the risk of thrombosis, particularly in patients who undergo longer surgeries or experience more intraoperative bleeding. As the duration of surgery increases, tissue and vascular damage can occur, which can result in extended postoperative pain, fatigue, and other adverse effects. Patients may then be hesitant to resume activities due to physical discomfort, further increasing the risk of blood clots in the legs and feet. Furthermore, elevated intraoperative bleeding disrupts the balance of coagulation and anticoagulation factors in the patient's blood, triggering the body's procoagulant response.³⁴ Therefore, healthcare professionals should closely monitor patients who undergo longer operative times and experience more intraoperative bleeding and take necessary precautions when thrombosis signs manifest to mitigate the risk of postoperative mortality.³⁵

The findings of this study suggest that patients with more aggressive and invasive gastric cancer have a higher likelihood of developing postoperative VTE, which is likely related to the extent of surgery required to remove the tumor.

Advanced gastric cancer typically necessitates more extensive resection and lymph node dissection, making it challenging to prevent postoperative thrombosis.³⁶ Additionally, malignancy can increase the levels of procoagulant substances and inflammatory factors in the blood, as supported by prior research.^{37,38} Furthermore, clinicians often use a combination of treatments, including adjuvant radiotherapy and chemotherapy, to eradicate advanced gastric cancer, which can increase coagulation and further elevate the risk of thrombosis. In conclusion, postoperative VTE in patients with advanced gastric cancer is a multifactorial phenomenon. Therefore, clinicians should monitor the preoperative blood profile, liver and kidney function of patients, minimize surgical duration and bleeding volume, and encourage patients to engage in physical activity as soon as possible after surgery to prevent thrombus formation.

This study examines the application of three samples to determine the rationale for models predicting VTE. Analysis of disease prediction in sample II identified the use of CVCs as a significant risk factor. The placement of a CVC in a larger venous vessel can lead to vein trauma and contribute to thrombosis.³⁹ Furthermore, the use of CVC may cause adverse complications such as blood stagnation and bacterial colonization of peripheral veins, increasing the risk of thrombosis. Importantly, patients with CVCs require rest and recuperation, which we believe is the main mechanism of thrombosis.

In recent years, several predictive models for VTE have been constructed successfully by clinical investigators.^{38,40} However, these studies usually only include basic clinical characteristics of the patient. The hemagglutination process often triggers the expression of proinflammatory cytokines and signaling molecules, leading to systemic inflammation.⁴¹ Inflammatory cells can be an important predictor of postoperative thrombosis, as suggested by the results of the current study, which indicate that patients with a higher NLR index are more likely to experience VTE in the postoperative period. Neutrophils, a type of white blood cell, play an important role in the inflammatory and immune response. They can cause oxidative stress that damages vascular endothelial cells and inhibits their function while also releasing tissue factors and other procoagulant molecules, increasing the risk of postoperative VTE.⁴² Similarly, lymphocytes have similar functions to neutrophils, and their reduced number usually indicates poorer immune function and a stronger postoperative trigger of infection and inflammation, increasing the risk of VTE in the postoperative period.

Limitations

This study provides a comprehensive evaluation of the discrimination, calibration, and clinical utility of the model, but it also has some limitations. While the study considered multiple risk factors, it did not focus on laboratory test indicators. Additionally, although machine learning algorithms were more accurate, their models were more complex and less interpretable. The entire computational and decision-making process of the model runs in a black box, which is not as intuitive and clear as the logistic regression model.^{43,44} Furthermore, the current study was a retrospective study, which has inherent limitations such as selection bias and retrospective bias. Therefore, future studies should incorporate multicenter prospective studies to further improve the reliability of the findings.

Conclusion

This study employs the XGBoost machine learning algorithm to construct a model that predicts the risk of postoperative VTE. The model exhibits high accuracy in prediction and clinical utility, providing surgeons with a means to promptly identify high-risk patients. Notably, the model identifies VTE as a significant challenge for patients undergoing radical gastrectomy, with risk factors including BMI, adjuvant radiotherapy and chemotherapy history, use of CVC, surgical time, intraoperative bleeding, T-stage, N-stage, and NLR.

Data Sharing Statement

The original data presented in the study are included in the Raw Data/[Table S1](#), and further inquiries can be directed to the corresponding author (shenweijis@outlook.com).

Ethics Statement

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Wuxi People's Hospital, with approval number KY22085. The review committee waived the requirement for written

informed consent because of the retrospective nature of the study. Prior to analysis, confidential patient information was deleted from the entire data set.

Funding

This work was supported by the Top Talent Support Program for young and middle-aged people of Wuxi Health Committee (Grant No. HB2020007).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34. doi:10.3322/caac.21551
2. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in China: good or bad news from the 2018 global cancer statistics? *Cancer Commun*. 2019;39(1):22. doi:10.1186/s40880-019-0368-6
3. Adachi Y, Shiraiishi N, Shiromizu A, Bandoh T, Aramaki M, Kitano S. Laparoscopy-assisted Billroth I gastrectomy compared with conventional open gastrectomy. *Arch Surg*. 2000;135(7):806–810. doi:10.1001/archsurg.135.7.806
4. Caruso S, Patriti A, Roviello F, et al. Laparoscopic and robot-assisted gastrectomy for gastric cancer: current considerations. *World J Gastroenterol*. 2016;22(25):5694–5717. doi:10.3748/wjg.v22.i25.5694
5. Braumann C, Jacobi CA, Menenakos C, Ismail M, Rueckert JC, Mueller JM. Robotic-assisted laparoscopic and thoracoscopic surgery with the da Vinci system: a 4-year experience in a single institution. *Surg Laparosc Endosc Percutan Tech*. 2008;18(3):260–266. doi:10.1097/SLE.0b013e31816f85e5
6. Wang JB, Zheng CH, Li P, et al. Effect of comorbidities on postoperative complications in patients with gastric cancer after laparoscopy-assisted total gastrectomy: results from an 8-year experience at a large-scale single center. *Surg Endosc*. 2017;31(6):2651–2660. doi:10.1007/s00464-016-5279-x
7. Song W, Yuan Y, Peng J, et al. The delayed massive hemorrhage after gastrectomy in patients with gastric cancer: characteristics, management opinions and risk factors. *Eur J Surg Oncol*. 2014;40(10):1299–1306. doi:10.1016/j.ejso.2014.03.020
8. Yamashita Y, Morimoto T, Kimura T. Venous thromboembolism: recent advancement and future perspective. *J Cardiol*. 2022;79(1):79–89. doi:10.1016/j.jcc.2021.08.026
9. Skeik N, Westergard E. Recommendations for VTE prophylaxis in medically ill patients. *Ann Vasc Dis*. 2020;13(1):38–44. doi:10.3400/avd.ra.19-00115
10. Awano N, Okano T, Kawachi R, et al. One-year incidences of venous thromboembolism, bleeding, and death in patients with lung cancer (cancer-VTE subanalysis). *JTO Clin Res Rep*. 2022;3(9):100392. doi:10.1016/j.jtocrr.2022.100392
11. Gould MK, Garcia DA, Wren SM, et al. Prevention of VTE in nonorthopedic surgical patients: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2Suppl):e227S–e77S. doi:10.1378/chest.11-2297
12. Nicholson M, Chan N, Bhagirath V, Ginsberg J. Prevention of venous thromboembolism in 2020 and beyond. *J Clin Med*. 2020;9(8):2467. doi:10.3390/jcm9082467
13. Heidar A, Ravanfar P, Namazi G, Nikseresh T, Niakan H. Determinants of successful non-operative management of intra- peritoneal bleeding following blunt abdominal trauma. *Bull Emerg Trauma*. 2014;2(3):125–129.
14. Matthiessen P, Hallböök O, Rutegård J, Simert G, Sjødahl R. Defunctioning stoma reduces symptomatic anastomotic leakage after low anterior resection of the rectum for cancer: a randomized multicenter trial. *Ann Surg*. 2007;246(2):207–214. doi:10.1097/SLA.0b013e3180603024
15. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol*. 2001;54(10):979–985. doi:10.1016/S0895-4356(01)00372-9
16. Katiyar A, Mohanty A, Hua J, et al. A Bayesian approach to determine the composition of heterogeneous cancer tissue. *BMC Bioinform*. 2018;19(Suppl 3):90. doi:10.1186/s12859-018-2062-0
17. Cypko MA, Stoehr M, Oeltze-Jafra S, Dietz A, Lemke HU. A guide for constructing bayesian network graphs of cancer treatment decisions. *Stud Health Technol Inform*. 2017;245:1355.
18. Wang L, Wang X, Chen A, Jin X, Che H. Prediction of type 2 diabetes risk and its effect evaluation based on the XGBoost model. *Healthcare*. 2020;8(3):1.
19. Hao PY, Chiang JH, Chen YD. Possibilistic classification by support vector networks. *Neural Netw*. 2022;149:40–56. doi:10.1016/j.neunet.2022.02.007
20. Lange NW, Salerno DM, Berger K, Cushing MM, Brown RS Jr. Management of hepatic coagulopathy in bleeding and nonbleeding patients: an evidence-based review. *J Intensive Care Med*. 2021;36(5):524–541. doi:10.1177/0885066620903027
21. Bulut Y, Sapru A, Roach GD. Hemostatic balance in pediatric acute liver failure: epidemiology of bleeding and thrombosis, physiology, and current strategies. *Front Pediatr*. 2020;8:618119. doi:10.3389/fped.2020.618119
22. Lijfering WM, Timp JF, Cannegieter SC. Predicting the risk of recurrent venous thrombosis: what the future might bring. *J Thromb Haemost*. 2019;17(9):1522–1526. doi:10.1111/jth.14534
23. Shaughnessy G, Blackburn C, Ballestin A, Akelina Y, Ascherman JA. Predicting thrombosis formation in 1-mm-diameter arterial anastomoses with transit-time ultrasound technology. *Plast Reconstr Surg*. 2017;139(6):1400–1405. doi:10.1097/PRS.0000000000003350
24. Jeong O, Park YK, Ryu SY, Kim DY, Kim HK, Jeong MR. Predisposing factors and management of postoperative bleeding after radical gastrectomy for gastric carcinoma. *Surg Today*. 2011;41(3):363–368. doi:10.1007/s00595-010-4284-2

25. Gregson J, Kaptoge S, Bolton T, et al. Cardiovascular risk factors associated with venous thromboembolism. *JAMA Cardiol.* 2019;4(2):163–173. doi:10.1001/jamacardio.2018.4537
26. Rogers MA, Levine DA, Blumberg N, Flanders SA, Chopra V, Langa KM. Triggers of hospitalization for venous thromboembolism. *Circulation.* 2012;125(17):2092–2099. doi:10.1161/CIRCULATIONAHA.111.084467
27. Lim S, Halandras PM, Bechara C, Aulivola B, Crisostomo P. Contemporary management of acute mesenteric ischemia in the endovascular era. *Vasc Endovascular Surg.* 2019;53(1):42–50. doi:10.1177/1538574418805228
28. Fernandes CJ, Morinaga LTK, Alves JJJ, et al. Cancer-associated thrombosis: the when, how and why. *Eur Respir Rev.* 2019;28(151). doi:10.1183/16000617.0119-2018
29. Jarosz-Biej M, Smolarczyk R, Cichoń T, Kułach N. Tumor microenvironment as A “game changer” in cancer radiotherapy. *Int J Mol Sci.* 2019;20(13):3212. doi:10.3390/ijms20133212
30. Pulito C, Cristaudo A, Porta C, et al. Oral mucositis: the hidden side of cancer therapy. *J Exp Clin Cancer Res.* 2020;39(1):210. doi:10.1186/s13046-020-01715-7
31. Wijerathne H, Langston JC, Yang Q, et al. Mechanisms of radiation-induced endothelium damage: emerging models and technologies. *Radiother Oncol.* 2021;158:21–32. doi:10.1016/j.radonc.2021.02.007
32. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol.* 2006;24(23):3726–3734. doi:10.1200/JCO.2005.04.7985
33. Canonico ME, Santoro C, Avvedimento M, et al. Venous thromboembolism and cancer: a comprehensive review from pathophysiology to novel treatment. *Biomolecules.* 2022;12(2):259. doi:10.3390/biom12020259
34. Weisel JW, Litvinov RI. Red blood cells: the forgotten player in hemostasis and thrombosis. *J Thromb Haemost.* 2019;17(2):271–282. doi:10.1111/jth.14360
35. Farge D, Frere C, Connors JM, et al. 2019 international clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer. *Lancet Oncol.* 2019;20(10):e566–e81. doi:10.1016/S1470-2045(19)30336-5
36. Cronin-Fenton DP, Søndergaard F, Pedersen LA, et al. Hospitalisation for venous thromboembolism in cancer patients and the general population: a population-based cohort study in Denmark, 1997–2006. *Br J Cancer.* 2010;103(7):947–953. doi:10.1038/sj.bjc.6605883
37. Wun T, White RH. Epidemiology of cancer-related venous thromboembolism. *Best Pract Res Clin Haematol.* 2009;22(1):9–23. doi:10.1016/j.beha.2008.12.001
38. Chew HK, Wun T, Harvey D, Zhou H, White RH. Incidence of venous thromboembolism and its effect on survival among patients with common cancers. *Arch Intern Med.* 2006;166(4):458–464.
39. Citla Sridhar D, Abou-Ismaïl MY, Ahuja SP. Central venous catheter-related thrombosis in children and adults. *Thromb Res.* 2020;187:103–112. doi:10.1016/j.thromres.2020.01.017
40. Lin Z, Mi B, Liu X, et al. Nomogram for predicting deep venous thrombosis in lower extremity fractures. *Biomed Res Int.* 2021;2021:9930524. doi:10.1155/2021/9930524
41. Carobbio A, Vannucchi AM, De Stefano V, et al. Neutrophil-to-lymphocyte ratio is a novel predictor of venous thrombosis in polycythemia vera. *Blood Cancer J.* 2022;12(2):28. doi:10.1038/s41408-022-00625-5
42. Gasparyan AY, Ayvazyan L, Mukanova U, Yessirkepov M, Kitas GD. The platelet-to-lymphocyte ratio as an inflammatory marker in rheumatic diseases. *Ann Lab Med.* 2019;39(4):345–357. doi:10.3343/alm.2019.39.4.345
43. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1(5):206–215. doi:10.1038/s42256-019-0048-x
44. Nohara Y, Iihara K, Nakashima N. Interpretable machine learning techniques for causal inference using balancing scores as meta-features. *Annu Int Conf IEEE Eng Med Biol Soc.* 2018;2018:4042–4045. doi:10.1109/EMBC.2018.8513026

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.