BMC Cancer

RESEARCH

Open Access

Enhancing prediction and stratifying risk: machine learning and bayesian-learning models for catheter-related thrombosis in chemotherapy patients

Tao An[1†], Han Han[2†], Junying Xie[3†], Yifan Wang[2], Yiqi Zhao[2], Hao Jia[2*] and Yanfeng Wang[4*]

Abstract

Background Catheter-related thrombosis (CRT) is a serious complication in cancer patients undergoing chemotherapy, yet existing risk prediction models demonstrate limited accuracy. This study aimed to evaluate the clinical utility of machine learning (ML) and Bayesian-learning models for CRT prediction in a large cohort of breast cancer patients undergoing catheterization.

Methods A total of 3337 breast cancer patients with central venous catheters (Cohort 1) were included to develop and test ML models. Given the suboptimal clinical feasibility of ML models, the Bayesian-learning model was constructed using odds ratio analysis and Gaussian distribution. The hazard ratio for the high-risk and low-risk groups was calculated using Cox proportional hazards regression analysis, and the model was validated in an independent cohort of 1274 patients (Cohort 2).

Results In Cohort 1, 246 patients (7.37%) developed CRT. Among the eight ML algorithms tested, WeightedEnsemble model exhibited relatively stable performance, achieving area under the receiver operating characteristic curves of 0.89 in the training set and 0.69 in the test set. WeightedEnsemble improved generalization by integrating multiple base models. The odds ratio analysis and Bayesian-learning modeling identified 4 independent risk factors: hemoglobin (threshold point [TP]: 134.63 g/L), activated partial thromboplastin time (TP: 31.71 s), total cholesterol (TP: 11.19 mmol/L), and catheterization approach (TP: peripherally inserted central catheters). A simplified risk stratification system was developed, categorizing patients into low-risk (0–1 factors) and high-risk (2–4 factors) groups. This system exhibited strong CRT risk discriminative ability, as confirmed through survival analysis ($P < 0.001$ in both cohorts). In Cohort 1, cox regression analysis showed that the high-risk group had hazard ratio (HR) of 1.60 (95% confidence interval [CI], 1.15–2.22) for both catheter indwelling time and catheter use duration. In Cohort 2, the system maintained stable discriminative ability, with an HR of 5.63 (95% CI, 3.46–9.21) for catheter indwelling time and 5.62 (95% CI, 3.46–9.12) for catheter use duration.

†Tao An, Han Han and Junying Xie contributed equally to this work.

*Correspondence:
Hao Jia
fuwai_jiahao@163.com
Yanfeng Wang
wangyf@cicams.ac.cn
Full list of author information is available at the end of the article

Conclusions While ML models demonstrated high predictive performance, their clinical applicability was limited due to complexity. The Bayesian-learning-based risk stratification model provided a simplified yet robust alternative, effectively predicting CRT risk and offering a clinically feasible tool for risk assessment in breast cancer patients with chemotherapy. Further validation in diverse cancer populations is warranted to refine its generalizability.

Keywords Breast cancer, Chemotherapy, Catheter-related thrombosis, Machine learning model, Bayesian-learning algorithm

Background

Breast cancer is the most common malignancy among women worldwide, with approximately 2.3 million new cases diagnosed in 2020, accounting for 11.7% of all cancer cases [1]. In China, the annual incidence of breast cancer reaches 357,200 cases, ranking second among female malignancies, and 60%–80% of patients receive chemotherapy during treatment [2, 3]. Catheter-related thrombosis (CRT), a life-threatening complication in cancer patients undergoing chemotherapy, occurs in 7.4%–13.9% of breast cancer patients with peripherally inserted central catheter (PICC) or central venous catheter (CVC) [4, 5]. CRT not only increases the risk of pulmonary embolism but is also associated with catheter dysfunction, chemotherapy delays, and prolonged hospitalization [6, 7]. Despite the widespread use of existing risk assessment tools, their predictive efficacy for CRT remains significantly limited.

The Khorana score, a classic predictive tool for chemotherapy-associated thrombosis, incorporates variables such as tumor type, platelet count, hemoglobin level, white blood cell count, and body mass index (BMI) [8]. The score has been widely validated in various types of cancer patients; however, it exhibited poor capability (pooled C-index < 0.7) in accurately discriminating risk for thrombosis, resulting in missed preventive anticoagulation opportunities for high-risk patients [9]. The primary reason for this limitation is its failure to account for the dynamic changes in blood parameters during chemotherapy (e.g., hemoglobin fluctuations). The COMPASS-CAT model has made partial improvements by integrating CVC, time since cancer diagnosis, cardiovascular risk factors, tumor staging, chemotherapy regimens, and a history of prior thrombosis, offering enhanced capabilities for dynamic marker application [10]. Its external validation shows an area under the curve (AUC) of 0.62 [11], which is slightly better than the Khorana score (AUC = 0.56) [12]. The shared limitations of these two models highlight the deficiencies in current CRT prediction tools: manual feature selection relied on logistic regression did not capture critical clinical parameters (e.g., tumor molecular markers).

Machine learning (ML) offers a promising approach for CRT risk prediction, yet its clinical application remains constrained by three key limitations: susceptibility to overfitting in imbalanced datasets (thromboembolism incidence < 10%), poor interpretability of "black-box" models, and reliance on manual feature selection for identifying critical clinical variables [13, 14]. To address these challenges, we developed a two-phase predictive framework. In Phase I, AutoGluon was employed to systematically screen 26 clinical, laboratory, and molecular variables, not only constructing a robust predictive model but also identifying novel risk factors, including human epidermal growth factor receptor 2 (HER2), estrogen receptor (ER), progesterone receptor (PR), and Ki-67 positive. In Phase II, we established a binary risk stratification system based on four independent predictors derived from Cohort 1 and validated in Cohort 2. The system demonstrated consistent discriminative ability across both catheter indwelling time and catheter use duration settings ($P < 0.001$), enabling effective identification of high-risk patients. This study provides a clinically feasible tool for individualized CRT risk assessment, offering new evidence to guide thromboprophylaxis strategies in breast cancer patients.

Methods

Patients and treatment

This retrospective study included breast cancer patients treating with or without chemotherapy at the National Canter-National Clinical Research Center for the Cancer-Cancer Hospital, Chinese Academy of Medical Sciences from August 1, 2012 to March 31, 2021. A total of 3337 patients (Cohort 1) were eligible according to the following criteria: (1) age ≥ 18 years, (2) pathological diagnosis of breast cancer, (3) accepted CVC or PICC in the hospital and treated with systemic therapy, and (4) underwent vascular Doppler ultrasound examination during catheter placement. Patients who were treated with anticoagulant therapy during CVCs or PICCs placement, failure to acquire complete basic information, and pregnant or lactating were excluded.

The venous access devices were placed by the modified Seldinger technique with ultrasound guidance. The direction of catheter and position of catheter tip were confirmed by anterior–posterior chest X-rays. All patients were provided with routine catheter therapy once or

twice each week by a professional team. The main outcome was the onset of CRT which referred to thrombotic events occurring in the vein draining the catheter. CRT was diagnosed by vascular Doppler ultrasound and color imaging (GE LOGIQTM E9; Philips), which showed a low-echo area in the lumen of vasculature, presenting as a mass, and the lumen still appear after local pressure application without blood flow signal [15, 16]. The complete baseline characteristics are provided in Table 1.

This study was approved by the National Canter/National Clinical Research Center for the Cancer-Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College (22/444–3646). The institutional review boards waived need for informed consent because the patient data were identified in the dataset.

Synthetic minority over-sampling technique

We employed the synthetic minority over-sampling technique (SMOTE) in the training dataset of Cohort 1. SMOTE is a widely used oversampling method that balances data by increasing the number of minority-class samples without modifying the majority class [17]. Specifically, SMOTE creates synthetic samples through linear interpolation based on differences between each minority-class instance and its nearest neighbors, thereby enhancing the model's ability to recognize minority-class patterns. This approach has been extensively adopted in medical research and proven to be an effective resampling strategy [5, 18].

In this study, SMOTE was applied exclusively to the training dataset to balance the minority class (thrombosis group). Meanwhile, the validation dataset maintained its original distribution to preserve the natural outcome frequency, ensuring that the assessment of the model's performance remained objective and clinically relevant.

ML algorithms

Cohort 1 was split 70/30 into training and testing groups, respectively, using standard stratified splitting method provided by the Caret package in R.2. A fixed random seed (88) was used to ensure reproducibility of the split. AutoGluon is an open-source automated machine learning framework designed to streamline model training, hyperparameter tuning, and ensembling [19]. By stacking multiple machine learning algorithms into a single ensemble classifier, it leverages diverse model architectures to improve predictive performance. AutoGluon also incorporates sophisticated techniques—such as regularization on individual models within the stacked ensemble and automated hyperparameter search—to minimize overfitting and reduce the burden of manual tuning. Through this combination of methods, AutoGluon consistently demonstrates strong predictive accuracy across various datasets with minimal user intervention. AutoGluon was run with the following parameter settings: time_limit = 720, num_bag_folds = 5, num_bag_sets = 5, num_stack_levels = 30, the use_bag_holdout option enabled, and verbosity = 2.

We selected eight ML methods within the framework of AutoGluon—random forest entropy (RandomForestEntr), random forest gini (RandomForestGini), categorical boosting (CatBoost), extra trees entropy (ExtraTreesEntr), neural net fast ai (NeuralNetFastAI), extreme gradient boosting (XGBoost), linear model, and weighted ensemble learning (WeightedEnsemble)—because they represent a broad spectrum of well-established modeling paradigms. This diversity spans bagging-based ensemble trees, gradient boosting, deep learning, linear modeling, and a second-level weighted ensemble, allowing the final classifier to leverage each algorithm's strengths while minimizing overfitting through stacking and regularization. Moreover, all eight methods are seamlessly integrated within AutoGluon, facilitating automated hyperparameter tuning and model selection with minimal manual intervention, which is essential for ensuring both high accuracy and reproducibility.

Random forest algorithm constructs numerous decision trees and amalgamate their predictions for a consolidated result. It employs entropy or Gini importance to optimize tree splits, aiming to maximize information gain—the disparity between the parent node's entropy or Gini impurity and the weighted mean of the child nodes' impurities [20, 21]. CatBoost, extra trees, and XGBoost all use multiple decision trees to perform classification or regression tasks and each tree is trained on a random subset of features, and the split points at each node are randomly selected [22]. Both XGBoost and CatBoost are gradient boosting algorithms but differ in how they handle categorical variables, gradient updates, and overfitting control. CatBoost preserves data order and automatically processes categorical features with ordered boosting, reducing target leakage [23]. XGBoost, by contrast, generally requires numerical or one-hot encoding [22]. Statistically, CatBoost's emphasis on data-order protection and category-optimized strategies enables more effective overfitting control in certain datasets, whereas XGBoost's streamlined structure can excel in speed-oriented or predominantly numerical feature settings. ExtraTreesEntr is an extreme version of random tree algorithm, constructing multiple decision trees on randomly chosen feature subsets and employing entropy to ascertain information gain. The linear model posits a direct correlation between independent and dependent variables [24]. The WeightedEnsemble operates as a second-level ensemble model, aggregating the predictions

Table 1 Baseline characteristics of Cohort 1

	Without CRT (*n* = 3091)	CRT (*n* = 246)	Shapiro–Wilk test (*P* value)	*p* Value
Age (years)	49.00 (42.00–56.00)	50.50 (44.00–58.00)	< 0.001/0.187	0.080[a]
Height (m)	1.59 (1.55–1.63)	1.59 (1.56–1.63)	< 0.001/0.167	0.174[a]
Weight (kg)	62.50 (57.00–69.00)	63.00 (55.00–69.00)	< 0.001/0.043	0.552[a]
Body mass index	24.75 (22.55–27.29)	24.66 (22.19–26.94)	< 0.001/0.837	0.247[a]
Karnofsky Performance Status	90.00 (90.00–90.00)	90.00 (90.00–90.00)	< 0.001/ < 0.001	0.855[a]
Catheter length (cm)	16.00 (16.00–38.00)	17.00 (16.00–39.00)	< 0.001/ < 0.001	**0.013**[a]
Time from diagnosis to catheterization (hours)	59.00 (42.00–241.00)	58.00 (43.00–257.00)	< 0.001/ < 0.001	0.367[a]
Indwelling time (days)	99.00 (63.00–114.00)	91.00 (61.00–109.00)	< 0.001/ < 0.001	0.251[a]
Duration of catheter use (days)	24.00 (16.00–33.00)	25.00 (16.00–32.00)	< 0.001/ < 0.001	0.689[a]
Leukocyte (10^9/L)	5.40 (4.43–6.73)	5.66 (4.42–7.02)	< 0.001/ < 0.001	0.279[a]
Neutrophil (10^9/L)	3.17 (2.44–4.23)	3.37 (2.43–4.37)	< 0.001/ < 0.001	0.310[a]
Lymphocyte (10^9/L)	1.63 (1.29–2.02)	1.65 (1.31–2.02)	< 0.001/0.056	0.571[a]
Neutrophil-to-lymphocyte ratio	1.96 (1.45–2.71)	1.98 (1.42–2.73)	< 0.001/ < 0.001	0.740[a]
Hemoglobin (g/L)	120.00 (110.00–129.00)	123.00 (115.00–131.00)	< 0.001/ < 0.001	**< 0.001**[a]
Platelet (10^9/L)	244.00 (199.00–297.00)	259.00 (204.00–310.00)	< 0.001/0.097	0.978[a]
Neutrophil-to-platelet ratio	0.01 (0.01–0.02)	0.01 (0.01–0.02)	< 0.001/ < 0.001	0.091[a]
Prothrombin time (seconds)	11.40 (10.90–11.90)	11.50 (10.90–11.90)	< 0.001/ < 0.001	0.190[a]
Fibrinogen (g/L)	2.88 (2.46–3.40)	2.98 (2.52–3.47)	< 0.001/ < 0.001	0.102[a]
Activated partial thromboplastin time (seconds)	25.70 (23.50–28.10)	25.40 (23.30–27.30)	< 0.001/0.003	**0.026**[a]
Albumin (g/L)	41.60 (39.60–43.70)	42.05 (39.80–43.80)	< 0.001/0.113	0.257[a]
Total cholesterol (mmol/L)	4.77 (4.17–5.42)	4.87 (4.21–5.75)	< 0.001/ < 0.001	**0.031**[a]
Creatinine (mmol/L)	53.00 (48.00–59.60)	54.00 (48.00–60.00)	< 0.001/ < 0.001	0.177[a]
D-Dimer (mg/L)	0.48 (0.24–0.89)	0.50 (0.28–0.96)	< 0.001/ < 0.001	0.079[a]
Catheterization approach				**0.007**[b]
CVC (%)	2057 (66.5)	143 (58.1)	/	
PICC (%)	1034 (33.5)	103 (41.9)	/	
Vein entry				**0.013**[d]
Left subclavian vein (%)	522 (16.9)	36 (14.6)	/	
Right subclavian vein (%)	1535 (49.7)	107 (43.5)	/	
Left basilic vein (%)	524 (17.0)	50 (20.3)	/	
Right basilic vein (%)	510 (16.5)	52 (21.1)	/	
Left median cubital vein (%)	0 (0.0)	1 (0.4)	/	
Catheter tip position				**0.013**[b]
Normal: T5-T8 (%)	3028 (98.0)	235 (95.5)	/	
Abnormal (%)	63 (2.0)	11 (4.5)	/	
Stage				0.260[b]
1 (%)	386 (12.5)	24 (9.8)	/	
2 (%)	1044 (33.8)	87 (35.4)	/	
3 (%)	920 (29.8)	84 (34.1)	/	
4 (%)	741 (24.0)	51 (20.7)	/	
Smoking (%)	95 (3.1)	5 (2.0)	/	0.357[b]
Alcohol use (%)	61 (2.0)	3 (1.2)	/	0.556[c]
Hypertension (%)	807 (26.1)	68 (27.6)	/	0.599[b]
Coronary artery disease (%)	36 (1.2)	5 (2.0)	/	0.374[c]
Atrial fibrillation (%)	22 (0.7)	3 (1.2)	/	0.614[c]
Diabetes mellitus (%)	167 (5.4)	19 (7.7)	/	0.127[b]
Hyperlipidemia (%)	60 (1.9)	4 (1.6)	/	0.916[c]
Stroke (%)	20 (0.6)	0 (0.0)	/	0.403[c]
Previous catheterization (%)	570 (18.4)	41 (16.7)	/	0.489[b]

Table 1 (continued)

	Without CRT ($n=3091$)	CRT ($n=246$)	Shapiro–Wilk test (P value)	p Value
History of venous thrombosis (%)	18 (0.6)	3 (1.2)	/	0.425[c]
Catheter occlusion (%)	16 (0.5)	2 (0.8)	/	0.876[c]
Chemotherapy (%)	2666 (86.3)	205 (83.3)	/	0.204[b]
Mediastinal radiotherapy (%)	4 (0.1)	1 (0.4)	/	0.318[d]
Antiangiogenic therapy (%)	36 (1.2)	1 (0.4)	/	0.437[c]
Immunotherapy (%)	4 (0.1)	1 (0.4)	/	0.318[d]
Antimicrobial therapy (%)	2 (0.1)	0 (0.0)	/	>0.99[d]
Nutrition (%)	24 (0.8)	1 (0.4)	/	0.792[c]
Targeted therapy (%)	74 (2.4)	4 (1.6)	/	0.443[b]
Lung radiotherapy (%)	1 (0.0)	0 (0.0)	/	>0.99[d]
Esophageal radiotherapy (%)	397 (12.8)	38 (15.4)	/	0.243[b]
Thoracic radiotherapy (%)	402 (13.0)	39 (15.9)	/	0.204[b]
ER positive (%)	2107 (68.2)	179 (72.8)	/	0.135[b]
PR positive (%)	2080 (67.3)	169 (68.7)	/	0.650[b]
HER2 positive (%)	848 (27.4)	55 (22.4)	/	0.085[b]
Ki-67 positive (%)	1461 (47.3)	125 (50.8)	/	0.284[b]
Luminal typing				0.464[b]
A (%)	1026 (33.2)	89 (36.2)	/	
B (%)	1258 (40.7)	101 (41.1)	/	
HER2 (%)	316 (10.2)	18 (7.3)	/	
TNBC (%)	491 (15.9)	38 (15.4)	/	

Depending on the normality, characteristics for the Without CRT and CRT groups were presented by Median (IQR)

P-value: [a]Mann-Whitney test; [b]Chi-square test; [c]Chi-square incorporating Yates' correction for continuity; [d]Fisher's Exact Test

from various first-level models—including tree-based algorithms (gradient boosting machine, XGBoost, CatBoost, random forest, extra treess), NeuralNetFastAI, and k-Nearest Neighbor—by assigning weights based on each model's performance. This weighted synthesis produces a final output that enhances overall predictive accuracy [25]. Linear model predicts outcomes by assuming a direct proportional relationship between the input variables and the target variable [26]. NeuralNetFastAI is a deep learning model architecture within the AutoGluon framework, built on the FastAI library—an API layer on top of PyTorch. By automating tasks such as data preprocessing, hyperparameter tuning, and adaptive learning rate scheduling, it offers a streamlined and efficient approach to deep learning model training.

Importance score

We used the WeightedEnsemble feature importance scores from AutoGluon, which are computed by combining 7 base models' importance scores, weighted according to that model's performance. A positive feature importance score indicates that removing the feature decreases the ensemble's performance, whereas a negative score suggests performance improvement if the feature is removed. Accordingly, variables with a positive score and p-value ≤ 0.05 were identified as candidate predictors of CRT.

Independent predictors, Bayesian-learning model, and threshold inflection point

We performed odds ratio (OR) analysis to select candidate features from baseline characteristics with a notable difference between CRT group and Without CRT group for predicting CRT. Only when the significance of both univariate-unadjusted and multivariate-adjusted OR analyses were less than 0.05, a feature could be defined as an independent predictor of CRT. These independent predictors were then used to construct a Bayesian-learning model and calculate threshold inflection point for CRT. The statistical analysis was performed by SPSS, version 26.00 (IBM Inc).

Bayesian learning is a probabilistic approach that updates prior knowledge with observed data using Bayes' theorem for more refined predictions. It typically involves specifying a prior distribution, defining a likelihood function, and computing a posterior distribution [27]. For continuous variables (hemoglobin, activated partial thromboplastin time [APTT], and total cholesterol [TC]) as well as categorical variable (catheterization approach), we used a Gaussian distribution as the

likelihood function to establish the correlation with probability of CRT events. In detail, for the derivation process, we assumed that the value of a variable was X, and the patient belongs to CRT was defined as event A1, given the variable, the probability of event A1 is $P(A_1|x) = P(A_1)P(x|A_1)/P(x)$. Similarly, if the patient belongs to a Without CRT was defined as event A0, given the variable, the probability of event A0 is $P(A_0|x) = P(A_0)P(x|A_0)/P(x)$. Given the variable, the probability of the patient belonging to either CRT or Without CRT is 1, which is $P(A_0|x) + P(A_1|x) = 1$, and we could assume that $P(A_1|x)/P(A_0|x) = \alpha$, eventually we can get the equation that $\alpha = \frac{P(A_1)P(x|A_1)}{P(A_o)P(x|A_o)}$. As X belongs to different Gaussian distributions in event A0 and A1, we could get $\alpha = \frac{P(A_1)}{P(A_0)} \cdot \frac{\sigma_0}{\sigma_1} \cdot exp[\frac{(x-\mu_0)^2}{2\sigma_0^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}]$, where μ_0 and μ_1 are the mean of the two Gaussian distributions, respectively, and σ_0^2 and σ_1^2 are the variance of the two Gaussian distributions, respectively. Lastly, we can obtain the probability of CRT event A1 as $P(A_1|x) = \frac{\alpha}{1+\alpha}$.

Furthermore, we obtained the inflection points of 4 variables for comparing patients' laboratory and clinical results with risk thresholds and fulfilling risk-dependent classification of chemotherapy patients. Specifically, the inflection point of X is determined from $P(A_1|x)$ using 2-order derivative approach, where the condition for the inflection point is given by $P(A_1|x)^{''} = 0$. The 2-order derivative of $y[i]$ is computed as $y^{''}[i] = \frac{2y[i]-y[i-1]-y[i+1]}{(\Delta x)^2}$ where $\Delta x = x[i+1] - x[i] = x[i] - x[i-1]$. This statistical analysis part was conducted by MATLAB software, version R2020b (Mathworks Corp).

Risk-dependent survival and model validation

We determined whether the survival varied among risk groups. Based on independent risk factors originated from OR analysis and inflection points derived from Bayesian-learning model, we categorized patients into 2 groups: low-risk patients with 0–1 risk factor and high-risk patients with 2–4 risk factors. We constructed 2 parallel assessment, overall survival (catheter indwelling time) was defined as time from the date of catheterization to CRT onset from any cause, and overall survival (duration of catheter use) was defined as cumulative time of catheter use from catheterization until the occurrence of CRT. Survival rates were estimated using the Kaplan–Meier method and compared using the log-rank test. $P<0.05$ (2-sided) was considered to be statistically significant.

To quantify the CRT risk in the high-risk group relative to the low-risk group, we performed Cox proportional hazards regression analysis, calculating the hazard ratio (HR) and corresponding 95% confidence interval (95% CI). HR values were computed separately for catheter indwelling time and duration of catheter use to compare survival risk differences between the two groups.

Following model development in Cohort 1, we applied the same methodology to evaluate model performance in an independent validation cohort (Cohort 2). This cohort included 1,274 breast cancer patients enrolled between January 1, 2022, and February 29, 2024, following the same inclusion and exclusion criteria as Cohort 1. The baseline characteristics of Cohort 2 are provided in Table S1. All statistical analyses were performed with MATLAB software, version R2020b (Mathworks Corp).

Results

Baseline characteristics

A total of 3337 female patients with breast cancer were included in the study, and 246 (7.37%) experienced a CRT event (Fig. 1). The baseline characteristics of the patients are summarized in Table 1. The median (interquartile range [IQR]) age of CRT group was 50.50 (44.80–58.00) years compared to 49.00 (42.00–58.00) in Without CRT group, and the p value was 0.080, suggesting a tendency of CRT in elderly patients. The patients had different stages of cancer, with stage 2 being the most prevalent (33.89%). Patients with longer catheter length (median [IQR] cm, 17.00 [16.00–39.00] vs 16.00 [16.00–38.00], $P=0.013$), higher hemoglobin level (median [IQR] g/L, 123.00 [115.00–131.00] vs 120.00 [110.00–129.00], $P<0.001$), shorter APTT (median [IQR] seconds 25.40 [23.30–27.30] vs 25.70 [23.50–28.10], $P=0.026$), and more TC (median [IQR] mmol/L, 4.87 [4.21–5.75] vs 4.77 [4.17,5.42], $P=0.031$) were more likely to experience CRT.

Risk of CRT

Figure 2A described the ROC curves of 8 machine leaning models for predicting CRT risk in breast cancer patients received chemotherapy. Except NeuralNetFastAI model (AUC, 0.83) and LinearModel model (AUC, 0.83), other 6 ML models exhibited superior performance (RandomForestEntr: AUC, 0.86; RandomForestGini: AUC, 0.85; ExtraTreesEntr: AUC, 0.88; WeightedEnsemble: AUC, 0.89; CatBoost: AUC, 0.86) of predicting CRT risk in training group. However, only WeightedEnsemble model maintained consistently good performance in testing group (Fig. 2B). Specifically, the AUC of WeightedEnsemble model was 0.69.

Table 2 delineates the area under the receiver operating characteristic curve (ROC-AUC), precision recall (PR)-AUC, sensitivity, specificity, accuracy, and precision of ML models within both the training and testing datasets. Notably, the WeightedEnsemble model demonstrated comparable efficacy across all parameters,

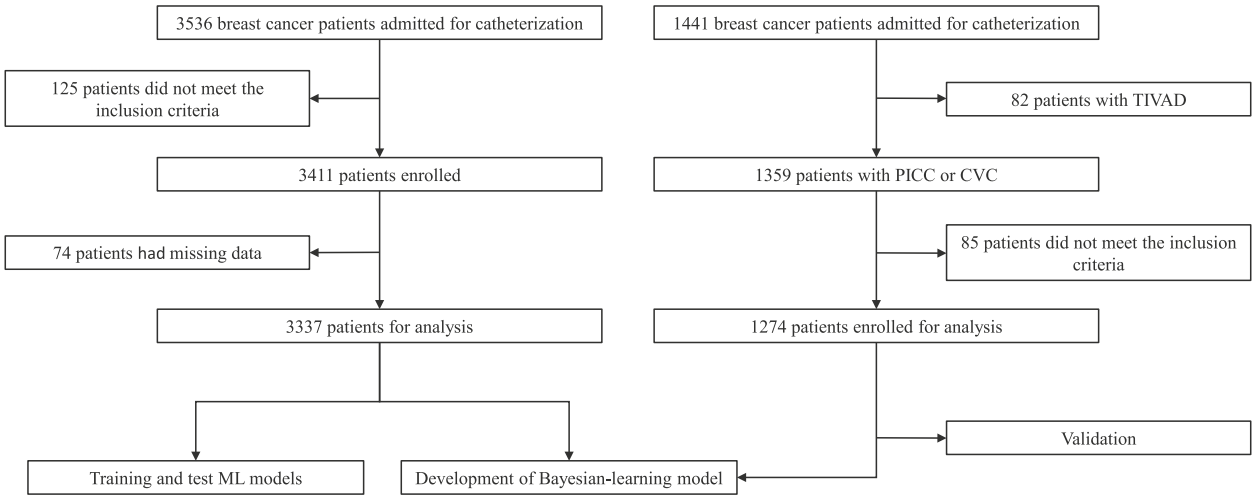


Fig. 1 The patient flowchart. CVC indicates central venous catheter; PICC, peripherally inserted central catheter; ML, machine learning; TIVAD, totally implantable venous access device

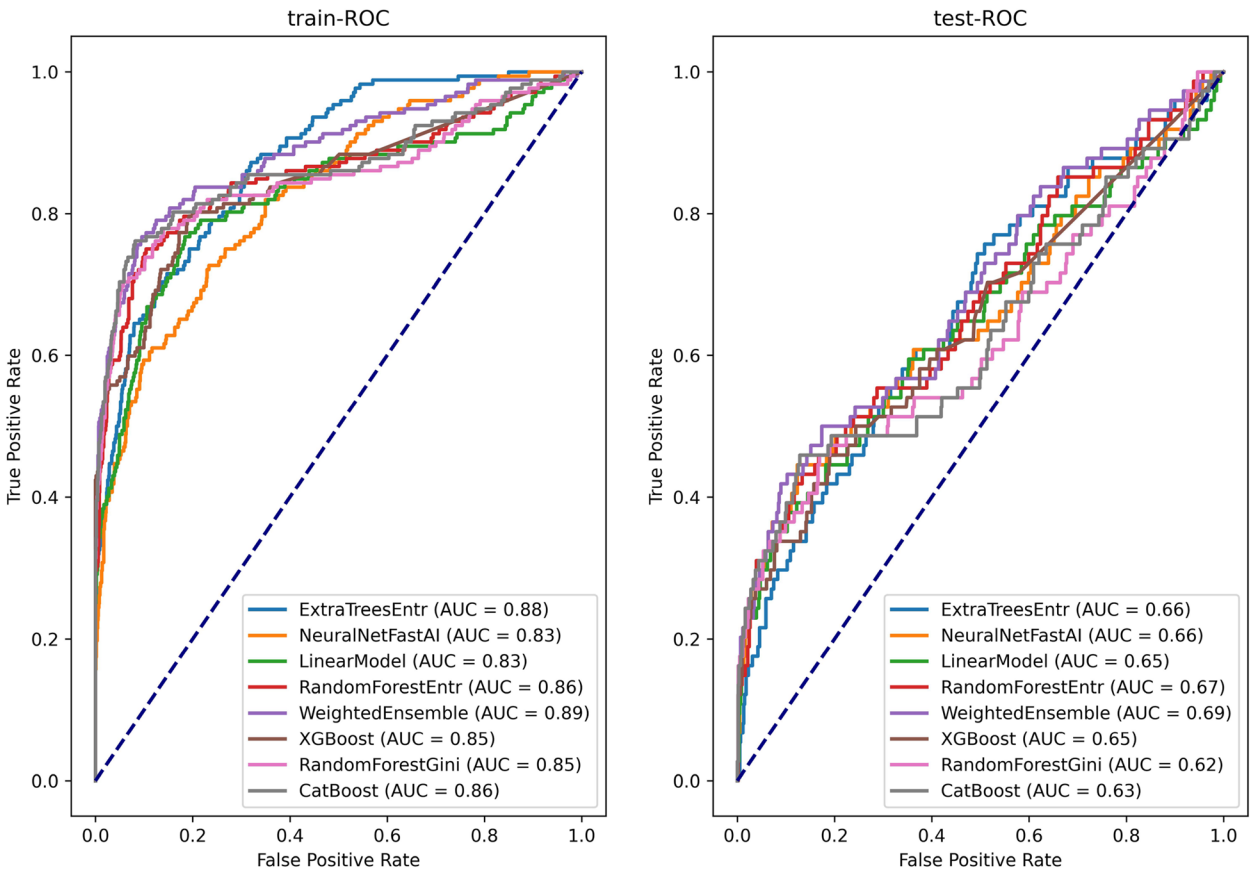


Fig. 2 Performance for Predicting Catheter Related Thrombosis in the Training and Testing Group. AUC indicates area under the receiver operating characteristic curve; RandomForestEntr, random forest entropy; RandomForestGini, random forest gini; CatBoost, categorical boosting; ExtraTreesEntr, extra trees entropy; NeuralNetFastAI, neural net fast ai; XGBoost, extreme gradient boosting; LinearModel, linear model; WeightedEnsemble, weighted ensemble learning

Table 2 Machine learning model evaluation

Model	ROC-AUC	PR-AUC	Sensitivity	Specificity	Accuracy	Precision
RandomForestEntr						
Training group	0.8563	0.6728	1	0.2733	0.9223	1
Testing group	0.6679	0.3196	0.9984	0.1081	0.9029	0.8889
RandomForestGini						
Training group	0.8512	0.6967	0.9993	0.3256	0.9272	0.9825
Testing group	0.6214	0.3394	0.9968	0.1622	0.9072	0.8571
ExtraTreesEntr						
Training group	0.8764	0.6131	1	0	0.893	0
Testing group	0.6631	0.2478	1	0	0.8928	0
WeightedEnsemble						
Training group	0.8882	0.7275	1	0.3081	0.926	1
Testing group	0.6879	0.3627	0.9984	0.1351	0.9058	0.9091
CatBoost						
Training group	0.8645	0.7164	0.9993	0.3372	0.9285	0.9831
Testing group	0.6308	0.3488	0.9968	0.1622	0.9072	0.8571
XGBoost						
Training group	0.8452	0.6642	1	0.2907	0.9241	1
Testing group	0.6467	0.3094	0.9984	0.1216	0.9043	0.9
NeuralNetFastAI						
Training group	0.8276	0.5366	1	0.1512	0.9092	1
Testing group	0.6562	0.3188	0.9984	0.027	0.8942	0.6667
LinearModel						
Training group	0.8266	0.5906	1	0.1802	0.9123	1
Testing group	0.6495	0.3039	0.9968	0.0541	0.8957	0.6667

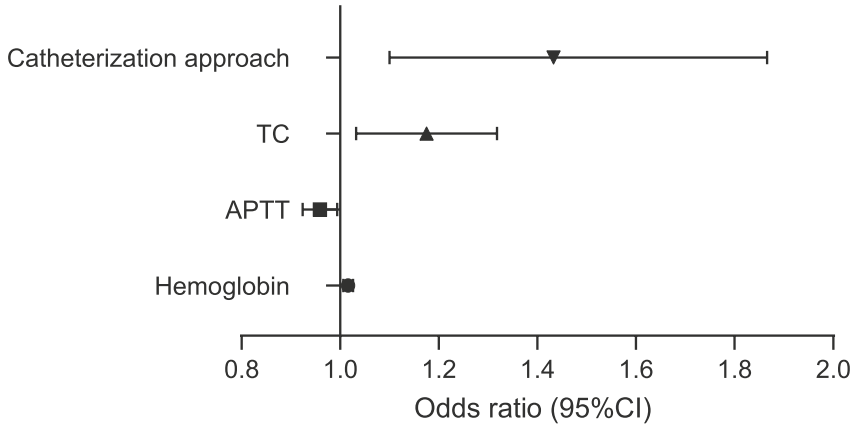
Abbreviations: ROC-AUC area under the receiver operating characteristic curve, *PR-AUC* area under the precision recall curve, *RandomForestEntr* random forest entropy, *RandomForestGini* random forest gini, *CatBoost* categorical boosting, *ExtraTreesEntr* extra trees entropy, *NeuralNetFastAI* neural net fast ai, *XGBoost* extreme gradient boosting, *LinearModel* linear model, *WeightedEnsemble* weighted ensemble learning

including cumulative gain, sensitivity, positive predictive value, in the testing cohort (Figure S1A-C). Further analysis of the calibration curves revealed that the model was well calibrated in the lower range of predicted probabilities, with predicted values closely aligned with actual observed frequencies. However, a slight overestimation of the actual incidence rates was observed at higher predicted probability ranges (Figure S1D). In contrast, other models (e.g., CatBoost, LinearModel) displayed less consistent calibration performance, with systematic overconfidence or underconfidence across different probability thresholds.

We listed the importance scores of features which positively affected ML model construction to assess impacts of different variables on prediction of CRT (Table S2). The variables with the highest importance scores were platelet count (0.159), APTT (0.144), age (0.129), TC (0.120), neutrophil-to-lymphocyte ratio (0.083), ER positive (0.032), catheterization approach (0.025), Ki-67 positive (0.014), and stage (0.010) in training group. Results were similar for the model in testing group (Table S3).

Independent predictors and development of Bayesian-learning model

Hemoglobin, APTT, TC, and catheterization approach were all statistically significant in OR analysis. The significance results and OR values were displayed in Fig. 3. We constructed predictive functions of 4 independent risk factors by integrating Batesian-learning model and Gaussian distribution, with the values of the Gaussian distribution parameters (μ_0, μ_1, σ_0^2, σ_1^2) provided in Table S4. These functions are as follows: For hemoglobin, the probability of CRT was $P(A_1|x) = \frac{exp[-0.00007x^2+0.03289x-2.91054]}{12.38573+\cdot exp[-0.00007x^2+0.03289x-2.91054]}$; for APTT, the probability of CRT was $P(A_1|x) = \frac{exp[-0.00522x^2+0.22552x-2.34363]}{11.76123+\cdot exp[-0.00522x^2+0.22552x-2.34363]}$; for TC, the probability of CRT was $P(A_1|x) = \frac{exp[0.04786x^2-0.31433x+0.38695]}{13.23577+\cdot exp[0.04786x^2-0.31433x+0.38695]}$ (Fig. 4A-C). Because catheterization approach was discrete variable, the probability of CRT was $P(A_1|x = \text{PICC}) = 0.09056$ for PICC and the probability of CRT was $P(A_1|x = \text{CVC}) = 0.06498$ for CVC.



	OR (95%CI)	*P* Value
Catheterization approach	1.43 (1.10-1.87)	.008
TC (mmol/L)	1.17 (1.04-1.32)	.012
APTT (s)	0.96 (0.92-0.99)	.023
Hemoglobin (mg/L)	1.02 (1.01-1.03)	<.001

Fig. 3 The Odds Ratio of Independent Risk Factors. APTT indicates activated partial thromboplastin time; TC, total cholesterol

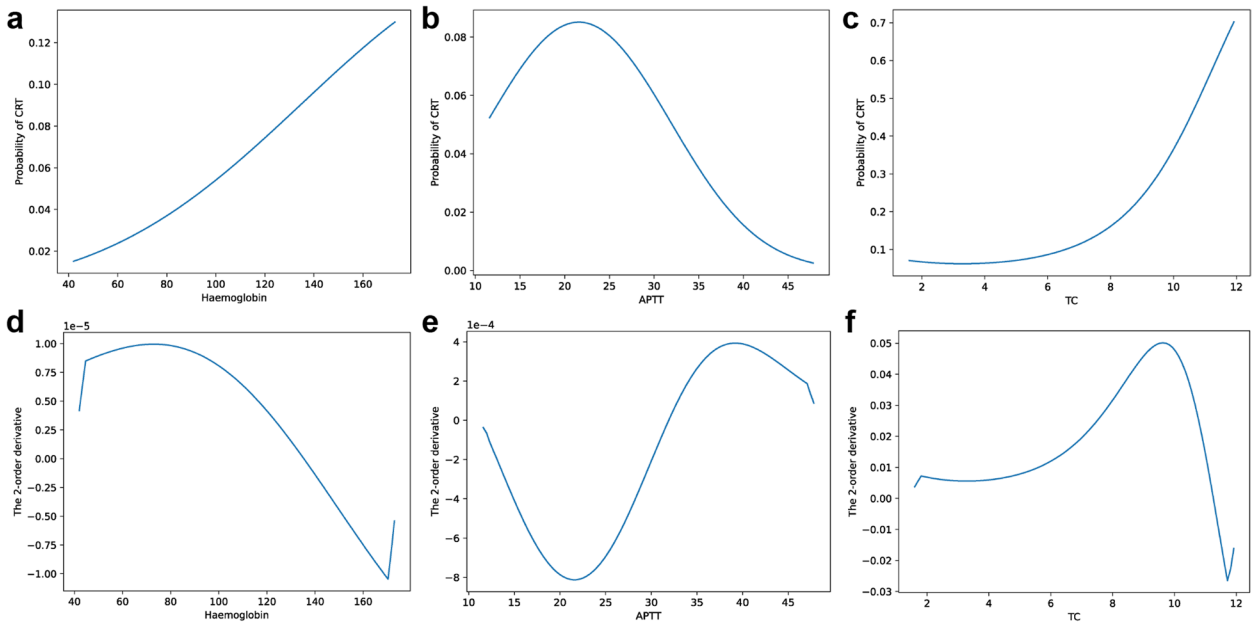


Fig. 4 The Threshold Inflection Point of Catheter-related Thrombosis. **a** Functional relationship between hemoglobin and probability of catheter-related thrombosis; **b** Functional relationship between activated partial thromboplastin time and probability of catheter-related thrombosis; **c** Functional relationship between total cholesterol and probability of catheter-related thrombosis; **d** 2-order derivative of the function in A; e. 2-order derivative of the function in B; f. 2-order derivative of the function in C

Utilizing 2-order derivative, an inflection point for the hemoglobin value was identified 134.63. This indicated that a hemoglobin below 134.63 acts as a protective factor, correlating with a lower probability of CRT event, whereas a hemoglobin above 134.63 sees a rapid incline in CRT probability. Concerning relative risk factors, APTT less than 31.71, TC above 11.19, or catheterization employed PICC leads to a swift increase in CRT incidence (Fig. 4D-F). Conversely, APTT above

31.71, TC below 11.19, or catheterization employed CVC is associated with a reduced CRT risk.

Evaluation of Bayesian-learning model

We divided the population into 2 risk categories based on above factors: low-risk (0–1 factor) and high-risk (2–4 factors) (Table S5). The P values of survival curve established by catheter indwelling days and duration of catheter use were both less than 0.001, indicating the good discriminative capacity of CRT (Fig. 5A and B). Cox regression analysis demonstrated that in Cohort 1, the high-risk group had a significantly higher CRT risk, with a hazard ratio (HR) of 1.60 (95% confidence interval [CI], 1.15–2.22) for both catheter indwelling time and catheter use duration.

The risk prediction model underwent validation in an independent cohort of 1274 patients, with 66 (5.18%) developing CRT (Fig. 1 and Table S1). Similarly, patients with 0–1 risk factor and 2–4 risk factors were categorized as low-risk group and high-risk group, respectively. The model's discriminative capacity remained significant, as indicated by P values less than 0.001 (Fig. 5C and D). In Cohort 2, the system maintained stable discriminative ability, with an HR of 5.63 (95% CI, 3.46–9.21) for catheter indwelling time and 5.62 (95% CI, 3.46–9.12) for catheter use duration.

Discussion

This study presents a ML-driven and Bayesian learning-based risk stratification framework for CRT prediction in breast cancer patients undergoing chemotherapy. By integrating advanced ML feature selection, OR analysis, and Bayesian modeling, we established a binary classification

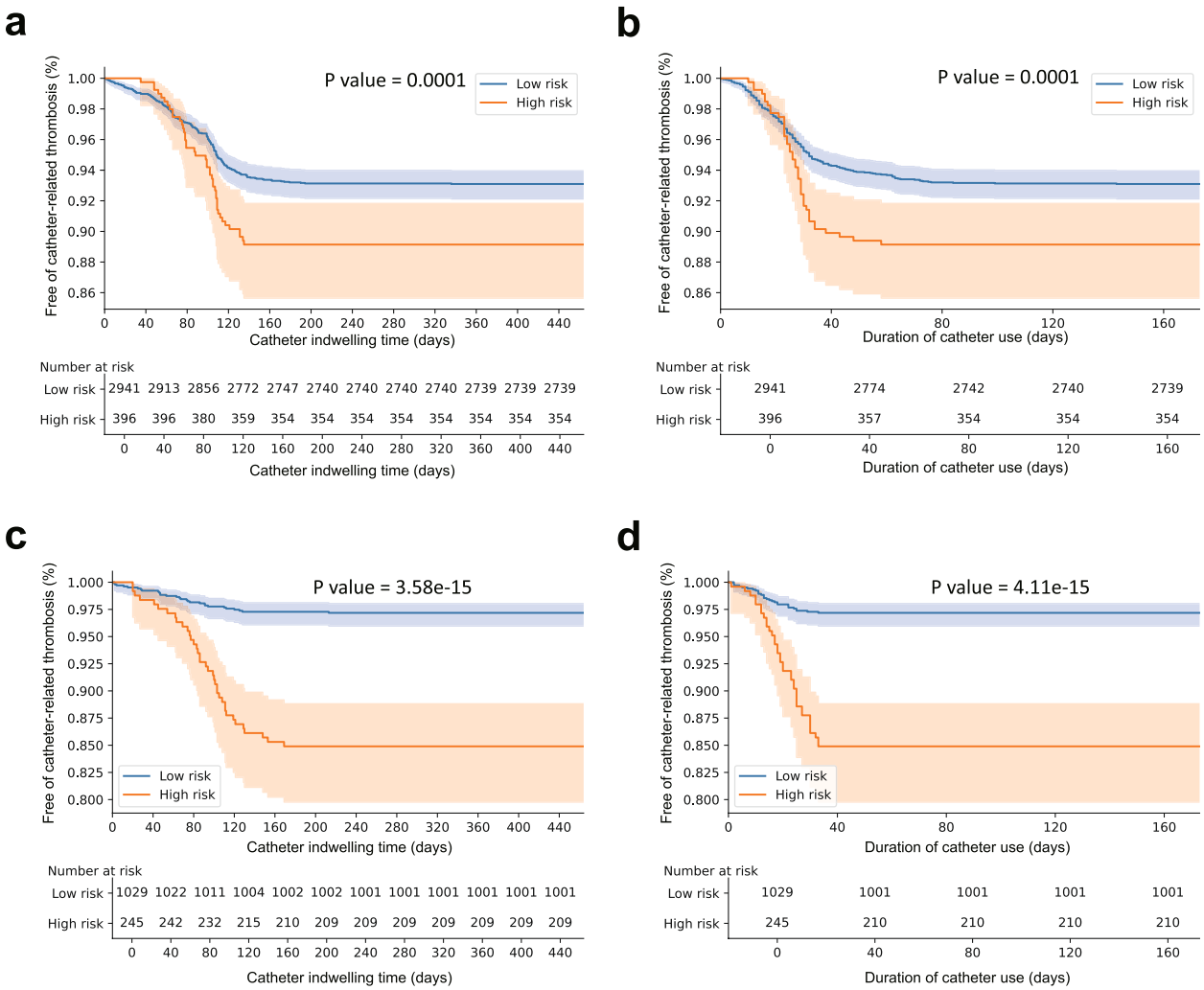


Fig. 5 Time to Catheter-related Thrombosis (CRT). **a** Time to CRT occurrence for patients in Cohort 1 calculated by catheter indwelling time; **b** Time to CRT occurrence for patients in Cohort 1 calculated by duration of catheter use; **c** Time to CRT occurrence for patients in Cohort 2 calculated by catheter indwelling time; **d** Time to CRT occurrence for patients in Cohort 2 calculated by duration of catheter use

predictive system that was based on hemoglobin, APTT, TC, and catheterization approach. Our findings confirm the relevance of established CRT risk factors while identifying novel predictors, particularly molecular features of tumor, that may refine risk stratification beyond traditional models.

The CRT incidence in Cohort 1 and Cohort 2 were 7.37% and 5.18%, respectively, consistent with previously reported rates in breast cancer populations (4.09%–13.9%), suggesting that despite the broad timeframe of this study, the baseline characteristics of patients remained relatively stable [5, 28, 29]. This consistency reinforces the external validity of our model. Notably, catheter management strategies for cancer patients remained largely unchanged throughout the study period [15, 30, 31]. All patients underwent consistent core management protocols, including ultrasound-guided catheter placement, standardized catheter care, routine prophylactic flushing, and infection surveillance. In addition, stringent inclusion and exclusion criteria were applied to maintain cohort homogeneity, and a data-driven approach was used to select the optimal model, minimizing potential biases arising from cohort heterogeneity and improving the robustness of the predictive performance.

Using the AutoGluon framework, eight ML algorithms were evaluated, with WeightedEnsemble demonstrating the most stable predictive performance in both the training (AUC = 0.89) and testing sets (AUC = 0.69). WeightedEnsemble leveraged stacked generalization to integrate multiple base models, thereby reducing variance and improving generalizability. Unlike previous studies that predominantly relied on logistic regression for feature selection, this study employed automated feature selection, reducing subjectivity and manual bias [32, 33]. Traditional CRT risk assessment models, such as the Khorana score and COMPASS-CAT, rely primarily on traditional clinical and laboratory factors, failing to capture tumor staging and molecular features [8, 10]. In contrast, the ML framework enabled the identification of tumor-related predictors, such as HER2, ER, PR, and Ki-67 positive, highlighting the potential contribution of tumor biology to CRT risk.

The AutoGluon framework identified traditional CRT or cancer associated thromboembolism risk factors, including platelet count, leukocyte count, BMI, age, hemoglobin, and PICC, all of which have been extensively reported in previous studies [34–36]. Beyond validating established CRT risk factors, this study also identified molecular features (HER2, Ki-67, PR, and ER) as novel predictors. The increased CRT risk in HER2-, PR-, or ER-positive patients may be attributed to the endothelial toxicity associated with targeted therapies. For instance, anti-HER2 therapies (such as trastuzumab and pertuzumab) have been linked to cardiovascular toxicities, including endothelial dysfunction, which can promote thrombosis [37]. Similarly, endocrine therapies (such as tamoxifen and aromatase inhibitors) used in PR-/ER-positive patients may activate the coagulation system, thereby increasing thrombotic risk [38]. Additionally, Ki-67 positivity may indicate a high proliferative state of tumor cells, stimulating tissue factor expression, further elevating thrombosis risk [39]. These findings highlight the importance of monitoring thrombotic risk in breast cancer patients undergoing chemotherapy combined with targeted or endocrine therapies.

OR analysis identified four independent predictors—hemoglobin, APTT, TC, and PICC—which aligned with the features computed by AutoGluon. This consistency underscores the statistical robustness of these predictors and reinforces their biological plausibility. A key finding was that CRT risk significantly increases when hemoglobin exceeds 134.6 g/L. While previous studies have primarily focused on anemia as a risk factor for thrombosis [8], our results suggest that elevated hemoglobin levels may enhance erythrocyte-platelet interactions, which in turn promote thrombus formation [40]. Furthermore, APTT < 31.71 s may indicate enhanced coagulation factor activity, reflecting a hypercoagulable state [41]. TC > 11.19 mmol/L was also associated with increased CRT risk, likely due to its role in vascular endothelial dysfunction and platelet hyperreactivity [42, 43]. The association between PICC and thrombosis is well-documented, attributed to mechanical trauma to venous intima caused by arm movements and catheter occupying a most portion of the venous lumen [36, 44].

Based on these four independent predictors, a low- (0–1 factors) and high-risk (2–4 factors) stratification system was developed to enhance individualized CRT risk assessment based. Our findings suggest that pre-catheterization assessment of hemoglobin, APTT, TC, and catheter type can effectively predict CRT risk, providing actionable insights for personalized anticoagulation strategies. Compared to previous studies that primarily focused on static CRT risk, our results demonstrated that both catheter indwelling time and duration of use are crucial considerations in the management of patient with chemotherapy. This finding emphasizes the need for a multidimensional approach to assessing CRT risk.

Limitation

Despite its strengths, this study has certain limitations. Being a single-center retrospective study, external validation in multicenter cohorts is necessary to further assess model applicability. Additionally, as this study focused solely on breast cancer patients, some predictors (e.g., tumor molecular features) may not be generalizable to

other malignancies, necessitating further investigation in diverse cancer populations. Moreover, as this study primarily relied on static laboratory data, future research should incorporate longitudinal laboratory measurements to refine CRT risk prediction and enhance the model's ability to capture dynamic changes in thrombotic risk.

Conclusion

By integrating ML and Bayesian learning, this study developed a CRT risk prediction model that balances predictive accuracy with clinical interpretability. In addition to confirming known risk factors, we incorporated tumor biology, addressing a critical gap in prior CRT models that primarily focused on coagulation physiology. Furthermore, the proposed low- and high-risk stratification system offers a practical tool for guiding personalized anticoagulation strategies, with future validation in multicenter cohorts needed to optimize the implementation of thrombosis prevention in clinical oncology.

Abbreviations

APTT	Activated partial thromboplastin time
AUC	Area under the curve
BMI	Body mass index
CatBoost	Categorical boosting
CI	Confidence interval
CRT	Catheter-related thrombosis
CVC	Central venous catheter
ER	Estrogen receptor
ExtraTreesEntr	Extra trees entropy
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hazard ratio
IQR	Interquartile range
ML	Machine learning
NeuralNetFastAI	Neural net fast ai
OR	Odds ratio
PICC	Peripherally inserted central catheter
PR	Progesterone receptor
PR-AUC	Precision recall-AUC
RandomForestEntr	Random forest entropy
RandomForestGini	Random forest gini
ROC-AUC	Area under the receiver operating characteristic curve
SMOTE	Synthetic minority over-sampling technique
TC	Total cholesterol
TP	Threshold point
WeightedEnsemble	Weighted ensemble learning
XGBoost	Extreme gradient boosting

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-13946-y.

Additional file 1.

Acknowledgements

Thanks to Siheng Xiong (Ph.D. of Georgia Institute of Technology, USA) for his suggestions on mathematical model construction.

Authors' contributions

T.A., J.X., H.J., and Y.W: Conceptualization, design, methodology. H.H. and H.J.: manuscript writing, data analysis.Y.W. and Y.Z.: data analysis.H.J. and Y.W.: supervision, professional suggestion, revision. Y.W. and Y.W. : Funding acquisition. All authors reviewed the manuscript.

Funding

This study was funded by the program of Beijing Hope Run Special Fund of Cancer Foundation of China (LC2020A17), the CAMS Innovation Fund for Medical Sciences (CIFMS) (supported by the Special Research Fund for Central Universities, Peking Union Medical College, 2022-I2M-C&T-B-069), and the National Natural Science Fundation of China (823B2007). The funder had no role in the study design; in the collection, analysis, and interpretation of data; and in the decision to submit the paper for publication.

Data availability

Data used to generate results of this study could be obtained from the corresponding author at reasonable request.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki and was approved by the Ethics Committee of National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (Approval number: 22/444–3646). As this research involved a retrospective analysis of previously collected data, the requirement for informed consent was waived by the committee. The ethics committee also waived the need for consent to participate in the study. All data were de-identified to maintain patient confidentiality and privacy. The retrospective nature of the study ensured that there was no direct contact with patients, and no additional risks were posed to individuals whose data were included in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Cardiology, Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [2]Department of Cardiac Surgery, Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [3]Department of Management Center, Cancer Hospital of Huanxing Chaoyang District, Beijing, China. [4]Department of Comprehensive Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Pekingunion Medical College, Beijing, China.

Received: 29 August 2024 Accepted: 17 March 2025
Published online: 27 March 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W, He J. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent. 2024;4(1):47–53.
3. Loibl S, André F, Bachelot T, Barrios CH, Bergh J, Burstein HJ, Cardoso MJ, Carey LA, Dawood S, Del Mastro L, et al. Early breast cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. Ann Oncol. 2024;35(2):159–82.
4. Redana S, Sharp A, Lote H, Mohammed K, Papadimitraki E, Capelan M, Ring A. Rates of major complications during neoadjuvant and adjuvant chemotherapy for early breast cancer: an off study population. Breast. 2016;30:13–8.
5. Fu J, Cai W, Zeng B, He L, Bao L, Lin Z, Lin F, Hu W, Lin L, Huang H, et al. Development and validation of a predictive model for peripherally inserted central catheter-related thrombosis in breast cancer patients based on artificial neural network: a prospective cohort study. Int J Nurs Stud. 2022;135: 104341.

6. Lee AY, Kamphuisen PW. Epidemiology and prevention of catheter-related thrombosis in patients with cancer. J Thromb Haemost. 2012;10(8):1491–9.
7. Ma G, Chen S, Peng S, Yao N, Hu J, Xu L, Chen T, Wang J, Huang X, Zhang J. Construction and validation of a nomogram prediction model for the catheter-related thrombosis risk of central venous access devices in patients with cancer: a prospective machine learning study. J Thromb Thrombolysis. 2025;58(2):220–31.
8. Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW. Development and validation of a predictive model for chemotherapy-associated thrombosis. Blood. 2008;111(10):4902–7.
9. Huang X, Chen H, Meng S, Pu L, Xu X, Xu P, He S, Hu X, Li Y, Wang G. External validation of the Khorana score for the prediction of venous thromboembolism in cancer patients: a systematic review and meta-analysis. Int J Nurs Stud. 2024;159: 104867.
10. Gerotziafas GT, Taher A, Abdel-Razeq H, AboElnazar E, Spyropoulos AC, El Shemmari S, Larsen AK, Elalamy I. A predictive score for thrombosis associated with breast, colorectal, lung, or ovarian cancer: the prospective COMPASS-cancer-associated thrombosis study. Oncologist. 2017;22(10):1222–31.
11. Spyropoulos AC, Eldredge JB, Anand LN, Zhang M, Qiu M, Nourabadi S, Rosenberg DJ. External validation of a venous thromboembolic risk score for cancer outpatients with solid tumors: the COMPASS-CAT venous thromboembolism risk assessment model. Oncologist. 2020;25(7):e1083–90.
12. Agnelli G, George DJ, Kakkar AK, Fisher W, Lassen MR, Mismetti P, Mouret P, Chaudhari U, Lawson F, Turpie AG. Semuloparin for thromboprophylaxis in patients receiving chemotherapy for cancer. N Engl J Med. 2012;366(7):601–9.
13. Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, Smola A. Autogluon-tabular: Robust and accurate automl for structured data. arXiv preprint arXiv:200306505 2020.
14. Mantha S, Dunbar A, Bolton KL, Devlin S, Gorenshteyn D, Donoghue M, Arcila ME, Soff GA. Machine learning for prediction of cancer-associated venous thromboembolism. Blood. 2020;136:37.
15. Baskin JL, Pui CH, Reiss U, Wilimas JA, Metzger ML, Ribeiro RC, Howard SC. Management of occlusion and thrombosis associated with long-term indwelling central venous catheters. Lancet. 2009;374(9684):159–69.
16. Wu C, Zhang M, Gu W, Wang C, Zheng X, Zhang J, Zhang X, Lv S, He X, Shen X, et al. Daily point-of-care ultrasound-assessment of central venous catheter-related thrombosis in critically ill patients: a prospective multicenter study. Intensive Care Med. 2023;49(4):401–10.
17. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Research. 2002;16:321–57.
18. Chen M, Zhou Q, Li Y, Lu Q, Bai A, Ruan F, Liu Y, Jiang Y, Li X. Association between pre-pregnancy maternal stress and small for gestational age: a population-based retrospective cohort study. BMC Med. 2025;23(1): 7.
19. Yu J, Peng X, Zhou R, Zhu T, Hao X. Development and validation of an interpretable machine learning model to predict major adverse cardiovascular events after noncardiac surgery in geriatric patients: a prospective study. Int J Surg. 2025;111(2):1939–49.
20. Liu X, Liu X, Lai Y, Yang F, Zeng Y. Random decision DAG: an entropy based compression approach for random forest. In: Database systems for advanced applications: 2019// 2019. Cham: Springer International Publishing; 2019. p. 319–323.
21. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics. 2009;10(1):213.
22. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63(1):3–42.
23. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc.; 2018. p. 6639–6649.
24. Huang M. Theory and Implementation of linear regression. In: 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL). 2020. p. 210–217.
25. Shahhosseini M, Hu G, Pham H. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. Machine Learning with Applications. 2022;7: 100251.
26. Curtis FE, Scheinberg K. Optimization methods for supervised machine learning: from linear models to deep learning. In: Leading developments from INFORMS communities. edn. Hanover, Md: INFORMS; 2017. p. 89–114.
27. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015;521(7553):452–9.
28. Meng F, Fan S, Guo L, Jia Z, Chang H, Liu F. Incidence and risk factors of PICC-related thrombosis in breast cancer: a meta-analysis. Jpn J Clin Oncol. 2024;54(8):863–72.
29. Peng SY, Wei T, Li XY, Yuan Z, Lin Q. A model to assess the risk of peripherally inserted central venous catheter-related thrombosis in patients with breast cancer: a retrospective cohort study. Support Care Cancer. 2022;30(2):1127–37.
30. Gallieni M, Pittiruti M, Biffi R. Vascular access in oncology patients. CA Cancer J Clin. 2008;58(6):323–46.
31. Timsit JF, Rupp M, Bouza E, Chopra V, Kärpänen T, Laupland K, Lisboa T, Mermel L, Mimoz O, Parienti JJ, et al. A state of the art review on optimal practices to prevent, recognize, and manage complications associated with intravascular devices in the critically ill. Intensive Care Med. 2018;44(6):742–59.
32. Decousus H, Bourmaud A, Fournel P, Bertoletti L, Labruyère C, Presles E, Merah A, Laporte S, Stefani L, Piano FD, et al. Cancer-associated thrombosis in patients with implanted ports: a prospective multicenter French cohort study (ONCOCIP). Blood. 2018;132(7):707–16.
33. Hu Z, Luo M, He R, Wu Z, Fan Y, Li J. Development and validation of a risk prediction model for PICC-related venous thrombosis in patients with cancer: a prospective cohort study. Sci Rep. 2025;15(1):4654.
34. Khorana AA, Mackman N, Falanga A, Pabinger I, Noble S, Ageno W, Moik F, Lee AYY. Cancer-associated venous thromboembolism. Nat Rev Dis Primers. 2022;8(1):11.
35. Ahlbrecht J, Dickmann B, Ay C, Dunkler D, Thaler J, Schmidinger M, Quehenberger P, Haitel A, Zielinski C, Pabinger I. Tumor grade is associated with venous thromboembolism in patients with cancer: results from the Vienna cancer and thrombosis study. J Clin Oncol. 2012;30(31):3870–5.
36. Chopra V, Anand S, Hickner A, Buist M, Rogers MA, Saint S, Flanders SA. Risk of venous thromboembolism associated with peripherally inserted central catheters: a systematic review and meta-analysis. Lancet. 2013;382(9889):311–25.
37. Zhang X, Gao Y, Yang B, Ma S, Zuo W, Wei J. The mechanism and treatment of targeted anti-tumour drugs induced cardiotoxicity. Int Immunopharmacol. 2023;117: 109895.
38. Cushman M, Kuller LH, Prentice R, Rodabough RJ, Psaty BM, Stafford RS, Sidney S, Rosendaal FR. Estrogen plus progestin and risk of venous thrombosis. JAMA. 2004;292(13):1573–80.
39. Unruh D, Horbinski C. Beyond thrombosis: the impact of tissue factor signaling in cancer. J Hematol Oncol. 2020;13(1):93.
40. Da Q, Teruya M, Guchhait P, Teruya J, Olson JS, Cruz MA. Free hemoglobin increases von Willebrand factor-mediated platelet adhesion in vitro: implications for circulatory devices. Blood. 2015;126(20):2338–41.
41. Sørensen B, Ingerslev J. Dynamic APTT parameters: applications in thrombophilia. J Thromb Haemost. 2012;10(2):244–50.
42. Saini HK, Arneja AS, Dhalla NS. Role of cholesterol in cardiovascular dysfunction. Can J Cardiol. 2004;20(3):333–46.
43. van der Stoep M, Korporaal SJ, Van Eck M. High-density lipoprotein as a modulator of platelet and coagulation responses. Cardiovasc Res. 2014;103(3):362–71.
44. Chopra V, Ratz D, Kuhn L, Lopus T, Lee A, Krein S. Peripherally inserted central catheter-related deep vein thrombosis: contemporary patterns and predictors. J Thromb Haemost. 2014;12(6):847–54.

Publisher's Note