BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Machine learning–based prediction model for hypofibrinogenemia after tigecycline therapy

Jianping Zhu[1], Rui Zhao[1], Zhenwei Yu[1], Liucheng Li[1], Jiayue Wei[2] and Yan Guan[1*]

Abstract

Background In clinical practice, the incidence of hypofibrinogenemia (HF) after tigecycline (TGC) treatment significantly exceeds the probability claimed by drug manufacturers.

Objective We aimed to identify the risk factors for TGC-associated HF and develop prediction and survival models for TGC-associated HF and the timing of TGC-associated HF.

Methods This single-center retrospective cohort study included 222 patients who were prescribed TGC. First, we used binary logistic regression to screen the independent factors influencing TGC-associated HF, which were used as predictors to train the extreme gradient boosting (XGBoost) model. Receiver operating characteristic curve (ROC), calibration curve, decision curve analysis (DCA), and clinical impact curve analysis (CICA) were used to evaluate the performance of the model in the verification cohort. Subsequently, we conducted survival analysis using the random survival forest (RSF) algorithm. A consistency index (C-index) was used to evaluate the accuracy of the RSF model in the verification cohort.

Results Binary logistic regression identified nine independent factors influencing TGC-associated HF, and the XGBoost model was constructed using these nine predictors. The ROC and calibration curves showed that the model had good discrimination (areas under the ROC curves (AUC) = 0.792 [95% confidence interval (CI), 0.668–0.915]) and calibration ability. In addition, DCA and CICA demonstrated good clinical practicability of this model. Notably, the RSF model showed good accuracy (C-index = 0.746 [95%CI, 0.652–0.820]) in the verification cohort. Stratifying patients treated with TGC based on the RSF model revealed a statistically significant difference in the mean survival time between the low- and high-risk groups.

Conclusions The XGBoost model effectively predicts the risk of TGC-associated HF, whereas the RSF model has advantages in risk stratification. These two models have significant clinical practical value, with the potential to reduce the risk of TGC therapy.

Keywords Tigecycline, Hypofibrinogenemia, Machine learning, Influencing factors, Prediction models, Survival model

*Correspondence:
Yan Guan
guanyan@zju.edu.cn
[1]Pharmacy Department, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou 310020, China
[2]Zhejiang Cancer Hospital, Hangzhou, Zhejiang 310022, China

Introduction

Infections caused by multidrug-resistant (MDR) and extensively drug-resistant (XDR) pathogens, which lead to elevated morbidity, increased mortality rates, and prolonged hospital stays, have become increasingly prevalent among patients, posing a significant global threat [1, 2]. Consequently, the quest for efficacious antimicrobial agents that target MDR and XDR strains has emerged as a paramount focus in clinical practice. Tigecycline (TGC) is a broad-spectrum parenteral glycylcycline antibiotic widely used in clinical practice. TGC is structurally similar to tetracyclines but has a 5-fold higher binding affinity due to the addition of an N-alkyl-glycinamide side chain on C9 to the main chain of minocycline, which has further broadened the antimicrobial spectrum of TGC, especially for MDR and XDR bacteria (e.g. vancomycin-resistant *enterococci*, carbapenem-resistant *Enterobacteriaceae*, and methicillin-resistant *Staphylococcus aureus*) [3]. TGC binds reversibly to the helical region (H30) on the 34 S subunit of the bacterial ribosome, preventing the addition of small amino acids to the peptide chain and protein translation, thereby inhibiting bacterial growth and reproduction [4]. For some patients with complex abdominal infections, lung infections, and those in the ICU, the broad-spectrum antibiotic TGC tends to show good results because these patients are immunocompromised, have more underlying diseases, and are more frequently subjected to invasive procedures [5]. Notwithstanding the disconcerting alert raised by the Food and Drug Administration (FDA) linking TGC to overall mortality, the extensive utilization of TGC in the management of infections stemming from MDR/XDR origins persists due to the lack of alternative efficacious antibiotics [6].

Assessment of drug-related adverse reactions is a pivotal criterion in the appraisal of drug safety. TGC, characterized by its limited drug interactions and patient-friendly tolerability, often manifests as common adverse events such as nausea, vomiting, diarrhea, and elevated transaminase and bilirubin levels in biochemical markers [7, 8]. However, recent reports showing that TGC appears to be associated with coagulation dysfunction have caught the attention of researchers. Prolongation of TGC treatment has been reported to be a major risk factor for HF, inducing a significant decrease in fibrinogen (Fib) and an increase in prothrombin time [9, 10]. In a cohort study, the incidence of HF was reported to be 50.5%, with 10.1% of patients experiencing bleeding after TGC treatment [11]. Fib, a crucial plasma glycoprotein synthesized and secreted by hepatic parenchymal cells, is a pivotal coagulation factor. Thrombin-mediated conversion of Fib into insoluble fibrin precipitates blood clot formation and plays an indispensable role in the hemostatic process [8]. The normal plasma Fib concentration is between 2 and 4 g/L. When the Fib level exceeds this range, it precipitates severe clotting abnormalities, markedly elevating the risk of conditions such as cerebral hemorrhage, gastrointestinal hemorrhage, atherosclerosis, thrombosis, and other related ailments. This perilous scenario poses a significant threat to the overall well-being and safety of patients [12, 13]. Although some studies have described the risk factors for TGC-associated HF, including treatment dose and duration, baseline Fib levels, and patient sex, conclusions are often controversial due to limitations such as small sample sizes or inadequate descriptions [11, 14–16].

Therefore, a well-performing model to predict the occurrence of TGC-associated HF in patients receiving TGC treatment is urgently needed. In this context, we conducted a single-center retrospective cohort study on the use of TGC in 222 infected patients. We developed and validated two models to predict the risk and timing of TGC-associated HF to improve the safety of TGC clinical medication.

Methods

Patients

This single-center retrospective cohort study was conducted at Sir Run Run Shaw Hospital, affiliated with the Zhejiang University School of Medicine, which is renowned as a comprehensive third -level 1st class hospital. We included patients who were hospitalized and received TGC treatment, with fibrinogen levels measured before and after treatment, from January 1, 2021, to December 31, 2021 ($n=390$). Patients with the following characteristics were excluded from the study: (1) Fib level < 2.0 g/L before administration; (2) TGC treatment duration < 3 days; (3) patients monitored for < 2 days after TGC; (4) patients who experienced recent bleeding complications; and (5) patients with data missing > 30%. Finally, 222 patients were included in the final analysis.

Data collection and definition

Using an electronic medical record system, we collected data on 222 eligible patients, encompassing basic patient information, clinical characteristics, and laboratory data. Routine monitoring of Fib levels was conducted during hospitalization (48 h after TGC administration). Within this monitoring timeframe, the nadir Fib levels were recorded, categorizing patients with Fib levels below 2.0 g/L as exhibiting HF and those above 2 g/L as non-HF. For survival analysis, the outcome variable was the time interval from the start of TGC treatment to HF (survival time) in patients with HF. However, the interval was censored as the TGC treatment time for patients without TGC-associated HF.

Model construction and verification

The samples ($n=222$) were initially assessed for multicollinearity to ensure the stability of subsequent models. Based on the Spearman's correlation analysis, two highly correlated features were identified and removed to mitigate multicollinearity. Following this preprocessing step, the samples were divided into two groups: the training cohort ($n=166$) and the verification cohort ($n=56$), at a 3:1 ratio, using R software (version 4.3.0, https://www.R-project.org/).

For the extreme gradient boosting (XGBoost) prediction model, univariate binary logistic regression was first conducted to identify potential influencing factors for TGC-associated HF in the training cohort. The strength and direction of the associations between these factors and outcome were quantified using odds ratios (OR) and 95% confidence intervals (CI). Factors identified with p-values less than 0.1 in the univariate analysis were subsequently included in a stepwise logistic regression, employing a forward selection method to determine the most significant predictors. Then, the predictors with p-values less than 0.05 in the stepwise regression were deemed as significant and employed for constructing the XGBoost prediction model. Prior to training the XGBoost model, we employed a Bayesian optimization technique for hyperparameter tuning [17–19]. Bayesian optimization is a global optimization method grounded in probabilistic models, wherein the optimal hyperparameter combination is sought through the establishment of a Gaussian process regression model of the objective function. In this study, we set a predefined search range for each hyperparameter and executed multiple iterations using the Bayesian optimization method to identify the optimal hyperparameter combination within the search space. The optimal hyperparameters for the XGB model are detailed in **Supplementary material**. Ultimately, we achieved an optimized XGBoost model that demonstrated an enhanced performance in the training cohort. The receiver-operating characteristic (ROC) curve, calibration curve, decision curve analysis (DCA), and clinical impact curve analysis (CICA) were used to evaluate the discrimination, calibration, and clinical practicability of the model. The model was internally validated using a five-fold cross-validation.

For the random survival forest (RSF) model, the partitioning of the training and verification cohorts was consistent with that of the XGBoost model. Prior to RSF modeling, we made specific adjustments to our data to meet the survival analysis prerequisites. For patients who developed HF, we designated the observation period as the time from the start of TGC treatment to the onset of HF. For patients without HF, the treatment time for TGC was considered as the observation period. This approach ensured an accurate representation of the time-to-event nature of HF in relation to the TGC treatment. To develop the RSF model, we utilized all the available features in the training dataset. Once the model is trained, it yields importance scores for every feature, allowing us to discern the most influential predictors. Subsequently, we conducted fivefold cross-validation for all features in the training cohort ten times to obtain stable feature selection results. Finally, the trained model was deployed on the verification cohort to predict the survival outcomes. The RSF model was constructed using 1,000 trees.

Statistical analysis

R software was used to support the grouping and statistical analysis of the data. Covariates featuring missing values exceeding 10% were eliminated, whereas those with values less than 10% were filled out by multiple imputation. Continuous data were expressed as mean ± standard deviation (mean ± SD) or median, interquartile range (IQR), and categorical data were expressed as numerical values and percentages (n, %); $P<0.05$, considered statistically significant. The t-test or Mann-Whitney U test was used to compare the two groups. The figures presented in this study were generated using R software and GraphPad Prism software (version 9.5.1, https://www.graphpad.com/).

Results

Study design

The patients ($n=222$) were randomly divided into training ($n=166$) and validation ($n=56$) cohorts. The training cohort was used to identify the independent influencing factors of TGC-associated HF and construct the XGBoost model. We conducted five-fold cross-validation for all features in the training cohort to obtain stable feature selection results to construct the RSF model. The validation cohort was then employed for internal validation of our models. Figure 1 presents a comprehensive summary of the patient selection process and study design, providing clear insights into the total registry population, excluded patients, reasons for exclusion, final study population, and group allocation.

Demographics, clinical characteristics, and laboratory examination of the study population

In this study, a comprehensive analysis was performed involving 222 patients satisfying predefined criteria. The mean age of all patients was 62.49 ± 16.93, comprising 149 (67.11%) males and 73 (32.98%) females. 99 (44.6%) developed HF after TGC treatment. The level of Fib, a crucial coagulation marker, significantly decreased after TGC treatment ($P<0.0001$) (Fig. 2A), suggesting a potential propensity for HF occurrence subsequent to TGC administration and an increased likelihood of bleeding complications. Before TGC prescription treatment, the

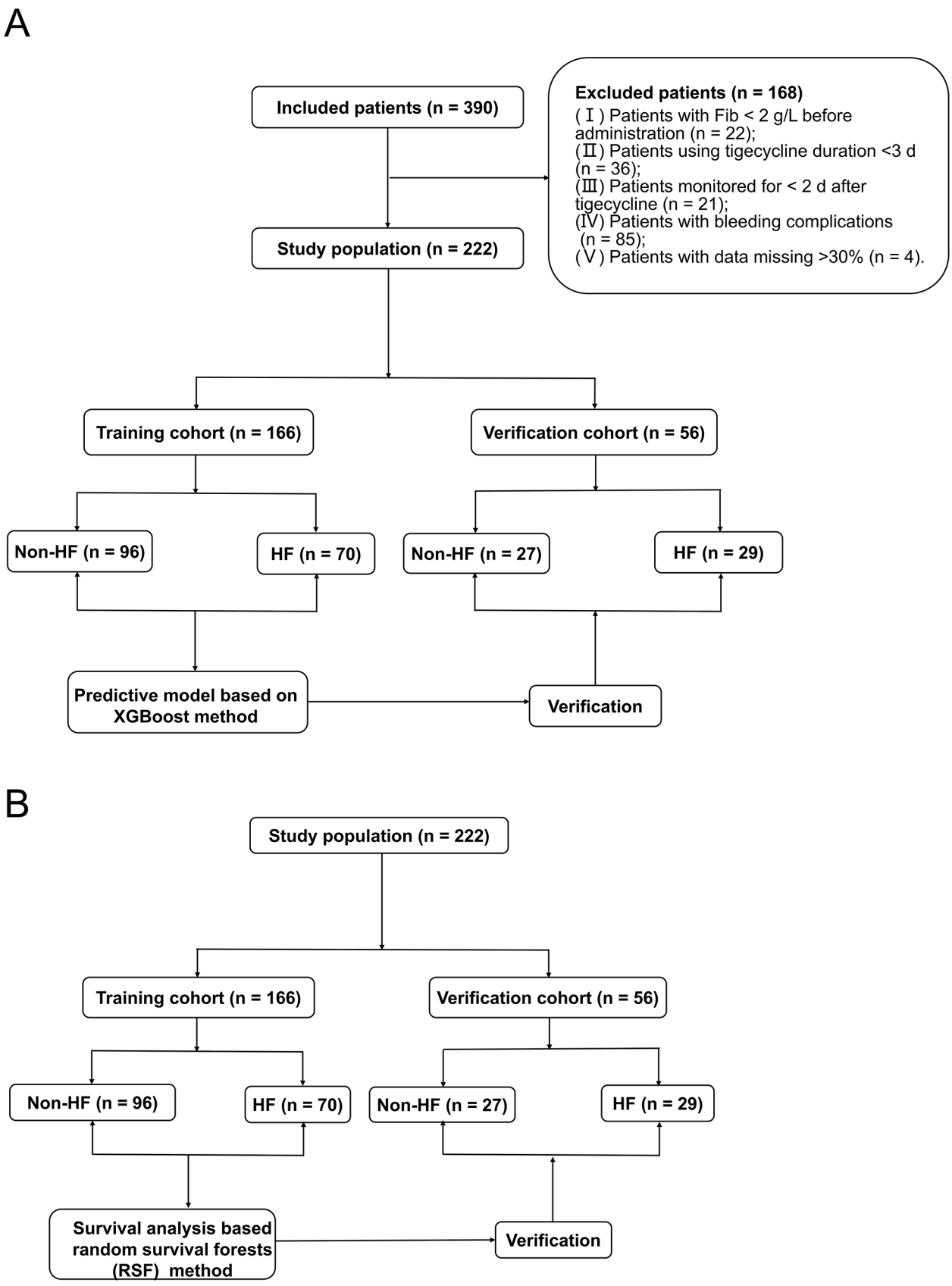

Fig. 1 Flowchart of study design. (**A**) Flowchart of patients selection process, development and verification of the XGBoost predictive model. (**B**) Flowchart of development and verification of the RSF survival analysis model

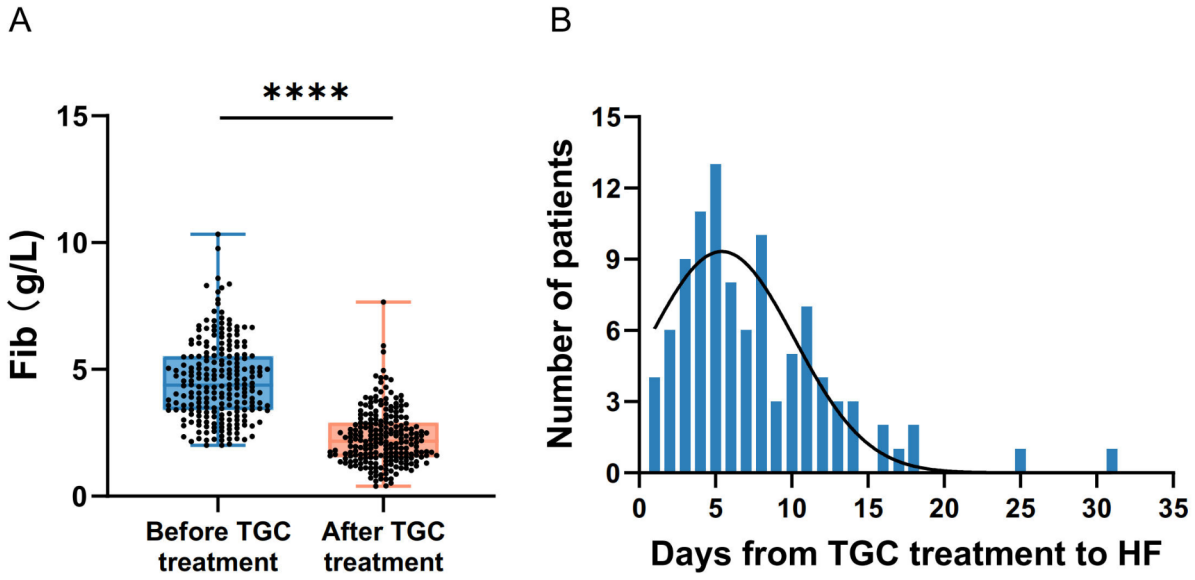

Fig. 2 The variation of Fib levels and distribution of HF occurrence time. (**A**) The variation of Fib before and after TGC prescription treatment. ****$P < 0.0001$. (**B**) The patients developed HF from 1 to 31 days after after TGC treatment, with a median (25-75%) of 6 (4–8) days

mean Fib level was 4.55 g/L, while the mean level of the lowest Fib we monitored decreased to 2.32 g/L after TGC treatment. Distinct disparities between the HF and non-HF groups were noted across 12 parameters, including factors such as age, SOFA score, duration of hospitalization, baseline fibrinogen levels, and PCT levels ($P<0.05$). Detailed demographic and clinical characteristics and laboratory examination results of the study population are shown in Tables 1 and 2. In the HF group, HF developed 1 to 31 days after TGC treatment, with a median (25-75%) of 6 (4–8) days (Fig. 2B). Table 3 presented the departmental distribution of patients subjected to TGC treatment and those who developed HF following treatment. The patients included in the present study were mainly distributed in the departments of Hematology, General surgery, and Respiratory medicine, and the incidence of TGC-associated HF was 40.00%, 50.82%, and 38.46%, respectively. Irrespective of the department, the likelihood of TGC-associated HF was consistently higher than that provided by drug manufacturers.

Univariate and multivariate binary logistic regression analysis for influencing factors of TGC-associated HF

We analyzed the demographic and clinical characteristics and laboratory data of the training cohort using univariate and multivariate binary logistic regression analyses. Table 4 displayed the results of univariate and multivariate binary logistic regression analyses. In the univariate analysis, 14 influencing factors were screened ($P<0.1$), including age, SOFA score, ICU, baseline fibrinogen level, treatment duration, TBIL, ALB, PCT, cardiovascular disease, anticoagulant drugs/antiplatelet drugs or nonsteroidal anti-inflammatory drugs, no combination with anticoagulant/procoagulant, combined with β-lactam antibiotics, abdominal infection, and sepsis. All the above factors were included in the multivariate analysis. In multivariate analysis, nine factors were identified as independent influencing factors of TGC-associated HF, including age, SOFA score, baseline fibrinogen level, treatment duration, PCT, combined with anticoagulant drugs, antiplatelet drugs, or nonsteroidal anti-inflammatory drugs, combined with β-lactam antibiotics, abdominal infection, and sepsis. Of these, age, SOFA score, treatment duration, and PCT level were the risk factors (adjusted OR>1).

Correlation analysis of influencing factors

To further examine the independence of the nine factors screened and ensure the stability of subsequent models, we performed a correlation analysis using the Spearman method. The p-values and correlation coefficients (r_s) between all the parameters are visually represented in the heatmap. Spearman's rank correlation coefficient was used to determine the correlation between the paired variables, thus forming a heat map composed of all correlation coefficients. As shown in Fig. 3, the maximum r_s value was only 0.28, observed between age and treatment duration, implying the absence of substantial correlations between the diverse factors. In other words, there was no multicollinearity between the nine influencing factors.

Prediction outcomes of the XGBoost model for TGC-associated HF

To determine whether the nine factors could predict TGC-associated HF, we built a prediction model using the XGBoost machine learning algorithm in the training

Table 1 Characteristics of the study population

Characteristics	Total (*n* = 222)	Non-HF (*n* = 123)	HF (*n* = 99)	*P*
HF	99 (44.59%)			
Male	149 (67.11%)	78 (63.41%)	71 (71.72%)	0.191
Age (Year)	62.49 ± 16.93	59.76 ± 18.44	65.89 ± 14.12	0.006
BMI	22.0 (20.2, 24.8)	22.0 (22.7, 25.1)	21.7 (19.6, 24.5)	0.141
Infection type				
Septic shock	3 (1.4%)	1 (0.81%)	2 (2.02%)	0.587
Sepsis	6 (2.7%)	4 (3.25%)	2 (2.02%)	0.694
Pulmonary infection	69 (31.08%)	43 (34.96%)	26 (26.26%)	0.164
Urinary tract infection	1 (0.45%)	0	1 (1.01%)	0.446
Skin and soft tissue infection	3 (1.4%)	2 (1.63%)	1 (1.01%)	1.000
Neutropenia with infection	4 (1.8%)	1 (0.81%)	3 (3.03%)	0.326
Abdominal infection	34 (15.32%)	24 (19.51%)	10 (10.10%)	0.053
Two or more infection types	102 (45.95%)	48 (39.02%)	54 (84.85%)	<0.001
SOFA score	4 (2, 7)	4 (2, 6)	4 (3, 8)	0.016
Hospitalization time	30 (20, 48)	30(20, 47)	32 (19, 48)	0.491
ICU(Yes)	146 (65.77%)	73 (59.35%)	73 (73.74%)	0.025
Shock (Yes)	54 (24.32)	23 (18.70%)	31 (31.31%)	0.029
Concomitant drugs				
Anticoagulant drugs/antiplatelet drugs/nonsteroidal anti-inflammatory drugs	86 (38.74%)	53 (43.09%)	33 (33.33%)	0.138
Hormone	11 (4.95%)	4 (3.25%)	7 (7.07%)	0.192
Combination with two types	110 (49.55%)	53 (43.09%)	57 (57.58%)	0.008
No combination	15 (6.76%)	13 (10.57%)	2 (2.02%)	0.012
Combined with other antimicrobials (Yes)	180 (%)	101 (82.11%)	79 (79.80%)	0.215
Treatment duration (Day)	11 (8, 17)	10 (7, 16)	12 (8, 19)	0.028
Surgery a month ago (Yes)	60	31	29	0.495

Notes: Continuous data are presented as mean ± SD or mean (IQR); categorical data are presented as n (%), and $P < 0.05$, which was considered to be statistically significant between the HF and non-HF groups

cohort. We then applied the model to predict TGC-associated HF in the verification cohort to evaluate its performance (Fig. 4A). We evaluated the performance of our model in terms of its discrimination, calibration, and clinical practicability. The ROC curves demonstrated good discrimination of our model. The areas under the ROC curves (AUC) were 0.812 (95%CI, 0.746–0.878) and 0.792 (95%CI, 0.668–0.915) in the training and verification cohorts, respectively (Fig. 4B and C). The calibration curve of the model was close to the ideal diagonal for both the training and verification cohorts, confirming the good degree of discrimination of the model (Fig. 4D and E). Furthermore, the XGBoost model showed a superior overall net benefit based on DCA and CICA results, substantiating its clinical practicability (Fig. 4F-I). Thus, we identified a set of predictors for TGC-associated HF and developed a well-performing prediction model.

Table 2 The results of laboratory examination of the study population

Characteristics	Total (*n* = 222)	Non-HF (*n* = 123)	HF (*n* = 99)	*P*
Baseline Fib level (g/L)	4.55 ± 1.56	4.90 ± 1.59	4.10 ± 1.40	<0.001
Lowest Fib level during treatment (g/L)	2.32 ± 1.08	3.05 ± 0.88	1.41 ± 0.38	<0.001
TBIL (g/L)	14.3 (10.1, 23.1)	13.9 (10.0, 19.4)	15.3 (10.4, 27.9)	0.062
ALT (U/L)	22.5 (13.8, 44.3)	22.0 (14.0, 43.0)	23.0 (13.0, 46.0)	0.899
AST (U/L)	26.0 (17.8, 43.0)	25.0 (18.0, 41.0)	27.0 (17.0, 52.0)	0.577
Crea (μmol/L)	61.5 (46.0, 92.3)	61.0 (46.0, 88.0)	62.0 (46.0, 106.0)	0.750
WBC (10^9/L)	7.3 (3.5, 12.5)	7.5 (2.3, 11.9)	7.2 (4.5, 13.3)	0.416
NE%	83.0 (70.9, 88.9)	83.0 (71.1, 86.8)	83.0 (70.5, 90.2)	0.297
ALB (g/L)	28.9 (26.4, 31.6)	29.3 (26.7, 32.3)	28.2 (25.7, 31.2)	0.047
CRP (mg/L)	99.3 (49.05, 160.75)	97.0 (45.6, 177.7)	101.6 (52.6, 157.6)	0.985
PCT (ng/mL)	0.40 (0.16, 1.11)	0.30 (0.13, 0.61)	0.61 (0.29, 1.69)	<0.001
Type of infected bacteria				
Acinetobacter baumannii	67 (30.18%)	32 (26.01%)	35 (35.35%)	0.132
Klebsiella pneumoniae	53 (23.87%)	32 (26.02%)	21 (21.21%)	0.404
Others	102 (45.95%)	59 (47.97%)	43 (43.43%)	0.523

Notes: Continuous data were presented as mean ± SD or mean (IQR), categorical data were presented as n (%), $P < 0.05$ was considered to be statistically significant between HF group and non-HF group

Prediction outcomes of the RSF model for the timing of TGC-associated HF

To further predict the timing of TGC-associated HF, we developed an RSF model in the training cohort and applied the model to predict the timing of TGC-associated HF in the verification cohort to measure the performance of our model (Fig. 5A). The evaluation of the feature importance revealed features that significantly contributed to the accuracy of the model. The features ranked in the top 10 in terms of importance for the RSF model are shown in Fig. 5B, including PCT, baseline

Table 3 The distribution of departments of patients

Department	Number of parents treated with TGC	Patients developed HF after TGC treatment	Incidence of HF (%)
Hematology	65	26	40.00
General surgery	61	31	50.82
Respiratory medicine	26	10	38.46
Emergency	12	7	58.33
Infectious liver diseases	10	4	40.00
Orthopaedics	9	3	33.33
Critical care medicine	6	2	30.00
Anorectal surgery	6	4	66.67
Neurosurgery	5	3	60.00
Cardiology	3	1	
Gastroenterology	3	2	
Neurology	2	1	
Cardiac surgery	2	0	
Nephrology	2	1	
Thoracic	2	1	
Urology surgery	2	1	
Oncology radiotherapy	2	1	
Medical oncologists	1	0	
Plastic surgery	1	1	
Endocrinolog	1	0	
Rehabilitation	1	0	
Total	222	99	

Notes: Departments with a small number of cases did not calculate the incidence

fibrinogen level, TBIL, ALB, cardiovascular disease, SOFA score, AST, ALT, age, and BMI index. Finally, we selected the top four features in terms of importance (PCT, baseline fibrinogen level, TBIL, ALB) by fivefold cross-validation, which showed good predictive power (concordance index (C-index)=0.746 [95%CI, 0.652–0.820]) for predicting the timing of TGC-associated HF (Fig. 5C). The estimated survival function of the patients in the verification cohort is shown in Fig. 5D. Therefore, we identified a potential predictive panel for the timing of TGC-associated HF.

The RSF risk stratification of patients

Patient stratification is of great significance for effective patient management. We utilized a RSF model to predict risk scores for each sample. For each calculated risk score, we iterated through them, treating each score as a potential threshold for dividing patients into high-risk and low-risk groups. Through this iterative process, we computed the log-rank test statistic for each attempt, seeking the threshold that yielded the maximum statistic. This threshold was considered the optimal risk score threshold (13.668), at which point the samples were categorized into high-risk and low-risk groups. As shown in Fig. 6, the mean survival time of the high-risk group was significantly shorter than that of the low-risk group in both cohorts. The mean survival time and occurrence of TGC-associated HF in different RSF stratifications are shown in Table 5.

Table 4 Univariate and multivariate binary logistic regression analysis for the influencing factors of TGC-Associated HF in the training cohort

Variables	Univariate		Multivariate	
	Crude OR (95%CI)	*P*-value	Adjusted OR (95%CI)	*P*-value
Age (Year)	1.026 (1.006–1.048)	0.010	1.048 (1.020–1.081)	0.001
SOFA score	1.087 (1.004–1.004)	0.044	1.140 (1.028–1.273)	0.016
Icu (Yes)	1.977 (1.019–3.936)	0.048		
Baseline Fib level (g/L)	0.750 (0.594–0.932)	0.010	0.502 (0.347–0.692)	<0.001
Treatment time (Day)	1.026 (0.997–1.059)	0.088	1.099 (1.055–1.150)	<0.001
TBIL	1.014 (1.000-1.031)	0.070		
ALB	0.934 (0.869–0.999)	0.056		
PCT	1.024 (0.998–1.057)	0.094	1.045 (1.005–1.091)	0.040
Cardiovascular disease	2.685 (1.269–5.844)	0.010		
Combined with anticoagulant drugs/ antiplatelet drugs/ nonsteroidal anti-inflammatory drugs	0.528 (0.272–1.005)	0.053	0.325 (0.133–0.747)	0.010
No combination with anticoagulant/ procoagulant	0.253 (0.038-1.000)	0.083		
Combined with β-lactam antibiotics	0.577 (0.289–1.124)	0.098	0.400 (0.161–0.957)	0.043
Abdominal infection	0.523 (0.204–1.237)	0.099	0.154 (0.037–0.541)	0.006
Sepsis	0.676 (0.092–3.569)	0.100	0.078 (0.007–0.661)	0.025

Notes: $P<0.05$ was considered statistically significant

Discussion

The current study first introduced machine learning algorithms, the XGBoost and RSF models, for the monitoring of patients following TGC combination therapy. The findings of this study are as follows: First, we identified 9

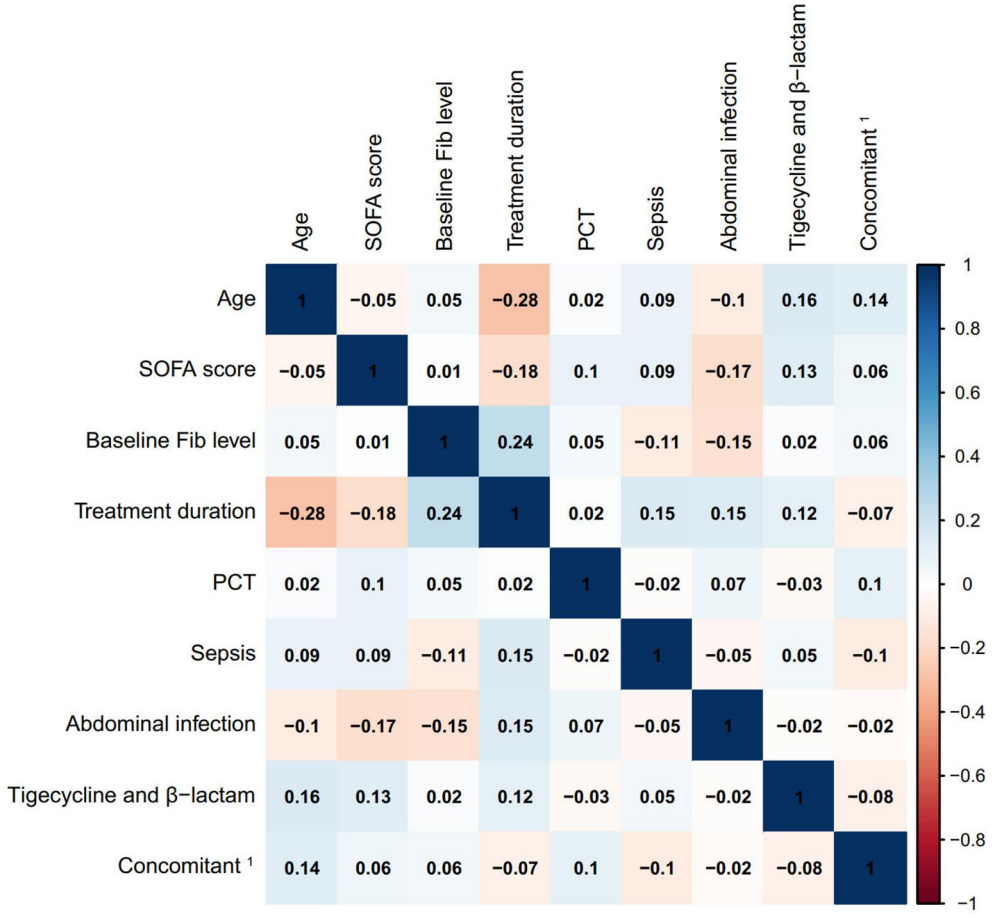

Fig. 3 The correlation coefficient heatmap of influencing factors. Concomitant[1] means the patients received treatment of TGC combined with anticoagulant drugs/antiplatelet drugs/nonsteroidal anti-inflammatory drugs

independent factors of TGC-associated HF, which were not completely consistent with those of previous studies. Second, these two models demonstrate superior performance in terms of discrimination, calibration, and clinical applicability. Third, the RSF model exhibited excellent performance in risk stratification of patients.

TGC, the first glycyl-tetracycline drug, is widely used for complex abdominal infections, complex skin and soft tissue infections, and community-acquired pneumonia [20]. TGC was approved by the FDA in June 2005 and entered the Chinese market in November 2011. The antibacterial spectrum of TGC covers both gram-negative and gram-positive bacteria, and it also exhibits strong antibacterial activity against MDR and XDR bacteria, leading to the continual expansion of its application domain [21]. With the widespread use of TGC, more attention has been paid to their adverse reactions.

Gastrointestinal symptoms, such as nausea and vomiting, are common adverse effects for drug manufacturers. Hemolymphatic symptoms, such as a decrease in fibrinogen level, prolonged prothrombin time (PT), and prolonged activated partial thromboplastin time (aPTT), are categorized as uncommon reactions [16]. However, since the first report of TGC-associated HF in 2010 [22], it has been increasingly reported more and more frequently [9, 23–25]. Previous studies have shown a probability of HF of 14–60% during TGC treatment, clearly highlighting that this adverse reaction is far from uncommon [13]. In many coagulopathic events, Fib is one of the pivotal factors that tends to precipitate a decline in the blood. Bleeding may occur when the fibrinogen level is below 1 g/L [26]. Zhu et al. suggested that prior TGC therapy within 1 month was an independent risk factor for hemorrhagic pneumonia [27]. In our study, 44.59% of patients developed HF after receiving TGC treatment, mirroring the results of Zhang et al., with a similarity rate of 50.50% [11]. HF developed at a median of 6 (4–10) days in our study, similar to the timing identified by Hu et al., although their study reported a higher HF incidence of 55.12% [14]. Based on the above studies, it can be concluded that the impact of TGC on coagulopathy is common in clinical applications and potentially leads to serious harm. Thus, identification of patients with an increased risk of TGC-associated HF and early

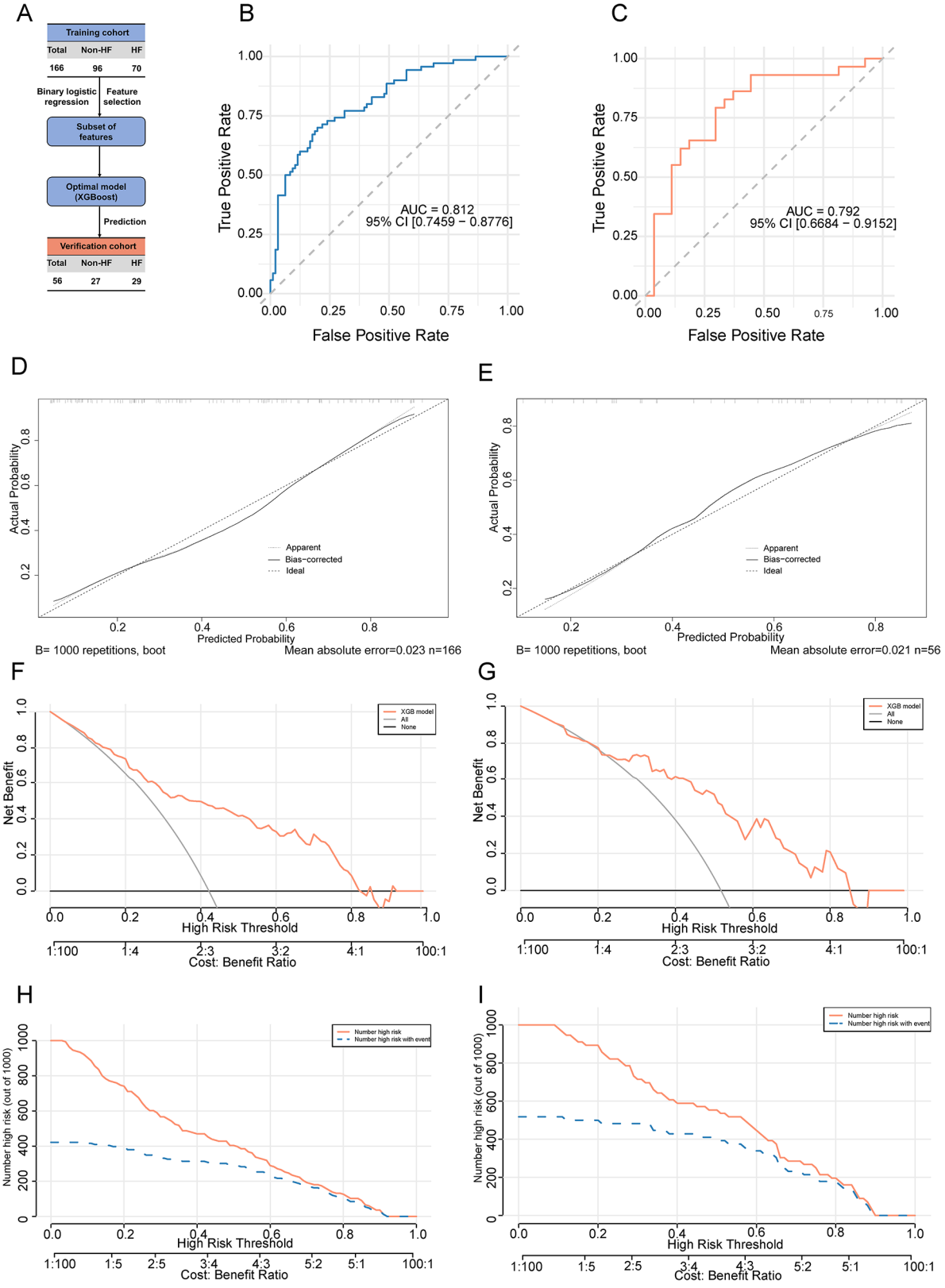

Fig. 4 Prediction outcomes of the XGBoost model for TGC-associated HF. (**A**) Schematic diagram of the dataset creation and analysis strategy for the prediction of TGC-associated HF. (**B** and **C**) The ROC curves of the XGBoost model in the training cohort and verification cohort. (**D** and **E**) The calibration curves of the XGBoost model in the training cohort and verification cohort. (**F** and **G**) The DCA of the XGBoost model in the training cohort and verification cohort. (**H** and **I**) The CICA of the XGBoost model in the training cohort and verification cohort

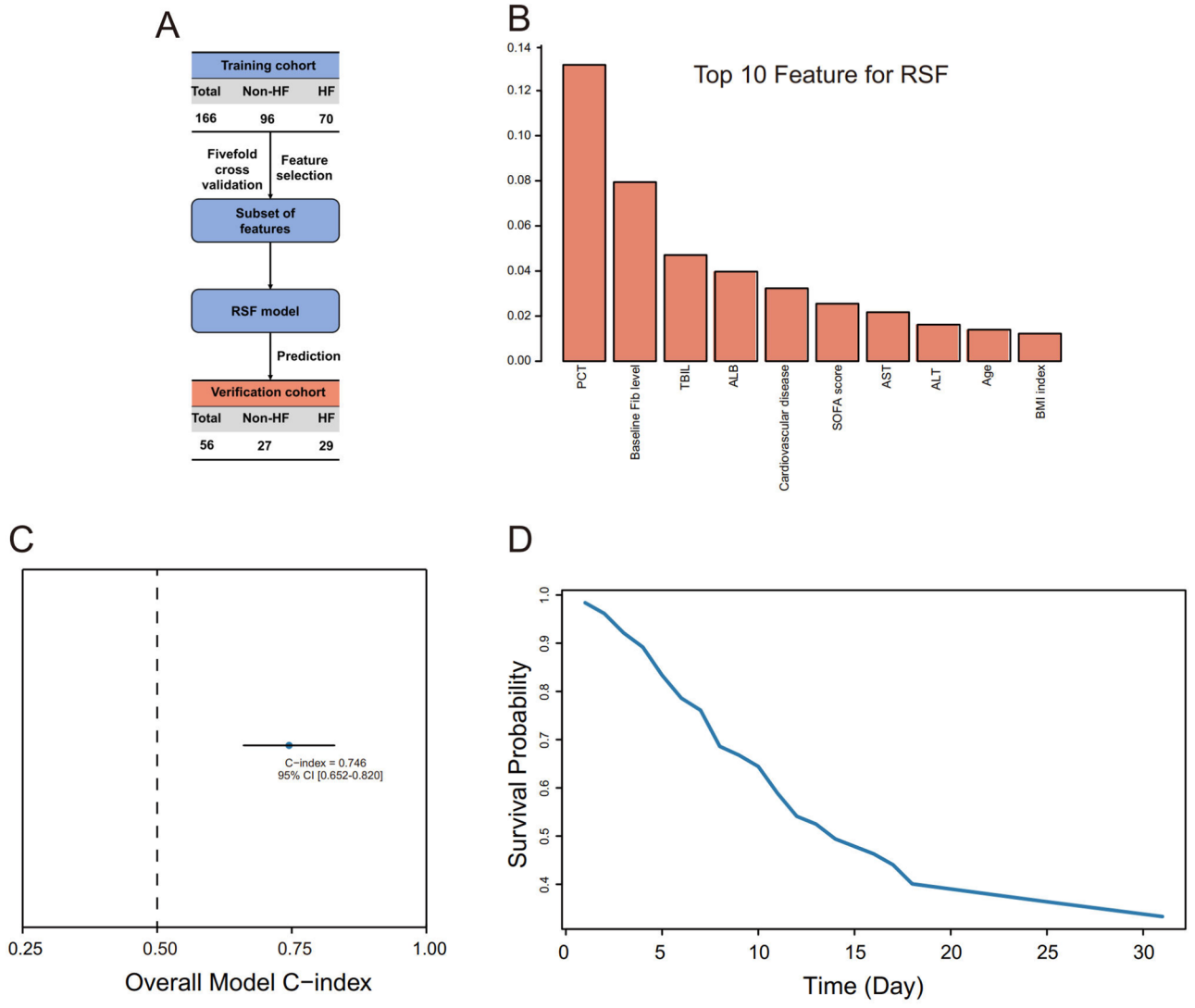

Fig. 5 Prediction outcomes of the RSF model for the timing of TGC-associated HF. (**A**) Schematic diagram of the dataset creation and analysis strategy for the prediction of the timing of TGC-associated HF. (**B**) The top 10 features in importance to the RSF model. (**C**) Overall C-index values of RSF model in the verification cohort. Error bars represent 95% confidence intervals. (**D**) The estimated survival function of patients in verification cohort

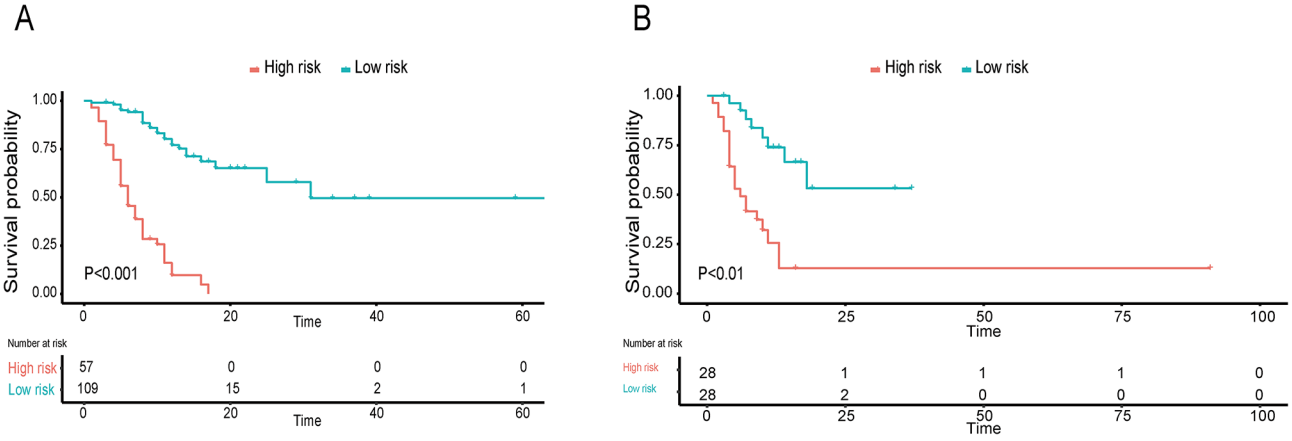

Fig. 6 Survival curves for patients after TGC treatment with different risks stratified using the RSF model. (**A**) Patients with low-risk vs. high-risk in the training cohort. (**B**) Patients with low-risk vs. high-risk in the verification cohort

Table 5 The mean survival time (days) of different risk stratifications in the training and verification cohort

Cohort	RSF Risk Stratification	Number	Events	Mean	95%LCI	95%UCI
Training	Low-risk	101	28	12.852	10.950	14.723
	High-risk	65	42	5.923	9.524	7.000
Verification	Low-risk	37	15	11.351	9.378	13.893
	High-risk	19	14	10.684	4.947	20.684

Notes: LCI means Low confidence interval, UCI means Up confidence interval

intervention and treatment is necessary in clinical practice.

At present, most studies on TGC-associated HF primarily focus on analyzing clinical characteristics and influencing factors, and the use of models for prediction is rarely reported. Therefore, it is imperative to acquire a more comprehensive understanding of the factors influencing TGC-associated HF and to construct a prediction model based on these factors to accurately forecast the likelihood of TGC-associated HF. In our study, we screened nine independent factors influencing TGC-associated HF using binary logistic regression analysis. Notably, our findings are not completely consistent with those of previous research. We found that age, SOFA score, treatment time, and PCT level were independent risk factors for TGC-associated HF. To our knowledge, there have been no reports that SOFA score and PCT are risk factors for TGC-associated HF, which is the first identified risk factor in this study. SOFA score is a simple method for evaluating and monitoring organ dysfunction in patients. A retrospective study conducted by Kato and Matsuura found that the SOFA score did not decrease in patients with mild coagulopathy but was significantly decreased in patients with severe coagulopathy [28]. The higher the score, the more severe the organ dysfunction is. Thus, the SOFA index is a risk factor for TGC-associated HF that is easily understood. PCT level is a commonly used indicator of infection in clinical practice. Under normal physiological conditions, the serum PCT levels are extremely low. However, elevated PCT levels indicate severe infection [29]. We speculate that severe infection may cause microvascular damage, subsequently triggering the release of inflammatory neurotransmitters and cytokines, thereby exacerbating coagulation system perturbations.

Among the five protective factors we found, the baseline fibrinogen level was consistent with numerous previous studies [13, 30]. Fib is an acute-phase protein and a marker of inflammation, which increases significantly when the body is inflamed [31], therefore it has a protective effect on TGC-associated HF. The main infection sites were the lung and abdominal cavity; however, intra-abdominal infection was an independent risk factor for TGC-associated HF [14]. However, our results were contrary; although infection also occurs mainly in the lungs and abdomen, abdominal infection is a protective factor. Liu et al. (2021) found that TGC-associated HF was not related to the site of infection [13]. We also found that combinations of anticoagulant drugs, antiplatelet drugs, nonsteroidal anti-inflammatory drugs, β-lactam antibiotics, and sepsis were protective factors. These inconsistencies may be due to differences in patient characteristics, operator self-awareness, and the lack of statistical samples. It is essential to acknowledge that our study sheds light on novel risk factors while also highlighting inconsistencies that warrant further investigation.

Independent influencing factors were used to develop a predictive model for TGC-associated HF. We tried several ways to build the model and finally chose the XGBoost machine-learning algorithm. The ROC and calibration curves proved that the model had good discrimination and calibration to identify individuals with a high risk of TGC-associated HF. Furthermore, DCA and CICA underscored the superiority of our model through net clinical benefit, which is a crucial asset for personalized evaluation. Our model represents a pioneering effort to predict TGC-associated HF. In the present study, both continuous and categorical variables were used to build a prediction model. However, the reliability of categorical variables may decrease because of the operator's self-awareness, potentially impacting the model's integrity. Therefore, we will consider adopting continuous variables and expanding the sample size to optimize the model further.

For the first time, we constructed an RSF model to analyze the timing of TGC-associated HF. Our results showed that baseline fibrinogen level and PCT not only served as influencing factors of TGC-associated HF but also played pivotal roles in predicting the timing of TGC-associated HF, as indicated by the high importance scores in the RSF model. In contrast, TBIL and ALB levels were not identified as risk factors for TGC-associated HF. However, they exhibited potential relevance in the timing of TGC-associated HF. When the concentration of TBIL, a crucial indicator of hepatic function, increases significantly, it may indicate a decline in liver transformation function [32]. ALB synthesized by the liver has a normal range of 40–55 g/L. Clinically, high ALB levels are mainly associated with increased blood concentrations, while low ALB levels are more commonly observed in malnutrition and liver function damage [33]. Among the 222 patients included in this study, the ALB level of 215

patients was < 40 g/L, which may be due to poor liver synthesis. Both TBIL and ALB levels suggest that there may be a relationship between the timing of TGC-associated HF and liver function. However, we did not find an association between TGC-associated HF and liver function. Previous studies have also debated whether liver function affects TGC-associated HF [13, 34]. We speculate that this may be due to the clinical application of TGC in accordance with the drug manufacturers' instructions to reduce the dose of severe liver function damage. Moreover, because Fib is synthesized by the liver, we recommend vigilan monitoring of Fib levels in patients with low liver function, although conclusive evidence of the connection between the liver and TGC-associated HF remains elusive. In addition, the RSF model showed excellent performance in patient stratification. Physicians guided by the RSF risk stratification can evaluate TGC-associated HF timing and pay more attention to high-risk individuals. Based on the above results, our models have the potential to help clinicians make better clinical decisions regarding TGC treatment.

This study has several limitations that should be mentioned objectively. First, this was a single-center retrospective cohort study, the sample size was not large enough, and the factors influencing screening may not be accurate and comprehensive. Second, potential selection and measurement biases were inevitable. In addition, it is pertinent to note that our dataset features certain instances of missing values. While we mitigated this concern by excluding factors with more than 10% missing values and applying multiple imputations for those with less than 10%, the possibility of residual impact remains. Finally, the model was not validated using external data. These inherent limitations collectively contribute to the potential attenuation of the accuracy of the prediction models.

Conclusion

In our study, as machine learning algorithms, the XGBoost and RSF models provided accurate predictions for TGC-associated HF and the timing of TGC-associated HF as well as remarkable risk stratification of patients. This study introduced the XGBoost and RSF models for the monitoring of patients following TGC combination therapy for the first time. These two models have clinical practical value and are worthy of further study, as they will contribute to reducing the risk of TGC therapy.

Abbreviations

MDR	multidrug-resistant
XDR	extensively drug-resistant
TGC	Tigecycline
FDA	Food and Drug Administration
HF	hypofibrinogenemia
Fib	fibrinogen
XGBoost	extreme gradient boosting
OR	odds ratios
CI	confidence intervals
ROC	receiver-operating characteristic
DCA	decision curve analysis
CICA	clinical impact curve analysis
RSF	random survival forest
PT	prothrombin time
aPTT	activated partial thromboplastin time
C-index	consistency index

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-024-02694-x.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We wish to acknowledge the support of Wenmin Wang, Ning Deng, and Hao Sun (Biological Medicine Research and Development Center, Yangtze Delta of Zhejiang, Hangzhou 314006, Zhejiang, China). Wenmin Wang and Ning Deng participated in the discussion during the writing process of the article and provided valuable insights and suggestions for revision. Hao Sun offered technical consultancy services during the process of constructing the machine-learning models.

Author contributions

JZ was responsible for the concepts, design, and data collection. YG and RZ were responsible for statistical analysis. ZY, LL, and JW were responsible for data organization. This manuscript was prepared by JZ. All the authors have reviewed and made critical revisions to the manuscript.

Funding

This work was funded by the Zhejiang Pharmaceutical Society Hospital Pharmacy Special Research Grant Project (grant number: 2022ZYY03).

Data availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with relevant laws and institutional guidelines and approved by the Ethics Committee of Sir Run Run Shaw Hospital (No. 2023 – 0518), affiliated with Zhejiang University School of Medicine.

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Received: 22 November 2023 / Accepted: 25 September 2024
Published online: 04 October 2024

References

1. Magiorakos AP, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect. 2012;18(3):268–81.
2. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. P t. 2015;40(4):277–83.

3. Rusu A, Buta EL. The Development of Third-Generation Tetracycline antibiotics and New perspectives. Pharmaceutics 2021; 13(12).
4. Yaghoubi S, Zekiy AO, Krutova M, et al. Tigecycline antibacterial activity, clinical effectiveness, and mechanisms and epidemiology of resistance: narrative review. Eur J Clin Microbiol Infect Dis. 2022;41(7):1003–22.
5. Dorn C, Petroff D, Kratzer A, et al. Tigecycline Soft Tissue Penetration in obese and non-obese Surgical patients determined by using in Vivo Microdialysis. Eur J Drug Metab Pharmacokinet. 2022;47(5):749–55.
6. Song Y, Hu L, Shu Q, et al. Tigecycline salvage therapy for critically ill children with multidrug-resistant/extensively drug-resistant infections after surgery. Int J Infect Dis. 2018;75:82–8.
7. Kadoyama K, Sakaeda T, Tamon A, et al. Adverse event profile of tigecycline: data mining of the public version of the U.S. Food and Drug Administration adverse event reporting system. Biol Pharm Bull. 2012;35(6):967–70.
8. Xie W, Ma K, Xu Z, et al. Risk factors of tigecycline-associated fibrinogen reduction in patients with renal transplantation: a case-control study. Transl Androl Urol. 2022;11(10):1410–8.
9. Fan Q, Huang W, Weng Y, et al. Hypofibrinogenemia induced by high-dose tigecycline-case report and review of literature. Med (Baltim). 2020;99(43):e22638.
10. Wang D, Lin C, Gu C, et al. Tigecycline-Associated Coagulopathy: a single-Center Retrospective Analysis. Pharmacology. 2022;107(9–10):524–36.
11. Zhang Q, Wang J, Liu H, et al. Risk factors for tigecycline-induced hypofibrinogenaemia. J Clin Pharm Ther. 2020;45(6):1434–41.
12. Kryczka KE, Kruk M, Demkow M et al. Fibrinogen and a Triad of Thrombosis, Inflammation, and the Renin-Angiotensin System in Premature Coronary Artery Disease in Women: A New Insight into Sex-Related Differences in the Pathogenesis of the Disease. Biomolecules. 2021; 11(7).
13. Liu J, Yan Y, Zhang F. Risk factors for Tigecycline-Associated Hypofibrinogenemia. Ther Clin Risk Manag. 2021;17:325–32.
14. Hu J, Xiao YH, Zheng Y, et al. Clinical characteristics and risk factors of tigecycline-associated hypofibrinogenaemia in critically ill patients. Eur J Clin Pharmacol. 2020;76(7):913–22.
15. Wu X, Zhao P, Dong L, et al. A case report of patient with severe acute cholangitis with tigecycline treatment causing coagulopathy and hypofibrinogenemia. Med (Baltim). 2017;96(49):e9124.
16. Zhang Q, Zhou S, Zhou J. Tigecycline treatment causes a decrease in fibrinogen levels. Antimicrob Agents Chemother. 2015;59(3):1650–5.
17. Frazier, PIJapa. A tutorial on Bayesian optimization. 2018.
18. Ma WJJN. Bayesian decision models: A primer. 2019; 104(1): pp. 164–175.
19. Gao L, Ding, YJBrn. Disease prediction via bayesian hyperparameter optimization and ensemble learning. 2020; 13: pp. 1–6.
20. Guirao X, Sánchez García M, Bassetti M, et al. Safety and tolerability of tigecycline for the treatment of complicated skin and soft-tissue and intra-abdominal infections: an analysis based on five European observational studies. J Antimicrob Chemother. 2013;68(Suppl 2):ii37–44.
21. Guo M, Liang J, Li D, et al. Coagulation dysfunction events associated with tigecycline: a real-world study from FDA adverse event reporting system (FAERS) database. Thromb J. 2022;20(1):12.
22. Pieringer H, Schmekal B, Biesenbach G, et al. Severe coagulation disorder with hypofibrinogenemia associated with the use of tigecycline. Ann Hematol. 2010;89(10):1063–4.
23. Zhang Q, Zhou J. Fibrinogenopenia caused by Tigecycline: a case report. Eur Rev Med Pharmacol Sci. 2015;19(6):915–7.
24. Wu PC, Wu CC. Tigecycline-associated hypofibrinogenemia: a case report and review of the literature. IDCases. 2018;11:56–7.
25. Sabanis N, Paschou E, Gavriilaki E, et al. Hypofibrinogenemia induced by tigecycline: a potentially life-threatening coagulation disorder. Infect Dis. 2015;47(10):743–6.
26. Bialkower M, Garnier G. Fibrinogen Diagnostics in Major Hemorrhage. Crit Rev Anal Chem. 2022;52(1):194–209.
27. Zhu L, Wang L, Zhang Y, et al. Fatal hemorrhagic pneumonia in patients with hematologic diseases and Stenotrophomonas maltophilia bacteremia: a retrospective study. BMC Infect Dis. 2021;21(1):723.
28. Kato T, Matsuura K. Recombinant human soluble thrombomodulin improves mortality in patients with sepsis especially for severe coagulopathy: a retrospective study. Thromb J. 2018;16:19.
29. Spoto S, Daniel Markley J, Valeriani E, et al. Active surveillance cultures and procalcitonin in Combination with Clinical data to guide empirical antimicrobial therapy in Hospitalized Medical patients with Sepsis. Front Microbiol. 2022;13:797932.
30. Li Z, Zeng Q, Xu S, et al. Development and validation of a Nomogram for Predicting Tigecycline-related Coagulopathy: a retrospective cohort study. Infect Drug Resist. 2023;16:423–34.
31. Sulimai N, Lominadze D. Fibrinogen and Neuroinflammation during Traumatic Brain Injury. Mol Neurobiol. 2020;57(11):4692–703.
32. Uemura S, Higuchi R, Yazawa T, et al. Level of total bilirubin in the bile of the future remnant liver of patients with obstructive jaundice undergoing hepatectomy predicts postoperative liver failure. J Hepatobiliary Pancreat Sci. 2020;27(9):614–21.
33. Asif A, Park SH, Soomro AM, et al. Microphysiological system with continuous analysis of albumin for hepatotoxicity modeling and drug screening. J Ind Eng Chem. 2021;98:318–26.
34. Huang YT, Yu CI, Chen PY, et al. Comparison of bleeding risk between colistin-tigecycline and colistin-carbapenem treatment regimens: a retrospective cohort study. Infect Drug Resist. 2021;14:4949–55.

Publisher's note