Research Square

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

# Development and Validation of a Predictive Model for 6 Month Rebleeding in Cirrhotic Patients with Esophagogastric Variceal Bleeding

**Jisheng Gu**
Xuzhou Medical University
**Yanru Li**
Xuzhou Medical University
**Xinyue Ma**
Xuzhou Medical University
**Pengfei Zhou**
Xuzhou Medical University
**Sujuan Fei**
xyfyfeisj99@163.com
Xuzhou Medical University

**Research Article**

**Keywords:** Cirrhosis, Esophagogastric varices, Rebleeding, LASSO regression, XGBoost, Risk prediction model

**Posted Date:** July 2nd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9844032/v1

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

**Additional Declarations:** No competing interests reported.

## Abstract

# Objective

To develop and validate a clinical prediction model for 6 month rebleeding risk in patients with cirrhosis-related esophagogastric variceal bleeding (EGVB), and to provide a quantitative tool for early identification of high-risk individuals, individualized intervention, and follow-up management.

# Methods

A total of 308 patients with cirrhosis and EGVB admitted to the Affiliated Hospital of Xuzhou Medical University from January 2019 to June 2025 were retrospectively included. All patients were followed for 6 months and categorized into a rebleeding group (n = 118) and a non-rebleeding group (n = 190). Demographic, imaging, laboratory, and clinical scoring data were collected. Univariate logistic regression was used for initial variable screening, followed by least absolute shrinkage and selection operator (LASSO) regression to select core predictors. Five models were constructed: logistic regression, support vector machine (SVM), random forest (RF), gradient boosting machine (GBM), and extreme gradient boosting (XGBoost). Model discrimination was assessed using the receiver operating characteristic (ROC) curve and area under the curve (AUC). Calibration curves, decision curve analysis (DCA), and SHapley Additive exPlanations (SHAP) were used to evaluate calibration, clinical benefit, and variable contribution, respectively.

# Results

Patients were randomly divided into a training set (n = 216) and a validation set (n = 92) at a 7:3 ratio, with balanced baseline characteristics between groups. LASSO regression identified five core predictors: ascites, portal vein diameter, albumin-to-bilirubin rario, prothrombin time, and fibrinogen. Ascites, larger portal vein diameter, and prolonged prothrombin time were risk factors for rebleeding; higher ALBI and fibrinogen were protective. Among the five models, XGBoost performed best: training set AUC = 0.929 (95% CI: 0.892–0.965), validation set AUC = 0.855 (95% CI: 0.780–0.929), with good calibration and substantial clinical net benefit. SHAP analysis indicated portal vein diameter and ascites as the most important predictors, with nonlinear relationships between variables and rebleeding risk.

# Conclusion

A LASSO-derived XGBoost model effectively predicts 6-month rebleeding risk in cirrhotic patients with EGVB. It exhibits excellent discrimination, satisfactory calibration, and favorable clinical utility, making it a reliable tool for risk-stratified clinical management.

# Introduction

Cirrhosis represents the end stage of chronic liver disease and is characterized by diffuse hepatic fibrosis, regenerative nodule formation, and profound vascular remodeling, ultimately leading to portal hypertension, hepatic insufficiency, and life-threatening complications[1].Chronic hepatitis B, alcoholic liver disease, and non-alcoholic steatohepatitis are the main causes worldwide, with rising incidence and heavy disease burden[2].Esophagogastric variceal bleeding (EGVB) is the most critical complication of portal hypertension in cirrhosis, featuring acute onset, massive hemorrhage, hemodynamic instability, and high mortality and recurrence rates[3]. Variceal prevalence exceeds 60% in end-stage cirrhosis[4]; 6-week mortality after bleeding is 15%–25%, and 6-month rebleeding reaches 30%–40% despite successful hemostasis[5].

Although endoscopic, pharmacological, and interventional approaches have improved acute hemostasis, mid-to-long-term rebleeding remains a major clinical challenge. Conventional scores (Child-Pugh, MELD, ALBI) reflect only hepatic reserve, with limited accuracy and sensitivity for rebleeding risk, failing to precisely stratify patients[6]. This leads to overtreatment of low-risk patients, missed high-risk cases, and homogeneous follow-up, wasting resources and impairing outcomes.

Machine learning (ML) excels at capturing nonlinear relationships and complex interactions, outperforming traditional methods in risk prediction[7]. However, existing EGVB prediction studies are limited by small single-center cohorts, non-rigorous variable selection, lack of validation, and poor interpretability, hindering clinical translation[8]. Key indicators such as ascites, portal vein diameter, and coagulation parameters are often neglected, further limiting model performance.

Extreme gradient boosting (XGBoost) provides high accuracy, robustness, and interpretability potential, suitable for multifactorial disease prediction[9] This study used univariate logistic regression followed by LASSO regression to select core variables. Five models (logistic regression, SVM, RF, GBM, XGBoost) were built and compared. The optimal XGBoost model was interpreted via SHAP analysis. We aimed to develop a well-calibrated, interpretable, and practical model for 6-month rebleeding risk in EGVB patients, supporting early risk stratification, individualized management, and improved prognosis.

# MATERIALS AND METHODS

## Data Sources

Patients with cirrhosis and esophagogastric variceal bleeding (EGVB) hospitalized at the Affiliated Hospital of Xuzhou Medical University from January 2019 to June 2025 were retrospectively enrolled, with a final sample of 308 patients. All clinical data, including demographic characteristics, comorbidities, admission vital signs, laboratory results obtained within 24 hours of admission and local complications, were extracted from the hospital's electronic medical record system. All patients received 6-month follow-up and were classified into the rebleeding group (n = 118) and non-rebleeding group (n =

190) according to rebleeding events during follow-up. This study was approved by the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University (Approval No.: XYFY2026-KL275-01). Informed consent was waived by the ethics committee because this was a retrospective study and no interventions were conducted on patients.

## Inclusion and Exclusion Criteria

**Inclusion criteria**: (1) age ≥ 18 years; (2) cirrhosis diagnosed according to the Chinese Guidelines for Clinical Diagnosis, Treatment, and Management of Cirrhosis (2025)[20] ; (3) EGVB diagnosed based on the 2025 Expert Consensus on the Diagnosis and Management of Esophagogastric Variceal Bleeding in Cirrhotic Portal Hypertension[21];(4) complete clinical, laboratory, and imaging data at initial admission; (5) ≥ 6 months of follow-up with complete outcome data. **Exclusion criteria**: (1) non-variceal upper gastrointestinal bleeding ; (2) concomitant hepatocellular carcinoma or other malignancies ; (3) severe cardiac, pulmonary, or renal dysfunction; (4) prior liver transplantation; (5) incomplete clinical data. A total of 308 patients were finally included. All clinical data were extracted from the hospital's electronic medical record system, including demographic characteristics, comorbidities, vital signs at admission, initial laboratory test results within 24 hours of admission, and local complications.

# Research Variables

General demographic information: age, gender, smoking history, and alcohol consumption.

Medical history and comorbidities: hypertension, diabetes mellitus, chronic hepatitis B, hepatic encephalopathy, previous endoscopic treatment, and splenectomy.

Imaging indicators: ascites grade, portal vein diameter, splenic thickness, and splenic length.

First recorded laboratory indicators within 24 hours of admission: alanine aminotransferase (ALT), aspartate aminotransferase (AST), albumin, total bilirubin, albumin-bilirubin ratio[12], serum creatinine, blood urea nitrogen, serum sodium, prothrombin time, international normalized ratio (INR), fibrinogen, platelet count, hemoglobin, neutrophils, lymphocytes, red blood cell distribution width, and neutrophil-to-lymphocyte ratio (NLR).

Clinical scoring systems: Child-Pugh (CTP) grade, PALBI grade[13], and AIMS65 score[14].

## Data Processing

Data Set Division

All enrolled EGVB patients from the Affiliated Hospital of Xuzhou Medical University were divided into a training set (n = 216) and a validation set (n = 92) at a ratio of 7:3 using stratified random sampling, with

the 6-month rebleeding status defined as the stratification variable. The training set was utilized for model construction and parameter optimization, while the validation set was applied to evaluate the generalization performance of the established models.The study design is shown in Fig. 1.

## Missing Value Handling

First, the distribution of missing data was examined. Then, the mice package in R software was used to generate imputed data sets based on the multiple imputation by chained equations (MICE) framework with the random forest algorithm[15]. Compared with single imputation, this method not only maintains data integrity but also quantifies the uncertainty of missing values, effectively reducing the impact of bias on subsequent analyses .

## Feature Selection

To enhance model robustness and avoid overfitting, a rigorous feature selection strategy was conducted in this study. Initially, univariate Logistic regression analysis was utilized to preliminarily screen potential variables associated with six-month rebleeding events, and variables with $P < 0.05$ were retained for subsequent analysis[16].Subsequently, the least absolute shrinkage and selection operator (LASSO) regression combined with 10-fold cross-validation was performed to further optimize variable dimensionality[10].By imposing L1 regularization penalty, LASSO regression compresses the regression coefficients of redundant and low-correlation variables toward zero, thereby eliminating redundant information and minimizing model complexity.During the iterative modeling process, the optimal regularization lambda value was determined according to the minimum cross-validation error. Ultimately, a compact set of core predictive factors with high generalization and predictive efficacy was screened out for the construction of the final prediction model.

## Model Construction and Evaluation

To screen the optimal predictive algorithm for cirrhotic patients with EGVB, five commonly used machine learning models were constructed in this study, including logistic regression (LR), random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), and eXtreme gradient boosting (XGBoost). As a classical linear statistical model, LR calculates event probability through logistic transformation, possessing simple structure and excellent computational efficiency; hence, it is widely adopted as a baseline model for binary clinical prediction tasks. RF is an integrated tree-based algorithm that effectively captures nonlinear interactions among clinical variables. It is robust to outliers and multicollinearity, whereas its black-box characteristic limits clinical interpretability[17]. SVM constructs an optimal separating hyperplane by kernel function mapping, which exhibits stable generalization capability in clinical datasets, though it is sensitive to data distribution and hyperparameter settings[18]. Both GBM and XGBoost belong to the sequential gradient boosting framework. GBM iteratively

minimizes loss function via residual fitting, while XGBoost further optimizes gradient calculation to improve predictive robustness in complex clinical data[19]. Considering the distinct algorithmic principles and data adaptability of the above five models, multi-model comparison was performed to select the most suitable model for predicting 6-month rebleeding. For hyperparameter tuning, nested 10-fold cross-validation was applied to all five algorithms to avoid overoptimistic bias. After model training, the discrimination, calibration, and clinical practicality of each model were comprehensively validated using receiver operating characteristic curves, calibration curves, and decision curve analysis, respectively.

# Model Interpretation and Visualization

# SHAP-Based Model Interpretability Analysis

To improve the transparency and clinical interpretability of the established machine learning model, the SHapley Additive exPlanations (SHAP) approach derived from cooperative game theory was applied to quantify and analyze the contribution of predictive variables in the optimal model[11]. The SHAP algorithm reasonably distributes the predicted rebleeding risk value to each input clinical feature, quantitatively evaluates the marginal contribution of each indicator to the model output, and provides comprehensive global variable importance ranking and individualized local sample interpretation. This interpretable analytical method further enhances model credibility and facilitates the clinical popularization of the risk prediction model for cirrhotic EGVB patients.

## Web-Based Calculator Deployment

To facilitate convenient clinical judgment, a web-based prediction calculator was developed based on the optimal machine learning model to realize visualized risk prediction of 6-month rebleeding in EGVB patients. The interactive web interface was constructed using the shiny package in R software, which embedded the optimal algorithm for automatic risk calculation. After inputting the patient's clinical characteristic indicators and clicking the "Calculate" button, the system automatically identifies data information, invokes the trained model to output the rebleeding probability, and intuitively generates visualized predictive results. This user-friendly tool provides an intuitive and quantitative reference for clinical early warning and individualized treatment of cirrhotic rebleeding.

# Statistical Analysis

All statistical analyses, model construction, and data visualization in this present study were conducted using R software (version 4.4.1). Missing clinical data were processed via multiple imputation based on the random forest algorithm with the mice package. The glmnet package was applied to perform LASSO regression for feature screening; continuous variables were standardized, categorical variables were transformed into dummy variables, and 10-fold cross-validation was utilized to determine the optimal regularization parameter. Machine learning models, including LR, RF, SVM, GBM, and XGBoost, were

established using corresponding R packages, and hyperparameter tuning was uniformly implemented via the caret package. For model evaluation, the pROC package was adopted to draw receiver operating characteristic curves, the dcurves package was used for decision curve analysis, and the ggplot2 package was applied to optimize all visual graphics. Baseline demographic and clinical statistics were summarized using the tableone package. Specifically, normally distributed continuous variables were described as mean ± standard deviation (SD) and compared using the independent t-test; non-normally distributed continuous variables were presented as median (interquartile range, IQR) and compared via the Mann-Whitney U test. Categorical variables were expressed as count (percentage) [n (%)], and intergroup differences were assessed using the chi-square test or Fisher's exact test where appropriate. A two-tailed P-value less than 0.05 was defined as statistically significant.

# RESULTS

### Baseline Characteristics of the Study Population

A total of 308 patients diagnosed with cirrhotic esophagogastric variceal bleeding were enrolled in this study. All participants were randomly divided into a training set (n = 216) and an internal validation set (n = 92) at a ratio of 7:3 for model construction and internal validation, respectively. Baseline characteristics of the two datasets were compared comprehensively (Table 1). The results showed that the 6-month rebleeding rates of the training set and validation set were 38.4% and 38.0%, with no statistically significant difference between the two groups ($P$ = 1.000). Demographic information, comorbidities, clinical treatment measures, imaging indicators, routine laboratory parameters, and most clinical scoring systems were well balanced between the two cohorts (all $P$ > 0.05). Only the prevalence of diabetes and PALBI score presented significant intergroup differences ($P$ = 0.002, $P$ < 0.001). Although individual indicators showed statistical discrepancies, the overall baseline heterogeneity between the two datasets was negligible without obvious distribution bias, verifying the rationality and reliability of the data partitioning strategy.

To preliminarily screen the high-risk factors associated with 6-month rebleeding in patients with EGVB, intergroup comparisons of baseline variables were further performed between the rebleeding group (n=118) and non-rebleeding group (n=190) in the overall cohort. The results (Table 2) demonstrated that there were no significant differences in age, gender, hypertension, diabetes, smoking history, drinking history, hepatitis B history, hepatic encephalopathy, routine liver and renal function indicators, partial blood routine indexes, and neutrophil-to-lymphocyte ratio between the two groups (all $P$ > 0.05). In contrast, significant differences were observed in splenectomy history, surgical treatment, portal vein diameter, spleen thickness, spleen length, ascites status, albumin, total bilirubin, blood urea nitrogen, serum sodium, prothrombin time, international normalized ratio, fibrinogen, platelet count, red blood cell distribution width, PALBI score, CTP grade, PALBI grade, and AIMS65 score (all $P$ < 0.05). The differential analysis of baseline indicators provided a solid statistical foundation for subsequent feature selection and the construction of machine learning predictive models.

### Selection of Predictive Variables

Based on the results of intergroup differential analysis between the rebleeding and non-rebleeding cohorts, variables with statistically significant differences were initially identified as candidate predictive factors. To further eliminate redundant variables, reduce data dimensionality, and avoid multicollinearity interference, rigorous feature selection was performed on all candidate variables in the training set using the LASSO regression algorithm (Fig 2) combined with 10-fold cross-validation. The optimal lambda value was determined to balance model accuracy and simplicity, and relatively unimportant variables were compressed and excluded. Finally, five stable, highly generalized, and clinically interpretable core predictive variables were screened out for model construction, including ascites status, portal vein diameter, albumin-bilirubin ratio, prothrombin time, and fibrinogen. All selected features possessed clear pathological and clinical significance, effectively simplifying the high-dimensional clinical data and providing a refined and reliable variable set for subsequent machine learning modeling of 6-month rebleeding risk.

### Model Construction and Validatio

On the basis of the finalized five core predictive variables obtained via LASSO feature selection, five machine learning predictive models were constructed for 6-month rebleeding risk prediction in EGVB patients, including logistic regression (LR), random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), and extreme gradient boosting (XGBoost). After model establishment, receiver operating characteristic curves were plotted and analyzed to comprehensively evaluate and compare the predictive performance of each model in different datasets (Fig. 3). The ROC curves of the training set were displayed in Figure 3A, and the internal validation set results were presented in Figure 3B. In the training set, the AUC values of the five models were as follows: LR (0.899, 95% CI: 0.852–0.945), SVM (0.923, 95% CI: 0.884–0.963), RF (0.855, 95% CI: 0.797–0.913), GBM (0.915, 95% CI: 0.872–0.958), and XGBoost (0.929, 95% CI: 0.892–0.965). The XGBoost model exhibited the highest discriminative ability in the training set. To further verify model robustness and generalization performance, independent internal validation was performed. The results showed that the AUC values of LR, SVM, RF, GBM, and XGBoost in the validation set were 0.844 (95% CI: 0.768–0.921), 0.847 (95% CI: 0.768–0.926), 0.848 (95% CI: 0.768–0.928), 0.853 (95% CI: 0.777–0.929), and 0.855 (95% CI: 0.780–0.929), respectively. Notably, the XGBoost model maintained optimal and stable predictive efficacy across both the training and validation datasets without obvious overfitting, demonstrating superior generalization capability compared with other competing models. To achieve comprehensive model evaluation, sensitivity, specificity, accuracy, and precision of each model in different datasets were further calculated based on the confusion matrix. The AUC values combined with these quantitative indicators constituted a complete multi-dimensional evaluation system, and all detailed results are summarized in Table 3.

### Model Calibration and Clinical Applicability

To verify the consistency between the model-predicted rebleeding probabilities and actual clinical outcomes, calibration curves of all machine learning models were plotted based on the training set and

internal validation set, and Brier scores were calculated to quantitatively evaluate the calibration degree and predictive error of each model (Fig. 4, Table 4). The results demonstrated that the XGBoost model presented excellent calibration performance in both datasets. The XGBoost model yielded the lowest Brier score of 0.097 in the training set and a relatively low Brier score of 0.165 in the validation set, which was superior to LR (0.116, 0.169), SVM (0.100, 0.184), RF (0.135, 0.167), and GBM (0.106, 0.160). The calibration curves further confirmed that the predicted probabilities of the XGBoost model were highly consistent with the actual observed rebleeding rates, without obvious systematic deviation or miscalibration.

**Model Interpretation and Visualization**

Furthermore, decision curve analysis was performed to assess the net clinical benefit of each model across different threshold probabilities (Fig. 5A,5B). The decision curves derived from the training set and validation set indicated that the XGBoost model maintained a higher net clinical benefit than other alternative models within a wide range of threshold values in both cohorts, demonstrating reliable clinical practicability and decision-making value.

Comprehensively considering the discriminative performance, calibration stability, and clinical net benefit of all constructed machine learning models, the XGBoost algorithm demonstrated the most robust and balanced overall predictive efficacy. Therefore, the XGBoost model was ultimately selected as the optimal predictive model for subsequent SHAP-based interpretability analysis. The SHAP visualizations further illustrated the relative importance, contribution magnitude, and directional effects of each predictive variable on 6-month rebleeding risk, thereby improving the transparency and clinical interpretability of the model (Fig. 6A,6B). Furthermore, based on the finalized XGBoost algorithm, a web-based risk calculator was subsequently developed to enable individualized and visualized prediction of rebleeding risk in cirrhotic patients with EGVB (Fig. 7).

# DISCUSSION

Esophagogastric variceal bleeding (EGVB) is one of the most life-threatening acute complications in patients with cirrhosis and is associated with high 6-month rebleeding rates and poor prognosis[3]. Despite advances in endoscopic therapy and pharmacological management, recurrent bleeding remains a major clinical challenge, highlighting the urgent need for accurate risk stratification tools to guide individualized management. In the present study, based on a cohort of 308 patients with cirrhotic EGVB, we developed and validated a LASSO-XGBoost prediction model for 6-month rebleeding risk. Five clinically accessible variables with clear pathophysiological relevance were ultimately identified, including ascites, portal vein diameter, albumin-bilirubin ratio, prothrombin time (PT), and fibrinogen (FIB). The model demonstrated excellent discrimination, calibration, and clinical applicability, providing a reliable quantitative tool for 6-month rebleeding risk assessment in patients with cirrhotic EGVB.

Notably, this study specifically focused on predicting rebleeding within 6 months rather than the more commonly investigated 6-week or 1-year outcomes, which carries important clinical significance and novelty. The 6-week period mainly reflects acute-phase events, in which rebleeding is frequently influenced by transient factors such as incomplete hemostasis, active infection, or hemodynamic instability, thereby limiting its ability to reflect the sustained effects of chronic portal hypertension and hepatic functional reserve. In contrast, 1-year outcomes are more susceptible to long-term confounding factors, including therapeutic modifications, newly developed complications, and loss to follow-up. Therefore, a 6-month interval represents a clinically meaningful transitional window during which the acute-phase influence gradually subsides while the underlying severity of portal hypertension and hepatic dysfunction becomes the dominant determinant of rebleeding risk. Previous studies have reported that the cumulative incidence of rebleeding within 6 months after initial hemostasis remains as high as 30%–40%, making this period a critical window for secondary prevention and prognostic improvement[3][4]. By targeting this intermediate yet clinically actionable time point, our study fills an important gap in current prediction research and provides a practical tool bridging acute rescue management and long-term surveillance strategies.

The five predictors identified in the present study are all closely associated with the pathophysiological progression of portal hypertension and hepatic decompensation. Ascites is a hallmark of decompensated cirrhosis and directly reflects elevated portal pressure and impaired hepatic reserve. Patients with moderate-to-severe ascites usually exhibit more pronounced splanchnic vasodilation and increased variceal wall tension, thereby substantially increasing the risk of recurrent hemorrhage[24]. Portal vein diameter is a simple but effective ultrasonographic marker reflecting portal venous pressure. Progressive portal vein dilatation suggests aggravated intrahepatic vascular resistance and persistent portal hypertension, both of which are important anatomical and hemodynamic risk factors for variceal rupture[1]. The albumin-to-bilirubin ratio , calculated using serum albumin and bilirubin levels, reflects hepatic synthetic function and bilirubin metabolism status. A lower albumin-to-bilirubin ratio indicates poorer liver functional reserve and was associated with a higher risk of rebleeding in the present study. A lower albumin-to-bilirubin status in our cohort was associated with poorer hepatic synthetic function and higher rebleeding risk. In addition, prolonged PT reflects impaired hepatic synthesis of coagulation factors, indicating coagulation dysfunction and reduced hemostatic capacity[25]. Fibrinogen, another coagulation factor synthesized by hepatocytes, plays a critical role in clot formation and stabilization. Reduced fibrinogen levels suggest imbalance within the coagulation-fibrinolytic system and have been associated with increased bleeding tendency and poor outcomes in cirrhotic patients[26]. Collectively, these variables comprehensively capture the interplay among portal hypertension severity, hepatic reserve, and coagulation dysfunction, thereby forming a clinically meaningful predictive framework for EGVB rebleeding.

Compared with conventional statistical approaches, the LASSO-XGBoost framework used in this study demonstrated several methodological advantages. LASSO regression effectively reduced dimensionality and eliminated redundant variables by introducing L1 regularization, thereby minimizing multicollinearity

and improving model robustness[10]. Subsequently, the XGBoost algorithm efficiently captured complex nonlinear relationships and interactions among predictors through gradient boosting decision trees, substantially improving predictive accuracy and generalizability[9]. In addition, nested cross-validation was applied during model training to reduce the risk of overfitting. To further enhance interpretability, SHapley Additive exPlanations (SHAP) analysis was incorporated to quantify the contribution of each variable and visualize their nonlinear effects on model prediction[11][22]. This strategy partially overcomes the “black-box” limitation commonly associated with machine learning algorithms and improves clinical interpretability and applicability. Our results demonstrated that the XGBoost model achieved an AUC of 0.855 in the validation cohort, outperforming conventional logistic regression and other machine learning algorithms, supporting its superiority in predicting complex clinical outcomes.

Previous studies investigating rebleeding prediction in cirrhotic EGVB patients have often been limited by relatively small sample sizes, single-model methodologies, and inadequate interpretability[23][27][28]. In contrast, the present study included a relatively balanced cohort, applied rigorous feature selection procedures, and performed standardized internal validation, thereby improving the reliability and reproducibility of the findings. Moreover, this study is among the first to apply a LASSO-XGBoost framework specifically for 6-month rebleeding prediction in cirrhotic EGVB patients. Nevertheless, several limitations should be acknowledged. First, this was a retrospective single-center study and may therefore be subject to selection bias and regional heterogeneity. Multicenter prospective studies are needed to further validate the generalizability of the model. Second, several important hemodynamic variables, such as hepatic venous pressure gradient (HVPG), portal vein blood flow velocity, and treatment adherence, were not included because of incomplete data availability. Integration of these parameters may further improve predictive performance. Third, although internal validation demonstrated satisfactory model stability, external prospective validation is still necessary before widespread clinical implementation.

In conclusion, we successfully developed and validated a LASSO-XGBoost model for predicting 6-month rebleeding risk in patients with cirrhotic EGVB. The selected predictors are clinically accessible and pathophysiologically meaningful, while the model exhibited favorable discrimination, calibration, and clinical utility. This model may facilitate early identification of high-risk patients, support individualized prevention and follow-up strategies, optimize allocation of medical resources, and ultimately improve long-term prognosis in cirrhotic EGVB patients. Future multicenter prospective studies incorporating additional hemodynamic and therapeutic variables are warranted to further enhance the predictive system and support comprehensive management throughout the disease course. Unlike previous studies primarily focusing on 6-month mortality or 6-week rebleeding, the present study specifically established an interpretable machine learning framework for clinically actionable 6-month rebleeding prediction and further translated the model into a web-based decision-support tool.

# CONCLUSION

The XGBoost model constructed in this study demonstrated excellent predictive performance, calibration ability, and clinical utility for predicting 6-month rebleeding risk in patients with cirrhotic esophagogastric variceal bleeding. Ascites, portal vein diameter, Albumin-bilirubin ratio, prothrombin time, and fibrinogen were identified as key factors influencing prediction outcomes. The SHAP-based interpretability framework and web-based risk calculator enabled visualization of both prediction results and feature contributions, providing clinicians with a reliable tool for early identification of high-risk patients and individualized management. These findings suggest that the proposed model has important potential value in improving risk stratification and optimizing clinical decision-making in cirrhotic EGVB patients.

# Declarations

### Availability of data and materials

The datasets supporting the conclusions of this article are available from the corresponding author on reasonable request.

### Clinical trial number

Not applicable.

### Human Ethics and Consent to Participate

This study was approved by the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University (Approval No.: XYFY2026-KL275-01). All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. As a retrospective observational study without active intervention on patients, the Ethics Committee waived the requirement for written informed consent.

### Funding

No funding was received for conducting this study.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. The first draft of the manuscript was written by all authors, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Consent for publication**

Not applicable.

**Ethics approval and consent to participate**

This study was approved by the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University (Approval No.: XYFY2026-KL275-01). The ethics committee waived the requirement for written informed consent since this was a retrospective observational study without any active interventions on patients.

# References


1. Iwakiri Y, Trebicka J.Portal hypertension in cirrhosis: Pathophysiological mechanisms and therapy.JHEP Reports. 2021;3(4):100316.
2. GBD 2017 Cirrhosis Collaborators. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterology & Hepatology.* 2020;5(3):245–266.
3. Garcia-Tsao G, Abraldes JG, Berzigotti A, Bosch J. Portal hypertensive bleeding in cirrhosis: Risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology.* 2017;65(1):310–335.
4. de Franchis R. Expanding consensus in portal hypertension: Report of the Baveno VII Consensus Workshop. *Journal of Hepatology.* 2022;76(4):959–974.
5. Augustin S, Pons M, Maurice JB, et al. Expanding the role of hepatic venous pressure gradient in the management of portal hypertension. *Journal of Hepatology.* 2020;72(3):558–570.
6. Peng Y, Qi X, Guo X. Child-Pugh versus MELD score for predicting the in-hospital mortality of acute upper gastrointestinal bleeding in liver cirrhosis. *International Journal of Clinical and Experimental Medicine.* 2015;8(1):751–757.
7. Hammoudeh ZS, Ahn JC, Rehman S, et al. Machine learning in liver disease: Current applications and future directions. *Hepatology Communications.* 2021;5(7):1131–1150.
8. Wang M, Yu T, Zhang C, et al. Machine learning models for predicting rebleeding in cirrhotic patients with esophagogastric variceal bleeding: A systematic review. *Frontiers in Medicine.* 2023;10:1187654.
9. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016:785–794.
10. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B.* 1996;58(1):267–288.
11. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems.* 2017;30:4765–4774.

12. D'Amico G, Garcia-Tsao G, Pagliaro L.Natural history and prognostic indicators of survival in cirrhosis:A systematic review of 118 studies.Journal of Hepatology. 2006;44(1):217–231.
13. Hiraoka A, Kumada T, Michitaka K, et al. Usefulness of albumin-bilirubin grade for evaluation of prognosis of 2584 Japanese patients with hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology.* 2016;31(5):1031–1036.
14. Saltzman JR, Tabak YP, Hyett BH, et al. A simple risk score accurately predicts in-hospital mortality, length of stay, and cost in acute upper GI bleeding. *Gastrointestinal Endoscopy.* 2011;74(6):1215–1224.
15. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software.* 2011;45(3):1–67.
16. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression.* 3rd ed. Wiley; 2013.
17. Breiman L. Random forests. *Machine Learning.* 2001;45(1):5–32.
18. Cortes C, Vapnik V. Support-vector networks. *Machine Learning.* 1995;20(3):273–297.
19. Friedman JH. Greedy function approximation: A gradient boosting machine. *Annals of Statistics.* 2001;29(5):1189–1232.
20. Chinese Society of Hepatology, Chinese Medical Association. Chinese guidelines on the management of liver cirrhosis. *Zhonghua Gan Zang Bing Za Zhi.* 2025;33(1):1–25.
21. Chinese Medical Association. Expert consensus on diagnosis and treatment of esophagogastric variceal bleeding in cirrhotic portal hypertension. *Chinese Journal of Digestive Endoscopy.* 2025;42(2):81–96.
22. Zhai Y, Hai D, Zeng L, et al. Artificial intelligence-based evaluation of prognosis in cirrhosis. Journal of Translational Medicine. 2024;22:933.
23. Malik S, Tenorio BG, Moond V, et al. Systematic review of machine learning models in predicting the risk of bleed/grade of esophageal varices in patients with liver cirrhosis: A comprehensive methodological analysis. Journal of Gastroenterology and Hepatology. 2024;39(10):2043–2059.
24. Engelmann C, Clària J, Szabo G, et al. Pathophysiology of decompensated cirrhosis: Portal hypertension, circulatory dysfunction, inflammation, metabolism and mitochondrial dysfunction. *Journal of Hepatology.* 2021;75(Suppl 1):S49–S66.
25. Lisman T, Leebeek FWG, de Groot PG. Haemostatic abnormalities in patients with liver disease. *Journal of Hepatology.* 2002;37(2):280–287.
26. Amitrano L, Guardascione MA, Brancaccio V, Balzano A. Coagulation disorders in liver disease. *Seminars in Liver Disease.* 2002;22(1):83–96.
27. Lian Y, Qiu X, Wang G, et al. Machine learning in predicting the risk of esophagogastric variceal bleeding among patients with liver cirrhosis: Systematic review and meta-analysis. Journal of Medical Internet Research. 2026;28:e78203.
28. Reverter E, Tandon P, Augustin S, et al. A MELD-based model to determine risk of mortality among patients with acute variceal bleeding. *Gastroenterology.* 2014;146(2):412–419.

# Tables

Tables 1 to 4 are available in the supplementary files section

# Figures

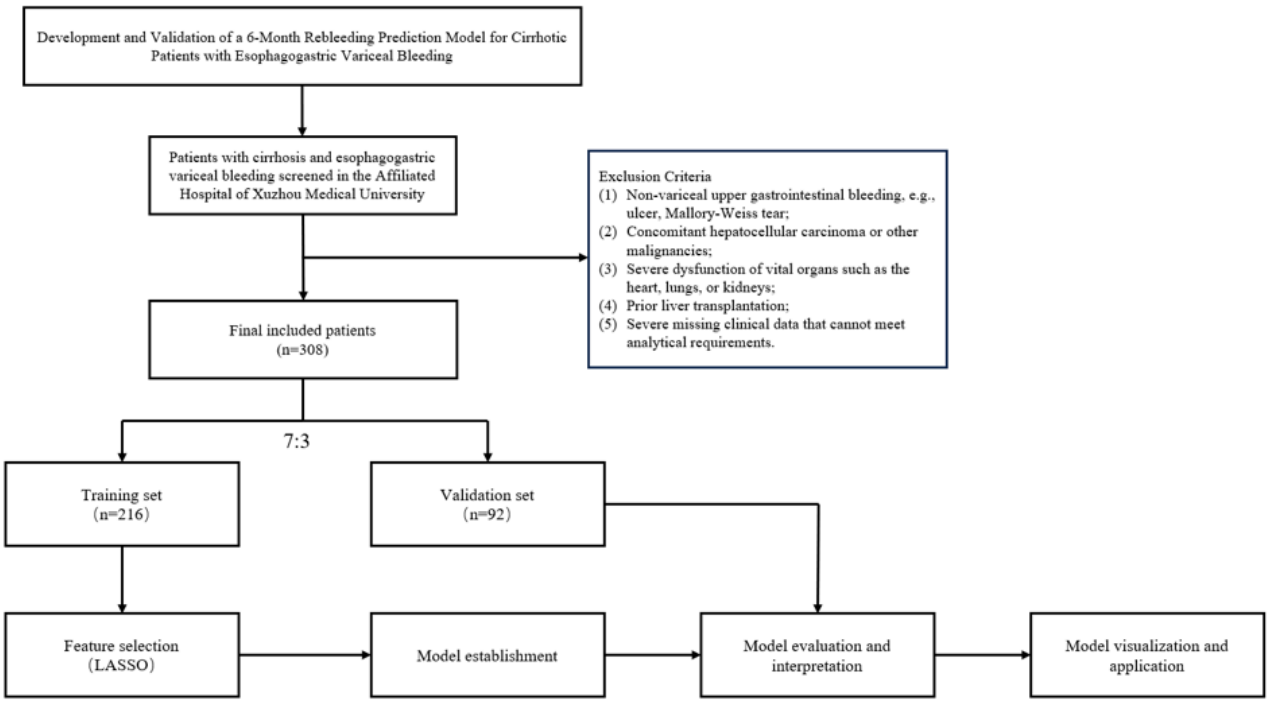
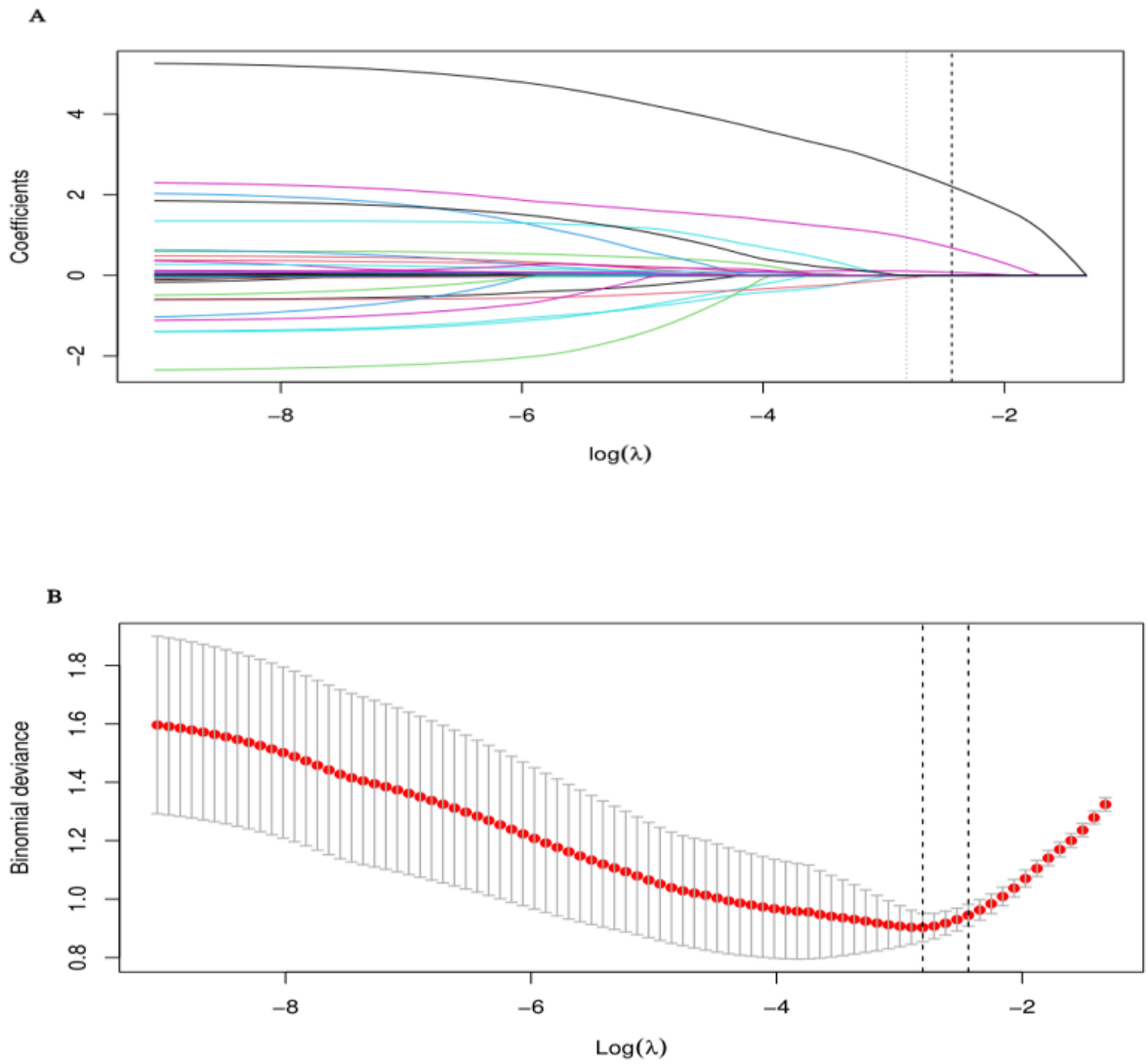



**Figure 1**

Flowchart of patient selection and study design.




## Figure 2

Process of feature variable selection. A, LASSO regression coefficient path plot. B, LASSO binomial deviance plot.

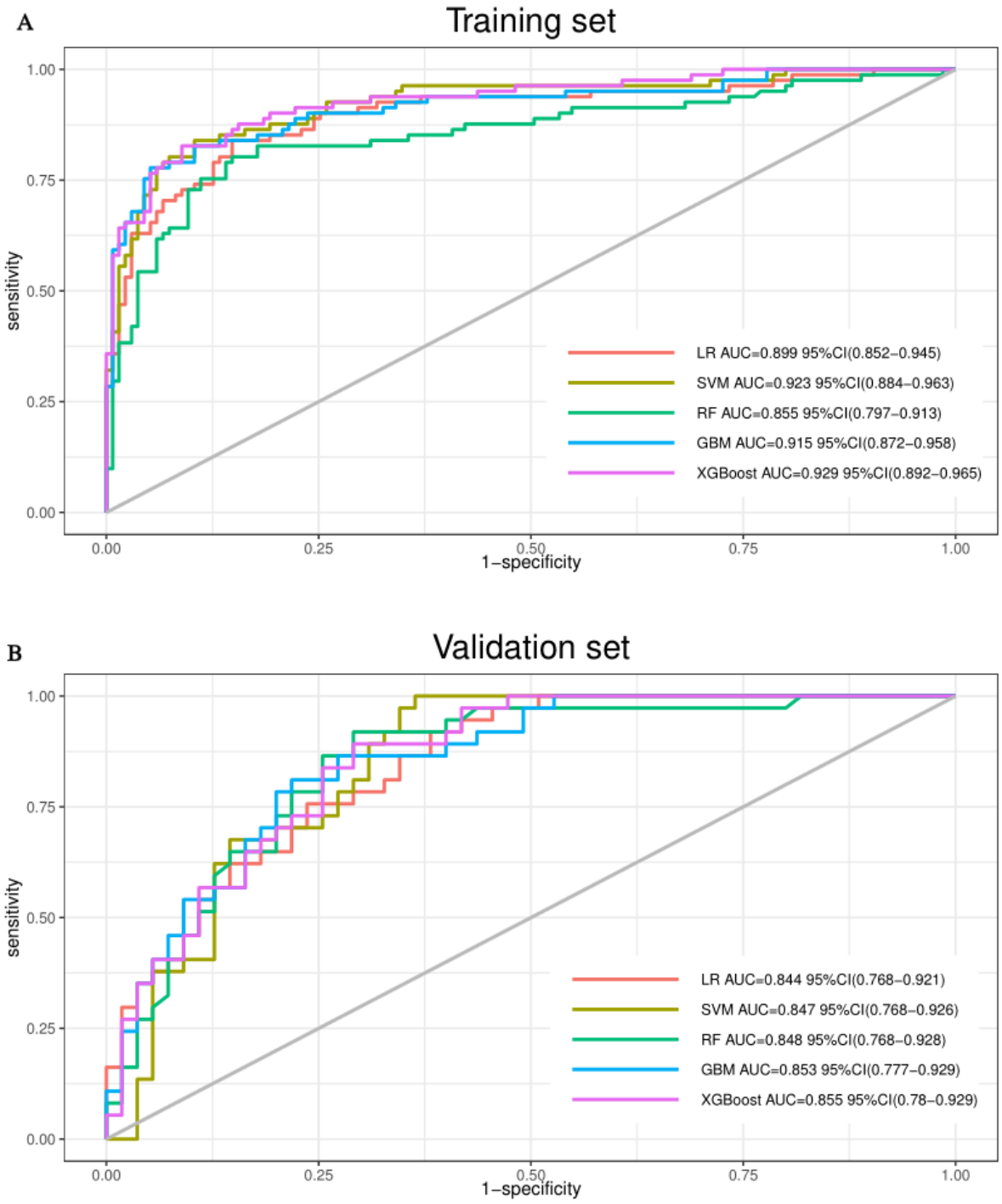



**Figure 3**

ROC curves of various machine learning models in the training set and validation set. Higher AUC values indicate better discriminative ability of the models.

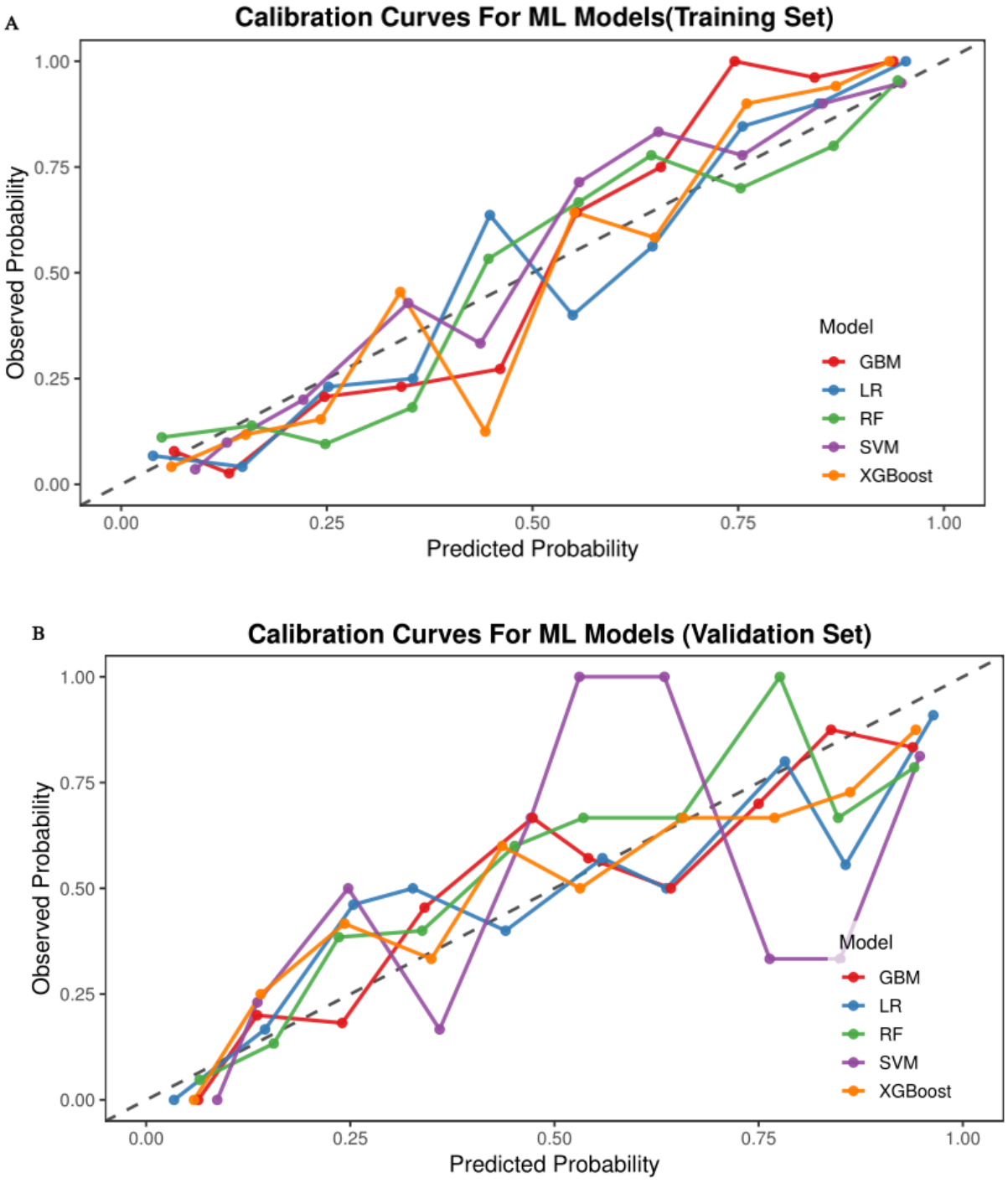



**Figure 4**

Calibration curves of each model in the training set and Validation set.

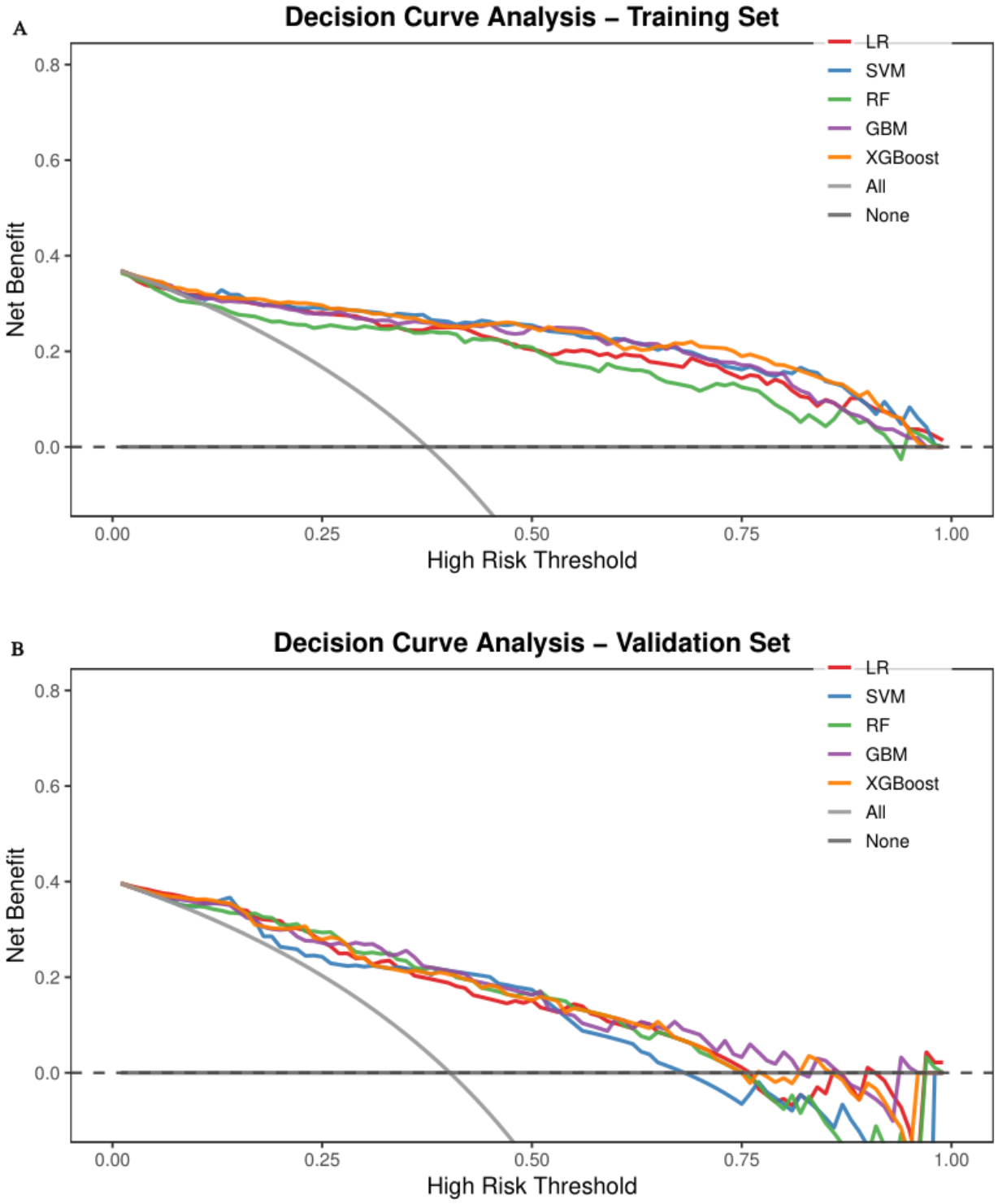



**Figure 5**

The net benefit of the ML models at different threshold probabilities for predicting 6-month rebleeding risk in cirrhotic EGVB patients.

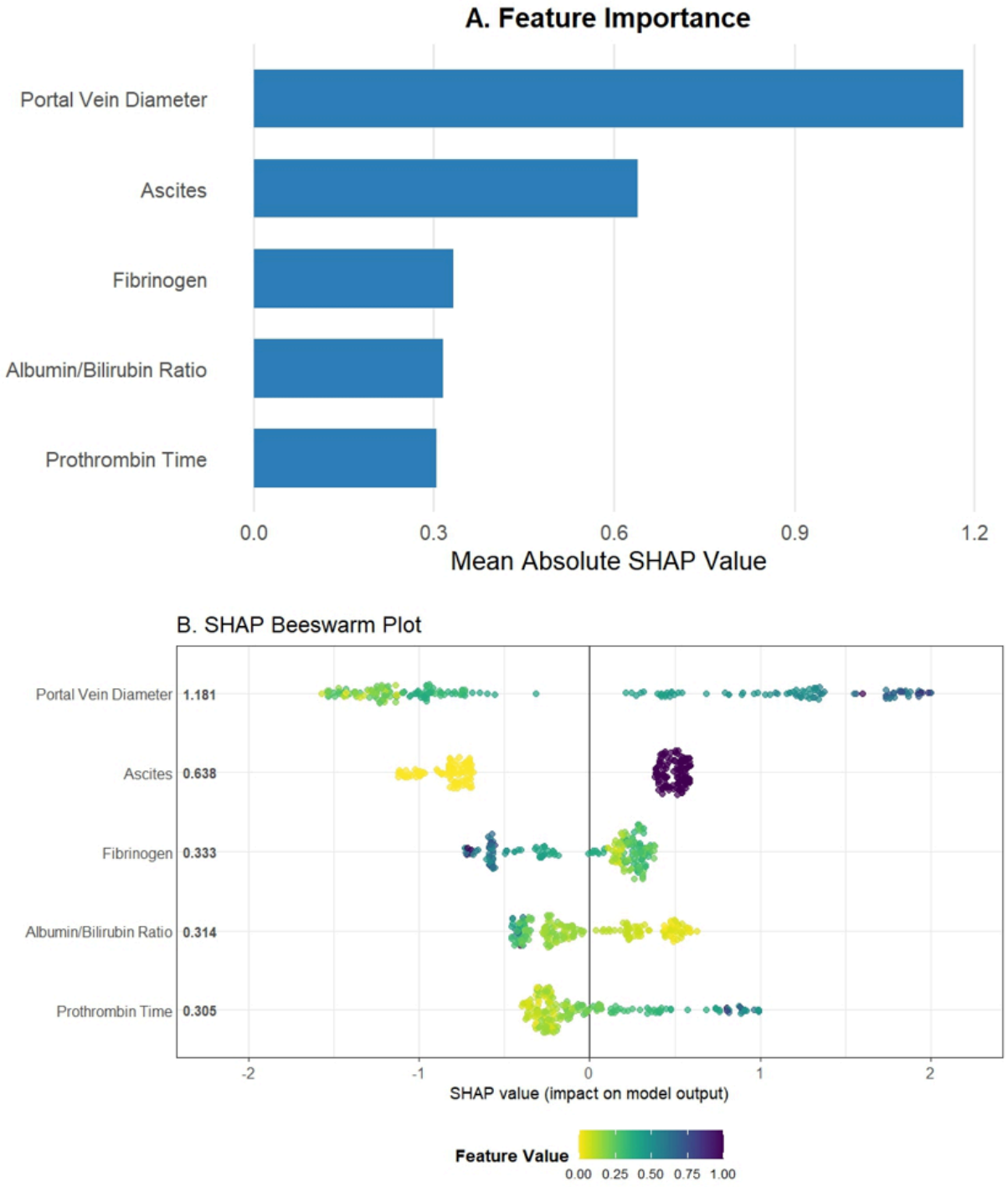



**Figure 6**

SHAP-based interpretability analysis of the XGBoost model for predicting 6-month rebleeding risk in cirrhotic EGVB patients.(A) SHAP feature importance plot showing the mean absolute contribution of each predictor to model output.(B) SHAP beeswarm plot illustrating the distribution, magnitude, and directional effects of individual feature values on prediction outcomes.

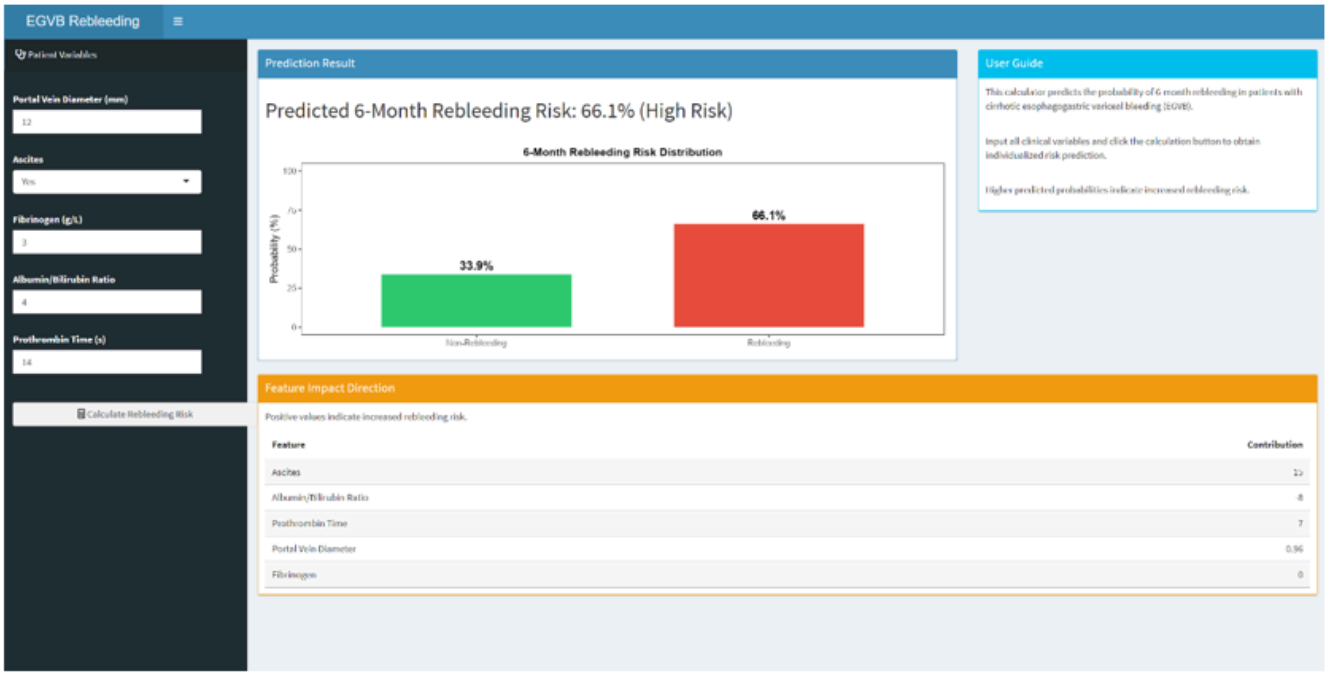



**Figure 7**

Web-based calculator for predicting 6-month rebleeding risk in patients with cirrhotic esophagogastric variceal bleeding (XGBoost model). It includes an input panel for clinical variables, a prediction result panel displaying the rebleeding probability and risk stratification, and a model explanation panel presenting the contribution weights of core predictors.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx