Journal of Cancer Research and Clinical Oncology (2025) 151:285
https://doi.org/10.1007/s00432-025-06345-0

RESEARCH

Machine learning-based hepatocellular carcinoma risk prediction model for patients with HBV-related compensated advanced chronic liver disease

Yanqiu Li[1] · Zihang Qiao[2] · Yongqi Li[1] · Ying Feng[1] · Xianbo Wang[1]

Received: 31 August 2025 / Accepted: 23 September 2025
© The Author(s) 2025

Abstract

Purpose Patients with hepatitis B virus (HBV)-related compensated advanced chronic liver disease (cACLD) demonstrate significant liver fibrosis and portal hypertension, further increasing their hepatocellular carcinoma (HCC) risk. This study aimed to develop and validate machine learning-based HCC risk prediction models.

Methods We retrospectively enrolled 1051 patients with HBV-related cACLD, randomly allocated patients to training (n = 736) and validation (n = 315) cohorts. Feature selection was performed using least absolute shrinkage and selection operator regression, random forest (RF), and support vector machine (SVM). Based on the selected key features, five machine learning models were constructed: SVM, RF, logistic regression, extreme gradient boosting, and Naive Bayes. Model performance was evaluated using area under the curve (AUC), accuracy, sensitivity, and specificity, etc. The Shapley additive explanations (SHAP) method was employed for model interpretability analysis.

Results During a median follow-up time of 35 (20–55) months, 103 patients (9.8%) developed HCC. Feature selection analysis identified five key predictors: liver stiffness measurement (LSM), age, platelet, bile acid, and white blood cell count. The RF model demonstrated superior performance with an AUC of 0.979, an accuracy of 0.977, and a sensitivity of 0.808. SHAP interpretability analysis identified LSM as the most influential predictor (mean SHAP value 1.2), followed by age and other indicators. Feature interaction analysis revealed significant synergistic effects between LSM, platelet, and bile acid.

Conclusion Machine learning-based HCC risk prediction models, particularly the RF algorithm, demonstrated excellent predictive performance in patients with HBV-related cACLD. LSM emerged as the most critical predictive factor.

Keywords Hepatitis B virus · Compensated advanced chronic liver disease · Hepatocellular carcinoma · Machine learning · Risk prediction

Introduction

Chronic hepatitis B virus (HBV) infection is a major public health problem worldwide (Kim 2025; Li et al. 2025a). Approximately 296 million people are infected with HBV worldwide (Hsu et al. 2023). In Asia and Africa, chronic HBV infection rates are more than 8%, making it the world's highest number of chronic hepatitis B patients (Kim 2025). HBV infection can lead to chronic hepatitis, cirrhosis, and ultimately hepatocellular carcinoma (HCC) (Hsu et al. 2023). HCC is the sixth most common malignancy and the third leading cause of cancer-related death worldwide (Rumgay et al. 2022). In the Asia–Pacific region, more than 50% of HCC cases are associated with HBV infection, making HBV-related HCC the leading type of liver cancer in the region (Lin and Kao 2023). This is particularly true in the compensated advanced chronic liver disease (cACLD) stage, where patients, despite not exhibiting overt signs of liver decompensation, already have significant liver fibrosis and portal hypertension, further increasing the risk of HCC.

✉ Ying Feng
fengying@ccmu.edu.cn

✉ Xianbo Wang
wangxb@ccmu.edu.cn

[1] Center of Integrative Medicine, Beijing Ditan Hospital, Capital Medical University, Beijing, China

[2] Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China

Published online: 07 October 2025

Therefore, HCC risk prediction and stratified management in patients with HBV-related cACLD are of great clinical significance, which can help detect HCC early, improve patient prognosis, and optimize medical resource allocation.

Currently, HCC risk prediction tools primarily include clinical scoring systems and noninvasive imaging-based indicators, such as fibrosis-4 index (FIB-4), albuminbilirubin (ALBI) score, and liver stiffness measurement (LSM). In patients with chronic liver disease, ALBI, age, gender, and viral liver disease were associated with progression to HCC from a compensated state (Shearer et al. 2023). John BV et al. found that LSM was associated with HCC risk in patients with metabolic dysfunction associated steatotic liver disease (MASLD), and that for every 5 kPa increase in LSM, the risk of HCC increased by 18% (John et al. 2025). MASLD patients with FIB-4 ≥ 3.25 or LSM ≥ 20 kPa had an increased risk of HCC (Lai et al. 2025). In addition to the prediction models of the above-mentioned individual indicators, the ALBI combined with FIB-4, diabetes and alpha-fetoprotein (AFP) prediction model has certain value in the risk stratification of HCC in patients with compensated cirrhosis (Wang et al. 2023). These methods reflect the degree of liver fibrosis and liver functional reserve, but lack the integration of multidimensional clinical information, especially in patients with HBV-related cACLD.

In recent years, machine learning methods have been increasingly used in medical research (Chong et al. 2025; Ho and Asrani 2025). They can model high-dimensional data and explore nonlinear relationships between complex variables, providing new tools for disease risk prediction. They have also been widely used in the field of liver disease. Huang X et al. used deep learning to establish a prognostic tool for HBV-related acute-on-chronic liver failure based on longitudinal liver ultrasound images with an AUC of 0.851 (Huang et al. 2025). LSM-based random forest (RF) model can effectively identify cirrhosis in patients with chronic hepatitis B (Bai et al. 2025). The RF model can also be used to predict acute liver injury occurring within one month after immune checkpoint inhibitor treatment in pan-cancer patients with an AUC of 0.81 (Wang et al. 2025). Machine learning models, such as lightweight gradient boosting and categorical boosting, have shown strong diagnostic performance for metabolic dysfunction-related steatohepatitis and liver fibrosis (Zhang et al. 2025). However, there are few machine learning prediction models for HCC risk in patients with HBV-related cACLD, and they lack systematic feature selection and model interpretation analysis.

Therefore, this study retrospectively analyzed a clinical cohort to construct machine learning-based HCC risk prediction models for patients with HBV-related cACLD. By comparing the predictive performance of different models, the optimal model was selected. This study aimed to provide a new tool for personalized HCC risk assessment in patients with HBV-related cACLD and provide a scientific basis for clinical follow-up and early intervention strategies.

Methods

Participants

This study retrospectively enrolled patients with HBV-related cACLD who visited Beijing Ditan Hospital between January 2016 and April 2024. The subjects were randomly divided into a training cohort and a validation cohort in a 7:3 ratio. Inclusion criteria were (1) age ≥ 18 years; (2) meeting the diagnostic criteria for cACLD (LSM ≥ 10 kPa (Franchis et al. 2022) and no signs of decompensation such as ascites, variceal bleeding, hepatic encephalopathy, or jaundice); (3) meeting the diagnostic criteria for chronic hepatitis B ([Guidelines for the prevention and treatment of chronic hepatitis B (version 2022)]. 2022): currently HBsAg positive, or HBsAg negative and anti-HBc positive, with a clear history of chronic hepatitis B infection (HBsAg positive history >6 months), and other causes have been excluded; (4) complete clinical data and follow-up data; (5) receiving regular antiviral treatment for at least 6 months. Exclusion criteria included: (1) patients with other liver diseases (other viral hepatitis, MASLD, alcoholic liver disease, drug-induced liver injury, autoimmune liver disease, etc.); (2) patients diagnosed with HCC before enrollment; (3) patients with severe organ dysfunction such as heart, lung, or kidney; (4) patients who had undergone liver transplantation, splenectomy, or transjugular intrahepatic portosystemic shunt; (5) pregnant or lactating women; and (6) patients with incomplete clinical data. The study protocol was approved by the Ethics Committee of Beijing Ditan Hospital (DTEC-KY2024-069–01). All procedures adhered to the Declaration of Helsinki.

Data collection and follow-up endpoints

Baseline clinical data were systematically collected from all patients, including demographic characteristics (age, gender), liver function indicators [aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TB), direct bilirubin (DB), serum albumin (ALB), gamma-glutamyltransferase (GGT), alkaline phosphatase (ALP), bile acids], electrolytes (sodium, chloride), renal function (creatinine), coagulation function [prothrombin time (PT), international normalized ratio (INR)], complete blood count [white blood cell count (WBC), neutrophil-to-lymphocyte ratio (NLR), hemoglobin (Hb),

platelet count (PLT)], lipid profile [total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C)], AFP, uric acid, and blood glucose. Imaging studies included LSM by transient elastography and ultrasound measurements of portal vein width, spleen diameter and thickness.

The primary endpoint of this study was the development of HCC. The diagnostic criteria for HCC were based on the American Association for the Study of Liver Diseases guidelines (Singal et al. 2023) and included: (1) histopathological confirmation; (2) typical imaging findings: arterial phase hyperenhancement and washout on portal venous or delayed phases; or (3) significantly elevated AFP (>400 ng/mL) combined with typical imaging findings.

Feature selection strategy

To identify predictors closely associated with HCC development, this study employed three machine learning feature selection methods in the training cohort only: least absolute shrinkage and selection operator (LASSO) regression, RF, and support vector machine (SVM). All feature selection procedures were performed exclusively in the training dataset to prevent data leakage and ensure unbiased performance evaluation in the independent validation cohort. LASSO regression incorporates an L1 regularization term to achieve feature selection and model simplification. 10-fold cross-validation was used to determine the optimal regularization parameter λ. The λ value that minimized the cross-validation error was used to select feature variables with non-zero regression coefficients. The RF method assessed feature importance based on the mean Gini impurity reduction (MeanDecreaseGini), with larger values indicating higher feature importance. The SVM method assessed feature importance using average rank (AvgRank), with smaller values indicating higher feature importance. Finally, venn diagram analysis was performed to determine the intersection of the features selected by LASSO regression with the top 20 features selected by RF and SVM methods to identify key predictors for subsequent model construction.

Machine learning model construction

Based on the selected key features, five machine learning prediction models were constructed: SVM, RF, logistic regression (LR), extreme gradient boosting (XGBoost), and Naive Bayes (NB). The SVM model used a radial basis function kernel, and the hyperparameters C and γ were optimized through grid search. The RF model set the number of decision trees to 100, and used bootstrap aggregation to reduce overfitting. The LR model used maximum likelihood estimation for parameter estimation, and added an L2 regularization term to prevent overfitting. The XGBoost model used the gradient boosting decision tree algorithm, and hyperparameters, such as the learning rate and maximum depth, were optimized through cross-validation. The NB model made probabilistic predictions based on Bayes' theorem and the conditional independence assumption of features. All models were trained on the training cohort, and model performance was evaluated on an independent validation cohort.

Model performance evaluation and validation

Model performance was comprehensively evaluated using multiple metrics, including accuracy, recall, F1 score, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Brier score, and area under the curve (AUC). Receiver operating characteristic (ROC) curves were plotted, and the AUC was calculated to assess the discriminative ability of the model. Decision curve analysis (DCA) was used to assess the clinical utility of the model. The model's clinical value was determined by comparing the net benefit predicted by the model at different risk thresholds with the net benefits of the "all treatment" and "no treatment" strategies. To translate model predictions into actionable clinical recommendations, we performed threshold optimization analysis of DCA. We systematically evaluated model performance across probability thresholds ranging from 0.05 to 0.50. At each threshold, we calculated sensitivity, specificity, PPV, NPV, and net benefit. The optimal threshold was defined as the one providing the highest net benefit while maintaining clinically acceptable sensitivity for HCC detection. Model calibration was assessed using calibration plots. The model with the best overall performance was selected as the final prediction model. To address the time-to-event nature of HCC development and account for varying follow-up durations, we additionally performed survival analysis using Cox proportional hazards regression and random survival forest (RSF) models. The Cox model was fitted using the key features identified through feature selection. The RSF model, an ensemble extension of survival trees, was constructed with 100 trees to handle complex non-linear relationships while incorporating censoring information. Model performance was evaluated using Harrell's concordance index (C-index) and time-dependent AUC at 12, 24, 36, and 48 months.

Model interpretability and feature interaction analysis

The Shapley additive interpretation (SHAP) method was used to analyze the interpretability of the optimal model and quantify the contribution of each feature to the model's predictions. The SHAP value for each feature was calculated, and feature importance ranking plots and SHAP bee swarm plots were plotted to visually demonstrate the specific impact of each feature value on the model output. SHAP interaction plots were used to analyze the interaction effects between key features and explore the synergistic mechanisms of feature combinations in HCC risk prediction. To maintain methodological rigor and prevent data leakage, SHAP analysis was performed as a post-hoc interpretability assessment on the final optimized models rather than being integrated into the iterative model construction process.

Statistical analysis

Continuous variables are presented as medians (interquartile ranges), and group comparisons were performed using the Mann–Whitney U test. Categorical variables are presented as frequencies (percentages), and group comparisons were performed using the chi-square test or Fisher's exact test. All statistical analyses were performed using Python 3.10. All statistical tests were two-sided, and differences were considered statistically significant when p values < 0.05.

Result

Baseline characteristics

This study enrolled 1051 patients with HBV-related cACLD (Fig. 1), of whom 588 were male (55.9%) with a median age of 57.3 (50.3–65.9) years. The median follow-up was 35 (20–55) months (Table 1). Patients were randomly divided into a training cohort (n = 736) and a validation cohort (n = 315). Based on the development of HCC during follow-up, patients were divided into an HCC group (n = 103, 9.8%) and a non-HCC group (n = 948, 90.2%). Patients in the HCC group were significantly older than those in the non-HCC group [63.0 (55.2–70.0) years vs 57.0 (49.7–65.4) years, $p < 0.001$]. In terms of auxiliary examination indicators, ALT, TB, DB, GGT, bile acid, PT, INR, spleen thickness, and LSM of HCC group were higher than those of the non-HCC group, while the ALB, WBC, and PLT were significantly lower than those of the non-HCC group. There were no significant differences in other indicators between the two groups.

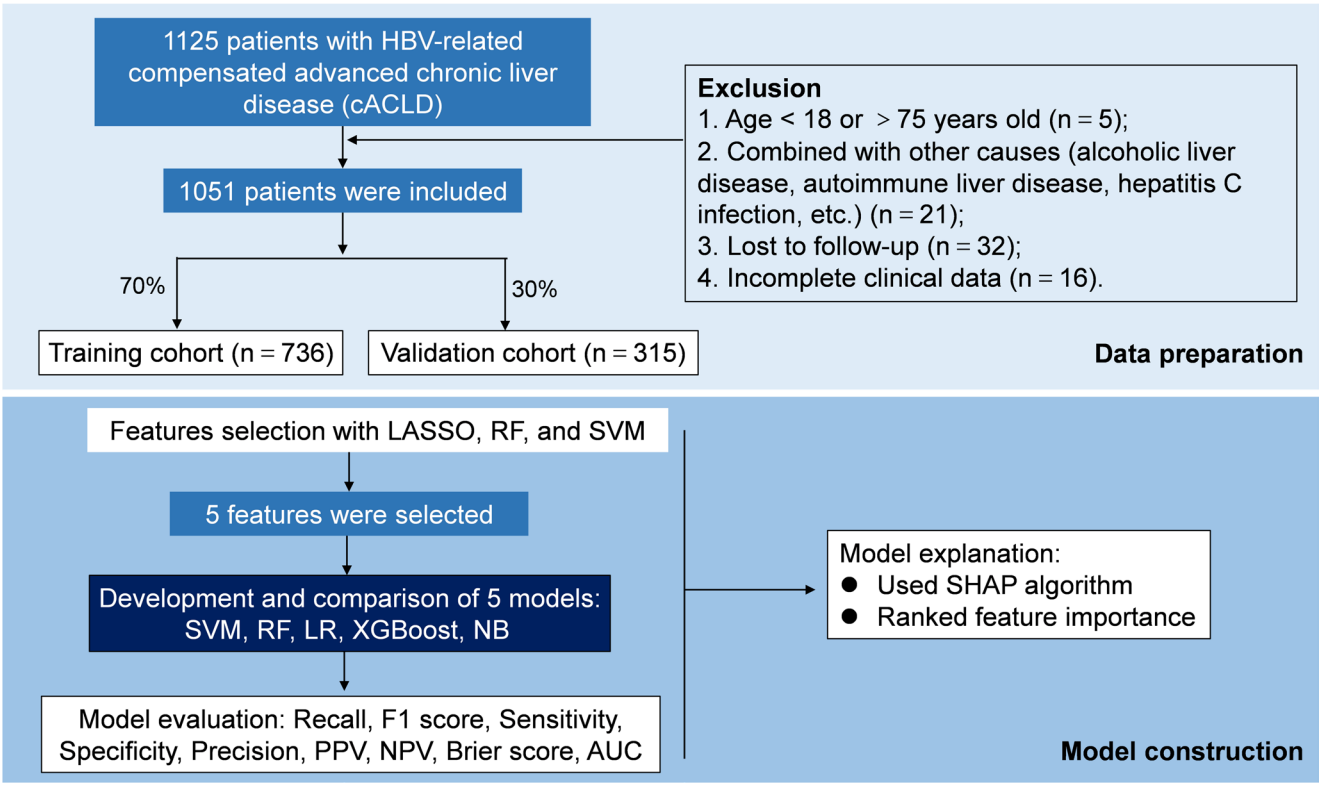


Fig. 1 Flowchart

Table 1 Baseline characteristics of HBV-related cACLD patients

	Overall cohort (n = 1051)	No HCC (n = 948)	HCC (n = 103)	*p* value
Age, years	57.3 (50.3–65.9)	57.0 (49.7–65.4)	63.0 (55.2–70.0)	< 0.001
Gender (%male)	588 (55.9)	527 (55.6)	61 (59.2)	0.481
Follow-up time, months	35 (20–55)	34 (20–55)	42 (22.5–62)	0.069
AST, U/L	32.2 (24.1–49.8)	31.9 (24.1–49.8)	34.4 (24.7–52. 6)	0.350
ALT, U/L	28.0 (19.8–44.9)	23.5 (16.4–33.9)	28.0 (20.1–45)	0.005
TB, μmol/L	19.0 (13.5–27.5)	18.8 (13.3–27)	22.0 (14.1–32.4)	0.005
DB, μmol/L	6.8 (4.7–10.3)	6.7 (4.7–10)	8.0 (5.2–14.2)	0.004
ALB, g/L	41.6 (35.0–46.4)	42.1 (35.3–46.8)	40.3 (33.3–44.4)	0.007
GGT, U/L	30.8 (18.6–59.7)	29.8 (18.6–58.8)	40.3 (22.8–83.2)	0.024
ALP, U/L	86.8 (66.8–113.2)	86.5 (66.9–112.0)	89.9 (65.4–110.6)	0.392
Bile acid, μmol/L	13.4 (5.8–35.6)	12.1 (5.5–33.9)	24.1 (12.0–52.6)	< 0.001
Sodium, mmol/L	141.4 (139.3–143.3)	141.4 (139.3–143.2)	141.6 (138.9–143.6)	0.880
Chloride, mmol/L	105.2 (103.6–107.2)	105.2 (103.6–107.1)	105.5 (103.8–107.8)	0.256
Creatinine, μmol/L	65.4 (56.3–74.5)	65.9 (56.4–74.6)	63.7 (54.9–72.1)	0.206
PT, s	12.7 (11.7–13.4)	12.6 (11.7–13.2)	12.7 (12.3–14.4)	< 0.001
INR	1.17 (1.08–1.26)	1.17 (1.08–1.25)	1.18 (1.11–1.29)	0.003
WBC, 10^9/L	4.21 (3.14–5.48)	4.25 (3.17–5.5)	3.64 (2.8–4.9)	0.011
NLR	1.75 (1.26–2.48)	1.75 (1.26–2.49)	1.72 (1.32–2.38)	0.932
Hb, g/L	137 (117–152.9)	137.4 (117–153)	135 (113.8–150)	0.241
PLT, 10^9/L	92 (58–138)	98 (60–144)	71 (45.5–102.5)	< 0.001
TC, mmol/L	3.84 (3.20–4.53)	3.86 (3.30–4.54)	3.79 (3.13–4.44)	0.216
TG, mmol/L	0.9 (0.66–1.29)	0.92 (0.66- 1.29)	0.87 (0.67–1.24)	0.595
HDL-C, mmol/L	1.12 (0.92–1.36)	1.13 (0.93–1.36)	1.07 (0.89–1.32)	0.115
LDL-C, mmol/L	2.16 (1.71–2.70)	2.17 (1.72–2.71)	2.14 (1.66–2.54)	0.228
Uric acid, mmol/L	306 (248–364.5)	306 (248.5–366)	292 (243.5–350.5)	0.245
Glucose, mmol/L	5.83 (5.34–6.63)	5.8 (5.3–6.6)	5.9 (5.4–7.1)	0.183
AFP, ng/mL	3.0 (2.1–4.4)	3.0 (2.1–4.4)	3.3 (2.4–4.5)	0.163
LSM, kPa	17.3 (12.7–28.3)	17.1 (12.5–26.3)	46.4 (21.6–51.0)	< 0.001
Portal vein Width, mm	12 (11–13)	12 (11–13)	12 (11–13)	0.134
Spleen diameter, mm	134 (117–154)	133 (117–154)	134 (121–156)	0.550
Spleen thickness, mm	42 (34–51)	42 (34–50.3)	45 (37–56)	0.023

HCC, hepatocellular carcinoma; HBV, hepatitis B virus; cACLD, compensated advanced chronic liver disease; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TB, total bilirubin; DB, direct bilirubin; ALB, albumin; GGT, gamma glutamyltransferase; ALP, alkaline phosphatase; PT, Prothrombin time; INR, international normalized ratio; WBC, white blood cell; NLR, Neutrophil-to-lymphocyte ratio; Hb, hemoglobin; PLT, platelets; TC, total cholesterol; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; AFP, alpha-fetoprotein; LSM, liver stiffness measurement.

Feature selection

To develop a predictive model for HCC development in patients with HBV-related cACLD, three machine learning methods were used for feature selection, including LASSO regression, RF, and SVM. LASSO regression analysis determined the optimal regularization parameter λ through 10-fold cross-validation. The regression coefficient path diagram (Fig. 2a) shows that the regression coefficients of each variable gradually shrink toward zero with increasing log(λ). The cross-validation error plot (Fig. 2b) shows that the model achieves minimum binomial deviance when log(λ) = -4.4. LASSO regression identified five feature variables with non-zero regression coefficients, including LSM, age, PLT, WBC, and bile acid. The RF model ranked feature importance based on the mean Gini impurity reduction (Fig. 2c). The results showed that LSM was the most important of all features. The SVM model used average rank to assess feature importance (Fig. 2d). LSM also ranked first. Venn diagram analysis (Fig. 2e) was performed to identify the intersection of the variables selected by LASSO regression with the top 20 variables from RF and SVM. Five key predictors were identified: LSM, age, PLT, bile acid, and WBC.

Machine learning model performance evaluation

Five machine learning models were constructed to predict the risk of HCC in patients with HBV-related cACLD, including SVM, RF, LR, XGBoost, and NB models. All five models demonstrated good predictive performance in the training cohort (Table 2). The RF model performed best, with accuracy, recall, F1 score, sensitivity, specificity, PPV, NPV, and AUC of 0.977, 0.808, 0.857, 0.808, 0.993, 0.913, 0.982, and 0.979, respectively. The XGBoost model performed worse than the RF model, with an AUC of 0.943, accuracy of 0.974, and sensitivity and recall of 0.808. In contrast, the LR model had lower sensitivity and recall (both 0.077), while the SVM and NB models had sensitivities of 0.135 and 0.250, respectively. The RF and XGBoost model also achieved the lower Brier scores

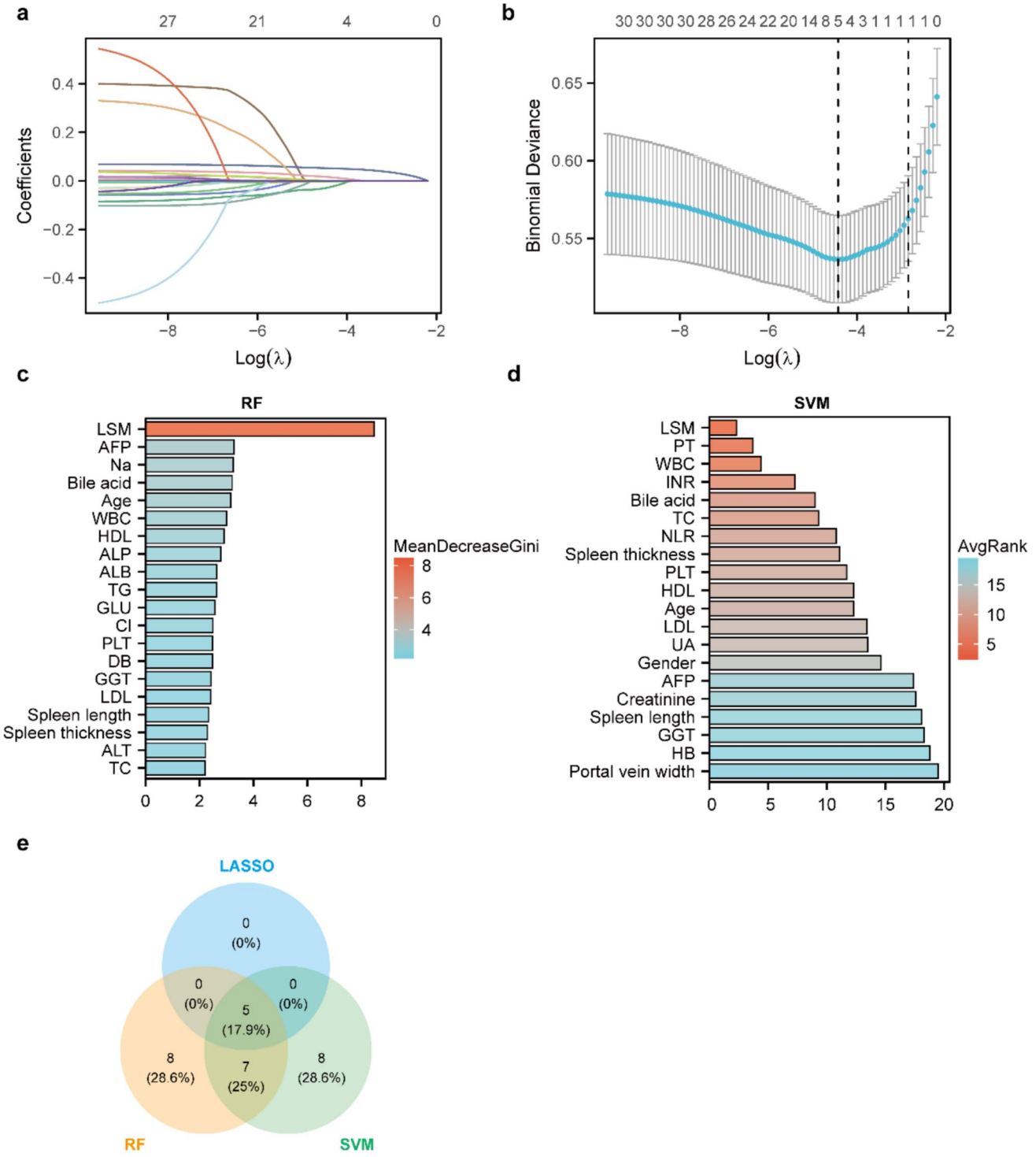


Fig. 2 Feature Selection Analysis. **a** LASSO regression coefficient path diagram. **b** Cross-validation error plot. **c** RF feature importance ranking. **d** SVM feature importance ranking. **e** Intersection of LASSO regression, RF, and SVM important features. LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; RF, random forest

in both training and validation cohorts. In addition, to evaluate the added value of machine learning, we used LSM alone for evaluation, and its AUC in the training and validation cohorts reached 0.745 and 0.826, respectively, which were lower than those of the RF and XGBoost models. ROC curve analysis (Fig. 3a) showed that in the training cohort, the RF model achieved the highest AUC (0.979), followed by XGBoost (0.943), NB (0.790), LR

Table 2 Performance of the machine learning models for predicting HCC

	Accuracy	Recall	F1 score	Sensitivity	Specificity	PPV	NPV	AUC	Brier score
Training cohort									
SVM	0.923	0.135	0.230	0.135	0.996	0.778	0.925	0.764	0.064
RF	0.977	0.808	0.857	0.808	0.993	0.913	0.982	0.979	0.052
LR	0.916	0.077	0.136	0.077	0.995	0.571	0.920	0.777	0.070
XGBoost	0.974	0.808	0.840	0.808	0.989	0.875	0.982	0.943	0.031
NB	0.913	0.250	0.329	0.250	0.975	0.481	0.933	0.790	0.074
Validation cohort									
SVM	0.892	0.078	0.143	0.078	0.9927	0.800	0.893	0.750	0.072
RF	0.957	0.647	0.776	0.647	0.997	0.971	0.956	0.956	0.063
LR	0.889	0.098	0.169	0.098	0.992	0.625	0.894	0.826	0.080
XGBoost	0.953	0.647	0.759	0.647	0.992	0.917	0.956	0.915	0.046
NB	0.901	0.294	0.405	0.294	0.979	0.652	0.914	0.815	0.082

HCC, hepatocellular carcinoma; XGBoost, extreme gradient boosting; SVM, support vector machine; LR, logistic regression; RF, random forest; NB, Naive Bayes; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve

(0.777), and SVM (0.764). Model performance was well validated in the validation cohort (Fig. 3b), with the RF model maintaining the best performance (AUC = 0.956) and the XGBoost model achieving an AUC of 0.915. DCA results (Fig. 3c, 3d) showed that in both the training and validation cohorts, the RF and XGBoost models provided positive net benefits across most risk thresholds, significantly outperforming the "all-treatment" and "no-treatment" strategies. Threshold optimization analysis revealed that the probability threshold of 0.15 provided the optimal balance between clinical sensitivity and specificity (Table S1). At this threshold, the RF model achieved perfect sensitivity (1.00), specificity of 0.91, PPV of 0.54, NPV of 1.00, and the highest net benefit of 0.0761. Calibration analysis demonstrated excellent agreement between predicted and observed probabilities for the RF model (Fig. 3e, 3f). Therefore, the RF model performed best in predicting the risk of HCC in patients with HBV-related cACLD.

Survival analysis results are presented in Table S2. In the training cohort, RSF achieved a C-index of 0.976 versus 0.821 for Cox regression. Time-dependent AUCs for RSF ranged from 0.908 at 12 months to 0.990 at 48 months in the training cohort, while Cox regression showed more modest performance (0.726–0.836). In the validation cohort, RSF maintained robust performance with a C-index of 0.984 and time-dependent AUCs ranging from 0.866 to 0.966.

Model interpretability analysis

To further elucidate the predictive mechanism of the RF model and the contribution of each feature to HCC risk prediction, we used the SHAP method to conduct model interpretability analysis. Feature importance ranking (Fig. 4a) showed that LSM contributed most to model prediction, with an average SHAP value of 1.2, making it the most important factor affecting HCC risk. Age ranked second, with an average SHAP value of 0.8, reflecting the close association between increasing age and increased HCC risk. PLT, WBC, and bile acid levels were also important contributors to model prediction. SHAP bee swarm plot analysis (Fig. 4b) further revealed the specific impact of each feature value on the model output. High LSM values were primarily distributed in the positive SHAP value region, indicating that increased LSM significantly increases the risk of HCC. High age values were also concentrated in the positive SHAP value region, confirming that advanced age is an independent risk factor for HCC. Notably, low PLT and WBC values were primarily located in the positive SHAP value region, suggesting that decreased PLT and WBC are closely associated with a higher risk of HCC.

Feature interaction effect analysis

To further explore the interaction patterns between key features in the RF model, we further analyzed the interplay between key predictor variables. Interaction analysis between LSM and PLT (Fig. 4c) revealed that when LSM values were below 30 kPa, the contribution of PLT to SHAP was relatively small and the variation was limited. However, when LSM exceeded 30 kPa, the impact of PLT significantly increased, with lower PLT values corresponding to higher positive SHAP values. Interaction analysis between LSM and bile acids (Fig. 4d) revealed a similar pattern. In regions with lower LSM, bile acid levels had a relatively limited impact on model predictions. As LSM increased, especially above 30 kPa, higher bile acid levels were closely associated with positive SHAP values. Interaction analysis between PLT and WBC (Fig. 4e) revealed that lower WBC values tended

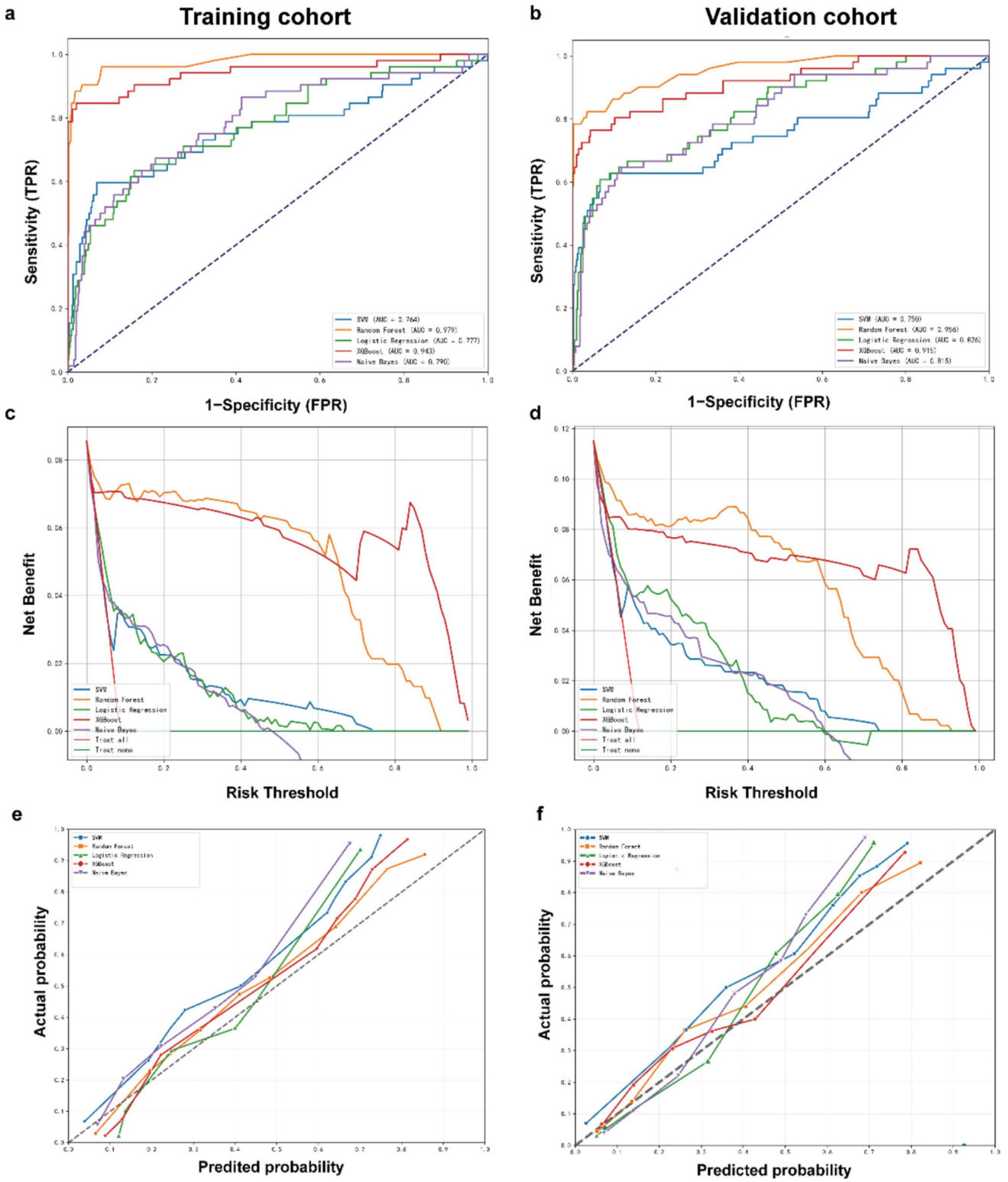


Fig. 3 Machine Learning Model Performance Evaluation. **a** Receiver operating characteristic (ROC) curve in the training set. **b** ROC curve in the validation set. **c** Decision curve analysis (DCA) in the training set. **d** DCA in the validation set. **e** Calibration curves in the training set. **f** Calibration curves in the validation set. SVM, support vector machine; RF, random forest; LR, logistic regression; XGBoost, extreme gradient boosting; NB, Naive Bayes

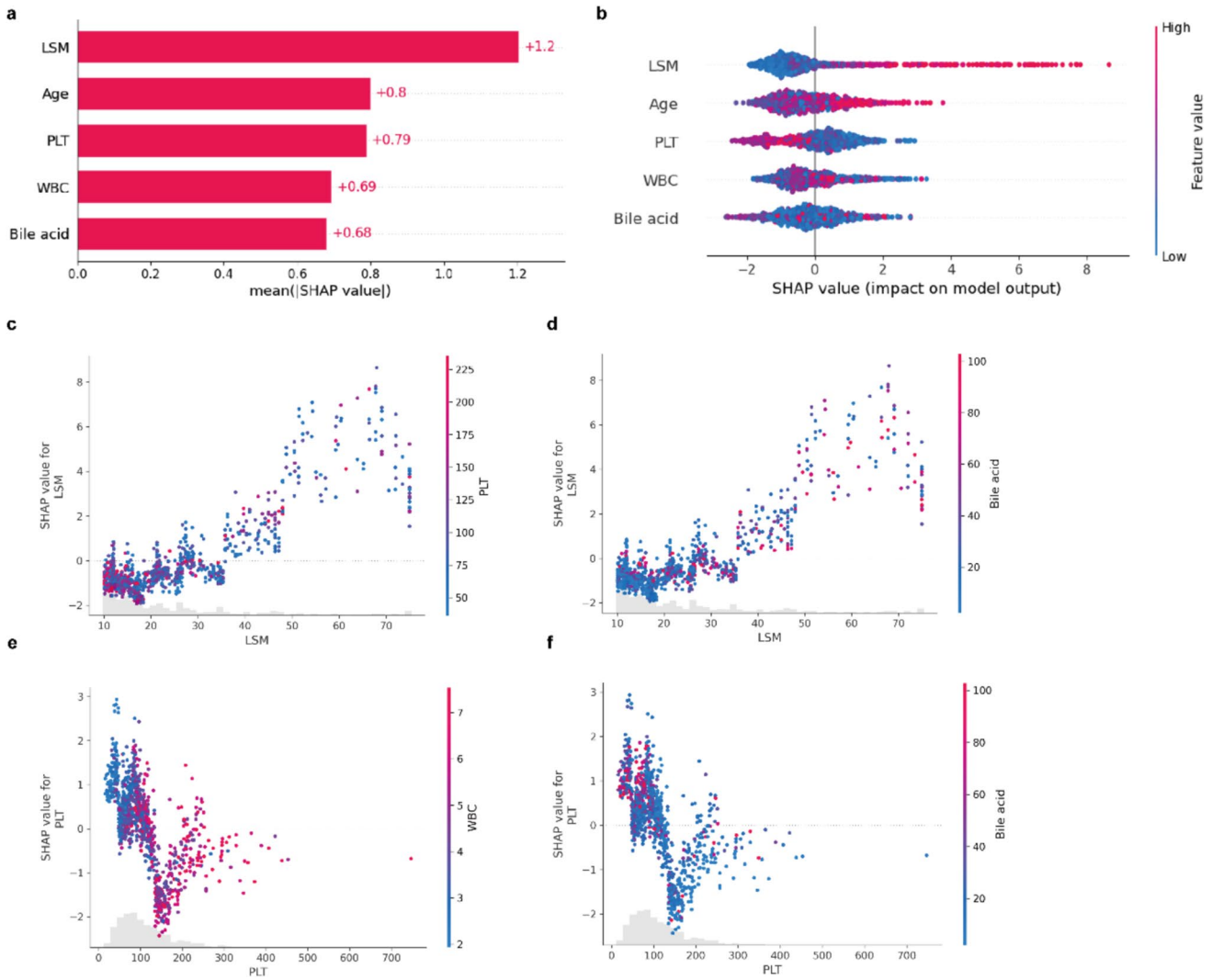


Fig. 4 Model Interpretability and Feature Interaction Analysis. **a** SHAP feature importance ranking plot. **b** SHAP bee swarm scatter plot. **c** Interactive analysis of LSM and PLT. **d** Interactive analysis of LSM and bile acid. **e** Interactive analysis of PLT and WBC. **f** Interactive analysis of PLT and bile acid. LSM, liver stiffness measurement; PLT, platelet; WBC, white blood cell; SHAP, Shapley additive interpretation

to produce higher positive SHAP values when PLT decreased. The interaction between PLT and bile acid (Fig. 4f) further confirmed that patients with high bile acid and low PLT showed the highest positive SHAP value.

Discussion

This study constructed a machine learning-based HCC risk prediction model for patients with HBV-related cACLD. Five key predictors, including LSM, age, PLT, bile acid, and WBC were identified. By comparing the predictive performance of five different algorithms, the RF model was identified as the optimal predictive model. This model demonstrated excellent predictive performance in both the training and validation cohorts, with AUCs of 0.979 and 0.956, respectively. SHAP interpretability analysis further elucidated the interaction mechanisms among these factors.

Traditional HCC risk prediction methods primarily rely on single or several clinical indicators, such as the FIB-4 and ALBI scoring systems (Lai et al. 2025; Wang et al. 2023). While these can reflect the degree of liver fibrosis and liver function reserve to a certain extent, they lack the integration of multidimensional clinical information. The machine learning method employed in this study is capable of processing high-dimensional data and exploring non-linear relationships between complex variables, providing a new tool for HCC risk prediction. The comparison with LSM alone performance provides compelling evidence for

the clinical utility of our machine learning approach. The RF model in this study showed excellent predictive performance with an AUC of 0.979. This is similar to the results of other studies. The RF model can effectively identify cirrhosis and significant liver fibrosis in patients with chronic hepatitis B (Bai et al. 2025; Song et al. 2025). The RF model can also identify high-risk patients for acute liver injury within one month after immune checkpoint inhibitor treatment in pan-cancer patients (Wang et al. 2025). RF analysis also improved the ability to predict in-hospital mortality in patients with cirrhosis better than traditional methods and is consistent across countries with different income levels (Silvey et al. 2025). While LSM emerged as the most critical single predictor in our SHAP analysis, the RF model's superior performance demonstrated that integrating multiple predictors through machine learning algorithms captures complex feature interactions that cannot be achieved through single-marker assessment. The particularly notable improvements in specificity and positive predictive value suggest that the machine learning model reduces false positive rates, which is crucial for clinical decision-making and resource allocation in HCC surveillance programs. The RF model's optimal performance stems primarily from its inherent advantages in ensemble learning. The algorithm constructs multiple decision trees through bootstrap sampling, with each tree trained on a different subset of data. It then randomly selects a subset of features for node splitting, effectively reducing the risk of overfitting and improving the model's generalization capabilities. More importantly, the RF algorithm can naturally handle complex interactions between features, automatically identifying the impact of variable combinations on outcomes through hierarchical decision-making within a tree structure. This is particularly important in complex medical problems such as HCC risk prediction. This performance difference between algorithms reveals the advantages of ensemble learning methods in medical predictive modeling, particularly when processing multidimensional biomedical data, which can effectively integrate pathophysiological information at different levels. Our RSF model's superior time-dependent performance demonstrated the advantage of incorporating temporal information in HCC risk prediction. The high time-dependent AUCs across different follow-up periods (12–48 months) indicate that the model can effectively stratify patients for surveillance intervals, potentially enabling more personalized monitoring strategies. This temporal precision is particularly valuable for resource allocation in HCC surveillance programs, allowing clinicians to intensify monitoring for high-risk patients while optimizing screening intervals for lower-risk individuals.

The threshold optimization analysis provided crucial guidance for translating our RF model into clinical practice. Based on the optimal probability threshold of 0.15, we propose a risk-stratified HCC surveillance protocol. Patients with RF model-predicted probability ≥ 0.15 should be classified as high-risk and warrant intensive surveillance with contrast-enhanced MRI or CT every 3–6 months, supported by the 54% positive predictive value at this threshold. Conversely, patients with predicted probability < 0.15 can safely continue standard ultrasound-based surveillance every 6–12 months, supported by the 100% negative predictive value. This approach would reduce healthcare costs and patient burden while maintaining optimal cancer detection rates. This strategy could optimize resource allocation, directing intensive surveillance toward patients most likely to benefit while reducing unnecessary testing in lower-risk individuals.

Among the five core predictors, LSM, ranked first with an average SHAP value of 1.2, becoming the most critical factor in predicting HCC risk. LSM quantitatively assesses the degree of liver fibrosis using transient elastography, and its value is positively correlated with the amount of liver collagen deposition and portal hypertension (Rockey et al. 2025). In the chronic HBV infection, persistent inflammatory responses activate hepatic stellate cells, promoting excessive deposition of extracellular matrix, leading to liver structural remodeling and increased portal venous pressure (You et al. 2023; Cheng et al. 2025). When LSM is significantly elevated, the liver microenvironment undergoes fundamental changes, including hypoxia, acidosis, and dysnutrient metabolism. These factors synergistically create a favorable environment that is favorable for tumor growth. Furthermore, the abnormal angiogenesis and impaired immune surveillance associated with fibrosis further promote malignant transformation. Elevated serum bile acid levels are a key marker of hepatocellular dysfunction and cholestasis. Recent studies have revealed that bile acids are not only important mediators of lipid metabolism but also crucial signaling molecules, regulating gene transcription and cellular metabolism through activation of the farnesoid X receptor and the G-protein-coupled bile acid receptor (Kydd-Sinclair et al. 2025; Zhou et al. 2025). Abnormal bile acid levels may promote malignant transformation of HCC by activating inflammatory pathways, oxidative stress, apoptosis and regulating tumor immune microenvironment (Tong and Lou 2025; Sun et al. 2025, 2024). Furthermore, certain bile acids, such as deoxycholic acid, have direct carcinogenic effects, inducing DNA damage and gene mutations (Ahmad et al. 2023). Changes in WBC reflect complex changes in the body's inflammatory state and immune function. In chronic HBV infection, persistent viral replication and immune responses lead to chronic liver inflammation and changes in WBC (Zhou et al. 2024a). SHAP analysis has shown that low WBC is associated with an increased risk

of HCC, which may reflect a decrease in immune surveillance, providing opportunities for tumor cell survival and proliferation. The integration of these 5 predictors reflects the multidimensional pathophysiological mechanisms underlying HCC development and progression, including progressive liver fibrosis, decreased liver function reserve, development of portal hypertension, immune dysfunction, and metabolic disturbances. These factors interact in complex ways, forming a comprehensive indicator system for HCC risk assessment.

The synergistic effects revealed by SHAP interaction analysis reflect the inherent interconnectedness of multisystemic and multi-level pathophysiological changes during HCC development and progression. The significant interaction between LSM and PLT has a profound pathophysiological basis. When LSM reaches a threshold of 30 kPa, clinically significant portal hypertension occurs. At this stage, liver sinusoidal endothelial cells dedifferentiate, capillarization increases, and portal vein resistance rises significantly (Li et al. 2025b). In this pathological state, the PLT-damaging effects of portal hypertension-induced splenic congestion and hypersplenism are significantly amplified, shortening the PLT lifespan. More importantly, when PLT drop to critical levels, the levels of cytokines decrease, such as platelet-derived growth factor and transforming growth factor-β (Su et al. 2008). These factors originally play a role in maintaining liver microenvironmental homeostasis and inhibiting abnormal angiogenesis. Therefore, thrombocytopenia in the context of high LSM not only reflects the severity of portal hypertension but also indicates a failure of liver microenvironmental regulatory mechanisms, creating favorable conditions for tumor angiogenesis and invasive growth. The interaction between LSM and bile acid levels also reveals complex molecular mechanisms. When LSM is significantly elevated, liver structural remodeling leads to obstruction of bile outflow pathways, while hepatocyte reduction and impaired function impair normal bile acid metabolism. Elevated serum bile acid levels are not only a marker of impaired liver function but activate a series of oncogenic signaling pathways. The interaction between PLT and WBC reveals a synergistic mechanism of the immune-coagulation system in HCC development. A low PLT not only impairs hemostasis but also affects the activation and function of immune cells. P-selectin on the platelet membrane can interact with neutrophils and monocytes to form PLT-WBC aggregates (Singbartl et al. 2001; Keating et al. 2008), which play a crucial role in anti-tumor immune responses. The interaction between PLT and bile acids further confirms this multisystem synergistic mechanism. High bile acid levels can directly activate PLT, leading to excessive platelet consumption (Qi et al. 2025; Zhou et al. 2024b). This consumption mechanism is further amplified in the setting of already low PLT. The synergistic effects revealed by these SHAP interaction analyses suggest that the development of HCC is not the result of a single factor but rather the result of the interaction and mutual amplification of multiple pathophysiological processes.

This study still has some limitations. First, this study is a single-center retrospective study and may have selection bias, which requires further verification in a multicenter prospective study. Second, the follow-up time was relatively short, with a median follow-up time of 35 months, and the long-term predictive efficacy needs further observation. Third, this study was conducted in a Chinese patient population, and the applicability of the model in populations of different races and regions needs further verification. Fourth, the class imbalance in our dataset (9.8% HCC incidence) may affect model performance, particularly sensitivity. The observed high AUC values are more likely to reflect the tendency of the RF algorithm to overfit. Additionally, while our SHAP analysis provided post-hoc interpretability insights, we did not implement an iterative feedback mechanism to incorporate these insights into model refinement. Future studies could conduct large-scale multicenter prospective validation studies to further confirm the external validation of the model, and explore methods to integrate interpretability analysis into the model development pipeline while maintaining validation integrity. At the same time, more biomarkers and genomic data should be included to construct a more comprehensive prediction model. Combining artificial intelligence technology to develop prediction tools and decision support systems that are easy to use in clinical practice will contribute to the clinical promotion of the model.

Conclusions

Machine learning-based HCC risk prediction models demonstrated good predictive performance in patients with HBV-related cACLD, with the RF algorithm showing superior capability. LSM emerged as the most critical predictive biomarker, underscoring its clinical significance. These models provide robust evidence-based frameworks for personalized risk stratification, targeted surveillance protocols, and resource-efficient healthcare delivery.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00432-025-06345-0.

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

the final manuscript.

Funding This research was funded by the High-level Chinese Medicine Key Discipline Construction Project (No. zyyzdxk-2023005, to Xianbo Wang); Capital's Funds for Health improvement and Research (No. 2024–1-2173, to Xianbo Wang); National Natural Science Foundation of China (No. 82474419, to Xianbo Wang and 82474426, to Ying Feng); Beijing Municipal Natural Science Foundation (No. 7232272, to Ying Feng); and Beijing Traditional Chinese Medicine Technology Development Fund Project (No. BJZYZD-2023–12, to Xianbo Wang).

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by by the Ethics Committee of Beijing Ditan Hospital (DTEC-KY2024-069–01).

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Ahmad F, Saha P, Singh V et al (2023) Diet as a modifiable factor in tumorigenesis: focus on microbiome-derived bile acid metabolites and short-chain fatty acids. Food Chem 410:135320. https://doi.org/10.1016/j.foodchem.2022.135320

Bai X, Pu C, Zhen W et al (2025) Identifying liver cirrhosis in patients with chronic hepatitis B: an interpretable machine learning algorithm based on LSM. Ann Med 57(1):2477294. https://doi.org/10.1080/07853890.2025.2477294

Cheng J-Y, Shan G-Y, Wan H et al (2025) MIF/CD74 axis in hepatic stellate cells mediates HBV-related liver fibrosis. Int Immunopharmacol 147:113929. https://doi.org/10.1016/j.intimp.2024.113929

Chong PL, Vaigeshwari V, Mohammed Reyasudin BK et al (2025) Integrating artificial intelligence in healthcare: applications, challenges, and future directions. Future Sci OA 11(1):2527505. https://doi.org/10.1080/20565623.2025.2527505

de Franchis R, Bosch J, Garcia-Tsao G et al (2022) Baveno VII - renewing consensus in portal hypertension. J Hepatol 76(4):959–974. https://doi.org/10.1016/j.jhep.2021.12.022

Ho CK, Asrani SK (2025) Ethics, bias, and governance in artificial intelligence for hepatology: toward building a safe and fair future. J Clin Exp Hepatol 15(6):102628. https://doi.org/10.1016/j.jceh.2025.102628

Hsu Y-C, Huang DQ, Nguyen MH (2023) Global burden of hepatitis B virus: current status, missed opportunities and a call for action. Nat Rev Gastroenterol Hepatol 20(8):524–537. https://doi.org/10.1038/s41575-023-00760-9

Huang X, Yuan S, Zhou A et al (2025) Predicting prognosis of patients with hepatitis B virus-related acute-on-chronic liver failure from longitudinal ultrasound images using a multi-task deep learning approach. Ann Med 57(1):2551819. https://doi.org/10.1080/07853890.2025.2551819

John BV, Bastaich DR, Deng Y et al (2025) Use of liver stiffness measurement for hepatocellular carcinoma risk stratification in metabolic dysfunction associated steatotic liver disease. Hepatology. https://doi.org/10.1097/HEP.0000000000001498.PubMedPMID:40810411;PubMedCentralPMCID:PMCQ1.eng

Keating FK, Fung MK, Schneider DJ (2008) Induction of platelet white blood cell (WBC) aggregate formation by platelets and WBCs in red blood cell units. Transfusion 48(6):1099–1105. https://doi.org/10.1111/j.1537-2995.2008.01692.x

Kim BK (2025) Future perspectives of liver research in the Asia-Pacific region: focus on hepatitis B and C. J Gastroenterol Hepatol 40(8):1855–1860. https://doi.org/10.1111/jgh.17028

Kydd-Sinclair D, Packer GL, Weymouth-Wilson AC et al (2025) Structural basis of novel bile acid-based modulators of FXR. J Mol Biol. https://doi.org/10.1016/j.jmb.2025.169383

Lai JC-T, Yang B, Lee HW et al (2025) Non-invasive risk-based surveillance of hepatocellular carcinoma in patients with metabolic dysfunction-associated steatotic liver disease. Gut. https://doi.org/10.1136/gutjnl-2025-334981

Li A, Yi Z, Ma C et al (2025a) Innate immune recognition in hepatitis B virus infection. Virulence 16(1):2492371. https://doi.org/10.1080/21505594.2025.2492371

Li Y, Zhu B, Shi K et al (2025b) Advances in intrahepatic and extrahepatic vascular dysregulations in cirrhotic portal hypertension. Front Med Lausanne 12:1515400. https://doi.org/10.3389/fmed.2025.1515400

Lin C-L, Kao J-H (2023) Development of hepatocellular carcinoma in treated and untreated patients with chronic hepatitis B virus infection. Clin Mol Hepatol 29(3):605–622. https://doi.org/10.3350/cmh.2022.0342

Qi Z, Zhang W, Zhang P et al (2025) The gut microbiota-bile acid-TGR5 axis orchestrates platelet activation and atherothrombosis. Nat Cardiovasc Res 4(5):584–601. https://doi.org/10.1038/s44161-025-00637-x

Rockey DC, Alsawas M, Duarte-Rojo A et al (2025) Noninvasive liver disease assessment to identify portal hypertension: systematic and narrative reviews supporting the AASLD practice guideline. Hepatology 81(3):1086–1104. https://doi.org/10.1097/HEP.0000000000000841

Rumgay H, Arnold M, Ferlay J et al (2022) Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol 77(6):1598–1606. https://doi.org/10.1016/j.jhep.2022.08.021

Shearer JE, Jones R, Parker R et al (2023) The natural history of advanced chronic liver disease defined by transient elastography. Clin Gastroenterol Hepatol. https://doi.org/10.1016/j.cgh.2022.03.015

Silvey S, Kamath PS, George J et al (2025) enhancement of inpatient mortality prognostication with machine learning in a prospective global cohort of patients with cirrhosis with external validation. Gastroenterology. https://doi.org/10.1053/j.gastro.2025.07.015.PubMedPMID:40712932;PubMedCentralPMCID:PMCQ1.eng

Singal AG, Llovet JM, Yarchoan M et al (2023) AASLD practice guidance on prevention, diagnosis, and treatment of hepatocellular

carcinoma. Hepatology 78(6):1922–1965. https://doi.org/10.1097/HEP.0000000000000466

Singbartl K, Forlow SB, Ley K (2001) Platelet, but not endothelial, P-selectin is critical for neutrophil-mediated acute postischemic renal failure. FASEB J. 15(13):2337–2344

Song J, Ma N, Aini R et al (2025) A machine learning model for non-invasive prediction of advanced liver fibrosis in patients with chronic hepatitis B. Am J Transl Res 17(7):4939–4951. https://doi.org/10.62347/KEVQ8263.PubMedPMID:40821076;PubMedCentralPMCID:PMCQ3.eng

Su CY, Kuo YP, Nieh H-L et al (2008) Quantitative assessment of the kinetics of growth factors release from platelet gel. Transfusion 48(11):2414–2420. https://doi.org/10.1111/j.1537-2995.2008.01862.x.PubMedPMID:18694465;PubMedCentralPMCID:PMCQ3.eng

Sun L, Yang N, Liu Z et al (2024) Cholestasis-induced phenotypic transformation of neutrophils contributes to immune escape of colorectal cancer liver metastasis. J Biomed Sci 31(1):66. https://doi.org/10.1186/s12929-024-01052-3

Sun Y, Sun K, Ling H et al (2025) Farnesoid X receptor-driven metabolic plasticity: bridging physiological adaptation and malignant transformation in lipid handling (Review). Int J Mol Med. https://doi.org/10.3892/ijmm.2025.5551

Tong Y, Lou X (2025) Interplay between bile acids, gut microbiota, and the tumor immune microenvironment: mechanistic insights and therapeutic strategies. Front Immunol 16:1638352. https://doi.org/10.3389/fimmu.2025.1638352

Wang H-W, Chen C-Y, Lai H-C et al (2023) Prediction model of hepatocellular carcinoma in patients with hepatitis B virus-related compensated cirrhosis receiving antiviral therapy. Am J Cancer Res. 13(2):526–537

Wang Y, Lei J, Jin Z et al (2025) Development and validation of a machine learning-based clinical prediction model for monitoring liver injury in patients with pan-cancer receiving immunotherapy. Int J Med Inform 203:106036. https://doi.org/10.1016/j.ijmedinf.2025.106036

You H, Wang X, Ma L et al (2023) Insights into the impact of hepatitis B virus on hepatic stellate cell activation. Cell Commun Signal 21(1):70. https://doi.org/10.1186/s12964-023-01091-7

Zhang H, Wu X, Ni W et al (2025) Application of machine learning and deep learning in metabolic dysfunction-associated steatotic liver disease: a systematic review and meta-analysis. J Adv Res. https://doi.org/10.1016/j.jare.2025.08.042.PubMedPMID:40850679;PubMedCentralPMCID:PMCQ1.eng

Zhonghua Gan Zang Bing Za Zhi. 2022. Guidelines for the prevention and treatment of chronic hepatitis B (version 2022). 30(12):1309-1331. https://doi.org/10.3760/cma.j.cn501113-20221204-00607

Zhou R-R, Song Y-H, Xu C-Y et al (2024a) Altered counts and mitochondrial mass of peripheral blood leucocytes in patients with chronic hepatitis B virus infection. J Cell Mol Med 28(12):e18440. https://doi.org/10.1111/jcmm.18440

Zhou X, Zhou X, Zhang Z et al (2024b) Mechanism of bile acid in regulating platelet function and thrombotic diseases. Adv Sci 11(32):e2401683. https://doi.org/10.1002/advs.202401683

Zhou H, Mu X, Hu H et al (2025) DHLCA alleviates diabetic kidney disease via TGR5/FXR activation and gut microbiota remodeling. Drug des Devel Ther 19:6469–6485. https://doi.org/10.2147/DDDT.S530823

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.